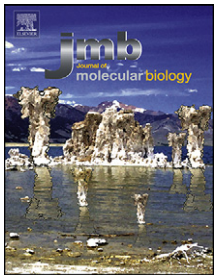


The Folding of a Family of Three-Helix Bundle Proteins: Spectrin R15 Has a Robust Folding Nucleus, Unlike Its Homologous Neighbours

Lee Gyan Kwa†, Beth G. Wensley†, Crispin G. Alexander, Stuart J. Browning, Benjamin R. Lichman and Jane Clarke

Department of Chemistry, *University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK*

Correspondence to Jane Clarke: *jc162@cam.ac.uk*
http://dx.doi.org/10.1016/j.jmb.2013.12.018
Edited by S. Radford

Abstract

Three homologous spectrin domains have remarkably different folding characteristics. We have previously shown that the slow-folding R16 and R17 spectrin domains can be altered to resemble the fast folding R15, in terms of speed of folding (and unfolding), landscape roughness and folding mechanism, simply by substituting five residues in the core. Here we show that, by contrast, R15 cannot be engineered to resemble R16 and R17. It is possible to engineer a slow-folding version of R15, but our analysis shows that this protein neither has a rougher energy landscape nor does change its folding mechanism. Quite remarkably, R15 appears to be a rare example of a protein with a folding nucleus that does not change in position or in size when its folding nucleus is disrupted. Thus, while two members of this protein family are remarkably plastic, the third has apparently a restricted folding landscape.

Introduction

In the two decades since the introduction of Φ-value analysis to investigate the structure of protein folding transition states (TSs), over 50 TSs have been investigated to some extent (recently reviewed in Ref. [1]). Although only providing a static snapshot of the folding process, these results have provided detailed insight into the nature of the TSs through which small proteins fold. This understanding has been aided by the use of perturbations of TSs, through the use of mutations, insertions, circularisation and circular permutations [2–12], allowing us to observe phenomena such as Hammond behaviour, where one might observe TS growth along the reaction coordinate [13,14], or anti-Hammond effects and parallel pathways [15–17]. Studying how a TS is affected by these various perturbations affords a more complete understanding of TS flexibility and mobility over the folding landscape.

More drastic differences can be observed when members of protein structural families are compared [1,18]. A seminal study of the three-helix-bundle homeodomain family allowed Fersht, Daggett and co-workers to propose that there is continuum of folding mechanisms between diffusion–collision and nucleation–condensation folding mechanisms [19–26]. This continuum has proved a useful framework for understanding both folding mechanism and pathway. They showed that the principal determinant of mechanism lies in the balance between secondary structure propensity and tertiary contacts. Where secondary structure (helical) propensity is high, early local structure formation is favoured, but where helical propensity is low, formation of long-range tertiary interactions is concomitant with secondary structure formation.

In addition to the homeodomain family, TSs seen for members of the well-studied PSBD, PDZ and spectrin domains all fall along this same continuum [14,27–35]. Interestingly, all-β proteins are less likely to fold using a framework-like mechanism, although even where all members fold by nucleation–condensation, significant movements in TS have been observed, in proteins such as the immunoglobulin-like [36] and SH3 [37] domains, as well as for the α/β protein S6 [8].

Spectrin domains are elongated three-helix bundle domains, and three have been studied in detail, R15,

R16 and R17. They are superficially similar with comparable structures, stabilities and β-Tanford values, but R15 folds and unfolds around 3 orders of magnitude faster than R16 and R17 [38–43]. The rate-limiting TSs of all three domains in water are topologically similar: all three involve only the A-helix and the C-helix [14,34,35]. However, the pattern of Φ-values differs between the R15 and the slow-folding domains R16 and R17, and this difference in pattern of Φ-values has led us to conclude that these domains use different folding mechanisms. The origin of the difference in folding rates appears to be due to this difference in folding mechanism: R15 folds via a nucleation–condensation mechanism with only the centres of the two helices structured and docked whereas both R16 and R17 fold via a framework-like mechanism involving the docking of partly preformed helices A and C along the entire length [44]. Framework-like mechanisms are commonly associated with faster rates of folding, but this is not the case for these spectrin domains. In fact, the folding landscapes of R16 and R17 appear to be frustrated having high internal friction that slows both folding and unfolding by altering the underlying roughness of the landscape rather than a more typical slowing due to barrier height. Such frustration has not been previously observed in domains of this size that fold on the millisecond-to-second timescale. This internal friction has been mapped along the landscape for these three proteins, and significant friction is observed only at the early, rate-determining TS of R16 and R17 [45]. This early TS is the point on the folding pathway where the topology is established through the correct alignment of the two terminal helices, and the presence of friction has been ascribed to a frustrated search for the correct docking of partly preformed helices. It is also, in part, responsible for the slower folding and unfolding kinetics exhibited by R16 and R17 [46]. R15 is not subject to high internal friction, as early formation of strong tertiary contacts between helices A and C establish the topology and both tertiary and secondary structure can develop around this nucleus, without forming non-native frustrated contacts.

Thus, despite these spectrin domains having very similar structures (pair-wise RMSD < 1 Å), the folding of R15 is distinct from that of R16 and R17. However, we have shown that R16 and R17 can be made more R15-like, in terms of folding speed, TS structure, folding mechanism and reduced landscape frustration. This was initially through the use of fully core swapped domains where we defined the core residues and engineered R16 and R17 to have the core of R15. Subsequently, we have shown that substitution of a "minimal core" of just seven residues (m7) or even just five core residues on the A-helix (m5) in both R16 and R17 to the amino acids seen at those positions in R15 is sufficient to shift the folding of R16 and R17 to become more R15-like [44,47]. That is, both folding and unfolding rates are increased, the folding mechanism is shifted towards nucleation–condensation and landscape frustration reduced. The folding landscapes of R16 and R17 are malleable enough for a small number of mutations to significantly shift the folding behaviour of these domains.

Using a similar core swap strategy, here we investigate the folding of R15, searching for slow-folding variants to ask whether slower folding in R15 is accompanied by a change in mechanism and/or an increase in landscape roughness. We find that, unlike R16 and R17, the TS of R15 is remarkably resilient to substitution, displaying an apparent lack of deformability that is rarely seen.

Results

R15 is resistant to core swapping

The full core swapped proteins with R15 as the major parent, R15o16c and R15o17c, express insolubly [44,47]. We were able to purify a small amount of R15o16c from inclusion bodies without removing the N-terminal His-tag, the presence of which does not affect the stability of wild-type (WT) spectrin proteins (data not shown). The protein is only partly folded, but a truncated chevron plot (Fig. 1) allows us to estimate that it has a $\Delta G_{D-N} <$ 1 kcal mol^{-1}, compared to 7.2 kcal mol^{-1} for WT R15. The destabilisation is mostly reflected in folding kinetics. It refolds significantly more slowly than R15, but crucially, it also unfolds more rapidly (we do not see the lower folding AND unfolding rate constants that characterise the intrinsically slow folding of R16 and R17). A "minimal core" version of R15, R15m7 (with just seven substitutions—five in the A-helix and one in each of the B- and C-helices—the only core residues that are identical in R16 and R15 but different in R15 [47]), could not be purified from inclusion bodies.

