



# HHS Public Access

Author manuscript

*J Mol Biol.* Author manuscript; available in PMC 2022 May 09.

Published in final edited form as:

*J Mol Biol.* 2014 January 09; 426(1): 21–35. doi:10.1016/j.jmb.2013.08.027.

## Mechanism of Assembly of the Non-Covalent Spectrin Tetramerization Domain from Intrinsically Disordered Partners

Stephanie A. Hill<sup>1,2</sup>, Lee Gyan Kwa<sup>1</sup>, Sarah L. Shammis<sup>1</sup>, Jennifer C. Lee<sup>2</sup>, Jane Clarke<sup>1</sup>

<sup>1</sup>University of Cambridge Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, UK

<sup>2</sup>Laboratory of Molecular Biophysics, Biochemistry and Biophysics Center, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA

### Abstract

Interdomain interactions of spectrin are critical for maintenance of the erythrocyte cytoskeleton. In particular, “head-to-head” dimerization occurs when the intrinsically disordered C-terminal tail of  $\beta$ -spectrin binds the N-terminal tail of  $\alpha$ -spectrin, folding to form the “spectrin tetramer domain”. This non-covalent three-helix bundle domain is homologous in structure and sequence to previously studied spectrin domains. We find that this tetramer domain is surprisingly kinetically stable. Using a protein engineering  $\Phi$ -value analysis to probe the mechanism of formation of this tetramer domain, we infer that the domain folds by the docking of the intrinsically disordered  $\beta$ -spectrin tail onto the more structured  $\alpha$ -spectrin tail.

### Graphical Abstract

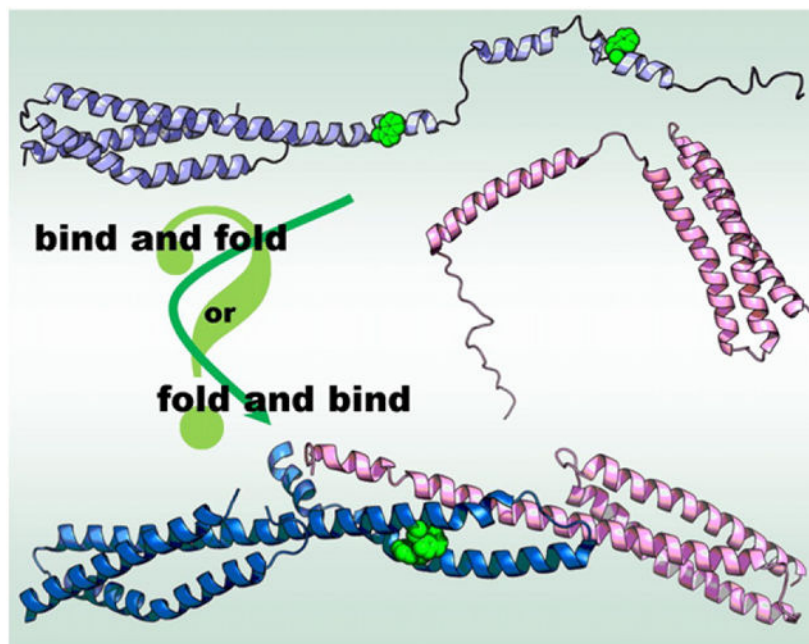
---

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Correspondence to Jane Clarke: jc162@cam.ac.uk.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.jmb.2013.08.027>.



**Legend:** Protein engineering methods may be used to investigate the mechanism of complex formation from disordered partners, discriminating binding-and-folding from folding-and-binding mechanisms. Simulated structure of human erythroid  $\beta$ -spectrin C-terminal domains previously published as Shamma, S. L., Rogers, J. M., Hill, S. A., & Clarke, J. (2012). Slow, reversible, coupled folding and binding of the spectrin tetramerization domain. *Biophys J* 103, 2203–14. N-terminal domains of human erythroid  $\alpha$ -spectrin PDB ID: 1OWA; erythroid spectrin tetramer domain PDB ID: 3LBX. Structures prepared in PyMol (DeLano Scientific).

## Keywords

protein folding; protein engineering;  $\Phi$ -value analysis; IDP; natively unfolded protein

## Introduction

Spectrin is an elongated structural protein common to all animal species and requisite for life [1,2]. Although non-structural functions of alternative isoforms in other tissues have now been reported [3–7], spectrin was originally known for its structural support properties in erythrocytes [8], where it forms extensive two-dimensional networks through interactions with itself and other proteins while binding elements of the three-dimensional cytoskeleton, such as ankyrin and adducin [2,9–11].

Spectrin networks have many strata of structure, described in Fig. 1. The basic unit of structure is the “spectrin domain”, a 106-residue, 5-nm-long three-helix bundle, with helices designated A, B, and C [12,13]. Spectrin domains are arranged as tandem repeating units with significant sequence homology, where Helix C of one domain is contiguous with Helix A of the next domain [14]. These tandem arrays of spectrin domains form two homologous multidomain proteins, subunits called  $\alpha$ -spectrin and  $\beta$ -spectrin. The two subunits found in

erythrocyte spectrin are of different lengths and compositions:  $\alpha$ -spectrin contains 20 full spectrin domains, the first of which is preceded by a partially helical N-terminal tail, while  $\beta$ -spectrin has only 16 full spectrin domains, followed by a C-terminal tail.  $\beta$ -Spectrin also has two additional CH domains at the N-terminus that bind actin, protein 4.1, adducin, and PIP<sub>2</sub> and function to assemble the spectrin network onto the erythrocyte membrane [2]. Single  $\alpha$ - and  $\beta$ -subunits associate laterally to form an antiparallel heterodimer, approximately 125 nm in length when extended and visible by electron microscopy [15]. Finally, two heterodimers associate laterally in what is termed a “head-to-head” manner to form a spectrin tetramer—the functional unit. Our study investigates the mechanism of formation of the interactions at this tetramer junction.

Two head-to-head interactions between the terminal tails of two heterodimers form the final stratum of structure, the erythrocyte spectrin tetramer. In each interaction, the N-terminal tail of  $\alpha$ -spectrin binds the C-terminal tail of  $\beta$ -spectrin from the other heterodimer, forming a new three-helix spectrin-like domain (Fig. 1). This non-covalent three-helix bundle is commonly known as the “spectrin tetramer domain” (even though it is, in fact, a heterodimer between just one  $\alpha$ -spectrin and one  $\beta$ -spectrin molecule) [16]. In the study of formation of this tetramer domain, truncated constructs that eliminate the possibility of heterodimer formation by removing the domains responsible for lateral interaction are frequently employed [17,18]. Since the adjacent complete spectrin domains,  $\alpha 1$  and  $\beta 16$ , are not involved in heterodimer formation [17,19,20], we use a truncated form of  $\alpha$ -spectrin that contains only the N-terminal tail and first full domain in  $\alpha 1$  ( $\alpha 0\alpha 1$ ) and the final full domain in  $\beta$ -spectrin with the C-terminal tail ( $\beta 16\beta 17$ ). Thus, we can study assembly of the tetramer domain in the absence of lateral interactions between heterodimers (Fig. 1b) [21,22]. We note that previous attempts to study this interaction using the unstructured tails alone have proved unsuccessful [23].

Two structures relevant to the erythrocyte spectrin tetramer exist: an NMR structure of  $\alpha 0\alpha 1$  [24] (PDB ID: 1OWA) and a crystal structure of the tetramer site [25] (PDB ID: 3LBX), which shows that the tails of each subunit fold together to form a new spectrin-like domain [26], as previously predicted from homology studies [27]. Comparing the structures of  $\alpha 0\alpha 1$  in isolation to the tetramer interaction, it is evident that the only significant structural change in  $\alpha 0\alpha 1$  is the formation of a linking helical region that joins Helix A in the first full domain to the helix in the tail region (Helix C of the tetramer) [28]. In contrast to the abundance of information regarding  $\alpha 0\alpha 1$ , little is known about the structure of  $\beta 16\beta 17$  in isolation. Circular dichroism (CD) has shown a significant increase in helicity upon formation of the tetramer interaction [22]; the majority of this increase can be attributed to increase in helical content of the  $\beta 17$  tail, as apparently only four additional residues become helical in  $\alpha 0\alpha 1$  upon tetramer formation. It is generally inferred that the C-terminal tail of  $\beta 16\beta 17$  is relatively unstructured in isolation.

Thus, the  $\beta 17$  tail has characteristics of an intrinsically disordered protein (IDP). IDPs feature an unstructured region that, in some cases, binds to structured proteins, adopting a folded structure (often  $\alpha$ -helical) in the process [29–31]. Mechanisms for IDPs associating with their partner protein are currently an area of significant interest, for which two extreme possible mechanisms have been proposed [32,33]. First, IDPs can bind their partner *via a*

conformational selection mechanism, where the IDP is fully folded before binding [34]. Alternatively, IDPs can bind to their partner in a fully disordered form and then fold using the partner protein as a scaffold [35]. Although not an example of an archetypal IDP, study of the mechanism of spectrin tetramer interaction may provide insights into IDP folding mechanisms more generally.

Interest in the spectrin tetramer has been driven by the identification of mutations in the spectrin tetramer region linked to hereditary elliptocytosis and spherocytosis, which are anemias characterized by anomalous erythrocyte morphology [36,37]. Previous investigations of the tetramer interaction have mainly used structural [38] and equilibrium [39] techniques, as well as molecular dynamics simulations [40]. Isothermal titration calorimetry (ITC) of the system has been performed by many groups, revealing a dissociation constant ( $K_d$ ) of about 0.4  $\mu\text{M}$  for wild type. ITC has also been performed on a small subset of mutants [41,42]. Two studies have investigated the kinetics of tetramer formation. A study utilizing surface plasmon resonance has reported the rate constant for association of the spectrin tetramer interaction ( $k_a$ ) to be 60  $\text{M}^{-1} \text{s}^{-1}$  [43]. More recently, our laboratory has developed a (solution-based) stopped-flow approach based on intrinsic tryptophan fluorescence, capable of extracting the association and dissociation rate constants for spectrin tetramerization [44]. We found the association rate constant (630  $\text{M}^{-1} \text{s}^{-1}$ ) to be an order of magnitude larger than that reported from the surface-based approach.

Protein engineering methods, particularly  $\Phi$ -value analysis, have played an invaluable role in elucidating folding mechanisms for a variety of proteins [45,46]. Of special relevance to this study are the previously published  $\Phi$ -value analyses of three individual spectrin domains from chicken brain  $\alpha$ -spectrin. These domains, called R15, R16, and R17, have high sequence and structural homology but fold at dramatically different rates. Through extensive mutation, it has been shown that R15, the fastest folder, features a strong nucleus during folding, while R16 and R17 lack such a nucleus and fold primarily through the docking of preformed elements of secondary structure [47–50]. Regardless of folding mechanism, the order of helix folding for each studied domain has proved to be the same: Helices A and C dock and fold first, followed by the docking of Helix B.

Here, we extend the  $\Phi$ -value analysis technique to study the mechanism of folding/assembly of the tetramer domain of spectrin.

