

The Folding of Spectrin Domains I: Wild-type Domains Have the Same Stability but very Different Kinetic Properties

Kathryn A. Scott, Sarah Batey, Karen A. Hooton and Jane Clarke*

MRC Centre for Protein
Engineering, University of
Cambridge Chemical
Laboratory, Lensfield Road
Cambridge CB2 1EW, UK

The study of proteins with the same architecture, but different sequence has proven to be a valuable tool in the protein folding field. As a prelude to studies on the folding mechanism of spectrin domains we present the kinetic characterisation of the wild-type forms of the 15th, 16th, and 17th domains of chicken brain α -spectrin (referred to as R15, R16 and R17, respectively). We show that the proteins all behave in a two-state manner, with different kinetic properties. The folding rate varies remarkably between different members, with a 5000-fold variation in folding rate and 3000-fold variation in unfolding rate seen for proteins differing only 1 kcal mol⁻¹ in stability. We show clear evidence for significant complexity in the energy landscape of R16, which shows a change in amplitude outside the stopped-flow timescale and curvature in the unfolding arm of the chevron plot. The accompanying paper describes the characterisation of the folding pathway of this domain.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: protein folding; helix-bundle; spectrin; residual structure; homologue

*Corresponding author

Introduction

How a protein folds to its functional state is a fundamental question of biology. This problem is hugely complex due to the large network of non-covalent interactions responsible for folding and the diversity of folds adopted. Studies have, however, highlighted the importance of topology in determining folding kinetics and mechanism.^{1–3} The comparison of folding pathways in proteins with the same structure but different sequence is thus one way to reduce the complexity of the folding problem. This approach has been successfully applied to the study of a number of folds, with differing results (many studies are reviewed by Gunasekaran *et al.*⁴). In particular, the folding mechanism of a family of helical proteins has recently been shown to be sensitive to the propensity to form local helical structure.⁵ Studies on the folding of other helical proteins are described in more detail in the accompanying paper.⁶

Spectrin domains are found as multiple tandem repeats in cytoskeletal proteins including spectrin, α -actinin and dystrophin. The 106 amino acid residue repeat unit folds into a coiled coil with three antiparallel helices (Figure 1).^{7–11} Uncommonly, there are very few highly conserved residues; only a tryptophan residue in helix A and a leucine two residues from the carboxyl terminus of helix C are conserved in more than 80% of sequences in the Pfam database.¹² Although found in multiple tandem repeats *in vivo*, isolated spectrin domains are folded in solution.^{8,13–15}

A number of structures are available, both of single domains and of domains linked in tandem.^{9,11,16,17} Part of the α actinin “rod domain” (four domains linked in a tandem array) and the 16th and 17th domains of chicken brain α -spectrin have been crystallised.^{16,17} The linker region between domains appears helical in both crystal structures. The properties of spectrin domains have been studied using atomic force microscopy, molecular dynamics simulations and chemical/thermal denaturation techniques.^{13–15,18–22} Under mechanical stress, unfolding events corresponding to both one and two domains unfolding occur.²³ Thus both independent and cooperative behaviour is seen, with this behaviour being tuneable with temperature.²⁴ Studies of spectrin domains in solution appear to

Present address: K. A. Scott, Department of Medicinal Chemistry, University of Washington, Seattle, WA 98195-7610, USA.

Abbreviations used: CD, circular dichroism.

E-mail address of the corresponding author: jc162@cam.ac.uk

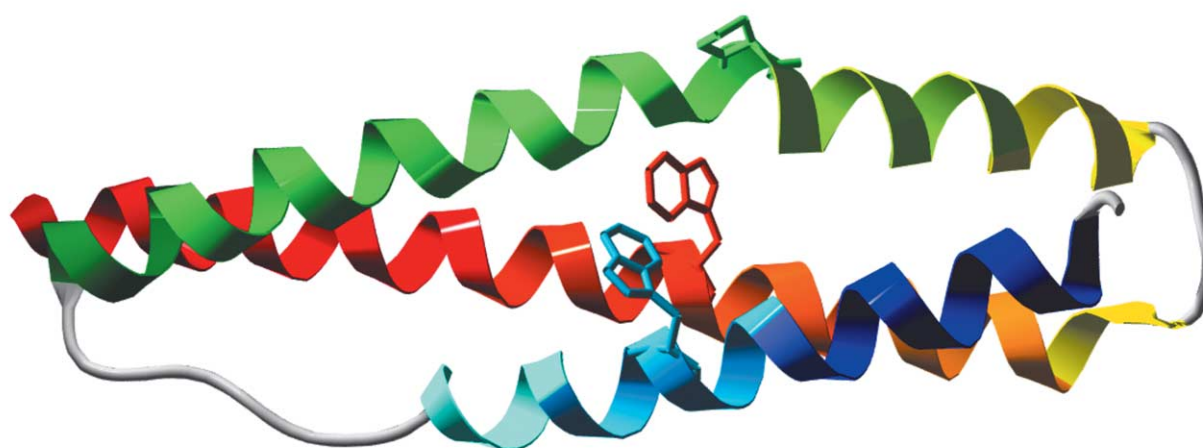


Figure 1. Ribbon diagram of R16. The structure of R16 is coloured from blue at the N terminus to red at the C terminus with the side-chains of the three residues that were mutated, W21, P60 and W94, displayed. The structure of R17 overlays well with that of R16; the backbone RMSD between domains R16 and R17 using coordinates taken from the crystal structure of R1617 (1CUN) is 1 Å. Comparison of these coordinates with the NMR structure of R16 (1AJ3) gives a backbone RMSD of 2.5 Å. The NMR structure is six residues shorter at the C terminus, helix C is therefore shorter in the NMR structure (helical structure is, however, seen to the end of the amino acid sequence). No structure of R15 is available. An alignment of the amino acid sequences of R15 and R16 reveals that, unlike R15 and R17, R16 has the semi-conserved proline residue in helix B and an additional tryptophan in helix C. The percentage pairwise sequence identities are 29%, 33% and 32% between R15 and R16, R16 and R17 and R15 and R17, respectively. These values are higher than the average value of 16% across the whole family; however it still represents a significant diversity in sequence. This Figure was drawn using DeepView/Swiss-PdbViewer v3.7 (<http://www.expasy.org/spdv/>) and POV-Ray v3.5 (www.povray.org).

show a similar behaviour. MacDonald and co-workers have carried out thermodynamic studies of seven two-domain α -spectrin constructs where the linker region showed a range of helical propensities. They report that the helical propensity of the linker region correlates with the stability of the domain pairs and suggest that the differences in structure of linker regions between different domains may have an important role in spectrin flexibility.¹³