Five individual mutations of helix A in R15

There are five core residues from R15 that, when substituted together into R16 and R17, promote faster folding, reduced internal friction at the TS and a more nucleation–condensation-like folding mechanism [47]. These residues are all found in the A-helix, in the region surrounding the highly conserved Trp21 that is key for nucleating folding (Fig. 2). The effect of substituting these residues in R15 with their equivalents from R16 and R17 was determined individually. Three of these substitutions, D19E, L22I and L29V (where the first residue is that found in R15 and the second is that found in R16 and R17), are extremely conservative; only two, F18E

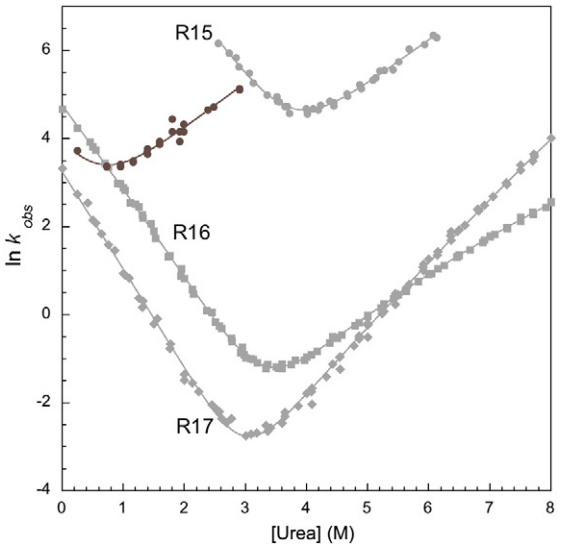


Fig. 1. The kinetics of R15o16c compared with the WT domains. R15o16c data are shown in brown, and R15, R16 and R17 are shown in grey. R15, R16 and R17 data are taken from Ref. [43].

and V25K, change the nature of the side chain significantly. All variants displayed similar fluorescence and CD profiles as WT R15.

Kinetic data were collected for all five mutant proteins (Fig. 3 and Table 1). Most have a lower thermodynamic stability (ΔG_{D-N}) and fold slower and unfold faster than WT R15. The stabilised D19E conversely folds faster and unfolds more slowly. Interpretation of such data may be complicated. A change in the thermodynamic stability of a domain may or may not affect TS barrier crossing (i.e., folding and unfolding speeds), depending on the effect of the mutation on the relative barrier heights. This relative change in barrier height upon mutation is what a Φ-value measures [48]; thus, we determined Φ-values for these five mutations and compared them to previously published Φ-values at the same sites, where a conservative deletion mutation had been made (Table 1) [35]. Where the

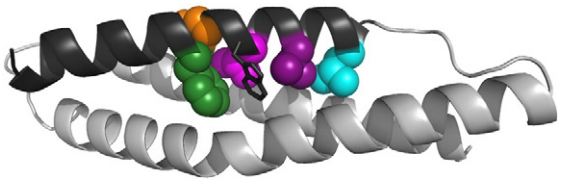


Fig. 2. The five nucleus residues in R15 that, when substituted into R16 or R17, alter their folding to be more R15-like. These five residues are shown as space-filling models: Phe18 (green), Asp19 (orange), Ile22 (pink), Val25 (purple) and Leu29 (cyan). These five residues pack round Trp21, which is at the centre of this cluster (sticks). All these residues are on the A-helix (black).

Φ-values correspond to the original reported Φ-values, it is reasonable to conclude that the TS is unchanged by this mutation. {Note that the F18E substitution was so destabilising that the data were fitted by fixing the kinetic *m*-values to those of WT, and the Φ-value of 0.4 must be considered to be an estimate. However, the Φ-value is the same as that for the corresponding F18A mutation as previously reported (0.3) [35]. Moreover, if we make the Phe-to-Glu mutation in the background of the stabilising D19E mutation, we get the same Φ-value (0.4) (see Supplementary Fig. S1).} The only Φ-value that is significantly different to the previous work is that of V25K (0.9 compared to 0.4 in V25A). In the core-swapped versions of R16 and R17, we found that changes in Φ-value pattern indicated a change in the structure of the TS and of the folding mechanism. This led us to ask whether V25K changes the folding mechanism and thus the overall Φ-value pattern of R15.

The results for these individual mutant proteins may shed some light on the fact that R15 does not seem to be able to tolerate multiple substitutions in the core to the same extent as R16 and R17, even though they have a high structural homology and have, in general, similar response to individual substitutions in the core [14,35]. Here we see that both mutations that introduce charged residues into the core, F18E and V25K, are quite destabilising (the total loss of stability from these two, non-interacting residues alone is 6.2 kcal mol^{-1}). Perhaps the structure of R15 does not allow the charges from these side chains to be fully solvated in the native state, unlike what is seen in R16 and R17. Interestingly, the only favourable substitution here (D19E) extends the hydrophobic portion of the side chain, perhaps allowing the charged moiety to be more solvent exposed. It seems likely that the various core-swapped versions of R15 were simply too destabilised to be produced successfully. An alternative possibility is that R15 cannot tolerate a number of substitutions of its folding nucleus.

Design of R15m2 (R15 D19E/V25K)

Thus, we found that the substitution that apparently alters the folding of R15 most significantly is V25K, which alters the Φ-value significantly and also has the largest effect on the folding rate constant, slowing the folding by about 50-fold. This protein is destabilised relative to WT R15. Therefore, to allow us to use protein engineering to investigate the reason for the slow folding of this mutant protein, we introduced a second, stabilising mutation, D19E, that does not affect the folding rate constant (note that these two residues do not interact in WT R15). This protein is termed R15m2 (R15 with *2* of the *m*inimal core residues substituted) and has a Φ-value of ~0 compared to V25K. The addition of D19E to V25K

