JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Topology is the Principal Determinant in the Folding of a Complex All-alpha Greek Key Death Domain from Human FADD

Annette Steward, Gary S. McDowell and Jane Clarke*

University of Cambridge Department of Chemistry, MRC Centre for Protein Engineering, Lensfield Road, Cambridge, CB2 1EW, UK

In order to elucidate the relative importance of secondary structure and topology in determining folding mechanism, we have carried out a phi-value analysis of the death domain (DD) from human FADD. FADD DD is a 100 amino acid domain consisting of six anti-parallel alpha helices arranged in a Greek key structure. We asked how does the folding of this domain compare with that of (a) other all-alpha-helical proteins and (b) other Greek key proteins? Is the folding pathway determined mainly by secondary structure or is topology the principal determinant? Our Φ-value analysis reveals a striking resemblance to the all-beta Greek key immunoglobulin-like domains. Both fold via diffuse transition states and, importantly, long-range interactions between the four central elements of secondary structure are established in the transition state. The elements of secondary structure that are less tightly associated with the central core are less well packed in both cases. Topology appears to be the dominant factor in determining the pathway of folding in all Greek key domains.

© 2009 Elsevier Ltd. All rights reserved.

Introduction

Comparative studies of the folding of homologous domains have proved to be valuable in understanding protein folding.[1,2] Two principal questions are asked in such studies: the first concerns the pathway; which elements of structure fold early and which fold late? The second is a question of mechanism; it has been suggested that there is a continuum of mechanisms,[3] from strictly framework, diffusion collision mechanisms characterised by polarised transition states,[4,5] to pure nucleation condensation mechanisms, characterised by diffuse transition states with strictly concomitant formation of secondary and tertiary structure.[6] Although in a few protein families, the folding pathway does not appear to be conserved (e.g. protein G and protein L[7]) in many protein families the transition states of homologous proteins are remarkably similar; even where the transition state of one protein is much more structured, the general structural features of the transition state are maintained. This has been observed, for instance, in the immunity proteins Im7 and Im9, which have 60% sequence identity,[8] and in the immunoglobulin-like (Ig-like) proteins including the Ig domain from human titin, TI I27, and a fibronectin type III domain from human tenascin, TNfn3, which have no significant sequence identity, although they share a common fold.[9–11] Folding mechanisms within a fold can vary significantly between homologous proteins, even where the transition states have similarly structured regions. In the homeodomain superfamily, for instance, folding mechanisms range from a strict framework/diffusion collision mechanism (engrailed homeodomain) to pure nucleation condensation (hTRF1).[3] In this case, the mechanism is apparently determined by secondary structure (helical) propensity. Thus, studies of the folding of related proteins

---

*Corresponding author. E-mail address: jc162@cam.ac.uk.
Present address: G.S. McDowell, Hutchison/MRC Research Centre, Hills Road, Cambridge, CB2 0XZ UK.

Abbreviations used: DD, death domain; FADD, Fas-associated death domain protein; FADD DD, FADD death domain; Ig, immunoglobulin; B1, bundle 1; B2, bundle 2; TNfn3, third fibronectin type III domain of human tenascin, an Ig-like domain; WT, wild type; TS, transition state; m_{k_f}, denaturant dependence of the folding rate constant; m_{k_u}, denaturant dependence of the unfolding rate constant.

suggest that both topology and the nature of the secondary structure content have a role in determining how a protein folds.[12]

We have studied the folding of two protein families extensively. The first is the complex, all-beta Greek key Ig-like fold.[9–11,13–15] Topology is the dominant factor influencing the folding of Ig-like domains. All these proteins fold via a nucleation condensation process, where key long-range interactions, between residues in the central BCEF strands, form early to nucleate folding. The transition state of these proteins is an expanded form of the native state and local interactions are not involved in determining the folding pathway.

In contrast to the complex Greek key topology of the Ig-like domains, spectrin domains are simple three-helix coiled-coil structures with up-down-up connectivity. Three spectrin domains have been studied in our laboratory (R15, R16 and R17 from chicken brain α-spectrin).[16–18] In all three of these proteins, two elements of secondary structure (helices A and C) form and interact early and the third (helix B) forms and packs after the rate-determining transition state. There is some suggestion of a change in mechanism, as is seen in the homeodomain family; R16 folds by a framework-like mechanism, with secondary structure (especially in helix C) preceding tertiary structure formation,[17] whereas R15 folds by a nucleation condensation mechanism with secondary and tertiary structure forming concomitantly (our unpublished results). R17 shows a mechanism similar to that of R16, but with higher Φ-values and more helical structure in helix A.[18]

The principal aim of this study was to compare the folding of a death domain (DD) from human FADD (FADD DD), an all-helical protein with a complex Greek key topology, with the two other classes of protein studied in depth in this laboratory: spectrin domains, simple all-alpha three-helix bundles, and Ig-like domains, all-beta proteins with complex Greek key topology.[19,20] The death domains provide an opportunity to study the interplay between formation of secondary structure and topology in protein folding. Death domains have six anti-parallel alpha helices, arranged in a Greek key topology, with helices 1, 5 and 6 grouped in an approximately orthogonal position above helices 2, 3 and 4 (Fig. 1).[21] Human

FADD comprises two domains from the DD super-family, each approximately 100 amino acids; the C-terminal DD is the one studied here.

In the Ig-like domains, two elements of structure, twisted beta sheets, pack together to form the hydrophobic core. Central to this structure is a four-strand motif consisting of two pairs of anti-parallel beta strands, one from each sheet (B-E and C-F, Fig. 1). Similarly, death domains can be thought of as two three-helix bundles, packed together via a central four-helix motif: bundle 1 (B1) is made from the packing of H1, H5 and H6, and forms the B1 core. Bundle 2 is made of contiguous helices H2, H3 and H4, which pack to form the B2 core. The central four-helix motif comprises two pairs of parallel helices (H1-H5 and H2-H4) packed together orthogonally, to form the central core. As is shown in Fig. 1, the three cores are clearly separated, so that different faces of H1, H2, H4 and H5 contribute to different cores. The peripheral helices H3 and H6 contribute only to the packing of the bundle cores (B2 and B1, respectively). In contrast, the Ig-like domains have a single hydrophobic core formed by packing of the twisted beta sheets. Thus, FADD DD can be thought of as two spectrin-like, three-helix bundles packed together to form a Greek key structure with central elements forming the central core in a manner that is reminiscent of Ig-like domains.

Results

Studies of wild type FADD DD

The equilibrium stability of wild type (WT) FADD was determined at pH 7.0, 25 °C in urea with unfolding monitored by changes in both intrinsic fluorescence and elipticity at 222 nm. Fluorescence and circular dichroism (CD) data overlaid (data not shown), indicating folding to be a completely reversible, co-operative, two-state transition. Data from five experiments gave an average $[D]_{50\%}$ (the concentration of denaturant at which half the molecules are unfolded) of 4.8 ± 0.1 M and a mean m-value of 1.4 ± 0.2 kcal mol^{-1} M^{-1}; this gives a mean free energy of unfolding for WT of 6.7 ± 0.15 kcal mol^{-1}. Kinetic rate constants obtained by fluorescence are in