## Results

### Design of the constructs

Four constructs were designed for this study:  $\beta 16$ ,  $\beta 16\beta 17$ ,  $\alpha 1$ , and  $\alpha 0\alpha 1$ . Domain boundary selection for the two  $\beta$ -spectrin constructs was guided by previous work by Nicolas *et al.* [23];  $\beta 16$  (residues 1898–2004) comprises the final full spectrin repeat in  $\beta$ -spectrin and  $\beta 16\beta 17$  (residues 1898–2083) is  $\beta 16$  plus the incomplete 17th domain, which has homology for one A- and one B-helix. The published NMR structure for the N-terminal end of  $\alpha$ -spectrin provided the domain boundaries for the first full spectrin domain, identifying this domain as S52-R156 [24]. However, since spectrin domains are defined as having 106 residues [51], we infer that the true domain C-terminal boundary is

L158. In addition, MacDonald and Pozharski have shown that additional (native) residues at both N- and C-terminal ends of single spectrin domains significantly improve domain stability [51]. Accordingly, we add four residues to the beginning of the domain, and five to the end, yielding our  $\alpha 1$  construct as residues 48–163. The helix in the first (incomplete) domain of  $\alpha$ -spectrin begins at residue 19; however, a construct containing residues 19–163 ( $\alpha 0\alpha 1'$ ) does not bind  $\beta 16\beta 17$  to yield the tetramer interaction (data not shown). Given this, we include much of the unstructured N-terminal tail, selecting residues 2–163 for construct  $\alpha 0\alpha 1$ .

Two residue identification systems are in common use to describe spectrin: R15, R16, and R17 are numbered individually as domains, while the human spectrin proteins are typically numbered in reference to the entire protein. With the use of domain-wise numbering, the tetramer covers residues 1–108, with residues 1–77 on  $\beta 16\beta 17$  and residues 79–108 on  $\alpha 0\alpha 1$ . Alternatively, with the use of protein-specific numbering, the  $\beta 16\beta 17$  contribution to the tetramer is residues 2004–2079;  $\alpha 0\alpha 1$  contributes residues 19–49. Unless otherwise noted, the domain-wise numbering scheme is used throughout the text, in order to facilitate comparison with previous  $\Phi$ -value analyses of spectrin.

### Wild-type spectrin

CD spectra of the extended constructs  $\alpha 0\alpha 1$  and  $\beta 16\beta 17$  were compared to those of the folded domains  $\alpha 1$  and  $\beta 16$  (Fig. 2). From the data, it is clear that all four constructs are predominantly  $\alpha$ -helical. Addition of the N-terminal tail to  $\alpha 1$  increases helical content, as one would expect, given the NMR structure of  $\alpha 0\alpha 1$ , which shows an isolated helix from residue 19 to residue 46 (protein-specific numbering). Further, there may be some helix fraying in Helix A in  $\alpha 1$ , which may be alleviated by the addition of residues at the domain's N-terminal boundary. With the use of the signal at 222 nm as a guide, the increase in helicity seen in  $\alpha 0\alpha 1$  does not appear great enough to indicate complete formation of helix in the  $\alpha 0$  domain despite  $\alpha 0$  looking fully structured by NMR techniques [24]. Similarly,  $\beta 16\beta 17$  has more helix than  $\beta 16$ , perhaps by around as much as a single helix. This implies some degree of native helix in the  $\beta 17$  domain, which is as of yet unstudied by NMR. Our results cannot identify the location of this helicity, and it may be residual and spread across the tail. This would be consistent with results of equilibrium unfolding curves performed by CD, which showed no cooperative unfolding of  $\alpha 0$  and  $\beta 17$  regions (data not shown). Although we did not include extra residues on the C-terminal end of  $\beta 16$ , it is unlikely that the C-helix in the  $\beta 16$  construct is significantly frayed, given the results of equilibrium and kinetic experiments, as discussed below.

The monomeric folding and unfolding of the wild-type constructs were studied by chemical denaturation experiments, utilizing intrinsic tryptophan fluorescence to monitor folding. Equilibrium results for the constructs indicate that both fold reversibly in a two-state manner (Fig. 3 and Table 1). We have previously observed that spectrin domains may be stabilized by unfolded neighboring domains [52]. Interestingly,  $\alpha 1$  is stabilized by the addition of the  $\alpha 0$  tail. The  $m$ -value of  $\alpha 0$  alone is also lower than that of  $\alpha 0\alpha 1$  and other previously studied spectrin domains (Table 1). This suggests that our  $\alpha 1$  construct is cut “too short” in the construct and, consequently, the A-helix is frayed. Notably,  $\alpha 1$  in the  $\alpha 1\alpha 0$  construct

has similar stability to R15, R16, and R17 [48,51]; however,  $\beta 16$  is extremely stable, with a free energy of unfolding ( $G_{D-N}$ ) of 14.2 kcal mol<sup>-1</sup> and a midpoint of >4 M guanidinium chloride (GdmCl). This stability is essentially unaffected by addition of the  $\beta 17$  tail. (Note that the equilibrium  $m$ -value of  $\beta 16$  is similar to that of R15, R16, and R17 in GdmCl.) Folding and unfolding kinetic data were well fitted by single-exponential functions. Plotting the natural logarithm of the observed rate constant,  $k_{obs}$ , against denaturant concentration yields typical chevron plots; the refolding and unfolding limbs for both  $\alpha 1$  and  $\beta 16$  are linear within the experimentally accessible range of denaturant. It can be seen that both fold as rapidly as R15; the stability of  $\beta 16$  arises from the decreased unfolding rate constant.

### $\Phi$ -Value analysis: 1. Choice of mutations

The selection of residues to mutate was guided by the residues studied in the  $\Phi$ -value analyses of R15, R16, and R17 [47,49,50] and the crystal structure of the tetramer interaction. Mutants were created in each of the three helices of the tetramer domain, at both core and solvent-exposed residues. Mutations of buried residues were made to delete interactions in the core. Each residue was mutated to Ala, except native Ala residues, which were mutated to Gly. Additionally, aromatic residues were mutated to Leu. Surface residues were mutated to both Ala and Gly in order to investigate the extent of secondary structure formation at the transition state. In all, 23 mutations were made to probe 14 sites in  $\alpha 0\alpha 1$ , and 40 mutations were made to probe 27 sites in  $\beta 16\beta 17$ . CD spectra were collected for each mutant to assess the effect of the mutations on secondary structure of the monomer. Only L91A, Y94A, and L108A significantly reduced the helicity of  $\alpha 0\alpha 1$  (Fig. S1);  $\Phi$ -values were not calculated for either L91A or L108A since L91A had an unusual  $m$ -value and L108A abolished formation of the tetramer domain. Y94A yielded a non-classical  $\Phi$ -value (see Discussion).

### $\Phi$ -Value analysis: 2. Collection of data

Mutant proteins were exclusively studied binding to their wild-type partner, never pair-wise with a mutant partner. Equilibrium data were collected by ITC, following the heat released as  $\alpha 0\alpha 1$  binds  $\beta 16\beta 17$  to form the tetramer domain. Fitting of these data yields association constants ( $K_a$  values), where  $K_a = \frac{1}{K_d}$  [44]. Eight mutations effectively abolished binding (Tables 2 and 3); the  $K_d$  for wild type was 0.4  $\mu$ M, and  $K_d$  values for mutant tetramer domains ranged from 0.15 to 170  $\mu$ M. Changes in free energy of dissociation upon mutation,  $G_{D-N}$ , are reported in Tables 2 and 3, with sample ITC traces in Fig. S2.

Kinetic data were collected with stopped-flow experiments, following the increase of intrinsic Trp fluorescence observed upon dissociation/unfolding of the preformed tetramer domain in the presence of urea.

### $\Phi$ -Value analysis: 3. Analysis of data

Although one might expect dissociation to be well represented by single-exponential kinetics, we have demonstrated previously that a second-order reversible model is needed to accurately represent the disassociation data despite the relatively weak association between  $\alpha 0\alpha 1$  and  $\beta 16\beta 17$  ( $K_a = 0.4 \mu$ M). All data were well fitted by this model [44].

$$F = F_0 + \Delta F \times \frac{b - z - (b + z) \left( \frac{2[AB]_0 + b - z}{2[AB]_0 + b + z} \right) \exp(zk_+t)}{2 \left( \left( \frac{2[AB]_0 + b - z}{2[AB]_0 + b + z} \right) \exp(zk_+t) - 1 \right)}$$

where  $b = -K_d - 2[A]_t$  and  $z = \sqrt{K_d^2 + 4K_d[A]_t}$ .  $F_0$  is the fluorescence at time zero,  $F$  is the fluorescence at time  $t$ , and  $k_+$  is the association rate constant.  $[AB]_0$ , the concentration of dimer in the reaction mixture at time zero, is calculated from the  $K_d$  obtained from ITC experiments and the initial concentration of  $\alpha 0\alpha 1$  ( $[A]_t$ ) in the protein solution prior to mixing with denaturant.

$$2[AB]_0 = \frac{2[A]_t + K_d - \sqrt{(K_d + 2[A]_t)^2 - 4[A]_t^2}}{2}$$

Dissociation rate constants for each mutant were plotted as a natural logarithm against the final urea concentration, yielding linear unfolding limbs for each mutant (Fig. 4). With only two exceptions, the unfolding limbs had very similar slopes; thus, these were fitted globally, yielding a shared  $m$ -value for disassociation,  $m_{k_-}$ , of  $1.20 \pm 0.02 \text{ M}^{-1}$ . This is similar to estimates of  $m_{k_-}$  from association and dissociation studies of the wild type reported previously of  $1.25 \pm 0.07 \text{ M}^{-1}$  and  $1.44 \pm 0.07 \text{ M}^{-1}$ , respectively [44]. The  $y$ -intercept of each fit represents the natural logarithm of the unfolding rate in the absence of denaturant,  $\ln(k_{-}^{\text{H}_2\text{O}})$ .

From these unfolding data, it is possible to calculate  $\Phi$  following the method of Fersht *et al.* [53].

$$\Phi = 1 - \frac{-RT \ln \left( \frac{k_{-}^{\text{WT, H}_2\text{O}}}{k_{-}^{\text{mut, H}_2\text{O}}} \right)}{\Delta\Delta G_{\text{D-N}}}$$

The calculated  $\Phi$ -values and dissociation rate constants ( $k_{-}$ ) are shown for the core and surface mutations in Tables 2 and 3, respectively.

## Discussion

### Characterization of spectrin domains: $\alpha 0\alpha 1$ , $\beta 16\beta 17$ , and the tetramer domain

In this work, we have characterized three new spectrin domains:  $\alpha 1$ ,  $\beta 16$ , and the tetramer domain. Of these,  $\alpha 1$  behaves most similarly to the previously described R15, R16, and R17. As shown in Table 1, the thermodynamic stability of  $\alpha 1$  ( $4.3 \text{ kcal mol}^{-1}$  or  $6.5 \text{ kcal mol}^{-1}$  when extended in  $\alpha 0\alpha 1$ ) fits well with the range of  $G_{\text{D-N}}$  for the three chicken brain spectrins. Further, visual inspection of the chevron curves in Fig. 2c reveals that the kinetic

behavior of  $\alpha 0\alpha 1$  is most similar to R15, a protein that folds by a nucleation mechanism with minimal landscape ruggedness.

$\beta 16$  has behavior that is distinct from the previously studied spectrin domains. First,  $\beta 16$  is not observed to unfold in the presence of urea, requiring 4 M GdmCl to unfold half of the population. Indeed,  $\beta 16\beta 17$  is twice as stable as  $\alpha 1$ , R15, R16, and R17, in terms of  $G_{D-N}$  (Table 1). This remarkable stability has been previously observed [54] and may have a stabilizing effect on the neighboring tetramer domain, as has been the case for other spectrin domain pairs [55]. As seen in Fig. 2d, projection of the refolding limb of  $\beta 16$  suggests that, in the absence of denaturant, the protein folds as rapidly as  $\alpha 1$  or R15, and projection of the unfolding limb of  $\beta 16$  reveals that the unfolding rate in water for the protein is much slower than anything previously seen for other members of the spectrin family.