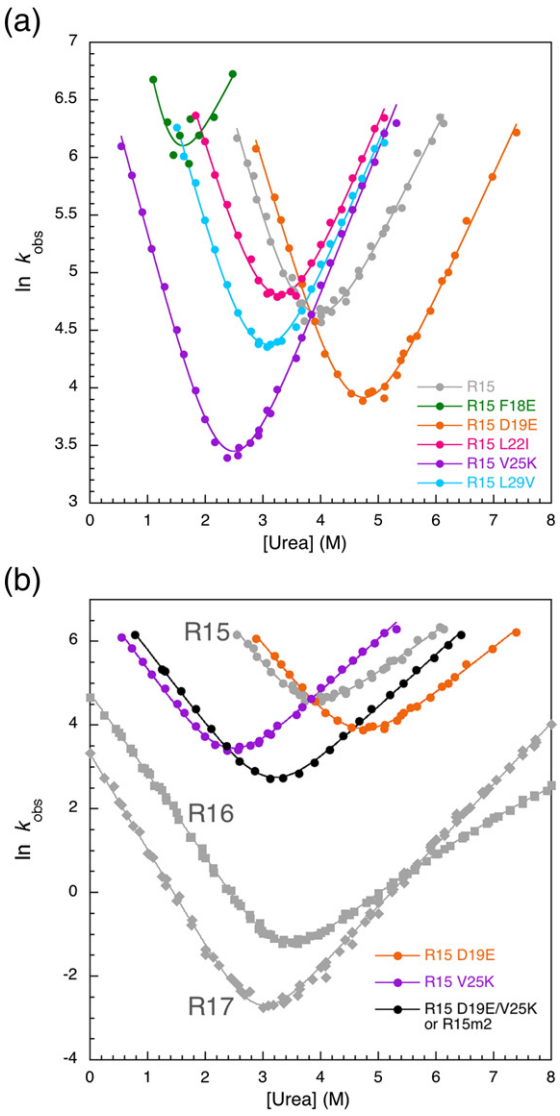


Fig. 3. The effect of mutating the five key residues in R15 to the side chain found at that position in R16 and R17. (a) Chevron plots of the five individual mutations. Continuous lines represent the fit of the data. (b) Chevron plots of R15m2, proteins containing the two constituent mutations D19E and V25K and the WT domains.

results in a stable R15 variant that has approximately the same stability as WT R15, thus allowing us to study the folding mechanism using Φ-value analysis. The Φ-value of R15m2 (compared to WT R15) is now highly anomalous, at 1.6, which, since the effects of V25K and D19E are essentially additive and D19E has a Φ-value of ~0, demonstrates the very large effect of the V-to-K mutation. R15m2 allows us to ask why the single point mutation V25K causes such an anomalous change in the folding of R15. Is the decrease in both folding and unfolding rate constants in R15m2 associated with a concomitant change in folding mechanism and/or increase in landscape roughness?

R15m2 viscosity dependence study

The rate constants for folding and unfolding of WT R15 have very strong solvent viscosity dependence, whereas WT R16 and R17 show only a weak dependence [44]. Weak solvent viscosity dependence reflects the rough energy landscapes in R16 and R17. Briefly, assuming that folding is a diffusive, Kramers-like process across the free-energy landscape, the rate constant for folding (and unfolding) can be described, empirically, by Eq. (1) where the folding or unfolding rate constant k is dependent on η (the solvent viscosity), σ (the "internal friction" of the protein), ΔG^{TS} (the height of the energy barrier) and C (a temperature- and solvent-independent term, including all components of the pre-exponential factor except the friction terms).

$$k = \frac{C}{\eta + \sigma} \exp\left(\frac{-\Delta G^{TS}}{RT}\right) \quad (1)$$

Thus, if internal friction σ is negligible, then the normalised folding (or unfolding) rate constant plotted against the inverse of the normalised solvent viscosity will give a slope of 1 (see Refs. [44] and [49–51] for detailed discussions). For most small proteins, and indeed for WT R15, such plots do give a slope close to unity. R16 and R17 are very rare exceptions, their plots having slopes ~0.2 (Fig. 4). Fast-folding, core-swapped versions of R16 and R17 have an increased solvent viscosity dependence [44,47]. We therefore asked: is the significant decrease in the rate constant for folding of R15m2 reflected in a concomitant decrease in solvent viscosity dependence?

The effect of solvent viscosity on the folding of R15m2 was determined as previously described [44]. Figure 4a shows the chevrons of R15m2 at various glucose concentrations, fitted individually. Since glucose, like most viscogens, stabilises the protein, we use an iso-stability approach [52–58] and determine the k_f at ΔG_{D-N} = 1.5 kcal mol^{-1} or $k_f = k_u$ at ΔG_{D-N} = 0.0 kcal mol^{-1} for each glucose concentration. The relative rate constants (k_0/k) were plotted against the relative solvent viscosity (η/η_0) (Fig. 4b and c). Slopes of these plots for R15m2 (mean slope of 1.00 ± 0.02) were similar to what had previously been observed for WT R15 (mean slope of 0.80 ± 0.05). Slow folding in R15m2 is not, therefore, related to any increase in the roughness of the free-energy landscape.

Φ-Value analysis of R15m2

We have shown that faster folding in the core-swapped versions of R16 is due in part to a change in the folding mechanism from a framework-like to a nucleation–condensation mechanism [44]. To

Table 1. Thermodynamic and kinetic parameters for R15 variants.

Mutant	$\Delta G_{D-N}^{H_2O}$ [a] (kcal mol^{-1})	$k_f^{H_2O}$ [b] (s^{-1})	m_{k_f} (kcal mol^{-1} M^{-1})	$k_u^{H_2O}$ (s^{-1})	m_{k_u} (kcal mol^{-1} M^{-1})	Φ_f^{2M}	"Classical" Φ[c]
R15[d]	7.2 ± (0.2)	60,000 ± (13,000)	1.9 ± (0.1)	1.3 ± (0.2)	1.0 ± (0.1)	—	—
R15 F18E[e]	3.5 ± (0.1)	3800 ± (600)	1.9 ± (0.1)	59.0 ± (6)	1.0 ± (0.1)	0.4	0.3
R15 D19E	8.9 ± (0.2)	48,000 ± (6400)	1.7 ± (0.1)	0.16 ± (0.02)	1.1 ± (0.1)	0.1	ND[f]
R15 L22I	6.0 ± (0.2)	9300 ± (1600)	1.7 ± (0.1)	1.55 ± (0.40)	1.2 ± (0.1)	0.6	0.5
R15 V25K	4.7 ± (0.1)	1340 ± (70)	1.9 ± (0.1)	0.83 ± (0.07)	1.2 ± (0.1)	0.9	0.4
R15 L29V	5.6 ± (0.2)	8000 ± (890)	1.8 ± (0.1)	1.36 ± (0.19)	1.2 ± (0.1)	0.7	0.5
R15m2 (D19E/V25K)	6.1 ± (0.2)	1500 ± (20)	1.4 ± (0.1)	0.05 ± (0.01)	1.9 ± (0.1)	1.6	—
R15 V25M	ND	55,000 ± (13,000)	1.7 ± (0.1)	2.6 ± (0.5)	0.9 ± (0.1)	—	—