Fig. 1. The structure of FADD DD compared to Ig-like and spectrin domains. (a) FADD DD (PDB code *1E41*) is composed of two three-helix bundles packed orthogonally. Bundle 1 (red) comprises helices 1, 5 and 6; bundle 2 (blue) comprises helices 2, 3 and 4. (b) An alternative view of FADD DD showing that the central region of the protein has two pairs of parallel helices packing against each other (green), H1 and H5 from bundle 1 and H2 and H4 from bundle 2. H3 and H6 are peripheral to the structure and make contacts only within their respective bundles. (c) FADD DD has three structurally distinct cores: two formed by each three-helix bundle (red and blue) and a central core formed by packing of the two pairs of central helices (green). (d) A diagram showing the 2-D topology of FADD DD. Five of the helices display classical Greek key topology.[19,20] The helices are coloured to distinguish between bundle 1 (red) and bundle 2 (blue). The helix that is not part of the Greek key motif is shown in pale red. (e) Greek key Ig-like domains (in this case, TNfn3, PDB code *1ten*) comprise two anti-parallel pairs of strands (one pair from each sheet, shown in green) which pack against each other to form the centre of the single hydrophobic core. Ig-like domains differ in the number and arrangement of peripheral strands (shown in red and blue to distinguish the sheets). (f) A diagram showing the 2-D topology of TNfn3. Five of the strands display classical Greek key topology. The two sheets are coloured in red and blue. Strands that are not part of the Greek key motif are shown in pale colours. (g) Spectrin domains (in this case spectrin R16 from PDB code *1cun*) are simple three-helix bundles with the same up-down-up arrangement as the bundles of FADD DD.
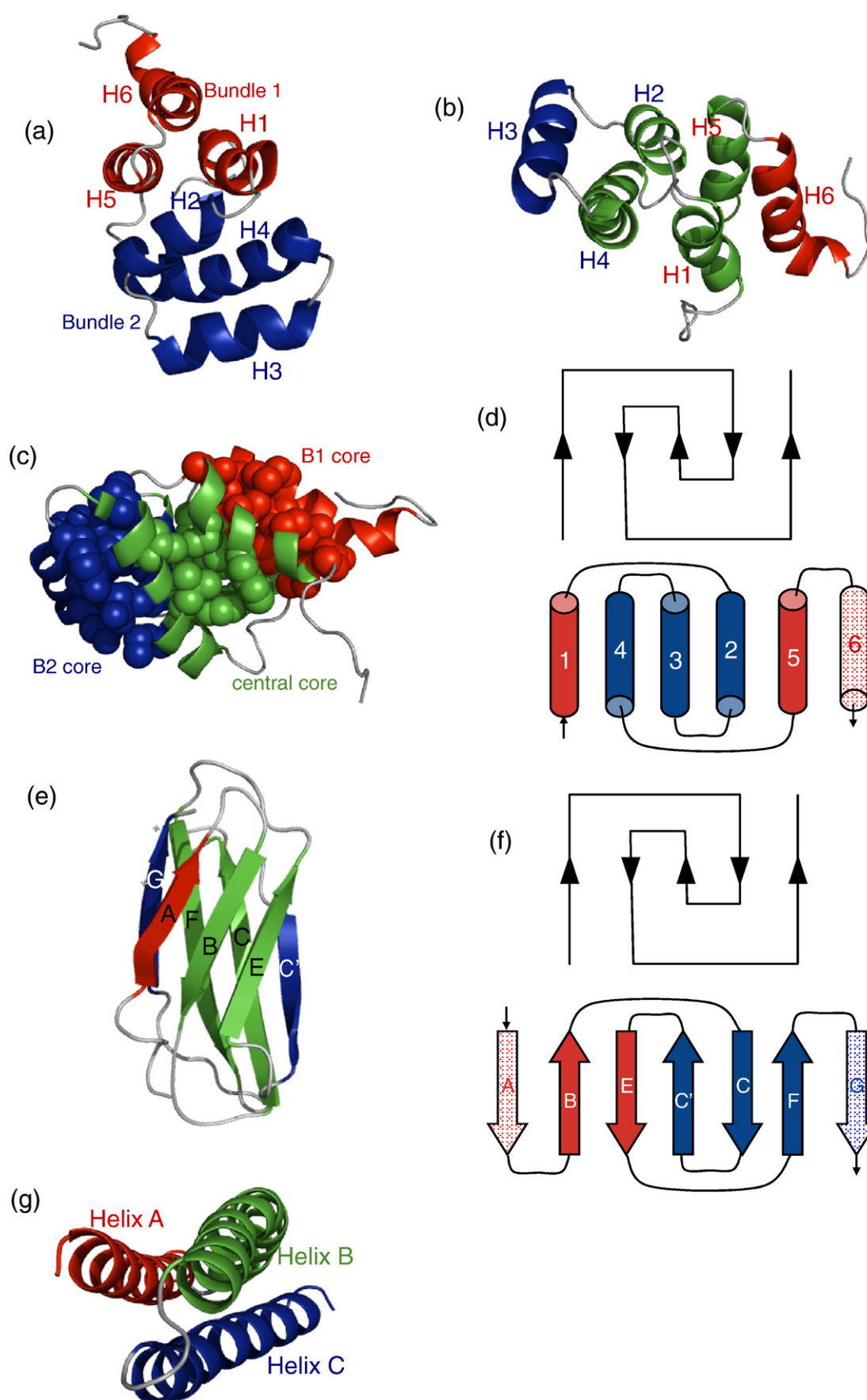
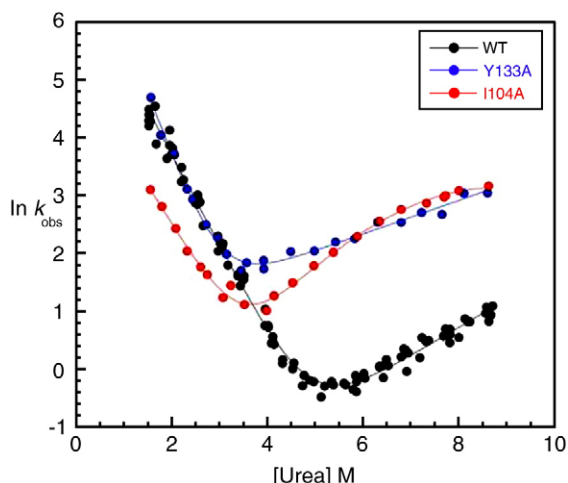**Fig. 1** (*legend on previous page*)

Fig. 2. Sample chevron plots for FADD DD. WT (black) and some mutants, such as Y133A (blue), have straight folding and unfolding limbs; other mutants, such as I104A (red), have clearly curved unfolding limbs. This is not a reflection of stability, i.e. the length of the unfolding limb; I104A and Y133A have similar [urea]$_{50\%}$ values. Chevron plots for all mutants are included in the Supplementary Data.

agreement with those obtained by CD. WT FADD DD contains one proline residue, but kinetic data fit well to a single exponential at all concentrations of urea. The chevron plot for WT FADD DD fit well to Eq. (2) (see Materials and Methods) with a linear dependence of both lnk_f and lnk_u (where k_f is the folding rate constant and k_u is the unfolding rate constant) on the concentration of urea (Fig. 2). No data were analysed below 1.5 M urea, because the refolding rate constants become dependent upon protein concentration. This behaviour has been ascribed to transient aggregation.[22] Linear extrapolation of chevron data to 0 M urea gives a $k_f^{H_2O}$ of 960±60 s^{-1} and a $k_u^{H_2O}$ of 0.04±0.01 s^{-1}; WT kinetic m-values give a β-Tanford (β$_T$) value of 0.76 (see Materials and Methods), indicating that the transition state is compact. The free energy of unfolding derived from equilibrium and kinetic data was compared at 2 M urea (to avoid long extrapolation to 0 M urea); the WT kinetic^{2M} stability is 3.5±0.1 kcal mol^{-1}, compared with that from equilibrium^{2M} experiments of 3.9±0.2 kcal mol^{-1}.

Choice of mutations

To obtain Φ-values for core residues, and residues within elements of secondary structure, 64 mutations were made at 42 positions throughout the six helices: 23 of these were mutations of core residues, and 19 were residues where alanine to glycine "scanning" was performed, to report on the extent of secondary structure formation.[23,24] Of the 23 core residues, most were non-disruptive replacements by alanine; however, in four cases this mutation was too destabilising and so non-conservative substitutions were made (Leu to Met

and Trp to Phe). The residues to which the contacts are deleted and the number of heavy-atom side-chain contacts within 6 Å deleted on mutation are given in Table 1, which illustrates that truncation of core residues either reports mainly on contacts within one of the bundle cores or probes contacts in the central core. Thus, for example, in H4, R140 contacts only residues within B2, whereas L145 only makes long-range contacts with residues in H1 (in B1), through the central core.