As it is composed of two chains, the stability of the spectrin tetramer domain cannot be directly compared to R15, R16, and R17 since the equilibrium between associated and dissociated forms is concentration dependent. The second-order association/folding rate is also concentration dependent ( $k_+$  has units of  $M^{-1} s^{-1}$ ), and hence,  $k_+$  cannot be compared with folding rate constants for other spectrin domains (where  $k_f$  has units of  $s^{-1}$ ). However, the first-order disassociation/unfolding rate constant,  $k_-$ , has units of  $s^{-1}$  and thus *can* be directly compared to the unfolding rate constants ( $k_u$ , units of  $s^{-1}$ ) of the other spectrin domains. The non-covalent spectrin domain unfolds remarkably slowly, like the slow unfolding R17. With a  $k_-$  about three orders of magnitude lower than the unfolding rate of  $\alpha 1$ , the tetramerization domain is actually significantly more kinetically stable than its direct (covalent) neighbor. It is interesting to note that the dependence of  $\ln(k_-)$  on urea concentration (the dissociation/unfolding  $m$ -value,  $m_{k_-}$ ) is very similar to the other spectrin domains (Fig. 3c and Table 1), suggesting that the transition state for unfolding of the tetramer domain is similarly expanded relative to the native state. We note that we have previously reported the dependence of the association rate constant,  $k_+$ , on urea concentration ( $m_{k_+}$ ). As expected, this has a weaker concentration dependence than for the other spectrin domains (Table 1) since  $\alpha 0$  contains significant residual structure in its dissociated state, while the covalent spectrin domains have essentially disordered denatured states [56].

### **$\Phi$ -Value analysis reveals transition state structure and suggests a folding mechanism**

We have previously found the association of the spectrin tetramer to be consistent with a two-state mechanism for association/dissociation, without populated intermediates; rate constants estimated from kinetic data collected at various subunit concentrations were consistent with each other, and equilibrium and kinetic estimates of complex stability were in agreement [44]. Applying this model allows us to infer the structure of the transition state from either folding or unfolding kinetics. The vast majority of the mutants (with few exceptions, see below) had the same  $m_{k_-}$  as the wild-type tetramer interaction, suggesting that these mutations did not alter the folding pathway. Some mutations effectively abolish binding, while others do not change the free energy of association sufficiently to allow us to determine a  $\Phi$ -value with confidence ( $G_{D-N} < 0.7 \text{ kcal mol}^{-1}$ ). Given this, we were able to determine  $\Phi$ -values for 26 positions in the tetramer domain. A  $\Phi$ -value of



1 indicates that the residue is in a region of the protein that is fully structured at the transition state; a  $\Phi$ -value of 0 indicates that the region of the protein is as unstructured in the transition state as it is in the starting (denatured, unbound) state. As is standard practice, we classify  $\Phi$ -values as high (largely structured), medium (partly structured), or low (largely unstructured), to give a qualitative picture of the structure of the transition state. Only three  $\Phi$ -values (mutants Y94A, Y94L, and F97A) were non-classical, that is, significantly greater than 1 or less than 0.

For the A-helix, we had good coverage: we determined ten  $\Phi$ -values for nine buried residues (which monitor tertiary structure formation) and four surface Ala-to-Gly  $\Phi$ -values (reporting on helix formation). The secondary structure  $\Phi$ -values are higher than the tertiary  $\Phi$ -values, and the  $\Phi$ -values are higher in the middle/C-terminal end of the helix than at the N-terminal end (Fig. 5a).

In Helix B, four buried and two surface residue mutations in this helix abolished binding, while two others had mutations with  $G_{D-N} < 0.7$  kcal mol<sup>-1</sup>. Thus, we could calculate only three surface and three buried  $\Phi$ -values in this helix. Nevertheless, we can say that there is an indication of more structure formation at the N-terminal half of the B-helix (high  $\Phi$ -values, 0.5–0.6 kcal mol<sup>-1</sup>) than for the C-terminal half, where  $\Phi$ -values are close to 0.

We determined six  $\Phi$ -values in the C-helix, three buried and three solvent exposed. The tertiary  $\Phi$ -values were, on average, higher than those in the Helices A and B and were consistently high along the helix. Helix C, which is apparently significantly structured in the unbound state, contained two residues whose mutation abolished binding, as well as two non-standard  $\Phi$ -values (Fig. 6). We note that both R87A and L108A abolished formation of the tetramer domain and that physiological mutations leading to red blood cell deformation have been attributed to both of these residues, suggesting the critical importance of these residues in the mechanism of folding [57–60]. Buried residue Y94 has  $\Phi$ -values of  $-2$  (Y94A) and  $+4$  (Y94L). F97A, in the next turn of helix, had a  $\Phi$ -value of  $-0.3$ . L91A, for which we did not determine a  $\Phi$ -value (low  $G_{D-N}$ ), had a slope that was inconsistent with the other mutants (Fig. S3). Each of these residues contacts the region of higher  $\Phi$ -values from Helix A. Taken together, we propose that this interaction is important for the assembly of the tetramer domain and that mutation of this region of the preformed C-helix disrupts assembly.

We mapped the  $\Phi$ -values onto the structure of the tetramer domain (Fig. 5b). It is clear that the two regions of higher  $\Phi$ -values in Helix A and Helix B do not contact each other; thus, at the transition state, Helices A and B do not, apparently, dock. However, we see that the two elements of structure from  $\beta$ 17 come into contact with the highly structured C-helix at the transition state. The middle-to-C-terminal part of the A-helix interacts with the N-terminal-to-middle part of Helix C, while the N-terminal end of the B-helix contacts the C-terminal end of the C-helix. Interestingly, three of the mutations in Helix B that disrupt structure formation are found in the N-terminal-to-middle region, which we propose binds Helix C during the transition state (Fig. 7).

These data are consistent with a mechanism whereby the preformed Helix C provides a template onto which Helices A and B dock (Fig. 7). There is some evidence that residual, possibly transient, helical structure in Helix A and/or B may exist, to which the C-helix may bind, thus stabilizing this residual structure. The docking of portions of Helices A and B onto Helix C establishes the correct topology and register of the helices, bringing the two Trp residues at the center of the core into contact (one from Helix A and one from Helix B), thus allowing folding to proceed efficiently.

## Conclusion

We have previously shown that spectrin domains can fold *via* nucleation condensation (R15) or by docking of partially preformed helices of A and B (R16 and R17). Here, we see yet another folding mechanism. The  $\Phi$ -value analysis of the spectrin tetramer domain suggests a templating mechanism wherein transient elements of secondary structure in  $\beta$ 17 dock onto the more highly structured C-helix. Our application of biophysical techniques and  $\Phi$ -value analysis to the spectrin tetramer domain clearly demonstrates that enhanced understanding of IDP folding mechanisms can be achieved through the use of such methods, provided that care and stringent controls are used.

## Materials and methods

### Buffer and general conditions

All experiments were carried out in PBS [50 mM sodium phosphate and 150 mM NaCl (pH 7.04)]. The temperature for each experiment was held constant at 25 °C.

### Protein expression and purification

Wild-type  $\alpha$ 1,  $\beta$ 16,  $\alpha$ 0 $\alpha$ 1, and  $\beta$ 16 $\beta$ 17 were expressed from synthetic genes (GenScript USA) in a modified pRSETA plasmid. All mutagenesis was carried out using a Stratagene QuikChange kit; identity of mutants was confirmed by DNA sequencing.

Protein expression and purification was carried out analogously to other spectrin domains [48]. The protein was expressed in *Escherichia coli* C41(DE3) cultured in 2 $\times$  TY media at 37 °C in the presence of 0.1 mg ml<sup>-1</sup> ampicillin. Cells were induced with 0.1 mg ml<sup>-1</sup> IPTG, and the temperature was reduced to 25 °C once culture optical density at 600 nm reached 0.4–0.6 AU. After overnight incubation, cells were collected by centrifugation, then lysed by sonication and centrifuged again to separate the soluble protein from insoluble cell debris. The protein of interest was purified from the soluble fraction by affinity chromatography on Ni<sup>+2</sup> agarose resin. Bound protein was released from the resin by thrombin cleavage and further purified with gel-filtration chromatography. Proteins were frozen in liquid N<sub>2</sub> and stored for long periods at –80 °C. Once thawed for use, proteins were stored at 4 °C. Typical yield for  $\alpha$ 0 $\alpha$ 1 and  $\beta$ 16 $\beta$ 17 is 50 and 35 mg l<sup>-1</sup> culture, respectively.

### Circular dichroism

Protein samples were prepared at 10  $\mu$ M in PBS buffer. Spectra were collected in a 4-mm-pathlength or 1-mm-pathlength cuvette as follows: core residues, Chirascan CD

Spectrometer (Applied Photophysics); Scan 250–208 nm, 1 nm bandwidth; surface residues and wild-type constructs, CD Spectropolarimeter (JASCO); Scan 260–200 nm, 0.5 nm bandwidth. Minor differences in spectra obtained from the two instruments were observed; however, no significant changes were observed between mutants.

### Stability of individual subunits

The stability of individual wild-type subunits was determined using urea- or GdmCl-induced denaturation. After excitation at 280 nm, intrinsic tryptophan fluorescence was collected from 300 to 450 nm on a PerkinElmer LS-55 fluorescence spectrophotometer. Protein concentration for these experiments was 5  $\mu$ M; temperature was 25 °C.

### Kinetics of individual subunits

The folding and unfolding kinetics of the individual wild-type proteins were monitored by change in fluorescence, using an Applied Photophysics SX.20 stopped-flow spectrophotometer at 25 °C. Trp residues were excited with 280 nm light, and fluorescence data were collected above 320 nm.  $\alpha$ 0 $\alpha$ .1 and  $\beta$ 16 $\beta$ 17 were unfolded in urea and GdmCl, respectively, at a final protein concentration of 1  $\mu$ M. Six traces were averaged for each concentration of denaturant and fitted to a single-exponential model, using KaleidaGraph (Synergy Software). In keeping with previous work using this instrument, we removed the first 2.5 ms prior to fitting to avoid mixing artifacts [47].

### Stability of the tetramer interaction

ITC experiments to determine the stability of the tetramer interaction were carried out on either an ITC200-Auto (core residues) or a VP-ITC (surface residues), both from MicroCal (GE).  $\beta$ 16 $\beta$ 17 at 25–50  $\mu$ M in the cell was titrated with 15–20 injections to a total of 2–2.5 eq of  $\alpha$ 0 $\alpha$ .1, at 300 or 600 s intervals for core and surface residues, respectively. Each injection profile was integrated, and the resulting data were fitted with a one-site binding model in Origin 7 for MicroCal (OriginLab).

### Kinetics of the tetramer interaction

Folding and unfolding kinetics were monitored by intrinsic tryptophan fluorescence on either an Applied Photophysics stopped-flow spectrophotometer model SX.18 or SX.20 or a Biologic rapid mixing device. For both instruments, Trp residues were excited with 280 nm light and fluorescence was collected above 320 nm. A solution of pre-mixed  $\alpha$ 0 $\alpha$ .1 and  $\beta$ 16 $\beta$ 17, equilibrated 2 h, was mixed 1:1 with a urea solution. For all mutant unfolding data, the initial protein solution was 40  $\mu$ M in each subunit, resulting in a final concentration in the cell of 20  $\mu$ M in each subunit. Experiments were conducted at 25 °C, and traces were fitted as described in Results, using a fitting script in MATLAB (MathWorks). Unfolding rates were plotted against denaturant concentration and fitted globally using Igor (WaveMetrics).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We gratefully acknowledge B.G. Wensley, K.A. Scott, and S. Batey for use of data describing R15, R16, and R17. We thank A. Steward for technical assistance and Christopher Johnson of the Laboratory of Molecular Biology, Cambridge, for helpful discussion and the use of instrumentation. A portion of the data was collected in the Biophysics Core Facility, National Heart, Lung, and Blood Institute, National Institutes of Health (NIH). This work was supported by the Intramural Research Program at the NIH, National Heart, Lung, and Blood Institute, and the Wellcome Trust (grant number WT095195). S.A.H. is supported by a NIH-Cambridge Partnership fellowship and the Cambridge Overseas Trust. J.C. is a Wellcome Trust Senior Research Fellow.