[a] All ΔG are calculated using mean m_{D-N}^{eqb} of 1.9 kcal mol^{-1}.
[b] All data fitted to a two-state model, except R15m2 that is fitted with a broad TS model (see the main text).
[c] All classical Φ-values were taken from Wensley et al. [35]. All the mutations are to alanine.
[d] R15 data are taken from Scott et al. [43].
[e] Substitution of F18E was so destabilising that very little kinetic data could be collected. In this case, data were fitted by fixing the m_{k_f} and m_{k_u} to those of R15. Φ_f^{2M} = 0.4 must be considered as an estimate.
[f] ND indicates that these data and values have not been collected or calculated.

determine whether the slow folding of R15m2 is also related to a change in folding mechanism, we carried out a Φ-value analysis for R15m2. Forty-five single point mutations were introduced into R15m2 at 36 sites to sample its entire structured region. For direct comparison with the Φ-value analysis on spectrin R15, the same mutations were made as those previously published for WT R15 [35]. Two classes of mutations, core and surface residues, were made as in previous studies [14,34,35]. The chevron plots are shown in Supplementary Figs. S2 and S3, and results are tabulated in Supplementary Tables S1 and S2.

Comparison of Φ-values

R15m2 and R15 Φ-values are compared in Figs. 5 and 6. Qualitatively, the Φ-value patterns are essentially the same (Fig. 5). We observe high Φ-values in specific regions of helices A and C indicating that both the helices are involved in early structure formation. These regions of the A- and C-helices pack together in the native structure. We have proposed that these regions form a folding nucleus around which the rest of the protein folds. In R15m2, we observe uniformly low Φ-values in helix B, suggesting that it is not involved in the early formation of structure.

We were surprised to also find a very good quantitative agreement between the Φ-values of the WT R15 and R15m2 (Fig. 6). The slope of the linear fit in Fig. 6 is close to 1 (0.93) with an intercept of 0.05 (R = 0.87). We conclude that the folding mechanism and indeed the structure of the TS for folding in R15m2 are unchanged.

Why then does R15 V25K fold so slowly?

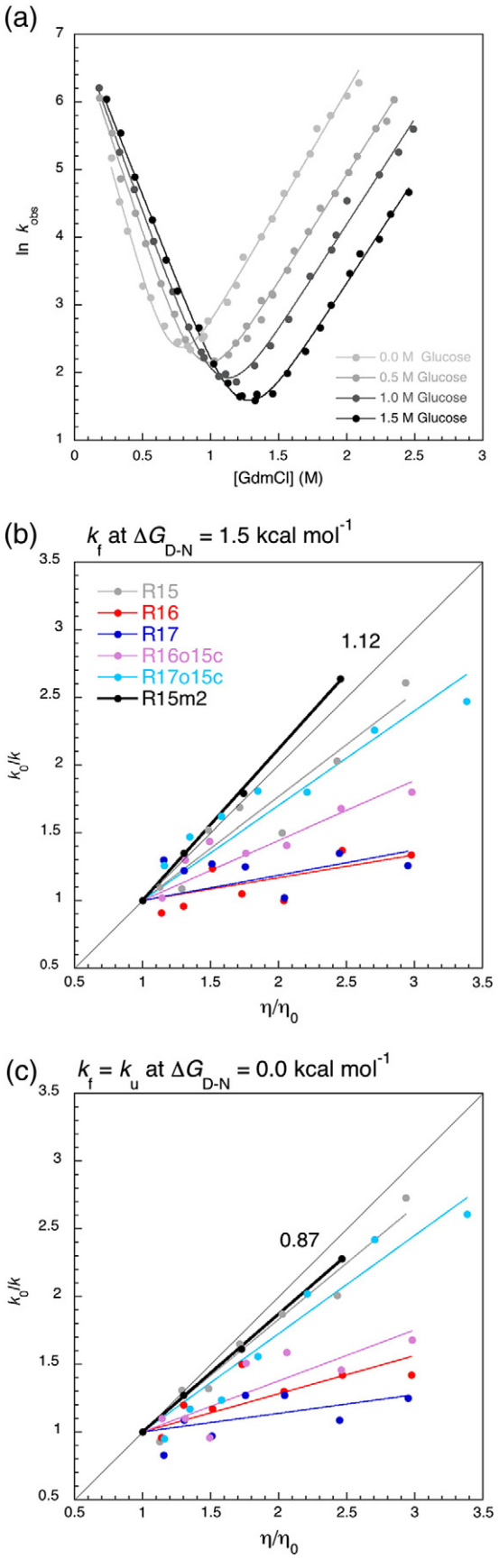
Since we have shown that R15m2 folds via an unaltered TS, we have to conclude that this TS is more destabilised by the Val-to-Lys mutation than the native state. Perhaps surprisingly, Lys25 is considered to be a core residue in R16 and R17 despite being charged. Analysis of the R16 and R17 structures shows that the hydrophobic part of K25 is well packed into the core while the charged moiety is solvent exposed. We have also shown that substitution of the Lys in R16 and R17 by hydrophobic residues speeds the folding. Thus, we investigated the folding of two further substitutions at this position in R15. V25A has been previously reported [35]. It slows folding very slightly with a Φ-value of 0.4. V25M has no effect on the stability of R15 but actually speeds both folding and unfolding (Fig. 7). Early formation of structure around this point may be disfavoured by packing the β-branched side chain of Val [59].

Discussion

The slow folding of R15m2

In this work, we describe an attempt to change the folding mechanism of R15 spectrin to match that of its slow-folding homologues R16 and R17. We have engineered a version of R15 that, upon initial inspection, displayed the reverse signature to that observed for the R16 and R17 core swaps, that is, a protein, R15m2, that both folds and unfolds significantly slower than the parent domain but has a similar stability. However, this is where the resemblance ends. Unlike the R16/17 core-swapped domains, neither TS frustration nor folding mechanism of R15m2 has been affected.