Why study spectrin domains? The helices have between 28 and 36 amino acid residues, giving an elongated, rod-like structure, in contrast with the globular shape of most other domains whose folding pathway has been studied to date. This unusual shape makes it an interesting target for a folding study (Figure 1). Further, the spectrin domains are well represented in the Pfam database, with 2132 sequences and an average percentage identity of 16%.¹² This allows the comparison of members of the spectrin family that are diverse in sequence. Our aim was to choose a system that would allow us to study both the folding of individual domains and to investigate whether the folding mechanism changes when the domains are linked in tandem. The 16th and 17th domains of α -spectrin have previously been shown to be suitable for a detailed folding study; they are sufficiently stable to withstand mutation and have no disulphide bridges or *cis*-proline residues that can complicate the folding analysis.^{14,21} As a preliminary to these studies the detailed characterisation of wild-type R15, R16 and R17 is presented here. Although the proteins have similar

thermodynamic stabilities, the folding rates vary over more than three orders of magnitude and there is evidence from kinetic data for complexity in the folding landscape of the spectrin domains. An in-depth investigation of the folding pathway of R16 is described in the accompanying paper.⁶

Results and Discussion

Equilibrium properties

All three domains unfold in a two-state manner at equilibrium as judged by the coincidence in the fluorescence and far-UV circular dichroism (CD) denaturation curves (Figure 2). The domains have

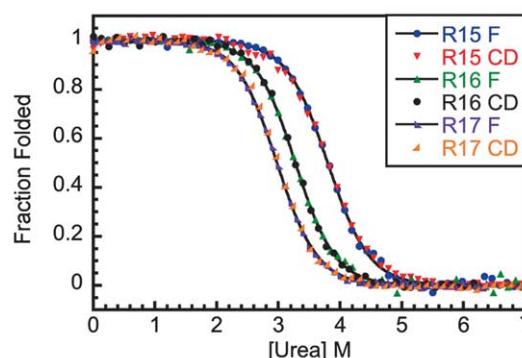


Figure 2. Equilibrium denaturation curves for R15, R16 and R17 measured using fluorescence (F) and circular dichroism (CD) displayed as fraction of folded protein. The line shows the fit of the data to a two-state equation.

Table 1. Kinetic and thermodynamic parameters

Protein	$m_{D-N}(\text{eq})^a$ (kcal mol ⁻¹ M ⁻¹)	$\Delta G_{D-N}^{\text{H}_2\text{O}}(\text{eq})$ (kcal mol ⁻¹)	$k_f^{\text{H}_2\text{O}}(\text{s}^{-1})$	$k_u^{\text{H}_2\text{O}}(\text{s}^{-1})$	m_{D-TS} (kcal mol ⁻¹ M ⁻¹)	m_{TS-N} (kcal mol ⁻¹ M ⁻¹)	$\Delta G_{D-N}^{\text{H}_2\text{O}}(\text{kin})$ (kcal mol ⁻¹)	$m_{D-N}(\text{kin})$ (kcal mol ⁻¹ M ⁻¹)
R15	1.8 (±0.1)	6.8 (±0.2)	60,000 ^b (±20,000)	1.3 ^b (±0.2)	1.12 (±0.04)	0.59 (±0.02)	~6.5	1.71
R15 (10 °C)	–	–	9800 ^b (±800)	0.021 ^b (±0.004)	1.04 (±0.02)	0.75 (±0.02)	~8	1.79
R16	1.9 (±0.1)	6.4 (±0.2)	125 (±3) ^c	2.6 × 10 ^{-3c} (±3 × 10 ⁻⁴)	1.19 ^d (±0.01)	0.71 ^d (±0.02)	6.3 (±0.3)	1.9 (±0.1)
R16 P60A	1.8 (±0.1)	6.8 (±0.2)	388 (±13) ^c	4.7 × 10 ^{-3c} (±3 × 10 ⁻⁴)	1.09 ^d (±0.01)	0.71 ^d (±0.02)	6.6 (±0.3)	1.8 (±0.1)
R16 W94F	1.8 (±0.1)	4.3 (±0.1)	64 (±2) ^c	4.3 × 10 ^{-2c} (±3 × 10 ⁻³)	1.26 ^d (±0.02)	0.71 ^d (±0.02)	4.4 (±0.2)	2.0 (±0.1)
R17	2.0 (±0.1)	6.0 (±0.2)	30 (±2) ^b	4.0 × 10 ^{-4b} (±3 × 10 ⁻⁵)	1.37 (±0.02)	0.87 (±0.01)	6.6 (±0.3)	2.2 (±0.1)

The errors in $k_f^{\text{H}_2\text{O}}$, $k_u^{\text{H}_2\text{O}}$, m_{D-TS} and m_{TS-N} are from the fit to the model as stated above. These errors were combined using standard methods to determine the error in the kinetic values of $\Delta G_{D-N}^{\text{H}_2\text{O}}$ and m_{D-N} .

^a The equilibrium (eq) m -value reported for R15 is the mean and standard deviation of three experiments on wild-type protein. Those for R16 and mutants and R17 are the mean and standard deviation over all mutants characterised in our laboratory.

^b Calculated using the two-state model.

^c The values quoted were calculated using a sequential transition state model.^{25,42} The same value within error is found if a model describing a broad transition state is used.^{37–41} The fit of the data to these models will be described in detail elsewhere. The program ProFit (Quantumsoft) was used in data fitting.

^d These values are for the transition state for folding in water, a global value of m_{TS-N} was determined, as will be described in detail elsewhere.

similar m -values, reflecting their similar structure, and they have similar thermodynamic stability (Table 1).

Kinetic studies

R16 and R17 were studied using both fluorescence and CD stopped-flow measurements at 25 °C. Rate constants measured using both techniques were the same, within experimental error, at each denaturant concentration. R15 folds and unfolds very rapidly, beyond the range of our CD instrument, hence (un)folding was monitored by following changes in fluorescence only. At 25 °C only a small range of denaturant concentrations could be studied for R15 (2.5–6.5 M urea); outside this range the rate constants were too rapid to be determined accurately. To compensate for this, R15 was also studied at 10 °C (increasing the accessible range of denaturant concentrations to 1.5–7.5 M urea).

R15 and R17 show simple kinetic behaviour

The folding kinetics of R15 and R17 were relatively simple to analyse; only one phase was apparent in both unfolding and refolding traces and no stable kinetic intermediates were detected in the measurable range of urea concentrations. Kinetic folding intermediates can be inferred from curvature (or “roll over”) in the refolding arm of a chevron plot. The folding arms of the chevron plots are linear for both R15 and R17. (For R15 it is important to note that, as rate constants at low concentrations of urea were too fast for us to measure, we cannot be certain that R15 does not populate a stable folding intermediate in the absence of denaturant.) The fluorescence and CD

amplitudes were also analysed in R17 to look for any dead-time loss or gain of signal. All the amplitude was accounted for in the kinetic traces. (Again, the data range for R15 is too small to be conclusive.)

Curvature in the unfolding limb of a chevron plot indicates some complexity in the folding landscape. Such curvature has been attributed to changes in the native state at high denaturant concentrations, the presence of high-energy intermediates, or to a moving transition state.^{25–29} As shown in Figure 3(a) it is possible that in R15 there is some curvature in the unfolding arm of the chevron at 10 °C. When all data points are used in the two-state fit (continuous line) there is a small systematic deviation about the line. If the last two points are excluded (broken line), then m_{k_u} increases and the last points fall below the line. The rate constants at high concentration of urea are in the region of 250 s⁻¹, inside the capabilities of our stopped-flow instrument, suggesting that this is not an artefact. This infers that R15 may show complexity in the folding landscape. The chevron for R17 is linear between 0 and about 8 M urea; above this value, the rate constants are slower than would be expected from a linear relation. The measured rate constants are similar to those seen for R15, but the deviation from linearity is much smaller†.