## Biographies



Stephanie A. Hill



Lee Gyan Kwa



Sarah L. Shammass

## Abbreviations used:

<b>IDP</b>	intrinsically disordered protein
<b>GdmCl</b>	guanidinium chloride
<b>ITC</b>	isothermal titration calorimetry
<b>NIH</b>	National Institutes of Health

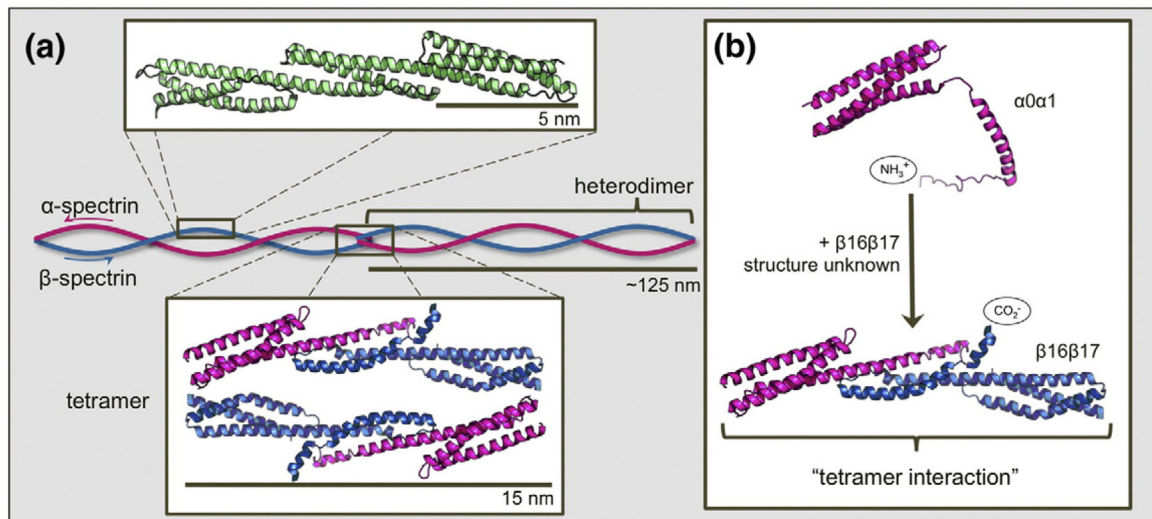
## References

- [1]. Salomao M, An XL, Guo XH, Gratzner WB, Mohandas N, Baines AJ. Mammalian alpha I-spectrin is a neofunctionalized polypeptide adapted to small highly deformable erythrocytes. *Proc Natl Acad Sci USA* 2006;103:643–8. [PubMed: 16407147]

- [2]. Baines AJ. Evolution of spectrin function in cytoskeletal and membrane networks. *Biochem Soc Trans* 2009;37:796–803. [PubMed: 19614597]
- [3]. Sridharan DM, McMahon LW, Lambert MW. alpha II-Spectrin interacts with five groups of functionally important proteins in the nucleus. *Cell Biol Int* 2006;30:866–78. [PubMed: 16889989]
- [4]. Gascard P, Mohandas N. New insights into functions of erythroid proteins in nonerythroid cells. *Curr Opin Hematol* 2000;7:123–9. [PubMed: 10698300]
- [5]. Young KG, Kothary R. Spectrin repeat proteins in the nucleus. *BioEssays* 2005;27:144–52. [PubMed: 15666356]
- [6]. Ayala-Grosso C, Tam J, Roy S, Xanthoudakis S, Da Costa D, Nicholson DW, et al. Caspase-3 cleaved spectrin colocalizes with neurofilament-immunoreactive neurons in Alzheimer's disease. *Neuroscience* 2006;141:863–74. [PubMed: 16750894]
- [7]. Voas MG, Lyons DA, Naylor SG, Arana N, Rasband MN, Talbot WS. alpha II-spectrin is essential for assembly of the nodes of Ranvier in myelinated axons. *Curr Biol* 2007;17:562–8. [PubMed: 17331725]
- [8]. Tillack TW, Marchesi SL, Marchesi VT, Steers E. A comparative study of spectrin. A protein isolated from red blood cell membranes. *Biochim Biophys Acta* 1970;200:125–31. [PubMed: 4983327]
- [9]. Bennett V, Chen LS. Ankyrins and cellular targeting of diverse membrane proteins to physiological sites. *Curr Opin Cell Biol* 2001;13:61–7. [PubMed: 11163135]
- [10]. van den Akker E, Satchwell TJ, Williamson RC, Toye AM. Band 3 multiprotein complexes in the red cell membrane; of mice and men. *Blood Cells Mol Dis* 2010;45:1–8. [PubMed: 20346715]
- [11]. Baines AJ. Evolution of the spectrin-based membrane skeleton. *Transfus Clin Biol* 2010;17:95–103. [PubMed: 20688550]
- [12]. Speicher DW, Marchesi VT. Erythrocyte spectrin is comprised of many homologous triple helical segments. *Nature* 1984;311:177–80. [PubMed: 6472478]
- [13]. 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. [PubMed: 9356261]
- [14]. Grum VL, Li DN, MacDonald RI, Mondragon A. Structures of two repeats of spectrin suggest models of flexibility. *Cell* 1999;98:523–35. [PubMed: 10481916]
- [15]. Sahr KE, Coetzer TL, Moy LS, Derick LH, Chishti AH, Jarolim P, et al. Spectrin cagliari. an Ala → Gly substitution in helix 1 of beta spectrin repeat 17 that severely disrupts the structure and self-association of the erythrocyte spectrin heterodimer. *J Biol Chem* 1993;268:22656–62. [PubMed: 8226774]
- [16]. Speicher DW, Desilva TM, Speicher KD, Ursitti JA, Hembach P, Weglarz L. Location of the human red-cell spectrin tetramer binding-site and detection of a related closed hairpin loop dimer using proteolytic footprinting. *J Biol Chem* 1993;268:4227–35. [PubMed: 8440706]
- [17]. Desilva TM, Peng KC, Speicher KD, Speicher DW. Analysis of human red-cell spectrin tetramer (head-to-head) assembly using complementary univalent peptides. *Biochemistry* 1992;31:10872–8. [PubMed: 1420200]
- [18]. Mehboob S, Luo BH, Patel BM, Fung LWM. alpha beta spectrin coiled coil association at the tetramerization site. *Biochemistry* 2001;40:12457–64. [PubMed: 11591167]
- [19]. Ursitti JA, Kotula L, DeSilva TM, Curtis PJ, Speicher DW. Mapping the human erythrocyte beta-spectrin dimer initiation site using recombinant peptides and correlation of its phasing with the alpha-actinin dimer site. *J Biol Chem* 1996;271:6636–44. [PubMed: 8636080]
- [20]. Begg GE, Harper SL, Morris MB, Speicher DW. Initiation of spectrin dimerization involves complementary electrostatic interactions between paired triple-helical bundles. *J Biol Chem* 2000;275:3279–87. [PubMed: 10652315]
- [21]. Luo BH, Mehboob S, Hurtuk MG, Pipalia NH, Fung LWM. Important region in the beta-spectrin C-terminus for spectrin tetramer formation. *Eur J Haematol* 2002;68:73–9. [PubMed: 12038451]
- [22]. Long F, McElheny D, Jiang SK, Park S, Caffrey MS, Fung LWM. Conformational change of erythroid alpha-spectrin at the tetramerization site upon binding beta-spectrin. *Protein Sci* 2007;16:2519–30. [PubMed: 17905835]