Analysis of mutant data

The change in stability upon mutation ($\Delta\Delta G_{D-N}$) was determined using equilibrium denaturation followed by changes in fluorescence. All equilibrium m-values were the same as WT within error, and so $\Delta\Delta G_{D-N}$ was determined using a mean m-value (<m>) of 1.4±0.21 kcal mol^{-1} M^{-1} as:[25]

$$\Delta\Delta G_{D-N} = <m>.\delta[\text{urea}]_{50\%} \qquad (1)$$

where δ [urea]$_{50\%}$ is the difference in the midpoint of denaturation between WT and mutant protein.

Fitting the kinetic data for FADD DD mutants proved to be complex. All kinetics were measured using intrinsic Trp fluorescence as a probe to allow data to be collected at a low concentration (≤1 μM) of protein As for WT, both refolding and unfolding data fit well to a single-exponential equation for all mutants. To avoid the possibility of complication from aggregation events observed in the WT protein, refolding data below 1.5 M urea were not used for fitting the chevron plots. An exception to this rule was made for highly destabilised mutants with a low [urea]$_{50\%}$, where the refolding arm was short (W112A, W148F, L161A and L165A). For S144A, all data below 2.5 M urea were omitted from chevron fitting, due to aggregation. All mutant chevron plots had linear folding arms with essentially the same slope (only three had refolding m_{kf} values that were significantly different from the mean value of 1.7 M^{-1}).

Some mutant chevron plots had linear unfolding limbs, as was seen in WT. However, many mutants exhibited some downward curvature in the unfolding arm of their chevron plot and for a few mutants this curvature was very significant (Fig. 2). (What curved chevron plots may mean is discussed later). It was therefore not possible to do any global fitting of the data. All chevrons were fit individually: first, the chevrons were fit to an equation with a quadratic term in the unfolding limb only (a "Hammond" fit,[26] see Eq. (3)); second, each chevron plot was fit individually to an equation describing a linear dependence of both lnk_f and lnk_u on the denaturant concentration (linear chevron fit, see Eq. (2)). To fit the kinetic data to the linear chevron fit, data points that were judged to be curving downwards were omitted from the fit. (All chevrons are shown in Supplementary Data

Table 1. Contacts deleted on mutation of core residues

	Number of contacts deleted to each helix[a]						Total[a]
Mutation	Helix 1	Helix2	Helix3	Helix4	Helix5	Helix6	B1/B2
Helix 1							96
F101A(H1)	36	0	0	43	17	0	17/43
V103A(H1)	16	0	0	0	6	20	42 26/0
I104A(H1)	15	13	0	25	9	0	62 9/38
Helix 2							194
W112A(H2)	5	54	64	71	0	0	5/135
L115M(H2)[b]	11	17	0	6	29	0	63 40/6
L119M(H2)[b]	1	24	0	27	15	0	67 16/27
Helix 3							69
I126A(H3)	0	50	19	0	0	0	0/50
I129A(H3)	0	21	26	23	0	0	70 0/44
Y133A(H3)	0	0	33	48	0	0	81 0/48
Helix 4							129
R140A(H4)	0	57	65	7	0	0	0/122
V141A(H4)	12	14	2	17	0	0	45 12/16
S144A(H4)	1	11	3	11	0	0	26 1/14
L145M(H4)[b]	24	0	0	41	1	0	66 25/0
W148F(H4)[b]	38	2	0	65	42	0	173 80/28
Helix 5							97
H160A(H5)	0	16	0	62	19	0	78/0
L161A(H5)	28	7	0	21	19	2	77 28/30
V162A(H5)	1	0	0	0	10	26	37 0/27
L165A(H5)	11	12	0	0	21	19	63 12/30
C168A(H5)	0	10	0	0	10	0	20 10/0
Helix 6							45
V173A(H6)	13	5	0	0	15	12	5/28
L176A(H6)	23	0	0	0	1	18	42 0/24
V177A(H6)	8	0	0	0	15	16	39 0/23
V180A(H6)	11	0	0	0	11	9	31 0/22

Data calculated for the FADD DD NMR structure *1e41*,[21] using the program InsightII (Accelrys Inc.).

[a] Total contacts made by sidechain (excluding Cβ) to other sidechain heavy atoms within 6 Å is shown in black. Contacts made to other helices within each bundle shown in red for B1 (H1, H5 and H6) and blue for B2 (H2, H3 and H4). Thus, for example, V103 (H1) makes long-range contacts only within B1, whereas I104 (H1) makes the majority of its long-range contacts with residues in B2 (through the central core).

[b] Note that for non-conservative replacements, these contacts are simply those made by the WT sidechain.

Figs. S1 and S2, fit to a linear equation, identifying the points omitted.)

Most importantly, all Φ-values in this work have been determined using refolding data only, to avoid any uncertainty that might arise from fits of the unfolding data. Also, to avoid a long extrapolation to 0 M denaturant[27] and to eliminate any possible effects of aggregation, all Φ-values were calculated at 2 M urea. We show (Supplementary Data Fig. S 4) that the Φ-values determined using either linear or Hammond fits of the data are essentially identical. Thus, despite the complexity of the analysis, the Φ-

Table 2. Folding data for FADD DD core mutants

FADD DD variant	Core probed by mutation	$\Delta G_{\text{D-N}}$ (kcal mol^{-1})	$\Delta\Delta G_{\text{D-N}}$ (kcal mol^{-1})	$m_{\text{D-N}}$ (kcal mol^{-1} M^{-1})	k_{f} $^{2\text{M}}$ (s^{-1})	$\Phi^{2\text{M}}$
WT		6.7		1.4	40	
Helix1						
Phe101Ala	Central	4.2	2.6	1.6	10	0.31
Val103Ala	B1	5.3	1.4	1.2	36	0.04
Ile104Ala	Central	4.3	2.4	1.1	12	0.29
Helix2						
Trp112Ala	B2	2.8	3.9	1.4	10	0.18
Leu115Met	Central	5.1	2.4	1.4	22	0.22
Leu119Met	Central	4.3	1.6	1.3	28	0.10
Helix3						
Ile126Ala	B2	5.6	1.2	1.2	36	0.05
Ile129Ala	B2	4.5	2.2	1.5	43	−0.02
Tyr133Ala	B2	4.5	2.2	1.1	39	0.01
Helix4						
Arg140Ala	B2	7.3	−0.6	1.1	72	0.60
Val141Ala	Central	5.5	1.3	1.3	14	0.51
Ser144Ala	B2	7.4	−0.7	2.1	114	0.86
Leu145Met	Central	4.5	2.3	1.4	23	0.15
Trp148Phe	Central	3.6	3.1	1.4	55	−0.06
Helix5						
His160Ala	Central	7.7	−1.0	1.3	80	0.40
Leu161Ala	Central	2.9	3.9	1.6	4	0.37
Val162Ala	B1	4.9	1.9	1.5	16	0.29
Leu165Ala	B1	2.8	3.9	1.6	7	0.26
Cys168Ala	Central	6.4	0.4	1.4	42	n.d.
Helix6						
Val173Ala	B1	5.2	1.1	1.4	29	0.17
Leu176Ala	B1	6.0	0.8	1.3	25	0.37
Val177Ala	B1	4.0	2.7	1.4	24	0.11
Val180Ala	B1	6.8	−0.1	1.2	47	n.d.

For clarity, errors are not shown in the table. The error in free energy measurements is in the range±0.1-0.2 kcal mol^{-1}, the error in k_{f} $^{2\text{M}}$ is±10%, the error in $\Phi^{2\text{M}}$<0.1. Note that Φ-values cannot be considered reliable if the $\Delta\Delta G_{\text{D-N}}$ is low[27], so Φ-values were not determined where $\Delta\Delta G_{\text{D-N}}$<0.6 kcal mol^{-1} (n.d).
We note that a condition of Φ-value analysis is that conservative mutations should be made[30]; where this is not possible, for example, Trp to Phe, these results should be interpreted with caution and in the context of the surrounding Φ-values obtained.

values obtained are reproducible. The results of the kinetic analysis are shown in Tables 2 and 3.