Since, for both R15 and R17 any deviation from linearity is small, both chevron plots were fitted assuming simple two-state behaviour; the kinetic

† It should be noted that a number of R17 mutant proteins show a change in the slope of the unfolding arm (data to be published elsewhere), indicating that the folding landscape of R17 may be more complex than inferred from studies of the wild-type alone, as has been seen in the protein S6.²⁷

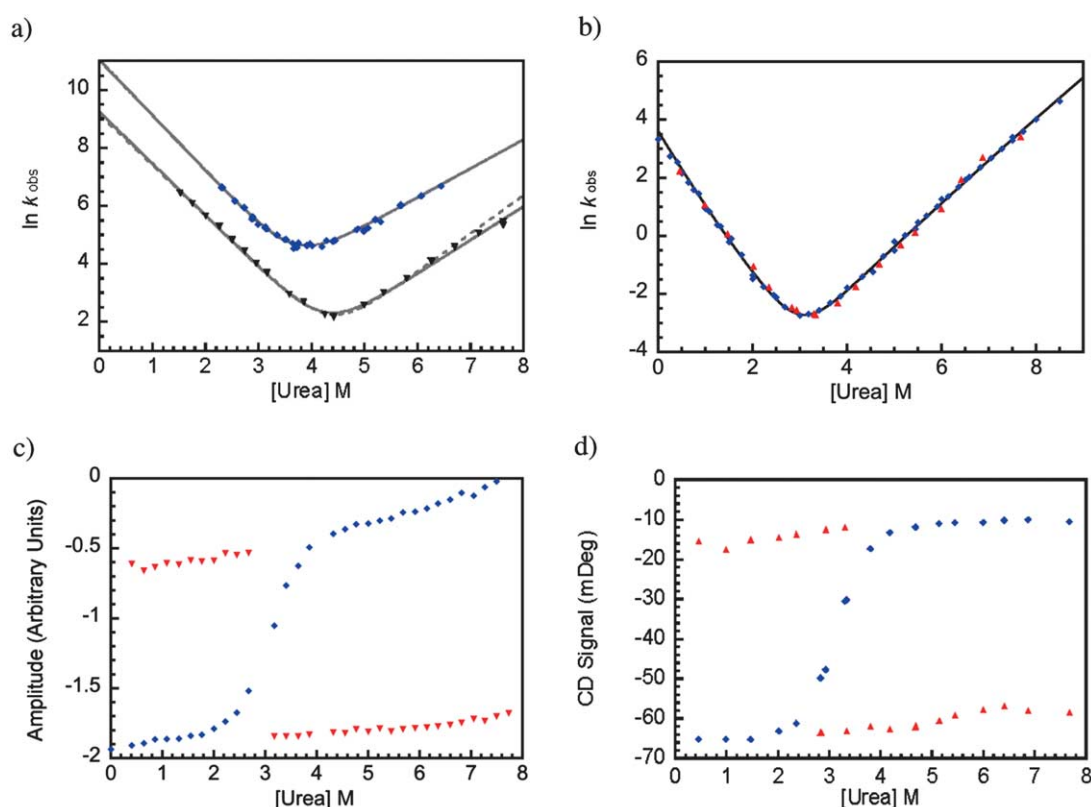


Figure 3. R15 and R17 kinetics. (a) Rate constants for the folding of R15 at 25 °C (blue diamonds) and at 10 °C (black triangle) measured using fluorescence are shown. The continuous line represents the fit to a two-state model of the whole data set for each temperature, to illustrate the maximum predicted folding rate constant for R15. The broken line for the data measured at 10 °C is the fit to a two-state model with the data points at urea concentrations above 7 M excluded. (b) Rate constants for folding of R17 measured using fluorescence (blue diamonds) and CD (red triangles) are shown. The line represents the fit to a two-state model. (c) R17 fluorescence amplitudes; the initial (red triangles) and final (blue diamonds) fluorescence measurements of kinetic traces are shown. (d) R17 CD amplitudes; the initial (red triangles) and final (blue diamonds) CD measurements of kinetic traces are shown.

parameters from this fit are listed in Table 1. (It should be noted that alternate methods of analysis, such as the inclusion of quadratic terms, made no significant difference to the kinetic m -value or $\Delta G_{\text{D-N}}$.) A common test for two-state behaviour is to compare kinetic and equilibrium parameters. For R15 both the kinetic $\Delta G_{\text{D-N}}$ and m -value are the same (within error) as those determined at equilibrium. For R17 the kinetic and equilibrium m -values are the same (within error) but the kinetic $\Delta G_{\text{D-N}}$ is higher than that determined at equilibrium. Although values for the free energy change for folding measured using equilibrium and kinetic methods differ slightly, there is no evidence of any other species on the folding pathway. In fact, the kinetic stability is apparently higher than the equilibrium stability, which would not be the case in the presence of a folding intermediate. We conclude that both R15 and R17 fold by a two-state mechanism under the observable conditions, i.e. the only stable, populated states are the denatured and native states.

R16 exhibits complex kinetic behaviour

The presence of multiple refolding phases. In R16, only one phase was apparent in unfolding kinetic

traces; however, in refolding, two phases were detected. The slower phase typically accounts for 10–15% of the total amplitude in both CD and fluorescence measurements and exhibits a dependence on urea concentration (Figure 4).

What is the second folding phase seen in R16? Possible causes of complex refolding kinetics include heterogeneity in the denatured state, the presence of an intermediate and folding to an alternate species.^{30–34} We carried out a series of experiments to distinguish these possibilities. Interrupted refolding experiments were undertaken to establish that both folding phases form the native species.³⁵ Unfolded wild-type protein was diluted to 1.5 M urea for refolding. After a variable delay the protein was then unfolded into a final concentration of 7.5 M urea. The unfolding traces showed the same rate constant for all delay times. The amplitude of the unfolding reaction increases in two phases (Figure 4(c)). The two phases appeared with same apparent rate constants as those measured for the refolding rate constants at 1.5 M urea, and had the same amplitudes (legend to Figure 4). These results indicate that both phases directly form the native protein.