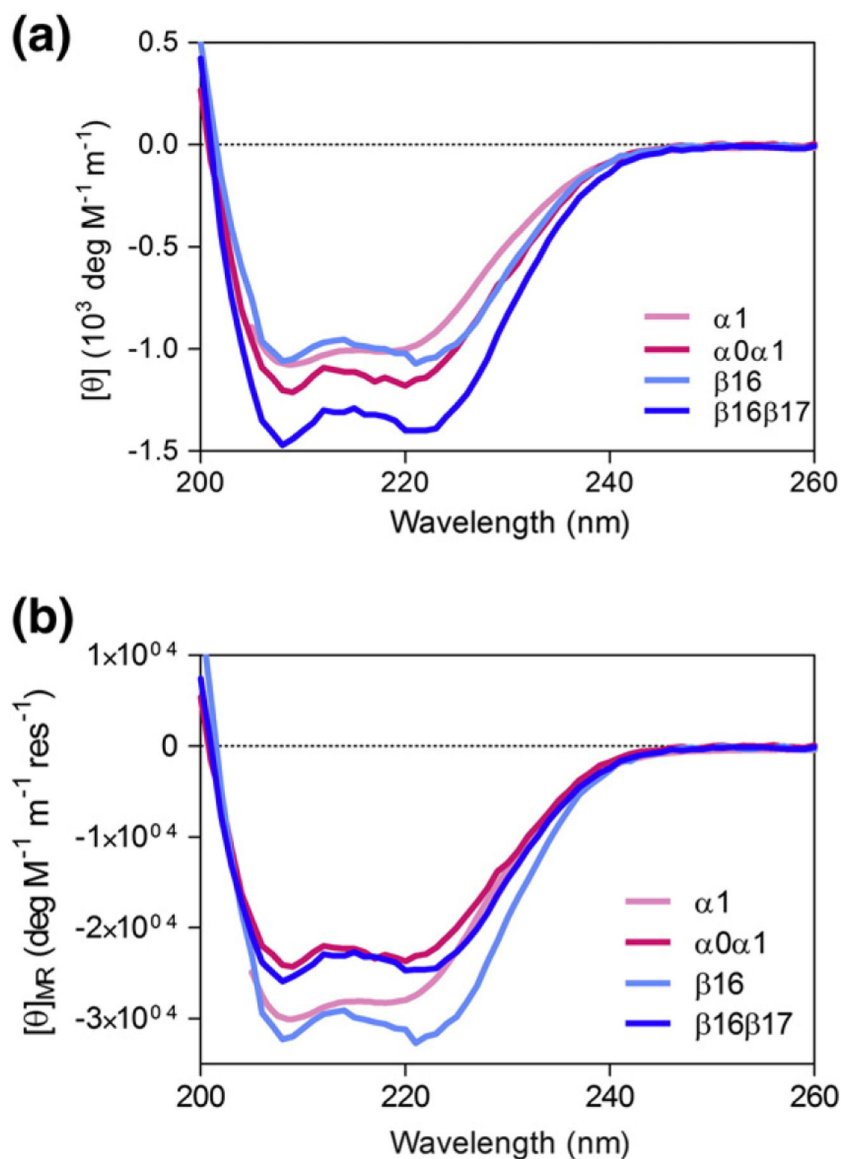
- [23]. Nicolas G, Pedroni S, Fournier C, Gautero H, Craescu C, Dhermy D, et al. Spectrin self-association site: characterization and study of beta-spectrin mutations associated with hereditary elliptocytosis. *Biochem J* 1998;332:81–9. [PubMed: 9576854]
- [24]. Park S, Caffrey MS, Johnson ME, Fung LWM. Solution structural studies on human erythrocyte alpha-spectrin tetramerization site. *J Biol Chem* 2003;278:21837–44. [PubMed: 12672815]
- [25]. Ipsaro JJ, Harper SL, Messick TE, Marmorstein R, Mondragon A, Speicher DW. Crystal structure and functional interpretation of the erythrocyte spectrin tetramerization domain complex. *Blood* 2010;115:4843–52. [PubMed: 20197550]
- [26]. Mehboob S, Song Y, Witek M, Long F, Santarsiero BD, Johnson ME, et al. Crystal structure of the nonerythroid alpha-spectrin tetramerization site reveals differences between erythroid and nonerythroid spectrin tetramer formation. *J Biol Chem* 2010;285:14572–84. [PubMed: 20228407]
- [27]. Zhang ZS, Weed SA, Gallagher PG, Morrow JS. Dynamic molecular modeling pathogenic mutations spectrin self-association domain. *Blood* 2001;98:1645–53. [PubMed: 11535493]
- [28]. Antoniou C, Lam VQ, Fung LWM. Conformational changes at the tetramerization site of erythroid alpha-spectrin upon binding beta-spectrin: a spin label EPR study. *Biochemistry* 2008;47:10765–72. [PubMed: 18783249]
- [29]. Uversky VN, Dunker AK. Understanding protein non-folding. *Biochim Biophys Acta Proteins Proteomics* 2010;1804:1231–64.
- [30]. Turoverov KK, Kuznetsova IM, Uversky VN. The protein kingdom extended: ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation. *Prog Biophys Mol Biol* 2010;102:73–84. [PubMed: 20097220]
- [31]. Teilum K, Olsen JG, Kragelund BB. Functional aspects of protein flexibility. *Cell Mol Life Sci* 2009;66:2231–47. [PubMed: 19308324]
- [32]. Wright PE, Dyson HJ. Linking folding and binding. *Curr Opin Struct Biol* 2009;19:31–8. [PubMed: 19157855]
- [33]. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* 2002;12:54–60. [PubMed: 11839490]
- [34]. Kjaergaard M, Teilum K, Poulsen FM. Conformational selection in the molten globule state of the nuclear co-activator binding domain of CBP. *Proc Natl Acad Sci USA* 2010;107:12535–40. [PubMed: 20616042]
- [35]. Oldfield CJ, Cheng YG, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* 2005;44:12454–70. [PubMed: 16156658]
- [36]. Maillet P, Alloisio N, Morle L, Delaunay J. Spectrin mutations in hereditary elliptocytosis and hereditary spherocytosis. *Hum Mutat* 1996;8:97–107. [PubMed: 8844207]
- [37]. Morle L, Morle F, Roux AF, Godet J, Forget BG, Denoroy L, et al. Spectrin Tunis (Sp-alpha-I/78), an elliptocytogenic variant, is due to the CGG-TGG codon change (Arg-Tryp) at position-35 of the alpha-I domain. *Blood* 1989;74:828–32. [PubMed: 2568861]
- [38]. Mehboob S, Jacob J, May M, Kotula L, Thiagarajan P, Johnson ME, et al. Structural analysis of the alpha N-terminal region of erythroid and nonerythroid spectrins by small-angle X-ray scattering. *Biochemistry* 2003;42:14702–10. [PubMed: 14661984]
- [39]. Li DH, Harper SL, Tang HY, Maksimova Y, Gallagher PG, Speicher DW. A comprehensive model of the spectrin divalent tetramer binding region deduced using homology modeling and chemical cross-linking of a mini-spectrin. *J Biol Chem* 2010;285:29535–45. [PubMed: 20610390]
- [40]. Song YL, Pipalia NH, Fung LWM. The L49F mutation in alpha erythroid spectrin induces local disorder in the tetramer association region: fluorescence and molecular dynamics studies of free and bound alpha spectrin. *Protein Sci* 2009;18:1916–25. [PubMed: 19593814]
- [41]. Kang JX, Song YL, Sevinc A, Fung LWM. Important residue (G46) in erythroid spectrin tetramer formation. *Cell Mol Biol Lett* 2010;15:46–54. [PubMed: 19756397]
- [42]. Gaetani M, Mootien S, Harper S, Gallagher PG, Speicher DW. Structural and functional effects of hereditary hemolytic anemia-associated point mutations in the alpha spectrin tetramer site. *Blood* 2008;111:5712–20. [PubMed: 18218854]

- [43]. Bignone PA, Baines AJ. Spectrin alpha II and beta II isoforms interact with high affinity at the tetramerization site. *Biochem J* 2003;374:613–24. [PubMed: 12820899]
- [44]. Shammas SL, Rogers JM, Hill SA, Clarke J. Slow, reversible, coupled folding and binding of the spectrin tetramerization domain. *Biophys J* 2012;103:2203–14. [PubMed: 23200054]
- [45]. Nickson AA, Clarke J. What lessons can be learned from studying the folding of homologous proteins? *Methods* 2010;52:38–50. [PubMed: 20570731]
- [46]. Shank EA, Cecconi C, Dill JW, Marqusee S, Bustamante C. The folding cooperativity of a protein is controlled by its chain topology. *Nature* 2010;465:637–40. [PubMed: 20495548]
- [47]. Wensley BG, Gartner M, Choo WX, 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. [PubMed: 19445951]
- [48]. 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. [PubMed: 15504411]
- [49]. Scott KA, Randles LG, Clarke J. The folding of spectrin domains II: Phi-value analysis of R16. *J Mol Biol* 2004;344:207–21. [PubMed: 15504412]
- [50]. 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. [PubMed: 16618492]
- [51]. MacDonald RI, Pozharski EV. Free energies of urea and of thermal unfolding show that two tandem repeats of spectrin are thermodynamically more stable than a single repeat. *Biochemistry* 2001;40:3974–84. [PubMed: 11300778]
- [52]. Randles LG, Batey S, Steward A, Clarke J. Distinguishing specific and nonspecific interdomain interactions in multi-domain proteins. *Biophys J* 2008;94:622–8. [PubMed: 17890397]
- [53]. 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. [PubMed: 1569556]
- [54]. An X, Guo X, Zhang X, Baines AJ, Debnath G, Moyo D, et al. Conformational stabilities of the structural repeats of erythroid spectrin and their functional implications. *J Biol Chem* 2006;281:10527–32. [PubMed: 16476728]
- [55]. Batey S, Randles LG, Steward A, Clarke J. Cooperative folding in a multi-domain protein. *J Mol Biol* 2005;349:1045–59. [PubMed: 15913648]
- [56]. Borgia A, Wensley BG, Soranno A, Nettels D, Borgia MB, Hoffmann A, et al. Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. *Nat Commun* 2012;3:9.
- [57]. Coetzer TL, Sahr K, Prchal J, Blacklock H, Peterson L, Koler R, et al. Four different mutations in codon 28 of alpha-spectrin are associated with structurally and functionally abnormal spectrin alpha-I/74 in hereditary elliptocytosis. *J Clin Invest* 1991;88:743–9. [PubMed: 1679439]
- [58]. Garbarz M, Lecomte MC, Feo C, Devaux I, Picat C, Lefebvre C, et al. Hereditary pyropoikilocytosis and elliptocytosis in a white French family with the spectrin-alpha-I/74 variant related to a CGT to CAT codon change (Arg to His) at position 22 of the spectrin-alpha-I domain. *Blood* 1990;75:1691–8. [PubMed: 2328319]
- [59]. Floyd PB, Gallagher PG, Valentino LA, Davis M, Marchesi SL, Forget BG. Heterogeneity of the molecular-basis of hereditary pyropoikilocytosis and hereditary elliptocytosis associated with increased levels of the spectrin alpha-I/72-kilodalton tryptic peptide. *Blood* 1991;78:1364–72. [PubMed: 1878597]
- [60]. Morle L, Roux AF, Alloisio N, Pothier B, Starck J, Denoroy L, et al. Two elliptocytogenic alpha-I/74 variants of the spectrin alpha-I domain: spectrin Culoz (GGT-GTT alpha-I-40 Gly-Val) and spectrin Lyon (CTT-TTT alpha-I-43 Leu-Phe). *J Clin Invest* 1990;86:548–54. [PubMed: 2384601]

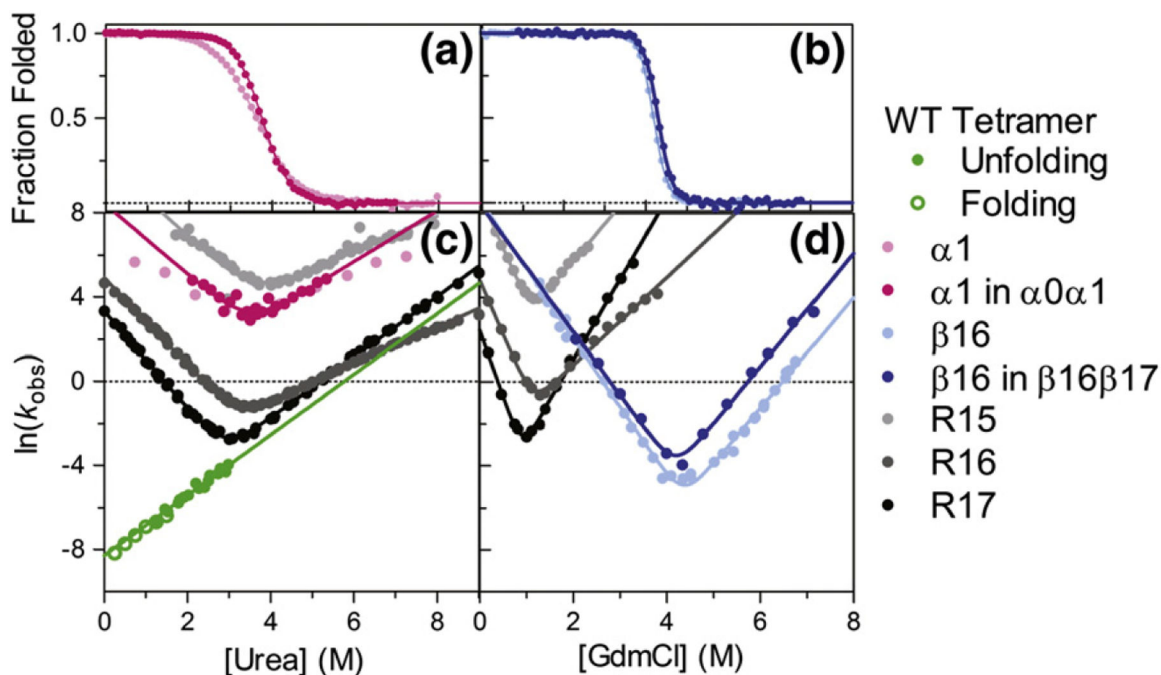


**Fig. 1.** Schematic depiction of the structure and assembly of erythrocyte spectrin. (a) Three contiguous spectrin domains (green, chicken brain  $\alpha$ -spectrin domains R15, R16, and R17; PDB ID: 1U4Q) show the three-helix-bundle structure characteristic of tandem spectrin domains. A continuous helix connects the C-helix of one domain to the A-helix of the following domain. Subunits  $\alpha$ -spectrin and  $\beta$ -spectrin (pink and blue, respectively) associate laterally to form heterodimers. Where two heterodimers meet “head on”, the C-terminal tail of  $\beta$ -spectrin binds and folds with the N-terminal tail of  $\alpha$ -spectrin of the opposing heterodimer to form a new, non-covalent three-helix bundle. This is termed the “tetramerization domain” (PDB ID: 3LBX). (b) Truncated constructs  $\alpha 0\alpha 1$  (PDB ID: 1OWA) and  $\beta 16\beta 17$  are employed to facilitate study of the tetramer domain as a dimerization process. Structures prepared using PyMOL (DeLano Scientific).