Φ-Value analysis

Φ-Values obtained were either zero or fractional, with no Φ-value of 1 (Tables 2 and 3), indicating no part of the protein is as fully formed in the transition state as it is in the native state. To interpret a Φ-value analysis, it is customary to consider Φ-value patterns, rather than to try to interpret individual Φ-values. This allows one to determine which regions of the protein are fully unfolded, partially folded or fully folded in the transition state.[28] The Φ-values obtained are thus generally classified into low, medium and high classes, with the boundaries chosen to reflect the overall Φ-value distribution (e.g. see Ref. [29]): in our case, the cut-offs used are: low, $\Phi \leq 0.15$; medium, $\Phi = 0.16$–0.5; high, $\Phi \geq 0.51$.

We will consider the structure of the transition state of FADD DD in terms of formation of the three structural cores, i.e. formation of the two

three-helix bundles B1 and B2, and the formation of the central core (the four-helix motif). (A detailed analysis of each helix is given in the Supplementary Data). The Φ-values were mapped on the native structure, to allow the patterns of Φ-values to be interpreted.

Formation of B1 (formed by H1, H5 and H6): (Fig. 3)

Formation of the B1 core was probed by six mutations: **H1**, V103A; **H5**, V162A, L165A; and **H6**, V173A, L176A, V177A. The extent of helix formation was probed by six Ala to Gly mutations: **H1**, A98G, A102G; **H5**, A166G, A167G; and **H6**, A175G, A178G.

All three helices are partially folded but the B1 core can be considered as only very weakly structured. Only H5 and H6 make contact through the B1 core, through interactions between L165 at the C-terminus of H5 and V173 at the N-terminus of H6. H1 is only connected to the B1 core in the native state via V103, but the Φ-value of V103A is zero, suggesting that H1 does not contribute to the formation of the B1 core in the transition state (TS).

Table 3. Folding data for FADD DD helix scanning mutants

FADD DD variant	ΔG_{D-N} (kcal mol^{-1})	$\Delta\Delta G_{D-N}$ (kcal mol^{-1})	m_{D-N} (kcal mol^{-1} M^{-1})	k_f^{2M} (s^{-1})	Φ^{2M}
Helix 1					
Cys98Ala	7.2		1.8	48	
Cys98Gly	6.2	1.0	1.6	26	0.38
Asn102Ala	6.9		1.3	58	
Asn102Gly	6.3	0.7	1.3	33	0.50
Helix2					
Arg113Ala	6.5		1.2	52	
Arg113Gly	5.4	1.1	1.1	38	0.17
Arg114Ala	6.4		1.6	46	
Arg114Gly	5.0	1.4	1.6	28	0.20
Arg117Ala	6.3		1.2	47	
Arg117Gly	5.2	1.1	1.1	31	0.23
Helix3					
Asp127Ala	7.2		1.5	49	
Asp127Gly	6.2	1.0	1.4	45	0.07
Ser128Ala	7.0		1.1	37	
Ser128Gly	6.6	0.43	1.4	47	n.d.
Asp131Ala	7.5		1.4	44	
Asp131Gly	6.5	1.0	1.4	46	−0.03
Arg132Ala	6.6		1.2	40	
Arg132Gly	5.5	1.1	1.1	43	−0.04
Helix4					
Thr138Ala	7.1		1.4	68	
Thr138Gly	6.4	0.7	1.5	35	0.56
Glu139Ala	6.4		1.4	29	
Glu139Gly	5.7	0.7	1.3	16	0.49
Arg142Ala	5.9		1.3	32	
Arg142Gly	4.8	1.1	1.3	12	0.51
Glu143Ala	6.6		1.4	34	
Glu143Gly	4.2	2.4	1.7	10	0.29
Arg146Ala	6.3			30	
Arg146Gly	5.1	1.2	20	20	0.19
Asn150Ala	6.8		2.0	46	
Asn150Gly	6.2	0.6	1.5	32	0.35
Helix5					
Arg166Ala	4.6		1.3	22	
Arg166Gly	3.7	0.9	1.6	13	0.30
Ser167Ala	7.6		1.7	69	
Ser167Gly	6.4	1.3	1.7	28	0.42
Helix6					
Asp175Ala	6.4		1.4	38	
Asp175Gly	5.4	1.0	1.5	23	0.12
Gln178Ala	6.6		1.2	49	
Gln178Gly	5.6	1.0	1.4	27	0.31

The $\Delta\Delta G_{D-N}$ and Φ-values are determined for the Ala to Gly mutation. For clarity, errors are not shown in the table. The error in free energy measurements is in the range ±0.1–0.2 kcal mol^{-1}; the error in k_f^{2M} is ±10% and the error in Φ is <0.1. Note that Φ-values cannot be considered reliable if $\Delta\Delta G_{D-N}$ is low,[27] so Φ-values were not determined when $\Delta\Delta G_{D-N}$ <0.6 kcal mol^{-1} (n.d.).

(Note, however, that H1 and H5 do interact in the TS, through sidechain contacts on the central core side of the helices).

Formation of B2 (formed by H2, H3 and H4): (Fig. 3)

Formation of the B2 core was probed by six mutations: **H2**, W112A; **H3**, I126A, I129A, Y133A; and **H4**, R140A, S144A. The extent of helix formation was probed by 12 Ala to Gly mutations: **H2**, A113G, A114G, A117G; **H3**, A127G, A131G, A132G; and **H4,** A138G, A139G, A142G, A143G, A146G, A150G.

Helix 4 is the most structured region of the entire protein, with high Ala to Gly Φ-values at the N-terminal end, becoming lower towards the C-terminal end. The two B2 core residues in this helix both have high Φ-values; W112 (the only hydrophobic residue from H2 which packs into this core) has a medium Φ-value. All the Φ-values of helix 3 are close to zero; this helix plays no role in the formation of B2 in the TS. Thus, we infer that the B2 core is loosely structured in the TS with H2 packing onto H4.

Formation of central core formed by H1, H2, H4 and H5 (Fig. 3)

Formation of the central core was probed by nine mutations: **H1**, F101A, I104A; **H2**, L115M, L119M; **H4**, V141A, L145M, W148F; and **H5**, H160A, L161A. The extent of helix formation was probed by 13 Ala to Gly mutations: **H1**, A98G, A102G; **H2**, A113G, A114G, A117G; **H4,** A138G, A139G, A142G, A143G, A146G, A150G; and **H5**, A166G, S167G. (Note that in native FADD DD, H1 and H5 of the central core run parallel with each other and are packed orthogonally onto the parallel helix pair H2 and H4; Fig. 1).

In the TS, the central core is partly formed, principally through interaction of F101 and I104 in H1, which contact residues from all the other three helices; H5 also contributes significantly to core packing, via H160 and L161 at the N-terminal end. The central core residue in H2, L115, which packs onto residues in both H1 and H5 in the native state, has a medium Φ-value. Notably, although H4 is apparently well structured, the only core residue that contributes structure in the TS is V141 at the extreme N-terminus of H4. It has a high Φ-value, and appears to pin this end of H4 to H1 via an interaction with I104. In the native state, the central core of FADD DD is dominated at one end by W148F from H4, which has a Φ-value of zero. Thus, one end of the central core appears to be largely unstructured, and H4 is essentially attached only via contacts with H2 (via the B2 core).