D19E was engineered into R15 V25K solely to import extra stability to allow a Φ-value analysis to be performed. Since D19E is essentially unstructured in



the TS of both WT R15 and R15 V25K (Φ-values of 0.1 and 0.0, respectively), we infer that all the effects on the TS that we observe in R15m2 are due to the presence of lysine at position 25. Our Φ-value results suggest that the side chain of residue 25 will be significantly buried in the TS: this region of the A-helix is at the centre of the folding nucleus and is packed onto the emerging C-helix (Fig. 5). In the native protein, which is more compact, the positively charged NH_3^+ group will be at least partly solvent exposed, as the long hydrophobic moiety of the side chain is packed in the small core of this three-helix bundle. Our results are consistent with the hypothesis that the charge of K25 will be more buried in the less compact TS than in the native state; thus, the TS will be more destabilised than the native state by substitution of a hydrophobic by a charged residue in this position. Thus, the barrier to both folding and unfolding will be raised, and both folding and unfolding rate constants will be lowered. In direct comparison, specific stabilisation of the TS subsequent to removal of charges from the core has been seen in R16 and R17. The reverse of all of the five individual mutations made here in R15 were made in R16 and R17 (E18F, E19D, I22L, K25V and V29L) [46]. In both R16 and R17, the two charge-to-neutral mutations have a significant effect on (un)folding, with a combined effect of the removal of the charges of about a 40-fold increase in the folding rate constant [46]. This is not dissimilar to the ~50-fold effect seen here upon the addition of Lys to the nucleus. The fact that the F18E does not have the same effect in R15 can be explained by the differences in the TS structures of the two domains. This residue is at the edge of the R15 nucleus (see Fig. 5; F18A has a Φ-value of 0.25 [35]); thus, the charge is not likely to be significantly buried at the TS. However, the TS in R16 is both frustrated and more framework-like, with partially preformed helices trying to correctly dock, and consequently, a more extended set of R16 core residues, including E18, are partly buried and thus destabilising when charged. Such a destabilisation of a TS by transient burial of a charge is not new [60,61].

Fig. 4. The dependence of the folding of R15m2 on viscosity is unchanged from WT R15. (a) R15m2 chevron plots collected in 0.0–1.5 M glucose. Data were fitted individually to a two-state model. (b) and (c) show the dependence of the relative rate constants (k_0/k) on the relative solvent viscosity (η/η_0). [Values for k were determined at both ΔG_{D-N} = 1.5 kcal mol^{-1} (k_f) (b) and ΔG_{D-N} = 0.0 kcal mol^{-1} ($k_f = k_u$) (c).] All data except those for R15m2 were taken from Ref. [44].

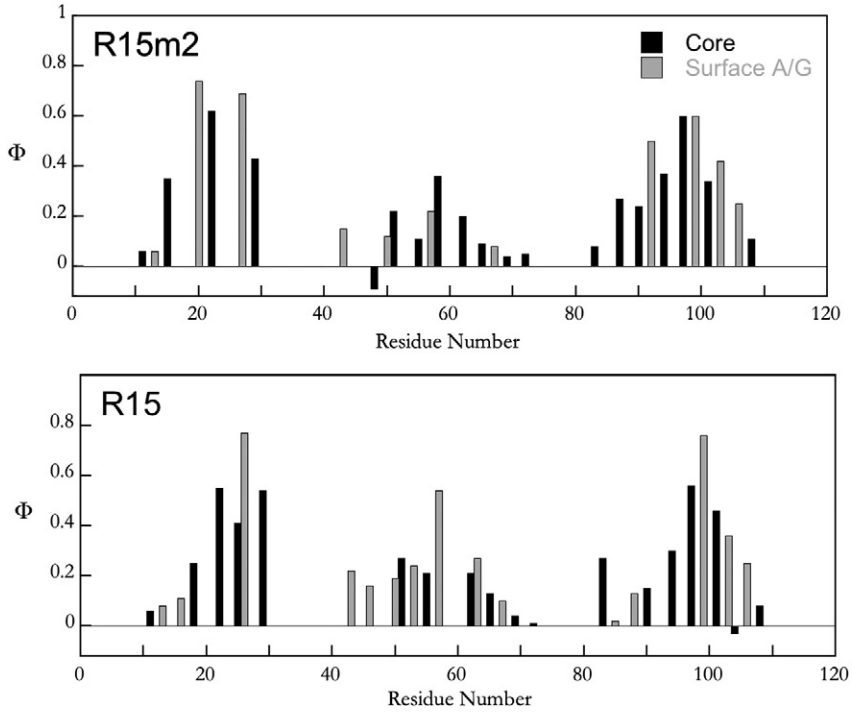


Fig. 5. The TS structure of R15m2 is qualitatively unchanged from that of R15. Bar charts for the Φ-values of R15m2 (top) and R15 (bottom) in water. Core mutations are shown in black, and surface-exposed Aly–Gly scanning mutations are shown in grey. Φ-Values are shown when $\Delta\Delta G^{eqb}_{D-N} \geq 0.70$ kcal mol^{-1} as determined by equilibrium denaturation. R15 data are taken from Ref. [35].

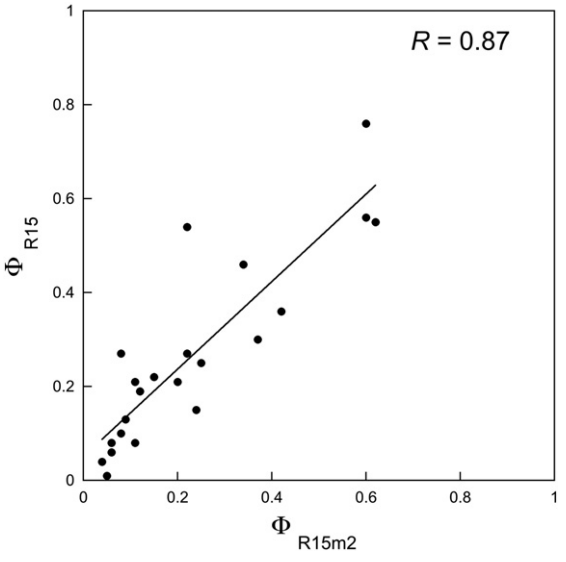


Fig. 6. The TS structure of R15m2 is quantitatively unchanged from that of R15. Comparison of the Φ-values that are available for both R15m2 and R15. The continuous line shows a linear fit of the data.

The lack of malleability in the R15 TS

What is remarkable about our results is that, quantitatively and qualitatively, the structure of the R15 TS is unaffected by destabilisation and R15m2 folds with the same nucleation–condensation mechanism as WT. Unlike most other systems that have been studied, where destabilisation of the TS results in an alteration of that TS, the TS in R15 appears to be very inflexible: folding and unfolding via a more destabilised WT TS is apparently more energetically favourable than to shift over the free-energy surface to an extended, partially altered or completely alternate TS structure. R15 appears to have no malleability in the folding landscape. This complete lack of malleability is unusual for domains that fold via a nucleation mechanism [18]. Studies that probe TS malleability are a sizable undertaking, the WT TS structure must be known (usually from Φ-value analysis), a perturbation must be made and any change in the TS structure must be probed for. Nevertheless, a number of domains have been studied in this way (reviewed in Ref. [18]). Using a stringent definition for TS inflexibity, that is, no shift in size or position of the nucleus nor of the mean Φ-value, the classic two-state folder CI2 and the