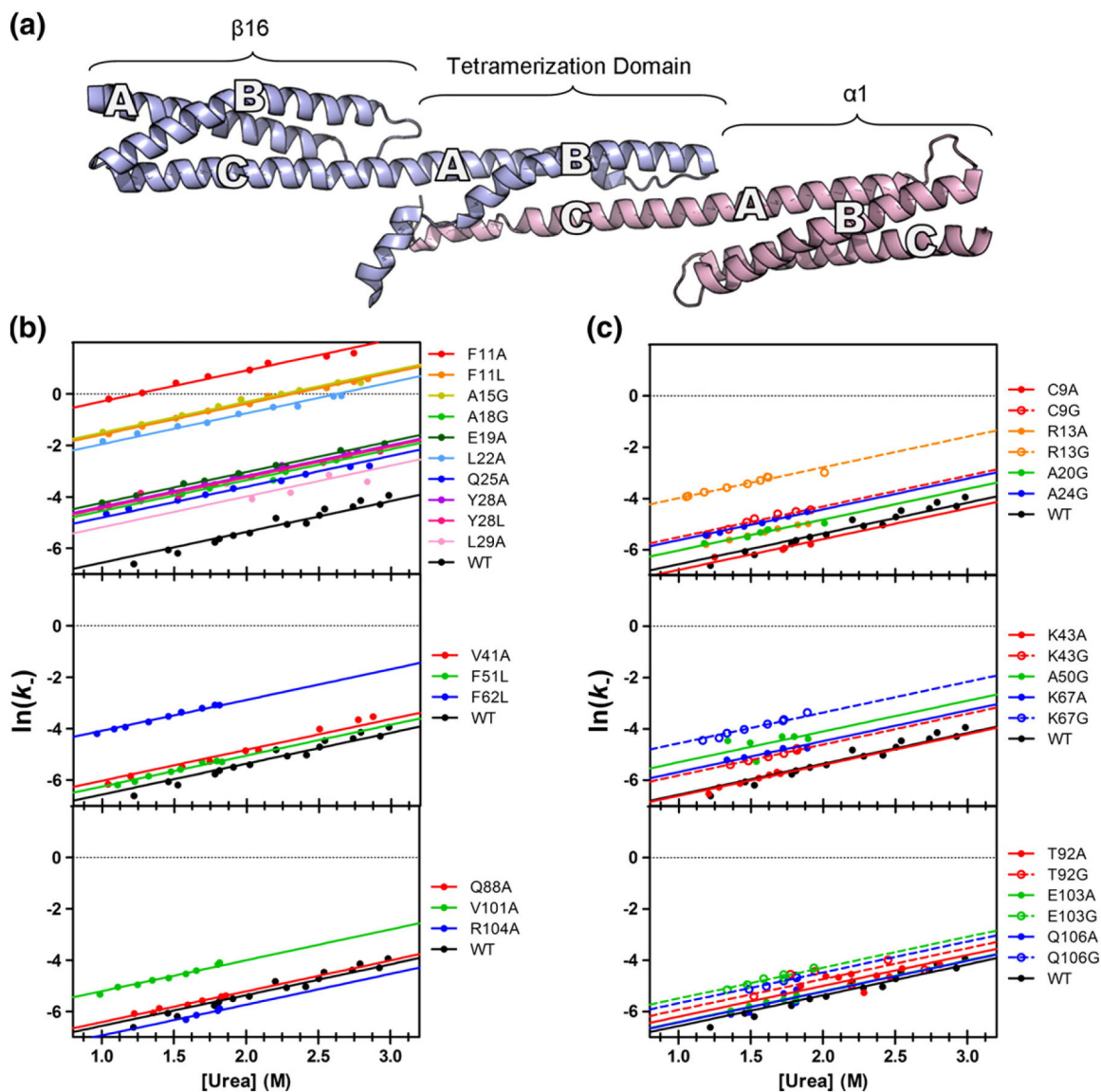


**Fig. 2.** CD profiles for erythrocyte spectrin constructs. (a) Equal concentrations of  $\alpha 1$  and  $\beta 16$  have similar CD spectra, as we might expect from 2 three-helix bundles of similar size. Addition of  $\alpha 0$  to  $\alpha 1$  and  $\beta 17$  to  $\beta 16$  shows an increase in overall helicity, suggesting that both  $\alpha 0$  and  $\beta 17$  have some helical structure. (b) However, when we compare mean residue ellipticity (MRE), it is clear that the MRE values for  $\alpha 0\alpha 1$  and  $\beta 16\beta 17$  are significantly lower in magnitude than those for folded  $\alpha 1$  and  $\beta 16$ , respectively. Both tail regions are also significantly unstructured.



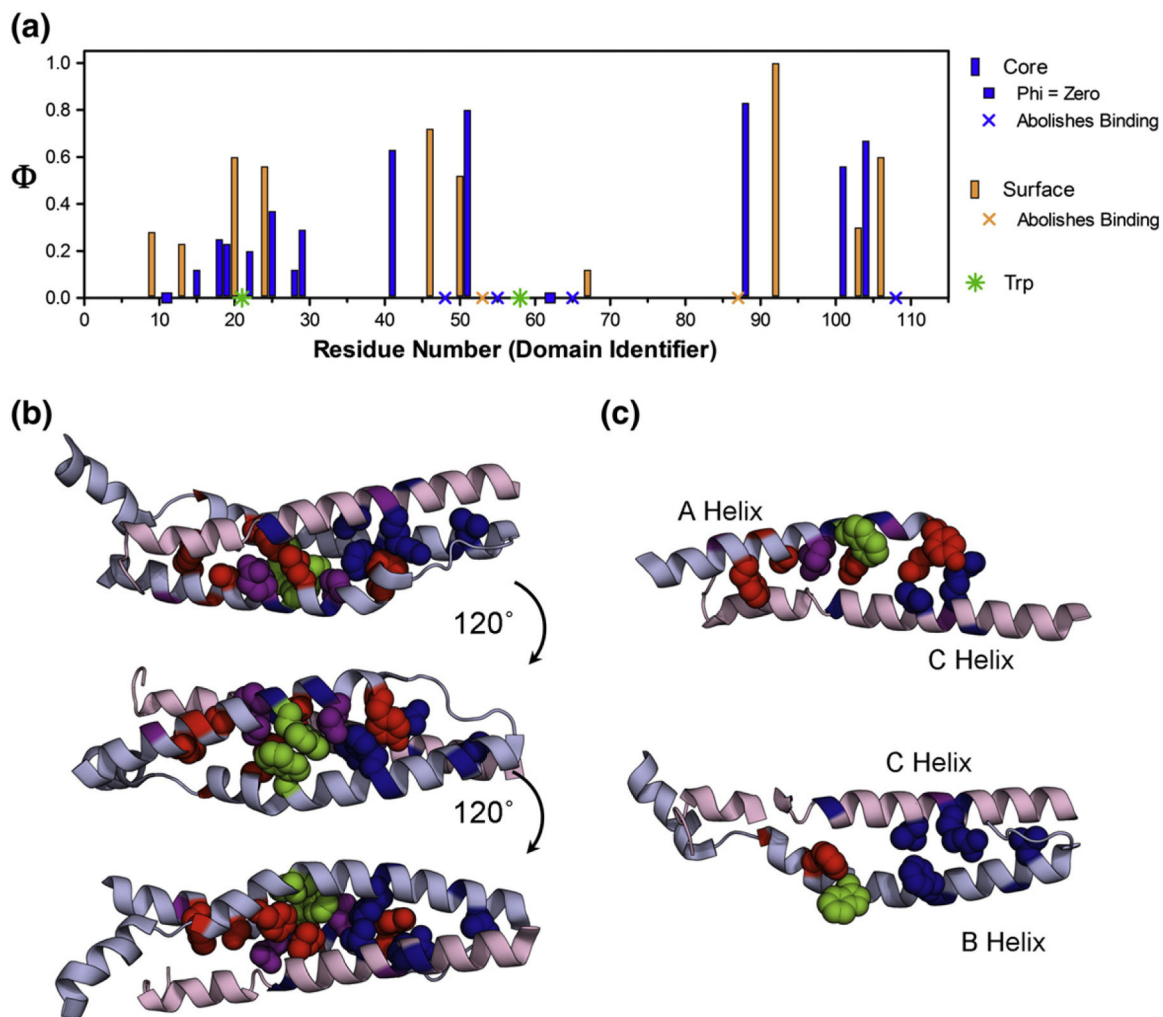
**Fig. 3.**

Characterization of  $\alpha 1$ ,  $\alpha 1$  in  $\alpha 0\alpha 1$ ,  $\beta 16$ ,  $\beta 16$  in  $\beta 16\beta 17$ , and the tetramer domain in relation to R15, R16, and R17. (a)  $\alpha 1$  in  $\alpha 0\alpha 1$  and  $\alpha 1$  alone have similar  $[\text{Den}]_{50\%}$  values, but different  $m_{\text{D-N}}$ . This probably reflects helix fraying in  $\alpha 1$ , in the absence of the N-terminal extension. (b) Addition of  $\beta 17$  does not alter the high stability of  $\beta 16$ , which unfolds only in the presence of high concentrations of GdmCl. (c) Chevron plots for  $\alpha 1$  and  $\alpha 1$  in  $\alpha 0\alpha 1$  are compared to R15, R16, and R17. Both  $\alpha$ -spectrin constructs fold fast, but the decrease in  $m_{\text{kin}}$  in  $\alpha 1$  is obvious in comparison to other spectrin domains (see the text). The observed rate constants for tetramer dissociation (green) are calculated from folding (open) and unfolding (closed) experiments and display a denaturant dependence ( $m_{k^-}$ ), which is consistent with the  $m_{k_u}$  for R17. Interestingly, the tetramer domain is much more kinetically stable than one of its neighboring domains,  $\alpha 1$ . (d) Chevrons for the  $\beta$ -spectrin constructs show that  $\beta 16$  alone and in  $\beta 16\beta 17$  folds as quickly as R15, but due to its inordinate stability, it has very low unfolding rate constants.