The slow refolding phase in R16 is only partly due to proline cis–trans isomerisation. Since the interrupted

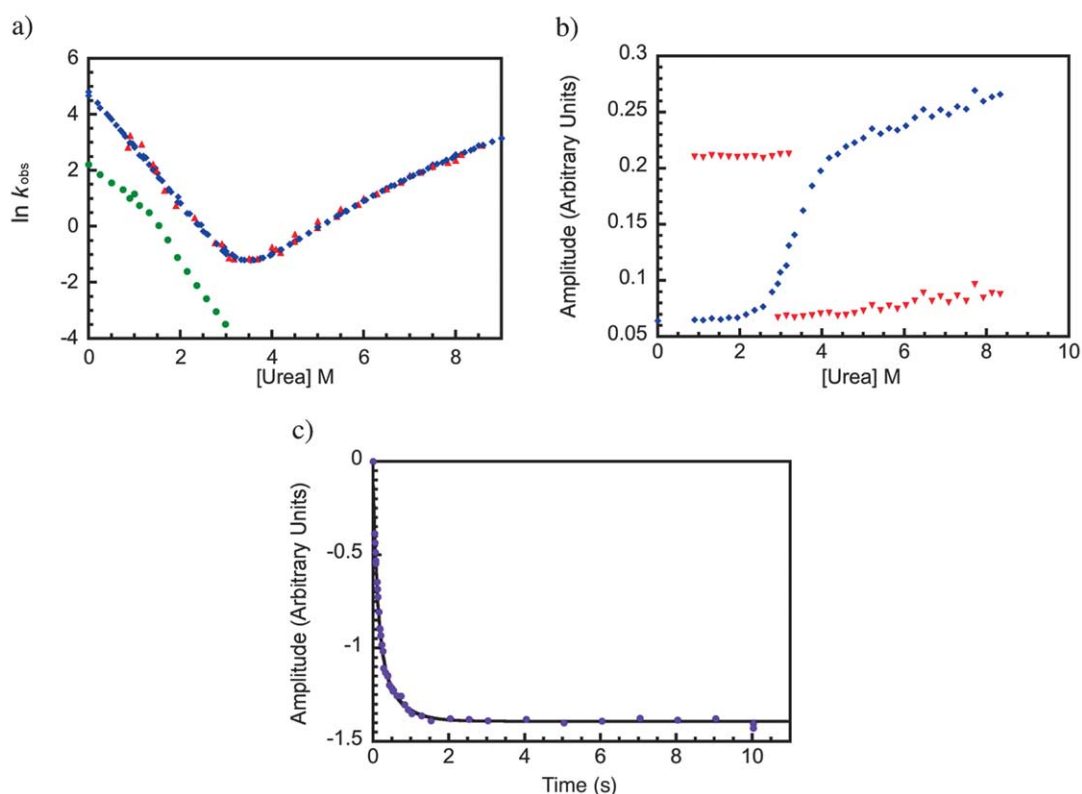


Figure 4. Wild-type R16 kinetics. (a) Folding and unfolding rate constants for R16. Rate constants for the major phase measured using fluorescence (blue diamonds), using CD (red triangles) and for the minor phase measured using fluorescence (green circles) are shown. (b) Fluorescence amplitudes; the initial (red triangles) and final (blue diamonds) fluorescence measurements of kinetic traces are shown. (c) The time-course of the appearance of the single unfolding phase of native R16 in interrupted refolding experiments. The fluorescence amplitude from interrupted refolding (purple circles) and the fit of the amplitudes to an equation describing a two-exponential process (black line) are shown. The appearance of the major phase has an apparent rate constant of $6.2(\pm 0.5) \text{ s}^{-1}$ and corresponds to $85(\pm 2)\%$ of the total amplitude, the minor phase appears with an apparent rate constant of $1.2(\pm 0.5) \text{ s}^{-1}$. These values are in good agreement with the rate constants for refolding measured directly by urea jump; in this case the rate constant of the major phase is $5.9(\pm 0.2) \text{ s}^{-1}$ and that of the minor phase $1.0(\pm 0.2) \text{ s}^{-1}$, with the major phase accounting for $87(\pm 3)\%$ of the total amplitude.

unfolding experiments demonstrate that both refolding phases form the native protein, heterogeneity in the denatured state is likely to be responsible for the second folding phase. As R15 and R17 folded with single-exponential kinetics, it seemed likely that the second refolding phase could be attributed to *cis-trans* proline isomerisation (neither R15 or R17 has a proline, but R16 has a *trans*-proline at position 60 causing a kink in the middle of the B-helix). However, on mutation of proline 60 to alanine (P60A) a second refolding phase is still observable, although the amplitude is reduced to approximately 3–5% of the total. This suggests that, whilst the presence of the proline residue is the major contributor to the second refolding phase, it is not solely responsible. Interrupted unfolding experiments were undertaken to investigate this further.

Pappenburger *et al.* have recently demonstrated that the isomerisation of non-prolyl *cis* peptide bonds can result in minor unfolding phases.³⁶ Interrupted unfolding experiments were carried out for the wild-type R16 and P60A to determine the number of equilibration processes in the

unfolded state (Figure 5). Unfolding was allowed to proceed for 50 ms–100 seconds before refolding was initiated and the amplitudes of the refolding reactions monitored with time.

In the wild-type protein, the refolding traces fit well to a double-exponential equation for delay times greater than 250 ms, for shorter delay times the data were well described by a single exponential. The faster of the two refolding phases appeared with an apparent rate constant of $3.4(\pm 0.5) \text{ s}^{-1}$ corresponding, within error, to the rate constant for unfolding at this concentration of urea. However, the appearance of the second phase was biphasic, indicating that two isomerisation processes were involved (Figure 5). A fit to a double-exponential equation gave values for the apparent rate constants for the appearance of these phases of $k_1 = 1.4(\pm 0.3) \text{ s}^{-1}$ and $k_2 = 0.02(\pm 0.005) \text{ s}^{-1}$ and at long delay times the phases account for $4(\pm 0.5)\%$ and $6(\pm 0.5)\%$ of the total amplitude, respectively.

In P60A the refolding traces also fit well to a double-exponential equation for longer delay times ($> 200 \text{ ms}$) and to a single exponential for shorter delays. This indicated that there is still