Discussion

There is no evidence for the presence of a refolding intermediate

Curvature or "roll-over" in the re-folding arm of a chevron is good evidence for the presence of a populated folding intermediate. However, it has been shown that curvature in the refolding arm can indicate the presence of transient aggregation at low concentrations of denaturant, as is observed in FADD DD. Any curvature present due to refolding from an intermediate will be masked by this aggregation. The non-concurrence of equilibrium and kinetic data is another indication of the presence of a refolding
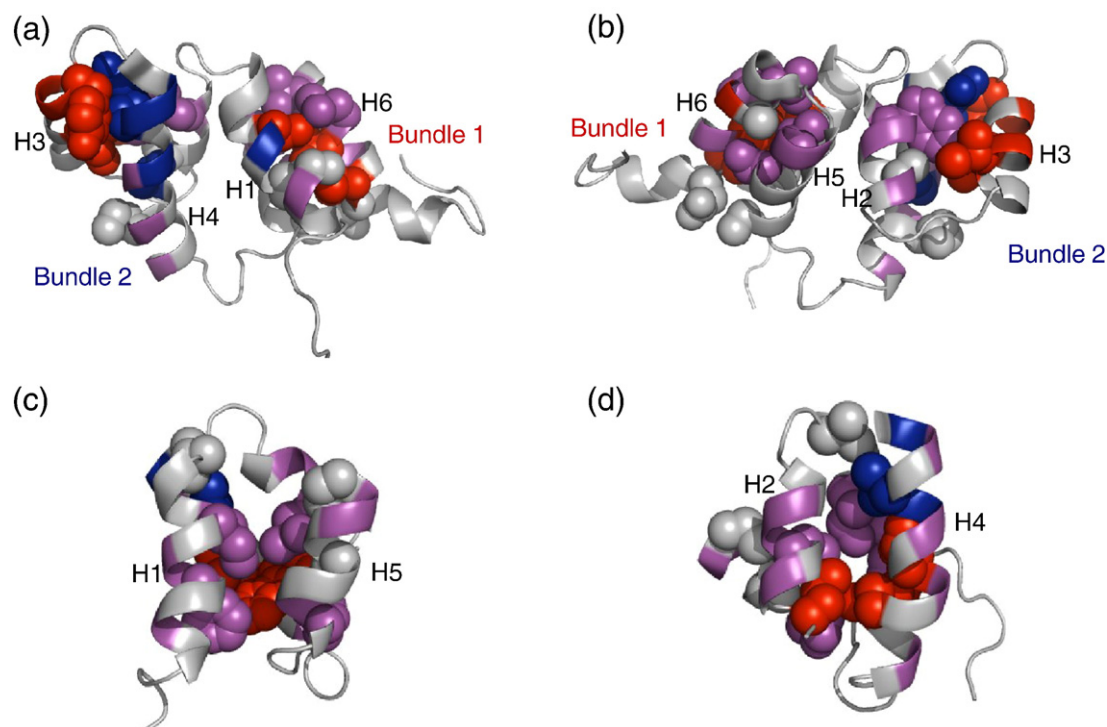
Fig. 3. The Φ-values mapped onto the native structure of FADD DD. High Φ-values (≥0.51) are shown in blue, medium Φ-values (0.16–0.5) are shown in magenta and low Φ-values (≤0.15) are shown in red. The Ala to Gly scanning surface mutations are shown by colouring the ribbon, the buried residues mutated are shown as spheres. (a and b) The Φ-values of the three helix bundles from different faces of the molecule. (c and d) The Φ-values of the central core from different faces. H3 and H6 have been removed to allow a clear view. In c, H6 would be facing the viewer; in (d) H3 would be facing the viewer.

intermediate accumulating on the pathway.[30] Where a Hammond fit is used, the agreement between kinetic and equilibrium free energies is good (Supplementary Data). Moreover, the average kinetic and equilibrium m-values, taken from WT and mutants, are the same within error (1.3±0.1 kcal mol^{-1} M^{-1} and 1.4±0.2 kcal mol^{-1} M^{-1}, respectively). Furthermore, an amplitude analysis by CD showed no evidence for a dead-time change in amplitude, which would suggest that no helical structure is formed in the dead-time of the stopped-flow experiments (data not shown). Thus, our experiments suggest that, at least above 1.5 M urea where we analyse the kinetic data, the folding of FADD DD is essentially two-state.

The transition state for folding in human FADD DD

In the TS of the human FADD DD, all three cores are partially structured. Helices H3 and H6, which are peripheral to the structure, are essentially detached from the protein. H3 is the only element of secondary structure that is completely unstructured, and H6 is only pinned to H5 via a helix-turn-helix interaction mediated through V173. In contrast, all four central elements of structure are already associated in the TS: three helices (H1, H2 and H5) contribute to central core packing, with the fourth helix (H4) pinned into the core at one end. However, H2 and H4 clearly interact with each other via residues that are packed in

the B2 core. Thus, in the TS, helices H1, H2, H4 and H5 are apparently aligned in two parallel pairs (H1/H5 and H2/H4) and the pairs are packed against each other through interactions between H2 and the H1/H5 pair. Long-range packing interactions (H1 – H5, H2 – H4 and H2 – H1/H5) have a predominant role in the formation of structure in the central four-helix bundle, and in B2. Only in the formation of the B1 core do short-range side-chain interactions appear to be significant (in the formation of the helix 5-turn-helix 6 motif).

Analysis of the secondary structure propensity of FADD DD (as determined using the program AGADIR[31]) suggests that the level of overall helical propensity is very low; only two regions of the protein, the N-terminal region of H4 and the C-terminal region of H6, show any significant helical propensity (>10%). Interestingly, neither of these regions is involved in forming the central topology-defining core, although the N-terminal region of H4 appears to be important in the folding of B2.

Curvature in the unfolding arm suggests consolidation of the central hydrophobic core in a late transition state

Downwards curvature in the unfolding arm can formally be caused by: (i) changes in the ground (native) state;[32] (ii) movement of the transition state along a broad energy barrier;[33,34] or (iii) a high-energy
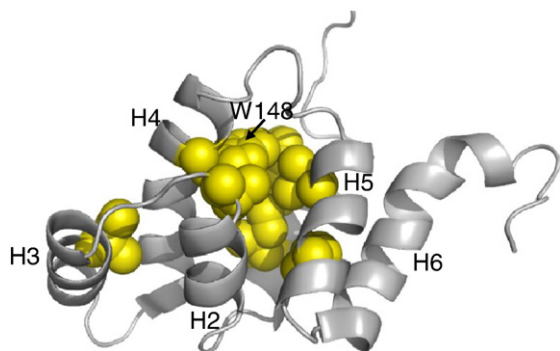
Fig. 4. Residues that show strong curvature in unfolding. Apart from a single residue in H3, all the residues that show strong curvature interact in the central core, in particular the residues that surround W148. We infer that the central core consolidates at this end as the protein traverses the TS region of the energy landscape.

intermediate separating two transitions states that switch due to, for example, mutation or an increase in the concentration of denaturant.[35] (See Refs [36] and [37] for detailed discussion of different explanations for curvature.) Note that in both models (ii) and (iii), mutation of a residue induces curvature when it is in a region of the protein that is more structured in a late transition state than in an early transition state. Thus, analysis of the chevron plots should, in principle, allow us to infer consolidation of the protein as it crosses the transition state barrier(s).[17,38]

There is no evidence for changes in the native state at high concentrations of denaturant. WT and some highly destabilised mutants show no evidence of curvature in the unfolding limbs of the chevron plots, and we see no obvious change in kinetic amplitude. To try to make some qualitative sense of our data, we examined chevron plots that had been fit to a linear equation to determine at which concentration of denaturant deviation from linearity occurred. For most mutants, the unfolding limb was either entirely linear (as in WT) or deviated from linearity only at very high concentrations of urea. A few mutants, however, show very clear deviation from linearity at concentrations of urea below 7 M: F101A and I104A in H1; L119M in H2; I129A in H3; W148F in H4; and L161A and L165A in H5 (Fig. 2, and Supplementary Data Fig. S1). These residues are shown in Fig. 4. Some mutants (L119M, I129A and W148F) have low Φ-values in the "early" transition state (determined at 2 M urea), suggesting that they are not significantly structured, and others (F101A, I104A, L161A and L165A) have medium Φ-values. The significant curvature suggests that the Φ-values of all of these mutants increase at higher concentrations of denaturant. Importantly, nearly all these residues form part of the central hydrophobic core, which is formed by the orthogonal parallel helix pairs: H1/H5 and H2/H4; in particular, they are found towards the C-terminal end of these helices (Fig. 4). In the early transition state structure, described above, the N-terminal part of this core is more structured than the

C-terminal part. Thus, the central hydrophobic core appears to consolidate as the protein traverses the transition state barrier. Significantly, buried residues in these same helices, H1, H2, H4 and H5, which pack on the opposite side of these helices into the hydrophobic cores formed by the three helix bundles, do not show similar curvature.