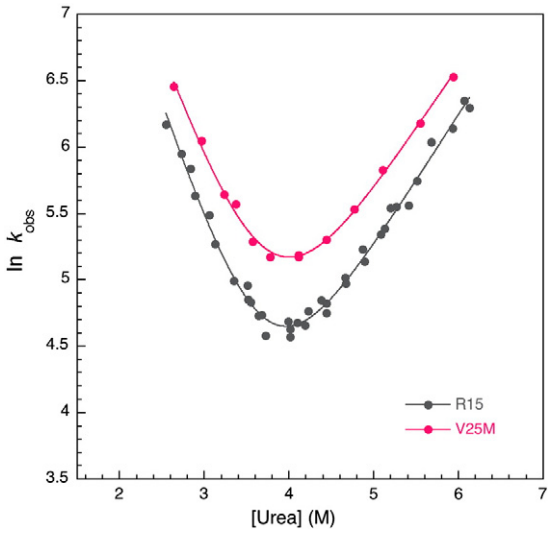


Fig. 7. R15 folds and unfolds even more rapidly when the β-branched Val25 is substituted by an unbranched methionine. This supports the hypothesis that the slow folding of V25K is due to the burial of the charged moiety of the lysine side chain.

small three-helix-bundle BdpA are the only other domains for which no experimental perturbations made have resulted in an altered TS structure. However, the nature of the perturbations made vary somewhat. The TS of CI2 has been quite severely altered via circularisation, circular permutation and bisection, but the TS structure, as judged by multiple Φ-values, remained unchanged [2,3]. A much less serious perturbation has been made of the BdpA TS: a single point mutation (L45A) designed to destabilise one of two potential nucleation motifs [4]. This has no effect on the population of the other motif, as judged by a Ψ-value. The TS of BdpA, as judged by Φ-value analysis, is also unchanged by increased temperature [62]. Surveying the literature, however, R15, CI2 and BdpA are exceptions [18]. It is important to note that, from a lack of observed malleability, it does not follow that there is no possibility of an alternative rather that none has been seen to date.

The apparent lack of malleability of the TS of R15 is especially surprising given its structure. Perhaps, for the more complex topology of CI2 and the very small fold of BdpA, the necessity of maintaining a specific folding nucleus to set up the topology may be a factor. However, it is easy to conceive of a landscape where the TS nucleus could shift along the long helices of R15 to other turns of helix, and indeed, Lindberg and Oliveberg suggested that spectrin domains will have overlapping foldons and, thus, will have alternative folding routes [17]. An analogous shift is seen in the TS of the fnIII superfamily members TNfn3 and CAfn2 [36]. Both are Ig-like domains, and TNfn3 displays the "classic" nucleus seen for this Greek-key fold, with well-conserved hydrophobic key nucleating residues that set up the topology of this complex fold [63]. These residues are not present in CAfn2, and Φ-value analysis shows that the nucleus of this family member has merely shifted along the β-strands to an adjacent set of hydrophobic residues. This is, however, not what we see for R15: the nucleus is apparently very inflexible.

Alternate origin of folding continuum shifts

The EnHD family proteins are another family of extensively studied three-helix bundles and are the family that was used to develop the hypothesis that the folding of small proteins sit somewhere on a continuum between true nucleation–condensation and true diffusion–collision (or framework) folding mechanisms [19–21,23–25]. In this family, the faster-folding domains (exemplified by EnHD itself) fold fast via a diffusion–collision mechanism where preformed helices dock to form the folded structure. The slower-folding domains lack the high helical propensity of EnHD and fold via true nucleation–condensation or a mixed mechanism. In this family, it is the *presence* of preformed helices that dictates the mechanism through which a domain folds. The evidence presented here suggests that the factor determining where a spectrin domain sits on this continuum is the presence or absence of a strong nucleus. It was relatively easy to engineer a nucleus into R16 and R17, to cause a switch towards a nucleation–condensation mechanism. We have been unable to engineer a form of R15 that folds by an alternative mechanism. Unfortunately, it is not possible to determine whether this is because R15 cannot access the alternative diffusion–collision mechanism used by to R16 and R17 or because the mutations needed to knock out the nucleus are key to the stability of R15. We have, for example, introduced the m5 residues into a version of R15 that has increased engineered helicity but could not produce any soluble, folded protein (results not shown). All these spectrin domains have similar values of intrinsic helical propensity; thus, differential helical propensity appears not to be critical.

For the EnHD family, secondary structure and a dominant framework mechanism dictate the folding landscape, whereas for the spectrin domains, it is the tertiary structure and the formation of a strong nucleus. Given that these two families share one of the simplest topologies seen in *Nature*, it is interesting to speculate on the relative importance of these two controlling properties across more complex fold space.

Conclusions

Spectrin domains have a simple fold, but, yet again, they have surprised us. We expected to find that we could disrupt the folding nucleus in R15 and find that it could fold by an alternative, framework-like mechanism or at least that the protein would use an alternative or expanded nucleus. Instead, we find that apparently R15 is a rare example of a protein with an inflexible TS. The energy landscape does not allow alternative routes to the native state. Why this should be true of R15 when R16 and R17 can be manipulated so readily is unclear. When we are unable to make predictions for such simple proteins, we are lead to conclude that the "protein folding problem" is not yet solved [64].

Methods

Full and minimal core swaps of R15 were synthesised (GenScript, USA) and cloned into pRSETA (Invitrogen) using BamHI and EcoRI. Standard mutagenesis, protein expression and purification methods have been described elsewhere, as have details of how the biophysical data are collected [35,43]. Sequences of the spectrin domains involved are shown in Supplementary Fig. S4.

The methodology used for the viscosity measurement and analysis of R15m2 was based on our previously established methods [44,46,47]. Kinetic data were collected and fitted individually to two-state models. These fits were then used to determine k_f at ΔG_{D-N} = 1.5 kcal mol^{-1} and $k_f = k_u$ at ΔG_{D-N} = 0.0 kcal mol^{-1}.

The R15m2 Φ-value analysis [48] was carried out in the same way as that on WT R15, except that all experiments were carried out at 25 °C and not at 10 °C [35]. The mean m_{D-N}^{eqb} was 1.90 ± 0.05 kcal mol^{-1} M^{-1}, which is very similar to the WT R15 value of 1.88 ±0.15 kcal mol^{-1} M^{-1} [35]. Equilibrium data were fitted to the two-state model used for WT R15. Chevron plots for all mutations are shown in Supplementary Figs. S2 and S3. Unlike WT R15, some mutants show observable curvature in the unfolding arm; thus, all kinetic data were fitted globally to a broad TS model with a shared quadratic term, m', to reduce the error [Eq. (2)]. The goodness of the fit was judged by comparing the kinetic free energy for unfolding and the kinetic m-values against the equilibrium data; this produced a good agreement (data not shown). Thermodynamic and kinetic parameters derived from the fits are shown in Supplementary Tables S1 and S2. Φ-Values were calculated at 2 M urea as this is within the experimental range and reduces the long extrapolation, using Eq. (3). Φ-Values were only determined where the change in free energy of unfolding on mutation, $\Delta\Delta G_{D-N}^{eqb}$, is ≥0.70 kcal mol^{-1}, as determined by equilibrium denaturation curves.

$$\ln k_{obs} = \ln\left(k_f^{H_2O}\exp\left(-m_{k_f}[\text{urea}] + m'[\text{urea}]^2\right) + k_u^{H_2O}\exp\left(m_{k_u}[\text{urea}] + m'[\text{urea}]\right)\right) \quad (2)$$

$$\Phi = \frac{\Delta\Delta G_{D-\ddagger}}{\Delta\Delta G_{D-N}}, \text{ where } \Delta\Delta G_{D-\ddagger} = \frac{\ln k_f^{wt}}{\ln k_f^{mut}} \quad (3)$$

Acknowledgements

This work was supported by the Wellcome Trust (WT095195). J.C. is a Wellcome Trust Senior Research fellow.

Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2013.12.018.

Received 24 September 2013;
Received in revised form 13 December 2013;
Accepted 17 December 2013
Available online 24 December 2013

Keywords:
protein folding;
Φ-value;
energy landscape;
internal friction

†L.G.K. and B.G.W. contributed equally to this work.

Present addresses: B. G. Wensley, MedImmune, Granta Park, Cambridge, CB21 6GH, UK; C. G. Alexander, School of Medicine and Medical Science, University College Dublin, Dublin 4, Ireland; B. R. Lichman, Department of Biochemical Engineering, University College London, Torrington Place, London WC1E 7JE, UK.

Abbreviations used:
TS, transition state; WT, wild type.

References

[1] Nickson AA, Clarke J. What lessons can be learned from studying the folding of homologous proteins? Methods 2010;52:38–50.
[2] Neira JL, Davis B, Ladurner AG, Buckle AM, de Prat Gay G, Fersht AR. Towards the complete structural characterization of a protein folding pathway: the structures of the denatured, transition and native states for the association/folding of two

complementary fragments of cleaved chymotrypsin inhibitor 2. Direct evidence for a nucleation–condensation mechanism. Folding Des 1996;1:189–208.
[3] Otzen D, Fersht A. Folding of circular and permuted chymotrypsin inhibitor 2: retention of the folding nucleus. Biochemistry 1998;37:8139–46.
[4] Baxa MC, Freed KF, Sosnick TR. Quantifying the structural requirements of the folding transition state of protein A and other systems. J Mol Biol 2008;381:1362–81.
[5] Viguera A, Serrano L, Wilmanns M. Different folding transition states may result in the same native structure. Nat Struct Biol 1996;4:874–80.
[6] Nauli S, Kuhlman B, Baker D. Computer-based redesign of a protein folding pathway. Nat Struct Biol 2001;8:602–5.
[7] Nickson AA, Stoll KE, Clarke J. Folding of a LysM domain: entropy–enthalpy compensation in the transition state of an ideal two-state folder. J Mol Biol 2008;380:557–69.
[8] Lindberg M, Tangrot J, Oliveberg M. Complete change of the protein folding transition state upon circular permutation. Nat Struct Biol 2002;9:818–22.
[9] Lindberg MO, Haglund E, Hubner IA, Shakhnovich EI, Oliveberg M. Identification of the minimal protein-folding nucleus through loop-entropy perturbations. Proc Natl Acad Sci U S A 2006;103:4083–8.
[10] Haglund E, Lindberg MO, Oliveberg M. Changes of protein folding pathways by circular permutation: overlapping nuclei promote global cooperativity. J Biol Chem 2008;283:27904–15.
[11] Grantcharova VP, Baker D. Circularization changes the folding transition state of the src SH3 domain. J Mol Biol 2001;306:555–63.
[12] Viguera AR, Serrano L. Loop length, intramolecular diffusion and protein folding. Nat Struct Biol 1997;4:939–46.
[13] Ternstrom T, Mayor U, Akke M, Oliveberg M. From snapshot to movie: phi analysis of protein folding transition states taken one step further. Proc Natl Acad Sci U S A 1999;96:14854–9.
[14] Scott KA, Randles LG, Clarke J. The folding of spectrin domains II: phi-value analysis of R16. J Mol Biol 2004;344:207–21.
[15] Matthews JM, Fersht AR. 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 1995;34:6805–14.
[16] Wright CF, Lindorff-Larsen K, Randles LG, Clarke J. Parallel protein-unfolding pathways revealed and mapped. Nat Struct Biol 2003;10:658–62.
[17] Lindberg MO, Oliveberg M. Malleability of protein folding pathways: a simple reason for complex behaviour. Curr Opin Struct Biol 2007;17:21–9.
[18] Nickson AA, Wensley BG, Clarke J. Take home lessons from studies of related proteins. Curr Opin Struct Biol 2012;23:67–74.
[19] Mayor U, Johnson CM, Daggett V, Fersht AR. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. Proc Natl Acad Sci U S A 2000;97:13518–22.
[20] Gianni S, Guydosh NR, Khan F, Caldas TD, Mayor U, White GW, et al. Unifying features in protein-folding mechanisms. Proc Natl Acad Sci U S A 2003;100:13286–91.
[21] Mayor U, Grossmann JG, Foster NW, Freund SM, Fersht AR. The denatured state of Engrailed Homeodomain under denaturing and native conditions. J Mol Biol 2003;333:977–91.
[22] Mayor U, Guydosh NR, Johnson CM, Grossmann JG, Sato S, Jas GS, et al. The complete folding pathway of a protein from nanoseconds to microseconds. Nature 2003;421:863–7.
[23] Religa TL, Markson JS, Mayor U, Freund SM, Fersht AR. Solution structure of a protein denatured state and folding intermediate. Nature 2005;437:1053–6.
[24] White GWN, Gianni S, Grossmann JG, Jemth P, Fersht AR, Daggett V. Simulation and experiment conspire to reveal cryptic intermediates and a slide from nucleation–condensation to framework mechanism of folding. J Mol Biol 2005;350:757–75.
[25] Religa TL, Johnson CM, Vu DM, Brewer SH, Dyer RB, Fersht AR. The helix-turn-helix motif as an ultrafast independently folding domain: the pathway of folding of Engrailed homeodomain. Proc Natl Acad Sci U S A 2007;104:9272–7.
[26] Banachewicz W, Religa TL, Schaeffer RD, Daggett V, Fersht AR. Malleability of folding intermediates in the homeodomain superfamily. Proc Natl Acad Sci U S A 2011;108:5596–601.
[27] Ferguson N, Schartau PJ, Sharpe TJ, Sato S, Fersht AR. One-state downhill *versus* conventional protein folding. J Mol Biol 2004;344:295–301.
[28] Ferguson N, Sharpe TD, Schartau PJ, Sato S, Allen MD, Johnson CM, et al. Ultra-fast barrier-limited folding in the peripheral subunit-binding domain family. J Mol Biol 2005;353:427–46.
[29] Sharpe TD, Ferguson N, Johnson CM, Fersht AR. Conservation of transition state structure in fast folding peripheral subunit-binding domains. J Mol Biol 2008;383:224–37.
[30] Neuweiler H, Sharpe TD, Johnson CM, Fersht AR. Downhill *versus* barrier-limited folding of BBL 2: mechanistic insights from kinetics of folding monitored by independent Trp probes. J Mol Biol 2009;387:975–85.
[31] Gianni S, Gierhaas CD, Calosci N, Jemth P, Vuister GW, Travaglini-Allocatelli C, et al. A PDZ domain recapitulates a unifying mechanism for protein folding. Proc Natl Acad Sci U S A 2007;104:128–33.
[32] Calosci N, Chi CN, Richter B, Camilloni C, Engström A, Eklund L, et al. Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. Proc Natl Acad Sci U S A 2008;105:19241–6.
[33] Ivarsson Y, Travaglini-Allocatelli C, Brunori M, Gianni S. Engineered symmetric connectivity of secondary structure elements highlights malleability of protein folding pathways. J Am Chem Soc 2009;131:11727–33.
[34] Scott KA, Randles LG, Moran SJ, Daggett V, Clarke J. 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 2006;359:159–73.
[35] Wensley BG, Gärtner M, Choo W, Batey S, Clarke J. Different members of a simple three-helix bundle protein family have very different folding rate constants and fold by different mechanisms. J Mol Biol 2009;390:1074–85.
[36] Lappalainen I, Hurley MG, Clarke J. Plasticity within the obligatory folding nucleus of an immunoglobulin-like domain. J Mol Biol 2008;375:547–59.
[37] Viguera AR, Serrano L, Wilmanns M. Different folding transition states may result in the same native structure. Nat Struct Biol 1996;3:874–80.