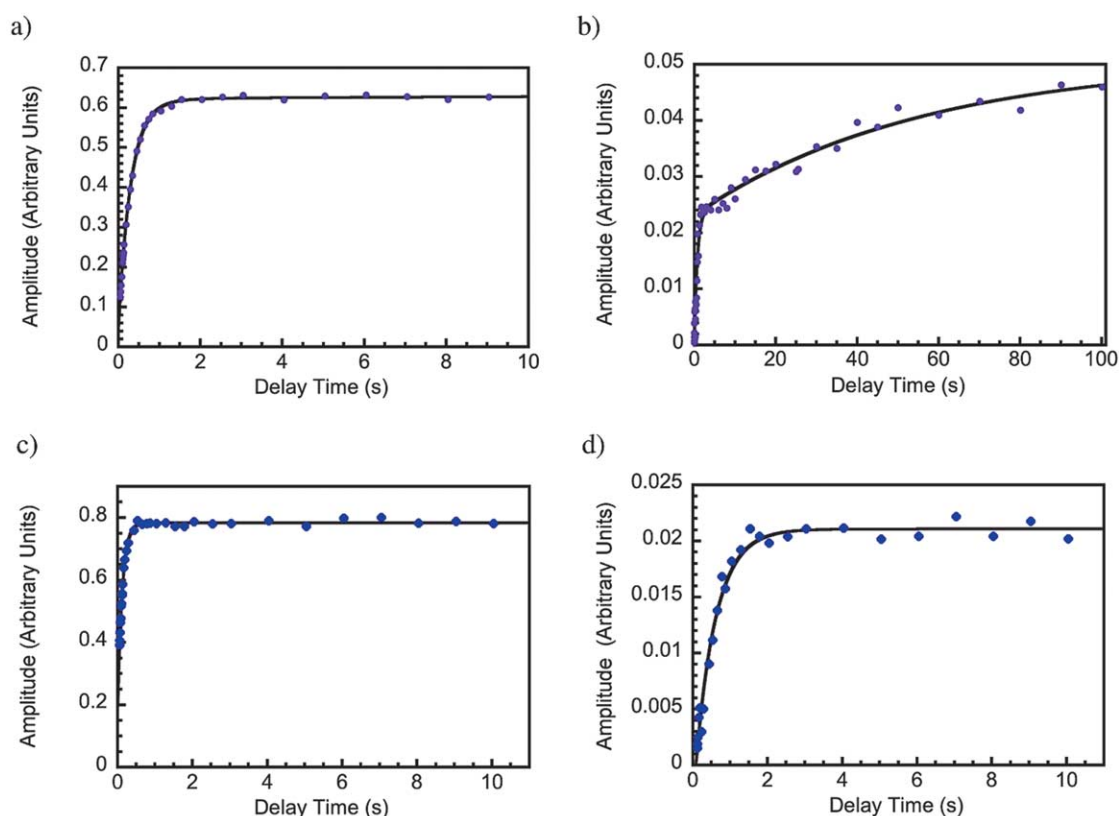


Figure 5. Interrupted unfolding experiments. (a) The time-course of the appearance of the major refolding phase in R16. The line shows a single-exponential fit with an apparent rate constant of $3.4(\pm 0.5) \text{ s}^{-1}$. (b) The time-course of the appearance of the minor refolding phase in R16. The line shows a double-exponential fit with apparent rate constants of $k_1 = 1.4(\pm 0.3) \text{ s}^{-1}$ and $k_2 = 0.02(\pm 0.005) \text{ s}^{-1}$. (c) The time-course of the appearance of the major refolding phase in P60A. (d) The time-course of the appearance of the minor refolding phase in P60A. (Note that for clarity this Figure shows the amplitude for the minor phase up to a delay time of ten seconds, however this amplitude remained the same within error up to 100 seconds.) The lines in (c) and (d) show single-exponential fits with apparent rate constants of $8.3(\pm 0.3) \text{ s}^{-1}$ and $1.8(\pm 0.2) \text{ s}^{-1}$, respectively.

heterogeneity in the denatured state in the absence of the proline residue. The major phase again appeared with a rate constant corresponding to that for unfolding, giving an apparent rate constant of $8.3(\pm 0.3) \text{ s}^{-1}$. For this mutant, however, the minor phase accounted for only $3.5(\pm 0.5)\%$ of the total amplitude and appeared as a single phase, whose apparent rate constant was $1.8(\pm 0.2) \text{ s}^{-1}$.

These data suggest that a second minor refolding phase in the folding of R16 reflects the folding of molecules limited both by the presence of a *cis* proline and by other *cis* peptide bonds in the denatured state. However, the rate constants for these two processes cannot be resolved in the wild-type protein.

Curvature in unfolding kinetics. The unfolding limb of the R16 chevron plot is distinctly curved (Figure 4(a)). Non-linear activation free energy relationships have been reported for a significant number of apparent two-state proteins (summarised by Sanchez & Kiefhaber²⁵). Two common interpretations have been suggested for such curvature. The first interpretation is that the free energy barrier is broad with a large number of energetically similar states; different regions of this barrier then become rate-limiting under different conditions, analogous to

Hammond behaviour in small-molecule reactions.^{28,37–41} Alternatively, non-linearity can be described using a sequential folding model.^{25,29,42} In this case, the free energy barriers are narrow and curvature arises due to a switch between two (or a small number of), transition states separated by obligatory on-pathway, high-energy intermediate(s). We have fitted the data using both of these models. The application of these different methods to the analysis of the kinetic data for R16 and mutants will be described in detail elsewhere. In summary, we find the same values for $k_f^{\text{H}_2\text{O}}$, $k_u^{\text{H}_2\text{O}}$ from both models, within error, resulting in the same value for $\Delta G_{\text{D-N}}^{\text{H}_2\text{O}}$ (Table 1). Most importantly, the kinetic $\Delta G_{\text{D-N}}$ and m -value are the same as those obtained from equilibrium measurements. This correspondence between kinetic and equilibrium parameters is clear evidence that the folding of R16 is two-state (i.e. the only populated states are the native and denatured states: there are no stable, populated intermediates). However, the curvature is indicative of some complexity in the folding landscape, suggesting either a broad transition state or a high-energy intermediate.

Dead-time changes in amplitude in refolding. Despite the clear evidence that there is no folding

intermediate (no roll-over in the refolding kinetics and correspondence between equilibrium and kinetic ΔG and m -values), the amplitudes for fluorescence stopped-flow measurements for refolding traces do not correspond to those expected from the observed amplitudes for unfolding. (Figure 4(b)). Unfortunately, we were unable to determine whether this is accompanied by a dead-time change in the CD signal, since at the low protein concentrations required to avoid aggregation artefacts ($> 5 \mu\text{M}$) the amplitude data were too noisy to be interpreted with confidence, even when a large number of traces were averaged.

The initial state in the refolding stopped-flow experiment is more fluorescent than expected from extrapolation of the unfolded amplitudes (Figure 4(b)). R16 has two tryptophan residues (Figure 1) at positions 21 (A helix) and 94 (C helix). Is one or are both tryptophan residues responsible for the observed initial increase in fluorescence? To answer this question, two mutants, W21F and W94F, were characterised. Both mutants were shown to be folded with a $\Delta G_{\text{D-N}}^{\text{H}_2\text{O}}$ of $4.7 \text{ kcal mol}^{-1}$ for W21F and $4.3 \text{ kcal mol}^{-1}$ for W94F (Table 1). W94F exhibits a similar fluorescence profile to the wild-type protein. Maxima are seen at 335 nm and 350 nm in the native and denatured states, respectively, and the observed fluorescence decreases on folding in stopped-flow experiments, as it does in wild-type. This suggests that the fluorescence of the remaining tryptophan, W21, is principally responsible for the fluorescence changes observed on R16 folding. As for wild-type, unfolding traces fit well to an equation describing a single-exponential process, whilst refolding had two phases, with the slower phase accounting for approximately 10–15% of the total amplitude. However, in this case, the two refolding phases seen on the stopped-flow time-scale account for all of the expected amplitude (Figure 6). This suggests that the dead-time change in amplitude does not come from changes in the region of W21.

In contrast, W21F has a very different fluorescence profile to wild-type. The maximal fluorescence in the native state is at 320 nm, shifting to 350 nm on denaturation and an increase in fluorescence is seen on folding in stopped-flow experiments. The kinetics of this mutant are complex, suggesting a possible change in folding behaviour, however, for present purposes it is sufficient to note that an increase in fluorescence is seen on folding when W94 is the only fluorescent probe.

Despite the discrepancy in observed amplitude, no significant curvature is seen in the refolding arm of the chevron and there is good agreement between the m -values and stabilities calculated from kinetic and thermodynamic experiments without postulating a populated intermediate. Thus, the most likely reason for the poor agreement between fluorescence amplitudes for folding and unfolding is the presence of residual structure in the denatured state in the region of a fluorophore. In the absence of W94, no burst phase in fluorescence is seen.