Importantly, the experiments cannot provide information about the order of structure formation. The Φ-values do not give sufficient resolution to determine whether the entire structure folds concomitantly, or if there is an order to the packing of the helices. Future simulations might allow us to detect early events.

Comparison with other members of the DD superfamily

Clark and co-workers have examined the folding of a number of WT DDs, all of which appear to have complex folding pathways characterised by multiphasic folding and unfolding.[39–42] Complex folding behaviour has been proposed, and they have suggested that this might be an intrinsic consequence of the complex all-alpha Greek key topology. This would, if true, be in stark contrast to the relatively simple, conserved folding mechanism observed in all-beta Greek key proteins, where populated intermediates are observed only in the more stable Ig-like domains.[43] Our results for FADD DD demonstrate clearly that these domains do not have intrinsically complex folding pathways. FADD DD folds in a simple two-state manner with no stable kinetic or equilibrium intermediate. Nonetheless, as is true for most of these other DDs, the observed rate constant of folding is somewhat lower than might be expected for such all-alpha proteins with relatively low contact orders.[44]

Comparison with the all-beta Greek key Ig-like domains

The folding of FADD DD can be compared with that of the all-beta Greek key Ig-like domains studied earlier, including TNfn3,[11,45] FNfn10,[13] CAfn2,[14] and TI I27[9,10] in our laboratory, and CD2d1.[15] In the Ig-like domains, tertiary structure is the dominant factor influencing the folding mechanism. These domains exhibit a near-classical nucleation-condensation mechanism where long-range key residues, in the central BCEF strands, interact in the TS to set up the complicated topology; the transition state is an expanded version of the native state, with the folding nucleus involving secondary and tertiary interactions, centred around the structural core. Peripheral regions pack late.

Unlike the Ig-like domains, FADD DD has three discrete hydrophobic cores. The central core is composed of residues from the central helices H1, H2, H4 and H5; the hydrophobic residues interacting in these helices can be regarded as the structural core of this four-helix bundle. The parallel helices H1-H5 and H2-H4 have to be organised relative to each other to form the Greek key topology; this is analogous to

the organisation of the anti-parallel beta strands B-E and C-F relative to each other, in the core of the Ig-like domains. In the Ig-like domains, we infer from the Φ-value analysis that the alignment of strands within the sheet and the packing of the two sheets together occurs concomitantly; these are the critical nucleating events for Ig-like domain folding. Although it is not possible to determine the order of events in the folding of the DD from the Φ-value analysis alone, we can show that helices H2 and H4 come together via the core of B2, whereas helices H1 and H5 require the formation of the central core to come together. In both Greek key structures, many long-range, tertiary interactions are involved in formation and stabilisation of the central core; however, local interactions important for helix formation are also involved in the DD. Interestingly, simulations of the TS for folding in the Ig-like protein TNfn3 suggested that the four strands form an "open horseshoe" structure, with all four central strands connected, but with the "ring" of structures incomplete.[45] This is reminiscent of what we observe in FADD DD with all four central helices in contact, but where the packing of H4 into the central core is only marginal. Another similarity with the Ig-like domains is the observation that the elements of secondary structure most peripheral to the central core are relatively unimportant in formation of the TS structure. In the Ig-like domains, the two beta strands at the N- and C-termini are not involved in forming the structural core, and these fold late. In the DD fold, the N-terminal helix H1 is involved in forming the central core, but H3 and the C-terminal helix H6 are both peripheral to the structure and not involved in packing the central hydrophobic core. The Φ-value analysis suggests that these peripheral helices are either completely unfolded (H3) or attached only loosely (H6) in the TS.

Although FADD DD and TNfn3 have similar stabilities, and appear to have similar folding mechanisms, FADD DD folds significantly faster than TNfn3 (~1000 s^{-1} compared to 6 s^{-1}). This is not unexpected; FADD DD is an all alpha-helical protein with a significantly lower relative contact order than the all-beta TNfn3, and it has been shown that proteins with low contact orders generally have higher folding rate constants.[44]

Comparison with other helix bundles

A number of helix-bundle proteins have been studied. FADD DD differs from all of these helical proteins in terms of structure, as three distinct hydrophobic cores can be identified. Here we compare the folding of FADD DD with simple helix bundle proteins. We largely ignore the cytochrome *c* proteins where the haem provides essential stability to the protein,[46] and the much larger globins, which have more complex hierarchical folding mechanisms.[47–50]

Formation of the central core of FADD involves packing of a four-helix bundle. Several other four-helix bundle proteins have been studied in detail by Φ-value analysis; two members of the ACBP family[51]

and apocytochrome b562[52] (which are up-down helical bundles) and Im7, the homologous Im9[8] and the FF domain of HYPA/FBP11[53] (which have three long helices plus a shorter helix). In all these cases, formation of the TS involves packing of three helices with one helix being essentially unstructured. It has been suggested that the early, obligatory stages in nucleation of folding will be formation of the interactions that are necessary to establish the overall topology.[54–56] Packing of three helices in a simple four-helix bundle is sufficient to establish the topology. It seems probable that the more complex Greek key topology of the DD requires all four central elements to be assembled, as was found in the Ig-like Greek key domains.

The folding of the two three-helix bundles of FADD DD can be compared with the folding of other three-helix bundle proteins. A number of these have been studied extensively, using Φ-value analysis. In some cases, one well-formed helix is observed in the TS with other elements of structure packed against it (as, for example, in protein A[27,57]), whereas in others, two elements of structure come together with the third helix relatively unstructured (e.g. spectrin domains[17,18] and peripheral subunit binding domains[58]). The three-helix bundles of FADD DD fall into this second category. Interestingly, as in the spectrin domains, the helices that are in contact are those that are separated in sequence (H1 with H5 and H2 with H4). This suggests that formation of these long-range interactions is the important step for folding these bundles.

Conclusion

Our Φ-value analysis of the complex all-alpha Greek key FADD DD reveals a striking resemblance to the all-beta Greek key Ig-like domains. Both fold via diffuse transition states and, importantly, long-range interactions between the four central elements of structure are established in the TS. This ensures the topology is established with the correct orientation and pairing of the helices/strands and correct packing of the helix/strand pairs that make up the central hydrophobic core of both proteins. The elements of secondary structure that are less tightly associated with the central core are less well packed in the TS in both cases. Topology appears to be the dominant factor in determining the pathway of folding in all Greek key domains studied to date. However, secondary structure also has a role: whereas in the all-beta Ig-like domains, long-range, tertiary interactions dominate the folding process, in the DD studied here, short-range, local interactions are also important in the folding process, reflecting the helical secondary structure. A recent theoretical study suggests that topology is likely to be the dominant factor in determining folding pathways in all-beta proteins where long-range interactions predominate.[12] However, in all-alpha proteins, where the majority of stabilising interactions are local, sequence changes are more likely to result in variable folding pathways

within a family. The Greek key DDs appear to be an ideal family for the investigation of this theoretical observation.

Materials and Methods

Protein expression and purification

A synthetic gene corresponding to residues 93–192 of the DD of human FADD, was obtained from overlapping primers using standard PCR techniques; the synthetic gene was inserted into a modified version of pRSETA (Invitrogen) that encodes an N-terminal histidine tag. Site-directed mutagenesis was performed using a Quik Change kit (Stratagene) and the identity of the mutants was confirmed by DNA sequencing.