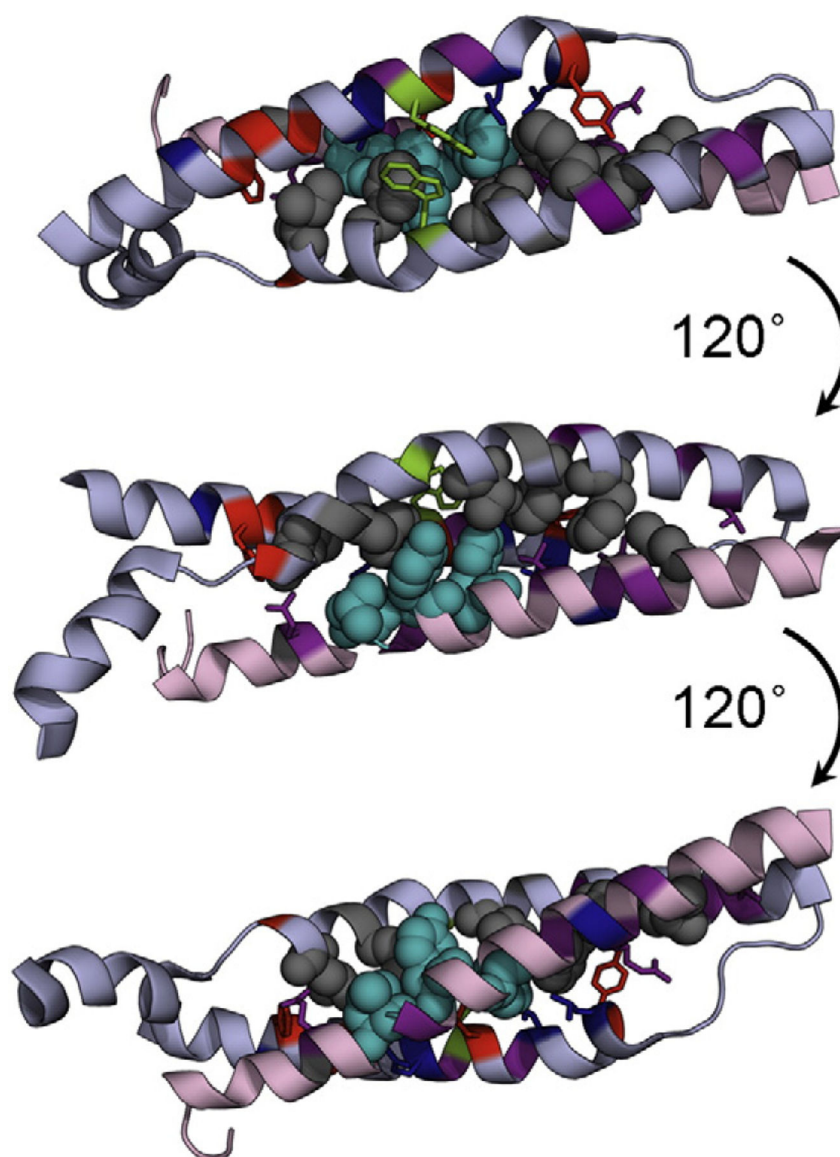


**Fig. 4.**

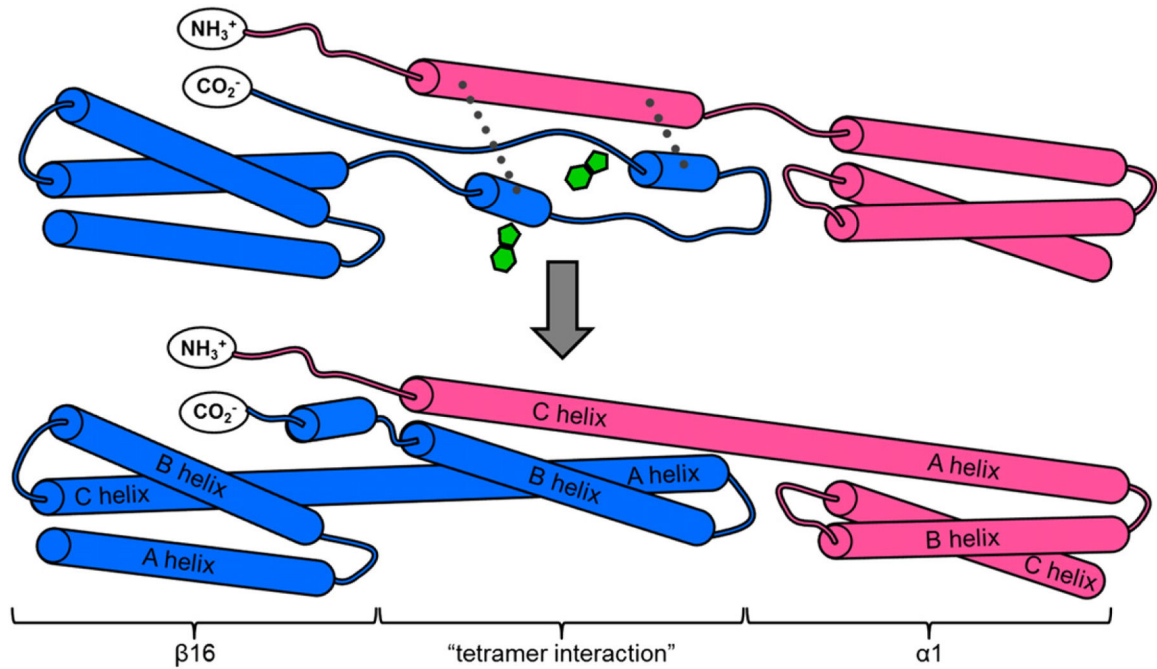
Structure of the tetramer domain and dissociation/unfolding kinetics for mutants. (a)  $\beta 16\beta 17$ , shown in blue, consists of a full three-helix domain ( $\beta 16$ ) followed by a tail ( $\beta 17$ ), forming the A- and B-helices of the tetramer domain.  $\alpha 0\alpha 1$  (pink) has a full domain ( $\alpha 1$ ), preceded by the N-terminal tail ( $\alpha 0$ ), which forms the C-helix of the tetramer domain. (b and c) The observed rates for dissociation/unfolding experiments, used to calculate  $\Phi$ -values for core and surface mutants, respectively. Mutants are divided into three groups: those found on Helix A (top), Helix B (middle), and Helix C (bottom). Each unfolding “limb” is well fitted with an  $m_{k_-}$  of  $1.2 \text{ kcal M}^{-1}$ . Mutations to the core of the A-helix show the greatest variation in observed dissociation/folding rate constants.



**Fig. 5.** Results of  $\Phi$ -value analysis. (a) The  $\Phi$ -values form a coherent pattern when plotted against the sequence, covering a large range of both core (blue) and surface (orange) residues. The mutations that abolish binding are presented as symbols (x) in the appropriate color, and the two Trp residues, in Helices A and B, are shown as green symbols (\*). Peaks in the  $\Phi$ -values for the A-helix (middle-to-C-terminal) and the B-helix (N-terminal-to-middle) indicate the most structured regions in  $\beta 17$  at the transition state. The  $\Phi$ -values in the C-helix are more uniform. (b)  $\Phi$ -values in the context of the structure of the spectrin tetramer are shown as spacefilling for core residues, while just the backbone is colored for surface residues ( $\alpha 1$  and  $\beta 16$  have been removed for clarity). The  $\Phi$ -values are divided as low ( $\Phi < 0.25$ , red), medium ( $0.25 \leq \Phi < 0.5$ , purple), and high ( $\Phi \geq 0.5$ , blue). (c) Top, the A-helix and the C-helix are shown with their  $\Phi$ -values in the absence of Helix B, which highlights the contact of high  $\Phi$ -values in C-terminal Helix A with N-terminal Helix C at the transition state. The bottom structure shows only Helices B and C, where the contact between high  $\Phi$ -values at the N-terminal end of Helix B with the C-terminal end of Helix C.



**Fig. 6.** Mutations that abolish tetramer formation (gray) are primarily located within the B-helix, but two also occur in Helix C. Mutations that cause  $m_k \approx 1.2 \text{ M}^{-1}$  or those that result in an unusual  $\Phi$ -value are shown in cyan. Interestingly, all of these mutants occur in the C-helix, which is highly structured at the transition state for assembly/folding. Trp residues on the A- and B-helices are shown in green.

**Fig. 7.**

Cartoon depiction of proposed folding mechanism for the spectrin tetramer domain.  $\beta 16\beta 17$  (blue) initially exists as a fully formed three-helix domain ( $\beta 16$ ) with a C-terminal tail with some transient helical structure.  $\alpha 0\alpha 1$  (pink) has a partially structured  $\alpha 0$  domain, followed by a fully structured  $\alpha 1$ . At the transition state, the middle-to-C-terminal end of the future A-helix comes into contact with the N-terminal end of the helical  $\alpha 0$  domain, while the N-to-middle portion of the future B-helix contacts the C-terminal end of  $\alpha 0$  (broken lines). Contact of  $\beta 17$  with the structured  $\alpha 0$  helix stabilizes these nascent areas of secondary structure, ensuring the correct orientation and register for the helices, allowing folding to proceed efficiently. The two Trp residues (green), located on the A-helix and on the B-helix, are at the ends of the regions in these helices that pack onto C at the transition state.

Table 1.

Equilibrium and kinetic data for  $\alpha.1$ ,  $\alpha.0\alpha.1$ ,  $\beta.16$ ,  $\beta.16\beta.17$ , and the tetramer domain

	$m_{D-N}$ (kcal mol <sup>-1</sup> M <sup>-1</sup> )	[Den] <sub>50%</sub> (M)	$G_{D-N}$ (kcal mol <sup>-1</sup> )	$k_f^{H_2O}$ (s <sup>-1</sup> )	$k_u^{H_2O} / k_{-}^{H_2O}$ (s <sup>-1</sup> )	$m_{kf}$ (M <sup>-1</sup> )	$m_{ku}$ ( $m_{k-}$ ) (M <sup>-1</sup> )
$\alpha.1^{a,c}$	1.38 ( $\pm 0.01$ )	3.47 ( $\pm 0.01$ )	4.78 ( $\pm 0.03$ )	—	—	—	—
$\alpha.1$ in $\alpha.0\alpha.1^a$	1.73 ( $\pm 0.03$ )	3.75 ( $\pm 0.01$ )	6.49 ( $\pm 0.03$ )	1500 ( $\pm 400$ )	0.08 ( $\pm 0.06$ )	1.3 ( $\pm 0.1$ )	1.4 ( $\pm 0.2$ )
$\beta.16$	3.44 ( $\pm 0.03$ )	4.19 ( $\pm 0.01$ )	14.4 ( $\pm 0.1$ )	5200 ( $\pm 800$ )	$3 \times 10^{-8}$ ( $\pm 2 \times 10^{-8}$ )	3.22 ( $\pm 0.06$ )	2.7 ( $\pm 0.1$ )
$\beta.16$ in $\beta.16\beta.17$	3.3 ( $\pm 0.1$ )	4.30 ( $\pm 0.01$ )	14.2 ( $\pm 0.1$ )	4000 ( $\pm 3000$ )	$1.7 \times 10^{-7}$ ( $\pm 1.5 \times 10^{-7}$ )	3.0 ( $\pm 0.2$ )	2.7 ( $\pm 0.1$ )
R15 <sup>ad</sup>	1.8 ( $\pm 0.1$ )	3.7 ( $\pm 0.1$ )	6.8 ( $\pm 0.2$ )	25,000 ( $\pm 8000$ )	2.0 ( $\pm 0.5$ )	1.85 ( $\pm 0.08$ )	1.00 ( $\pm 0.03$ )
R16 <sup>ad</sup>	1.9 ( $\pm 0.1$ )	3.6 ( $\pm 0.1$ )	6.4 ( $\pm 0.2$ )	136 ( $\pm 4$ )	$1.11 \times 10^{-2}$ ( $\pm 6 \times 10^{-4}$ )	2.08 ( $\pm 0.02$ )	0.885 ( $\pm 0.008$ )
R17 <sup>ad</sup>	2.0 ( $\pm 0.1$ )	3.0 ( $\pm 0.1$ )	6.0 ( $\pm 0.2$ )	27.9 ( $\pm 1.2$ )	$5.7 \times 10^{-4}$ ( $\pm 4 \times 10^{-5}$ )	2.29 ( $\pm 0.03$ )	1.46 ( $\pm 0.01$ )
Wild-type tetramer <sup>e</sup>	—	—	—	—	$2.5 \times 10^{-4}$ ( $\pm 3 \times 10^{-5}$ )	—	1.44 ( $\pm 0.07$ )

<sup>a</sup>[Den]<sub>50%</sub>,  $m_{D-N}$ ,  $m_{kf}$ ,  $m_{ku}$ , and  $m_{k-}$ , reported in molar units of urea.<sup>b</sup>Chevron not fitted.<sup>c</sup>[Den]<sub>50%</sub>,  $m_{D-N}$ ,  $m_{kf}$ , and  $m_{ku}$ , reported in molar units of GdmCl.<sup>d</sup>Data from Ref. [48].<sup>e</sup> $k_{-}^{H_2O}$  and  $m_{k-}$  reported in lieu of  $k_u^{H_2O}$  and  $m_{ku}$ .

Table 2.