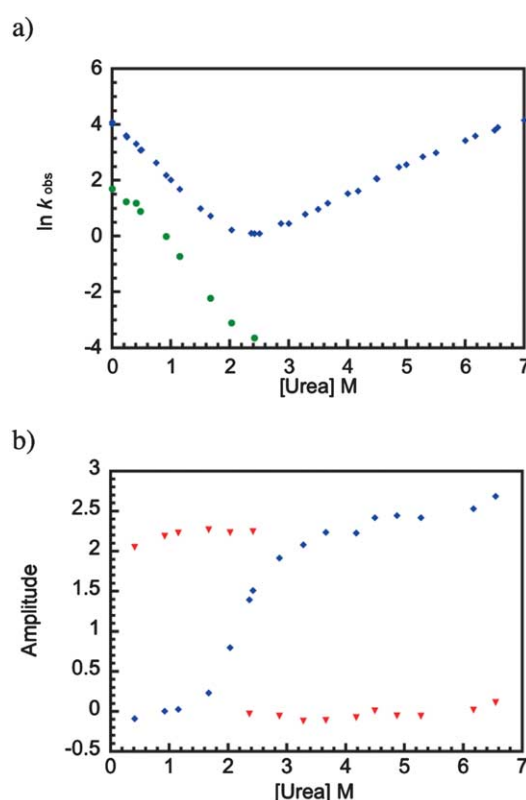


Figure 6. W94F kinetics. (a) Folding and unfolding rate constants for W94F. Rate constants for the major phase (blue diamonds) and minor phase (green triangles) measured using fluorescence. (b) W94F fluorescence amplitudes; the initial (red triangles) and final (blue diamonds) fluorescence measurements (arbitrary units) of kinetic traces are shown. No increase in initial amplitude is observed (compare with Figure 4(b)).

Interestingly, AGADIR predicts a helical propensity of 20–30% between residues 80–100 (Figure 7).^{43–46} Tryptophan 94, therefore, lies in a region of the protein that could be structured to some extent in the denatured state. The data are consistent with the presence either of helical structure in the region of W94 or some form of hydrophobic collapse about

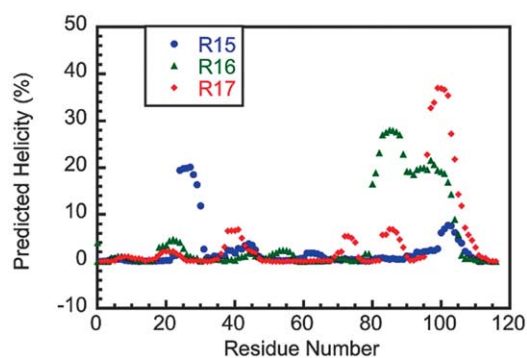


Figure 7. Predicted helical propensities.⁴³ Helical propensity of the R15, R16 and R17 sequences at 298 K predicted using AGADIR is shown.

tryptophan 94. The absence of increased fluorescence in the starting state in W94F is consistent either with structure still being present in the unfolded state (in this case we have simply removed the fluorophore that reports on its presence) or with mutation of tryptophan 94 to phenylalanine causing the structure to be lost. It is not necessary to distinguish between these possibilities, as both imply that the residues around W94 are structured at the start of the kinetic traces.

Comparison of domains

In addition to similar thermodynamic stability, the position of the transition state for folding in water is similar in all three proteins, with the Tanford β -values being in the region of 0.6. However, there is a remarkable difference in folding rates for these three domains. At 2 M urea (chosen to be well out of the transition region for all proteins, and only a small extrapolation from measured results for R15) the folding rate constants are 1000 s^{-1} , 3 s^{-1} and 0.2 s^{-1} for R15, R16 and R17, respectively, thus the folding rates show a 5000-fold variation, with unfolding rates showing a similar range. This is remarkable, given the similar thermodynamic stability and 30% sequence identity (Table 1). In immunoglobulin domains, where the largest variation in folding rates has been observed, the folding rate is strongly dependent on stability.^{47,48} A three-orders of magnitude difference in folding rates has also been observed for acylphosphatase and the homologous activation domain of procarboxipeptidase A2. In this case, however, the proteins are far more distantly related, being members of different superfamilies (as defined in SCOP) having a significantly lower ($\sim 13\%$) sequence identity, much more different structures and different contact orders.⁴⁹

Baker and co-workers have demonstrated a correlation between relative contact order (RCO) and folding rate for a number of two-state proteins.^{1,50} Whilst results from families of homologous proteins and mutant proteins show that topological parameters alone are not sufficient to predict folding rate exactly, it is clear that topology does play an important role. The RCO of R16 and R17 can be estimated from the appropriate region of the R1617 crystal structure,¹⁷ or in the case of R16 from the NMR structure,¹¹ yielding similar values. The structure of R15 has not been determined, but given the similar m -values and identical length of all three proteins it is not unreasonable to assume that R15 has the same RCO as R16 and R17. Interestingly, the rate constant for R15 extrapolated using a two-state model ($60,000\text{ s}^{-1}$) is close to that predicted using the relation between folding rate and RCO presented by Baker and co-workers. The RCO calculated from the structure of R16 is 8.3%, predicting a value for $k_f^{\text{H}_2\text{O}}$ of around $30,000\text{ s}^{-1}$. This suggests that we should perhaps raise the question: why do R16 and R17 fold so slowly?

One of the challenges in protein folding is to try to explain how such large changes in folding rates of

related proteins with similar thermodynamic properties can occur. For the spectrin domains, one of the contributing factors to the difference observed in folding rate may be the extent and/or location of residual structure in the denatured state. In the case of R16 we propose that residual (possibly helical) structure exists as far as the transition region. At least some of this structure exists in the region of tryptophan 94, allowing the residual structure to be detected using fluorescence measurements. The helical propensity for each of the proteins, predicted using AGADIR,⁴³ is shown in Figure 7. Both R16 and R17 have helical propensity above 10% in part of helix C. We cannot rule out the existence of residual structure in R17; where any residual structure is not in the region of a fluorophore it would not be detectable by fluorescence spectroscopy. Unfortunately CD amplitude measurements are not sensitive enough to detect this amount of structure. We would be looking for, at most, 30% helical structure in 10% of residues; given the error in determining real time zero and the noise in the kinetic traces, the change in amplitude expected is less than our experimental error. In contrast, R15 has no significant helical propensity in the C-helix, instead showing a propensity of up to 20% for residues 22 to 27 in helix A.

For helical proteins in particular, the faster folding proteins are often those with either residual helical structure, or a very high helical propensity (reviewed by Friel *et al.*⁵¹). In spectrin domains, intriguingly, the fastest folding domain has the lowest predicted helical propensity.⁴³

Conclusions

As has been seen with many other families, properties such as folding rates and the presence/population of intermediate species are not conserved within members of the family.⁴ For example, the family of all α -helical bacterial immunity (Im) proteins, studied by the Radford group, where at 10°C in a pH 7.0 buffer Im9 is two-state, but Im7 populates a folding intermediate.⁵² For spectrin domains, a remarkable variation in folding rate is seen and we have evidence for complexity in the folding landscape. In R16, we have evidence that residual structure exists as far as the transition region and there is clearly significant curvature in the unfolding arm of the chevron. There is also some curvature in the unfolding arm of the R15 chevron at 10°C , whilst the R17 chevron is linear in the range 0 to 8 M urea with only a small deviation from linearity above this urea concentration. The differences in folding rate, degree of curvature and conditions under which curvature is seen indicate differing degrees of "ruggedness" in the energy landscape of spectrin domains.