[38] Pascual J, Pfuhl M, Rivas G, Pastore A, Saraste M. The spectrin repeat folds into a three-helix bundle in solution. FEBS Lett 1996;383:201–7.
[39] Pascual J, Pfuhl M, Walther D, Saraste M, Nilges M. Solution structure of the spectrin repeat: a left-handed antiparallel triple-helical coiled-coil. J Mol Biol 1997;273:740–51.
[40] Winograd E, Hume D, Branton D. Phasing the conformational unit of spectrin. Proc Natl Acad Sci U S A 1991;88:10788–91.
[41] Yan Y, Winograd E, Viel A, Cronin T, Harrison SC, Branton D. Crystal structure of the repetitive segments of spectrin. Science 1993;262:2027–30.
[42] Kusunoki H, Minasov G, Macdonald RI, Mondragon A. Independent movement, dimerization and stability of tandem repeats of chicken brain alpha-spectrin. J Mol Biol 2004;344: 495–511.
[43] Scott KA, Batey S, Hooton KA, Clarke J. The folding of spectrin domains I: wild-type domains have the same stability but very different kinetic properties. J Mol Biol 2004;344: 195–205.
[44] Wensley BG, Batey S, Bone FAC, Chan ZM, Tumelty NR, Steward A, et al. Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. Nature 2010;463:685–8.
[45] Borgia A, Wensley BG, Soranno A, Nettels D, Borgia M, Hoffman A, et al. Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single molecule fluorescence spectroscopy. Nat Commun 2012;3. http://dx.doi.org/10.1038/ncomms2204.
[46] Wensley BG, Kwa LG, Shammas SL, Rogers RM, Browning S, Yang Z, et al. Separating the effects of internal friction and transition state energy to explain the slow, frustrated folding of spectrin domains. Proc Natl Acad Sci U S A 2012;109: 17795–9.
[47] Wensley BG, Kwa LG, Shammas SL, Rogers RM, Clarke J. Protein folding: adding a nucleus to guide helix docking reduces landscape roughness. J Mol Biol 2012;423: 273–83.
[48] Fersht AR, Matouschek A, Serrano L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol 1992;224:771–82.
[49] Ansari A, Jones CM, Henry ER, Hofrichter J, Eaton WA. The role of solvent viscosity in the dynamics of protein conformational changes. Science 1992;256:1796–8.
[50] Cellmer T, Henry ER, Hofrichter J, Eaton WA. Measuring internal friction of an ultrafast-folding protein. Proc Natl Acad Sci U S A 2008;105:18320–5.
[51] Kramers HA. Brownian motion in a field of force and the diffusion model of chemical reactions. Physica 1940;7:284–304.
[52] Chrunyk BA, Mathews CR. Role of diffusion in the folding of the alpha subunit of tryptophan synthase from *Escherichia coli*. Biochemistry 1990;29:2149–54.
[53] Jacob M, Schindler T, Balbach J, Schmid FX. Diffusion control in an elementary protein folding reaction. Proc Natl Acad Sci U S A 1997;94:5622–7.
[54] Plaxco KW, Baker D. Limited internal friction in the rate-limiting step of a two-state protein folding reaction. Proc Natl Acad Sci U S A 1998;95:13591–6.
[55] Bhattacharyya RP, Sosnick TR. Viscosity dependence of the folding kinetics of a dimeric and monomeric coiled coil. Biochemistry 1999;38:2601–9.
[56] Pradeep L, Udgaonkar JB. Diffusional barrier in the unfolding of a small protein. J Mol Biol 2007;366:1016–28.
[57] Qiu LL, Hagen SJ. Internal friction in the ultrafast folding of the tryptophan cage. Chem Phys 2004;312:327–33.
[58] Jas GS, Eaton WA, Hofrichter J. Effect of viscosity on the kinetics of α-helix and β-hairpin formation. J Phys Chem B 2001;105:261–72.
[59] O'Neil KT, DeGrado WF. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. Science 1990;250:646–51.
[60] Matthews CR, Cristanti MM, Manz JT, Gepner GL. Effect of a single amino acid substitution on the folding of the alpha subunit of tryptophan synthase. Biochemistry 1983;22:1445–52.
[61] Ladurner AG, Itzhaki LS, Daggett V, Fersht AR. Synergy between simulation and experiment in describing the energy landscape of protein folding. Proc Natl Acad Sci U S A 1998;95:8473–8.
[62] Sato S, Fersht AR. Searching for multiple pathways of a nearly symmetrical protein: temperature dependent Φ-value analysis of the B domain of protein A. J Mol Biol 2007;372: 254–67.
[63] Hamill SJ, Steward A, Clarke J. The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. J Mol Biol 2000;297:165–78.
[64] Service RF. Problem solved* (*sort of). Science 2008;321: 784–6.