Protein expression was carried out in *Escherichia coli* C41 (DE3) cells:[59] transformed cells were grown to an A_{600} of 0.4 – 0.6 at 37 °C before induction and growth overnight at 28 °C. The harvested cells were lysed by sonication and the protein purified from the soluble fraction, after centrifugation, by binding to Ni-NTA agarose (Qiagen). The bound protein was cleaved from the resin using thrombin (Sigma) and further purified using a Superdex G75 gel-filtration column (GE Healthcare). The purified protein was stored at 4 °C in 50 mM sodium phosphate, pH 7.0, 150 mM NaCl, 0.2 mM Tris(2-carboxyethyl)-phosphine (TCEP).

Determination of equilibrium stability

The stability of WT and mutant proteins was determined using urea-induced equilibrium denaturation followed by fluorimetry using a Cary Eclipse spectrometer (Varian); the excitation wavelength was 280 nm, the emission wavelength was 360 nm and the concentration of protein was ≤1–2 μM. The stability of the WT protein was also determined by CD using an Applied Photophysics Chirascan circular dichroism spectropolarimeter; data were collected at 222 nm and the concentration of protein was ≤5 μM. All experiments were performed at 25 °C in 50 mM sodium phosphate, pH 7.0, 150 mM NaCl, 5 mM DTT. The data were fit to a two-state transition as described.[60,61]

Kinetics

For WT and mutants, refolding and unfolding were monitored by changes in fluorescence, using an Applied Photophysics SX.18MV stopped-flow fluorimeter. An excitation wavelength of 280 nm was used with a 320 nm cut-off filter; the final concentration of protein was 1–2 μM. WT kinetics were also monitored by CD using an Applied Photophysics Π∗-180 instrument, with a final maximum concentration of protein of 5 μM. In both cases, the stopped-flow apparatus was maintained at 25 (±0.1) °C. Data collected from 8–12 experiments were averaged and traces were fit to a single-exponential function. Kinetic traces were analysed using Kaleidagraph (Synergy Software).

Φ-Value analysis

The variation of the logarithm of the observed rate constant with concentration of urea was analysed using Eq.

(2); those unfolding rate constants judged to be non-linear were omitted from the fit (see Supplementary Data):

$$\ln k_{\mathrm{obs}} = \ln\!\left(k_{\mathrm{f}}^{\mathrm{H_2O}}\exp\!\left(-m_{k_{\mathrm{f}}}.[\mathrm{urea}]\right) + k_{\mathrm{u}}^{\mathrm{H_2O}}\exp\!\left(m_{k_{\mathrm{u}}}.[\mathrm{urea}]\right) \right)$$

(2)

where k_{obs} is the observed rate constant, $k_{\mathrm{f}}^{\mathrm{H_2O}}$ and $k_{\mathrm{u}}^{\mathrm{H_2O}}$ are the folding and unfolding rate constants in 0 M denaturant, respectively, and $m_{k\mathrm{f}}$ and $m_{k\mathrm{u}}$ are the gradients of the folding and unfolding arms, respectively.

The variation of the natural logarithm of the observed rate constant with the concentration of urea was also analysed using Eq. (3), which includes a term to account for curvature in the unfolding limb (Hammond fit):

$$\begin{aligned}\ln k_{\mathrm{obs}} = \ln\Big(& k_{\mathrm{f}}^{\mathrm{H_2O}}\exp\!\left(-m_{k_{\mathrm{f}}}.[\mathrm{urea}]\right) \\ & + k_{\mathrm{u}}^{\mathrm{H_2O}}\exp\!\left(m_{k_{\mathrm{u}}}.[\mathrm{urea}] + m'_{k_{u}}.[\mathrm{urea}]^2\right)\Big)\end{aligned}$$

(3)

All Φ-values were calculated at 2 M urea using Eq. (4):

$$\Phi = \frac{\Delta\Delta G_{\mathrm{D-\ddagger}}}{\Delta\Delta G_{\mathrm{D-N}}}$$

(4)

where $\Delta\Delta G_{\mathrm{D-N}}$ is the equilibrium constant for unfolding determined from equilibrium denaturation experiments and $\Delta\Delta G_{\mathrm{D-\ddagger}}$ is the change in the free energy difference between the denatured (D) and transition (‡) states for folding upon mutation, calculated as:

$$\Delta\Delta G_{\mathrm{D-\ddagger}} = RT\ln\!\left(k_{\mathrm{f}}^{\mathrm{wt}}/k_{\mathrm{f}}^{\mathrm{mut}}\right)$$

(5)

$k_{\mathrm{f}}^{\mathrm{wt}}$ and $k_{\mathrm{f}}^{\mathrm{mut}}$ are the folding rate constants at 2 M urea for WT and mutant protein, respectively.

The β-Tanford (β_{T}) value was calculated using Eq. (6); $m_{k\mathrm{f}}$ and $m_{k\mathrm{u}}$ values were derived from chevrons fit to Eq. (2) only:

$$\beta_{\mathrm{T}} = \frac{m_{k\mathrm{f}}}{m_{k\mathrm{f}} + m_{k\mathrm{u}}}$$

(6)

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.04.004

References

1. Gunasekaran, K., Eyles, S. E., Hagler, A. T. & Gierasch, L. M. (2001). Keeping it in the family: folding studies of related proteins. *Curr. Opin. Struct. Biol.* **11**, 83–93.
2. Zarrine-Asfar, A., Larson, S. M. & Davidson, A. R. (2005). The family feud: Do proteins with similar structures fold via the same pathway? Curr. *Opin. Struct. Biol.* **15**, 42–49.
3. Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W. *et al.* (2003). Unifying features

in protein-folding mechanisms. *Proc. Natl Acad. Sci. USA*, **100**, 13286–13291.

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

5. Kim, P. S. & Baldwin, R. L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* **51**, 459–489.

6. Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995). 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. *J. Mol. Biol.* **254**, 260–288.

7. Kim, D. E., Fisher, C. & Baker, D. (2000). A breakdown of symmetry in the folding transition state of protein L. *J. Mol. Biol.* **298**, 971–984.

8. Friel, C. T., Capaldi, A. P. & Radford, S. E. (2003). Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: similarities and differences in the folding of homologous proteins. *J. Mol. Biol.* **326**, 293–305.

9. Fowler, S. B. & Clarke, J. (2001). Mapping the folding pathway of an immunoglobulin domain: structural detail from phi value analysis and movement of the transition state. *Structure*, **9**, 355–366.

10. Geierhaas, C. D., Paci, E., Vendruscolo, M. & Clarke, J. (2004). Comparison of the transition states for folding of two Ig-like proteins from different superfamilies. *J. Mol. Biol.* **343**, 1111–1123.

11. Hamill, S. J., Steward, A. & Clarke, J. (2000). The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. *J. Mol. Biol.* **297**, 165–178.

12. Cho, S. S., Levy, Y. & Wolynes, P. G. (2009). Quantitative criteria for native energetic heterogeneity influences in the prediction of protein folding kinetics. *Proc. Natl Acad. Sci. USA*, **106**, 434–439.

13. Cota, E., Steward, A., Fowler, S. B. & Clarke, J. (2001). The folding nucleus of a fibronectin type III domain is composed of core residues of the conserved immunoglobulin-like fold. *J. Mol. Biol.* **305**, 1185–1197.

14. Lappalainen, I., Hurley, M. G. & Clarke, J. (2008). Plasticity within the obligatory folding nucleus of an immunoglobulin-like domain. *J. Mol. Biol.* **375**, 547–559.

15. Lorch, M., Mason, J. M., Clarke, A. R. & Parker, M. J. (1999). Effects of core mutations on the folding of a beta-sheet protein: implications for backbone organization in the I-state. *Biochemistry*, **38**, 1377–1385.

16. Scott, K. A., Batey, S., Hooton, K. A. & Clarke, J. (2004). The folding of spectrin domains I: wild-type domains have the same stability but very different kinetic properties. *J. Mol. Biol.* **344**, 195–205.