However, this does not mean that there are no similarities in the folding pathway. For example, the immunoglobulin-like (Ig-like) domain titin I27

populates an intermediate species and unfolds *via* parallel pathways at high denaturant concentrations,^{53,54} whereas an Ig-like domain from tenascin does not populate an intermediate species.^{55,56} Despite these differences, the folding nucleus shows a high level of similarity in these proteins and other members of the same fold.^{47,53,56,57} Although the results presented indicate differences in the folding landscape of these domains, the folding pathways may still be similar.

In the accompanying paper⁶ we present a complete protein engineering analysis of R16, a first step in comparing the folding pathways in detail.

Materials and Methods

Protein expression and purification

MacDonald and co-workers previously described the phasing of R16 and R17.¹⁴ The R15 sequence was aligned with the R16 and R17 sequences, allowing suitable phasing to be determined. The constructs used in this work consist of the 106 residues defining the domains plus an additional four amino acid residues (natural sequence) at the N terminus and six amino acid residues at the C terminus. Synthetic genes for R15, R16 and R17 were obtained from overlapping primers using standard PCR techniques and inserted into a modified version of the pRSETA vector that encodes an N-terminal histidine tag with a thrombin cleavage site between the tag and the protein. Site-directed mutagenesis of R16 was performed using the Quick Change kit from Stratagene and the identity of the mutants confirmed by DNA sequencing and mass spectrometry.

Protein expression was carried out in *Escherichia coli* C41 cells⁵⁸ with the transformed cells grown to an absorbance at 600 nm 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 affinity chromatography on Ni²⁺-agarose resin. The bound protein was cleaved from the resin with thrombin and further purified by gel-filtration using a Pharmacia Biotech Superdex 75 column. R17 was stored in 5 mM DTT. Where not used immediately proteins were concentrated to approximately 100 μ M using a Sartorius Vivaspin concentrator, flash-frozen and stored at –80 °C.

Stability

The stability of mutant and wild-type proteins was determined using urea-induced equilibrium denaturation, followed by fluorescence using an Aminco Bowman series 2 spectrometer or by CD using a JASCO J720 circular dichroism spectropolarimeter. In the fluorescence studies protein concentration was 1–2 μ M, and the excitation wavelength 280 nm. The emission wavelength in fluorescence studies was 350 nm for all except W21F, where 320 nm was used. In the CD studies, protein concentration was 4 μ M. All experiments were carried out in 50 mM sodium phosphate buffer at pH 7.0. The data were fitted to a two-state transition as described.⁵⁹

Kinetics

Single and sequential mixing kinetic experiments monitoring fluorescence change were carried out using an Applied Photophysics SX.18MV stopped-flow fluorimeter. An excitation wavelength of 280 nm was used with a 320 nm cut-off filter for R15 and R16, and a 335 nm cut-off filter for R17. The final protein concentration in a typical fluorescence experiment was 1–2 μ M. Rate constants for R15 were measured using a 5 μ l cell, which gives a shorter dead-time than the 20 μ l cell. Kinetics monitored using CD were measured using an Applied Photophysics Π^* -180 instrument, with a final protein concentration of 3–4 μ M. In both cases the stopped-flow apparatus was maintained at a temperature of 25 ($\pm$ 0.1) °C. Between ten and 15 traces were averaged for typical measurements; however, at least 20 traces were averaged when measuring the CD profile of the unfolded species in buffer. Kinetic traces from single mixing experiments were analysed using Kaleidagraph (Synergy Software) and ProFit (Quantumsoft). Those from sequential mixing were fitted globally with shared rate constants using Prism (GraphPad Software). All rate constants were independent of protein concentration under the conditions of the experiment.

The variation of the logarithm of the observed rate constant with concentration of urea for R16 were analysed using both a sequential transition state model^{25,42} and a broad, smooth, transition state model;^{28,39–41} this will be described in detail elsewhere.

Acknowledgements

K.A.S. & S.B. were funded by the Medical Research Council. J.C. is a Wellcome Trust Senior Research Fellow.

References

1. 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.
2. Plaxco, K. W., Simons, K. T., Ruczinski, I. & David, B. (2000). Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry*, **39**, 11177–11183.
3. Lindberg, M., Tangrot, J. & Oliveberg, M. (2002). Complete change of the protein folding transition state upon circular permutation. *Nature Struct. Biol.* **9**, 818–822.
4. Gunasekaran, K., Eyles, S. J., Hagler, A. T. & Gierasch, L. M. (2001). Keeping it in the family: folding studies of related proteins. *Curr. Opin. Struct. Biol.* **11**, 83–93.
5. Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W. N. *et al.* (2003). Unifying features in protein-folding mechanisms. *Proc. Natl Acad. Sci. USA*, **100**, 13286–13291.
6. Scott, K.A., Randles, L.G., Clarke, J. (2004). The folding of spectrin domains II. Φ -value analysis of R16. *J. Mol. Biol.* **344** (this issue).
7. Speicher, D. W. & Marchesi, V. T. (1984). Erythrocyte spectrin is comprised of many homologous triple helical segments. *Nature*, **311**, 177–180.

8. Winograd, E., Hume, D. & Branton, D. (1991). Phasing the conformational unit of spectrin. *Proc. Natl Acad. Sci. USA*, **88**, 10788–10791.
9. Yan, Y., Winograd, E., Viel, A., Cronin, T., Harrison, S. C. & Branton, D. (1993). Crystal-structure of the repetitive segments of spectrin. *Science*, **262**, 2027–2030.
10. Pascual, J., Pfuhl, M., Rivas, G., Pastore, A. & Saraste, M. (1996). The spectrin repeat folds into a three-helix bundle in solution. *FEBS Letters*, **383**, 201–207.
11. Pascual, J., Pfuhl, M., Walther, D., Saraste, M. & Nilges, M. (1997). Solution structure of the spectrin repeat: a left-handed antiparallel triple-helical coiled-coil. *J. Mol. Biol.* **273**, 740–751.
12. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S. *et al.* (2004). The Pfam protein families database. *Nucl. Acids Res.* **32**, D138–D141.
13. MacDonald, R. I. & Cummings, J. A. (2004). Stabilities of folding of clustered, two-repeat fragments of spectrin reveal a potential hinge in the human erythroid spectrin tetramer. *Proc. Natl Acad. Sci. USA*, **101**, 1502–1507.
14. MacDonald, R. I. & Pozharski, E. V. (2001). Free energies of urea and of thermal unfolding show that two tandem repeats of spectrin are thermodynamically more stable than a single repeat. *Biochemistry*, **40**, 3974–3984.
15. Menhart, N., Mitchell, T., Lusitani, D., Topouzian, N. & Fung, L. W. M. (1996). Peptides with more than one 106-amino acid sequence motif are needed to mimic the structural stability of spectrin. *J. Biol. Chem.* **271**, 30410–30416.
16. Ylanne, J., Scheffzek, K., Young, P. & Saraste, M. (2001). Crystal structure of the alpha-actinin rod reveals an extensive torsional twist. *Structure*, **9**, 597–604.
17. Grum, V. L., Li, D. N., MacDonald, R. I. & Mondragon, A. (1999). Structures of two repeats of spectrin suggest models of flexibility. *Cell*, **98**, 523–535.
18. Rief, M., Pascual, J., Saraste, M. & Gaub, H. E. (1999). Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *J. Mol. Biol.* **286**, 553–561.
19. Lenne, P. F., Raae, A. J., Altmann, S. M., Saraste, M. & Horber, J. K. H. (2000). States and transitions during forced unfolding of a single spectrin repeat. *FEBS Letters*, **476**, 124–128.
20. Paci, E. & Karplus, M. (2000). Unfolding proteins by external forces and temperature: the importance of topology and energetics. *Proc. Natl Acad. Sci. USA*, **97**, 6521–6526.
21. MacDonald, R. I., Musacchio, A., Holmgren, R. A. & Saraste, M. (1994). Invariant tryptophan at a shielded site promotes folding of the conformational unit of spectrin. *Proc. Natl Acad. Sci. USA*, **91**, 1299–1303.
22. Mitchell, T., Menhart, N., Meerson, M., Topouzian, N. & Fung, L. W. M. (1996). Spectrin domain stability as measured solvent denaturation: Implications for minimum domain. *Biophys. J.* **70**, TU232–TU232.
23. Law, R., Carl, P., Harper, S., Daleheimer, P., Speicher, D. W. & Discher, D. E. (2003). Cooperativity in forced unfolding of tandem spectrin repeats. *Biophys. J.* **84**, 533–544.
24. Law, R., Liao, G., Harper, S., Yang, G., Speicher, D. W. & Discher, D. E. (2003). Pathway shifts and thermal softening in temperature-coupled forced unfolding of spectrin domains. *Biophys. J.* **85**, 3286–3293.
25. 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.
26. Seeliger, M. A., Breward, S. E. & Itzhaki, L. S. (2003). Weak cooperativity in the core causes a switch in folding mechanism between two proteins of the cks family. *J. Mol. Biol.* **325**, 189–199.
27. 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.
28. 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.
29. Oliveberg, M. (1998). Alternative explanations for “multistate” kinetics in protein folding: transient aggregation and changing transition-state ensembles. *Accs Chem. Res.* **31**, 765–772.
30. Brandts, J. F., Halvorson, H. R. & Brennan, M. (1975). Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis-trans* isomerism of proline residues. *Biochemistry*, **14**, 4953–4963.
31. Schmid, F. X. & Baldwin, R. L. (1978). Acid catalysis of the formation of the slow-folding species of RNAs A: evidence that the reaction is proline isomerization. *Proc. Natl Acad. Sci. USA*, **10**, 4764–4768.
32. Kim, P. S. & Baldwin, R. L. (1990). Intermediates in the folding reactions of small proteins. *Ann. Rev. Biochem.* **59**, 631–660.
33. Chiti, F., Taddei, N., Giannoni, E., van Nuland, N. A. J., Ramponi, G. & Dobson, C. M. (1999). Development of enzymatic activity during protein folding—detection of a spectroscopically silent native-like intermediate of muscle acylphosphatase. *J. Biol. Chem.* **274**, 20151–20158.
34. Lilie, H., Rudolph, R. & Buchner, J. (1995). Association of antibody chains at different stages of folding: prolyl isomerisation occurs after formation of quaternary structure. *J. Mol. Biol.* **248**, 190–201.
35. Schmid, F. X. (1983). Mechanism of folding of ribonuclease A. Slow refolding is a sequential reaction *via* structural intermediates. *Biochemistry*, **22**, 4690–4696.
36. Pappenberger, G., Aygun, H., Engels, J. W., Reimer, U., Fischer, G. & Kiefhaber, T. (2001). Nonprolyl *cis* peptide bonds in unfolded proteins cause complex folding kinetics. *Nature Struct. Biol.* **8**, 452–458.
37. Silow, M. & Oliveberg, M. (1997). High-energy channeling in protein folding. *Biochemistry*, **36**, 7633–7637.
38. 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**, 13656–13662.
39. 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.
40. 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.
41. 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.

42. Bachmann, A. & Kiefhaber, T. (2001). Apparent two-state tendamistat folding is a sequential process along a defined route. *J. Mol. Biol.* **306**, 375–386.
43. Munoz, V. & Serrano, L. (1997). Development of the multiple sequence approximation within the agadir model of alpha-helix formation. Comparison with Zimm–Bragg and Lifson–Roig formalisms. *Biopolymers*, **41**, 495–509.
44. Lacroix, E., Viguera, A. R. & Serrano, L. (1998). Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.* **284**, 173–191.
45. Munoz, V. & Serrano, L. (1995). Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. *J. Mol. Biol.* **245**, 297–308.
46. Munoz, V. & Serrano, L. (1995). Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. *J. Mol. Biol.* **245**, 275–296.
47. 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. *Struct. Fold. Des.* **7**, 1145–1153.
48. Scott, K. A., Steward, A., Fowler, S. B. & Clarke, J. (2002). Titin; a multidomain protein that behaves as the sum of its parts. *J. Mol. Biol.* **315**, 819–829.
49. Calloni, G., Taddei, N., Plaxco, K. W., Ramponi, G., Stefani, M. & Chiti, F. (2003). Comparison of the folding processes of distantly related proteins. Importance of hydrophobic content in folding. *J. Mol. Biol.* **330**, 577–591.
50. Baker, D. (2000). A surprising simplicity to protein folding. *Nature*, **405**, 39–42.
51. Ferguson, N. & Fersht, A. R. (2003). Early events in protein folding. *Curr. Opin. Struct. Biol.* **13**, 75–81.
52. 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.
53. 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.
54. Wright, C. F., Lindorff-Larsen, K., Randles, L. G. & Clarke, J. (2003). Parallel protein-unfolding pathways revealed and mapped. *Nature Struct. Biol.* **10**, 658–662.
55. Hamill, S. J., Meekhof, A. E. & Clarke, J. (1998). The effect of boundary selection on the stability and folding of the third fibronectin type III domain from human tenascin. *Biochemistry*, **37**, 8071–8079.
56. 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.
57. 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 immunoglobulin-like fold. *J. Mol. Biol.* **305**, 1185–1194.
58. 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.
59. Pace, C. N. (1986). Determination and analysis of urea and guanidinium hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266–280.

Edited by C. R. Matthews

(Received 31 May 2004; received in revised form 7 September 2004; accepted 17 September 2004)

Note added in proof: Kusunoki *et al.* have recently reported the crystal structure of R15 in tandem array with R16 and R16+R17. This confirms that the structure of R15 is the same as that of R16 and R17 and that it has the same contact order (see: Kusunoki, H., Minasov, G., MacDonald, R. I. & Mondrago, A. (2004) Independent movement, dimerization and stability of tandem repeats of chicken brain α -spectrin. *J. Mol. Biol.*, in press). We thank Dr. Ruby MacDonald for sending us the coordinates in advance of publication.