17. Scott, K. A., Randles, L. G. & Clarke, J. (2004). The folding of spectrin domains II: phi-value analysis of R16. *J. Mol. Biol.* **344**, 207–221.

18. Scott, K. A., Randles, L. G., Moran, S. J., Daggett, V. & Clarke, J. (2006). The folding pathway of spectrin R17 from experiment and simulation: using experimentally validated MD simulations to characterize states hinted at by experiment. *J. Mol. Biol.* **359**, 159–173.

19. Hutchinson, E. G. & Thornton, J. M. (1993). The Greek key motif: extraction, classification and analysis. *Protein Eng.* **6**, 233–245.

20. Richardson, J. S. (1977). β-Sheet topology and the relatedness of proteins. *Nature*, **268**, 495–500.

21. Berglund, H., Olerenshaw, D., Sankar, A., Federwisch, M., McDonald, N. Q. & Driscoll, P. C. (2000). The three-dimensional solution structure and dynamic properties of the human FADD death domain. *J. Mol. Biol.* **302**, 171–188.

22. Silow, M. & Oliveberg, M. (1997). Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc. Natl Acad. Sci. USA*, **94**, 6084–6086.

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

24. Scott, K. A., Alonso, D. O. V., Sato, S., Fersht, A. R. & Daggett, V. (2007). Conformational entropy of alanine versus glycine in protein denatured states. *Proc. Natl Acad. Sci. USA*, **104**, 2661–2666.

25. Serrano, L. & Fersht, A. R. (1989). Capping and alpha-helix stability. *Nature*, **342**, 296–299.

26. 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.

27. Sato, S., Religa, T. L. & Fersht, A. R. (2006). Phi-analysis of the folding of the B domain of protein A using multiple optical probes. *J. Mol. Biol.* **360**, 850–864.

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

29. Nickson, A. A., Stoll, K. E. & Clarke, J. (2008). Folding of a LysM domain: entropy-enthalpy compensation in the transition state of an ideal two-state folder. *J. Mol. Biol.* **380**, 557–569.

30. Fersht, A. R. (1998). *Structure and Mechanism in Protein Science: a Guide to Enzyme Catalysis and Protein Folding.* W.H. Freeman and Company, New York.

31. Munoz, V. & Serrano, L. (1997). Development of the multiple sequence approximation within the AGA-DIR model of alpha-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers*, **41**, 495–509.

32. Otzen, D. E. & Oliveberg, M. (2002). Conformational plasticity in folding of the split beta-alpha-beta protein S6: evidence for burst-phase disruption of the native state. *J. Mol. Biol.* **317**, 613–627.

33. Oliveberg, M., Tan, Y. J., Silow, M. & Fersht, A. R. (1998). The changing nature of the protein folding transition state: implications for the shape of the free-energy profile for folding. *J. Mol. Biol.* **277**, 933–943.

34. Otzen, D. E., Kristensen, O., Proctor, M. & Oliveberg, M. (1999). Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots. *Biochemistry*, **38**, 6499–6511.

35. Sanchez, I. E. & Kiefhaber, T. (2003). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325**, 367–376.

36. Oliveberg, M. (2001). Characterisation of the transition states for protein folding: towards a new level of mechanistic detail in protein engineering analysis. *Curr. Opin. Struct. Biol.* **11**, 94–100.

37. Scott, K. A. & Clarke, J. (2005). Spectrin R16: Broad energy barrier or sequential transition states? *Protein Sci.* **14**, 1617–1629.

38. Ternstrom, T., Mayor, U., Akke, M. & Oliveberg, M. (1999). From snapshot to movie: phi analysis of protein folding transition states taken one step further. *Proc. Natl Acad. Sci. USA*, **96**, 14854–14859.

39. Chen, Y. R. & Clark, A. C. (2003). Equilibrium and kinetic folding of an alpha-helical Greek key protein domain: caspase recruitment domain (CARD) of RICK. *Biochemistry*, **42**, 6310–6320.

40. Chen, Y. R. & Clark, A. C. (2004). Kinetic traps in the folding/unfolding of procaspase-1 CARD domain. *Protein Sci.* **13**, 2196–2206.

41. Chen, Y. R. & Clark, A. C. (2006). Substitutions of prolines examine their role in kinetic trap formation of the caspase recruitment domain (CARD) of RICK. *Protein Sci.* **15**, 395–409.

42. Milam, S. L., Nicely, N. I., Feeney, B., Mattos, C. & Clark, A. C. (2007). Rapid folding and unfolding of Apaf-1 CARD. *J. Mol. Biol.* **369**, 290–304.

43. Clarke, J., Cota, E., Fowler, S. B. & Hamill, S. J. (1999). Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway. *Structure*, **7**, 1145–1153.

44. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994.

45. Paci, E., Clarke, J., Steward, A., Vendruscolo, M. & Karplus, M. (2003). Self-consistent determination of the transition state for protein folding: application to a fibronectin type III domain. *Proc. Natl Acad. Sci. USA*, **100**, 394–399.

46. Garcia, P., Bruix, M., Rico, M., Ciofi-Baffoni, S., Banci, L., Ramachandra Shastry, M. C. *et al.* (2005). Effects of heme on the structure of the denatured state and folding kinetics of cytochrome b562. *J. Mol. Biol.* **346**, 331–344.

47. Eun, Y. J., Kurt, N., Sekhar, A. & Cavagnero, S. (2008). Thermodynamic and kinetic characterization of apoHmpH, a fast-folding bacterial globin. *J. Mol. Biol.* **376**, 879–897.

48. Nishimura, C., Dyson, H. J. & Wright, P. E. (2006). Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin. *J. Mol. Biol.* **355**, 139–156.

49. Nishimura, C., Dyson, H. J. & Wright, P. E. (2008). The kinetic and equilibrium molten globule intermediates of apoleghemoglobin differ in structure. *J. Mol. Biol.* **378**, 715–725.

50. Uzawa, T., Nishimura, C., Akiyama, S., Ishimori, K., Takahashi, S., Dyson, H. J. & Wright, P. E. (2008). Hierarchical folding mechanism of apomyoglobin revealed by ultra-fast H/D exchange coupled with 2D NMR. *Proc. Natl Acad. Sci. USA*, **105**, 13859–13864.

51. Teilum, K., Thormann, T., Caterer, N. R., Poulsen, H. I., Jensen, P. H., Knudsen, J. *et al.* (2005). Different secondary structure elements as scaffolds for protein folding transition states of two homologous four-helix bundles. *Proteins: Struct. Funct. Genet.* **59**, 80–90.

52. Feng, H., Vu, N. D., Zhou, Z. & Bai, Y. (2004). Structural examination of phi-value analysis in protein folding. *Biochemistry*, **43**, 14325–14331.

53. Jemth, P., Day, R., Gianni, S., Khan, F., Allen, M., Daggett, V. & Fersht, A. R. (2005). The structure of the major transition state for folding of an FF domain from experiment and simulation. *J. Mol. Biol.* **350**, 363–378.

54. Hubner, I. A., Lindberg, M., Haglund, E., Oliveberg, M. & Shakhnovich, E. I. (2006). Common motifs and topological effects in the protein folding transition state. *J. Mol. Biol.* **359**, 1075–1085.

55. Oliveberg, M. & Wolynes, P. G. (2005). The experimental survey of protein-folding energy landscapes. *Quart. Rev. Biophys.* **38**, 245–288.

56. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. (2001). Three key residues form a critical contact network in a protein folding transition state. *Nature*, **409**, 641–645.

57. Baxa, M. C., Freed, K. F. & Sosnick, T. R. (2008). Quantifying the structural requirements of the folding transition state of protein A and other systems. *J. Mol. Biol.* **381**, 1362–1381.

58. Sharpe, T. D., Ferguson, N., Johnson, C. M. & Fersht, A. R. (2008). Conservation of transition state structure in fast folding peripheral subunit-binding domains. *J. Mol. Biol.* **383**, 224–237.

59. Miroux, B. & Walker, J. E. (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298.

60. Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266–280.

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