Core mutants: thermodynamic and kinetic parameters and calculation of  $\Phi$ -values

Domain identifier	Protein identifier	$K_a$ ( $10^4$ M <sup>-1</sup> )	$G_{D-N}$ (kcal mol <sup>-1</sup> )	$k_c$ ( $10^{-4}$ M <sup>-1</sup> s <sup>-1</sup> )	$\Phi^a$
<i>Wild type</i>		207 ( $\pm 8$ )	—	4.7 ( $\pm 0.3$ )	—
$\beta 16\beta 17$					
Helix A					
F11A	F2014A	0.6 ( $\pm 0.6$ )	3.5 ( $\pm 0.6$ )	2500 ( $\pm 170$ )	-0.08 ( $\pm 0.17$ )
F11L	F2014L	1.6 ( $\pm 0.2$ )	2.93 ( $\pm 0.09$ )	680 ( $\pm 40$ )	-0.02 ( $\pm 0.04$ )
A15G	A2018G	0.6 ( $\pm 0.1$ )	3.51 ( $\pm 0.11$ )	740 ( $\pm 50$ )	0.14 ( $\pm 0.03$ )
A18G	A2021G	1.3 ( $\pm 3$ )	1.67 ( $\pm 0.13$ )	35 ( $\pm 3$ )	0.28 ( $\pm 0.07$ )
E19A	E2022A	9.3 ( $\pm 5$ )	1.9 ( $\pm 0.3$ )	48 ( $\pm 3$ )	0.25 ( $\pm 0.13$ )
L22A	L2025A	0.6 ( $\pm 0.1$ )	3.50 ( $\pm 0.13$ )	470 ( $\pm 30$ )	0.21 ( $\pm 0.03$ )
Q25A	Q2028A	12 ( $\pm 3$ )	1.80 ( $\pm 0.14$ )	27 ( $\pm 1$ )	0.39 ( $\pm 0.06$ )
Y28A	Y2031A	18 ( $\pm 2$ )	1.47 ( $\pm 0.08$ )	41 ( $\pm 3$ )	0.12 ( $\pm 0.06$ )
Y28L	Y2031L	17 ( $\pm 5$ )	1.50 ( $\pm 0.19$ )	40 ( $\pm 3$ )	0.15 ( $\pm 0.11$ )
L29A	L2032A	29 ( $\pm 2$ )	1.18 ( $\pm 0.05$ )	19 ( $\pm 2$ )	0.29 ( $\pm 0.07$ )
Helix B					
V41A	V2044A	42 ( $\pm 4$ )	0.96 ( $\pm 0.06$ )	8.3 ( $\pm 0.5$ )	0.65 ( $\pm 0.07$ )
H48A <sup>b</sup>	H2051A	—	—	—	—
F51A <sup>b</sup>	F2054A	—	—	—	—
F51L	F2054G	18 ( $\pm 3$ )	1.47 ( $\pm 0.11$ )	6.2 ( $\pm 0.4$ )	0.89 ( $\pm 0.04$ )
T55A <sup>b</sup>	T2057A	—	—	—	—
F62A	F2065A	1.3 ( $\pm 0.9$ )	3.0 ( $\pm 0.4$ )	640 ( $\pm 40$ )	0.3 ( $\pm 0.14$ )
F62L	F2065L	10 ( $\pm 2$ )	1.46 ( $\pm 0.04$ )	48 ( $\pm 3$ )	0.05 ( $\pm 0.06$ )
L65A <sup>b</sup>	L2068A	—	—	—	—
$\alpha 6\alpha 1$					
Helix C					
Q88A <sup>c</sup>	Q29A	100 ( $\pm 20$ )	0.42 ( $\pm 0.1$ )	5.3 ( $\pm 0.4$ )	—
L91A <sup>c,d</sup>	L32A	130 ( $\pm 30$ )	0.27 ( $\pm 0.14$ )	—	—



Domain identifier	Protein identifier	$K_a$ ( $10^4 M^{-1}$ )	$G_{D-N}$ (kcal mol $^{-1}$ )	$k_-$ ( $10^{-4} M^{-1} s^{-1}$ )	$\Phi^a$
Y94A <sup>c</sup>	Y35A	67 ( $\pm 2$ )	0.68 ( $\pm 0.03$ )	148 ( $\pm 10$ )	—
Y94L <sup>c,d</sup>	Y35L	440 ( $\pm 50$ )	-0.45 ( $\pm 0.07$ )	50 ( $\pm 4$ )	—
F97A	F38A	5.8 ( $\pm 0.7$ )	2.14 ( $\pm 0.07$ )	420 ( $\pm 3$ )	-0.27 ( $\pm 0.05$ )
F97L <sup>e</sup>	F38L	—	—	—	—
V101A	V42A	10 ( $\pm 2$ )	1.80 ( $\pm 0.13$ )	18 ( $\pm 12$ )	0.6 ( $\pm 0.2$ )
R104A	R45A	670 ( $\pm 190$ )	-0.70 ( $\pm 0.17$ )	3 ( $\pm 3$ )	0.7 ( $\pm 0.8$ )
L108A <sup>b</sup>	L49A	—	—	—	—

<sup>a</sup>  $\Phi$  calculated using  $\langle mK_- \rangle = 1.2 M^{-1}$ .

<sup>b</sup> Mutation abolished formation of the tetramer domain, as judged by ITC.

<sup>c</sup>  $G_{D-N} < 0.7$  kcal mol $^{-1}$ ;  $\Phi$  not calculated.

<sup>d</sup> Slope of unfolding limb 1.2 M $^{-1}$ .

<sup>e</sup> Protein insoluble.

Table 3.

Surface mutants: thermodynamic and kinetic parameters and calculation of  $\Phi$ -values

Domain identifier	Protein identifier	$K_a$ ( $10^4$ M <sup>-1</sup> )	$G_{D-N}$ (kcal mol <sup>-1</sup> )	$k_c$ ( $10^{-4}$ M <sup>-1</sup> s <sup>-1</sup> )	$\Phi^a$
<i>Wild type</i>		207 ( $\pm 8$ )	—	4.7 ( $\pm 0.3$ )	—
<i><math>\beta</math>16<math>\beta</math>17</i>					
C9A	C2012A	270 ( $\pm 50$ )	—	3.7 ( $\pm 0.3$ )	—
C9G	C2012G	47 ( $\pm 1.2$ )	1.05 ( $\pm 0.11$ )	1.3 ( $\pm 0.9$ )	0.28 ( $\pm 0.11$ )
R13A	R2016A	98 ( $\pm 5$ )	—	7.8 ( $\pm 0.5$ )	—
R13G	R2016G	7.1 ( $\pm 0.4$ )	1.57 ( $\pm 0.05$ )	60 ( $\pm 4$ )	0.23 ( $\pm 0.05$ )
S16A	S2019A	49 ( $\pm 4$ )	—	6.2 ( $\pm 0.5$ )	—
S16G	S2019G	36 ( $\pm 5$ )	0.18 ( $\pm 0.1$ )	11.6 ( $\pm 0.8$ )	—
A20G	A2023G	53 ( $\pm 8$ )	0.82 ( $\pm 0.09$ )	7.9 ( $\pm 0.6$ )	0.63 ( $\pm 0.09$ )
A24G	A2027G	24.2 ( $\pm 0.8$ )	1.29 ( $\pm 0.03$ )	11.7 ( $\pm 0.8$ )	0.58 ( $\pm 0.05$ )
Helix B					
D39A <sup>c</sup>	D2042A	47 ( $\pm 3$ )	—	6.4 ( $\pm 0.5$ )	—
D39G <sup>c</sup>	D2042G	28 ( $\pm 3$ )	0.3 ( $\pm 0.07$ )	10.3 ( $\pm 0.8$ )	—
K43A	K2046A	180 ( $\pm 17$ )	—	4.4 ( $\pm 0.3$ )	—
K43G	K2046G	32.8 ( $\pm 1.0$ )	1.02 ( $\pm 0.06$ )	9.7 ( $\pm 0.8$ )	0.54 ( $\pm 0.07$ )
K46A <sup>c</sup>	K2049A	52 ( $\pm 17$ )	—	12.7 ( $\pm 1.0$ )	—
K46G <sup>c</sup>	K2049G	23.0 ( $\pm 1.2$ )	0.5 ( $\pm 0.2$ )	16.0 ( $\pm 1.3$ )	—
A50G	A2053G	16.3 ( $\pm 0.6$ )	1.52 ( $\pm 0.04$ )	16.1 ( $\pm 1.0$ )	0.52 ( $\pm 0.04$ )
K53A	K2056A	150 ( $\pm 13$ )	—	2.2 ( $\pm 1.2$ )	—
K53G <sup>b</sup>	K2056G	—	—	—	—
S57A <sup>ce</sup>	S2060A	150 ( $\pm 15$ )	—	—	—
S57G <sup>c</sup>	S2060G	90 ( $\pm 30$ )	0.32 ( $\pm 0.18$ )	4 ( $\pm 4$ )	—
A63G <sup>b</sup>	A2066G	—	—	—	—
K67A	K2070A	50 ( $\pm 5$ )	—	12.0 ( $\pm 1.0$ )	—
K67G	K2070G	14 ( $\pm 2$ )	0.8 ( $\pm 0.1$ )	37 ( $\pm 3$ )	0.12 ( $\pm 0.16$ )
<i><math>\alpha</math>6a1</i>					

Domain identifier	Protein identifier	$K_a$ ( $10^4 M^{-1}$ )	$G_{D-N}$ (kcal mol $^{-1}$ )	$k_-$ ( $10^{-4} M^{-1} s^{-1}$ )	$\Phi^a$
Helix C					
E85A <sup>c,e</sup>	E26A	120 ( $\pm 0.4$ )	—	—	—
E85G <sup>c,e</sup>	E26G	83.0 ( $\pm 0.7$ )	0.221 ( $\pm 0.010$ )	—	—
R87A <sup>b</sup>	R28A	—	—	—	—
R87G <sup>d</sup>	R28G	400 ( $\pm 200$ )	—	4.1 ( $\pm 0.9$ )	—
T92A	T33A	67 ( $\pm 8$ )	—	6.0 ( $\pm 1.1$ )	—
T92G	T33G	330 ( $\pm 30$ )	—	7.8 ( $\pm 1.5$ )	1.16 ( $\pm 0.17$ )
Q95A <sup>c,e</sup>	Q36A	229 ( $\pm 9$ )	—	—	—
Q95G <sup>c,e</sup>	Q36G	200 ( $\pm 20$ )	0.08 ( $\pm 0.02$ )	—	—
E99A <sup>c</sup>	E40A	90 ( $\pm 9$ )	—	6.2 ( $\pm 1.3$ )	—
E99G <sup>c</sup>	E40G	80 ( $\pm 20$ )	0.04 ( $\pm 0.16$ )	7.7 ( $\pm 1.2$ )	—
E103A	E44A	81 ( $\pm 9$ )	—	4.9 ( $\pm 0.8$ )	—
E103G	E44G	22 ( $\pm 3$ )	0.78 ( $\pm 0.11$ )	12 ( $\pm 2$ )	0.3 ( $\pm 0.2$ )
Q106A	Q47A	130 ( $\pm 14$ )	—	4.8 ( $\pm 0.8$ )	—
Q106G	Q47G	20.0 ( $\pm 1.8$ )	1.12 ( $\pm 0.08$ )	10.2 ( $\pm 1.7$ )	0.60 ( $\pm 0.13$ )

<sup>a</sup>  $\Phi$  calculated using  $m_{K-} = 1.2 M^{-1}$ .

<sup>b</sup> Mutation abolished formation of the tetramer domain, as judged by ITC.

<sup>c</sup>  $G_{D-N} < 0.7$  kcal mol $^{-1}$ ;  $\Phi$  not calculated.

<sup>d</sup> Slope of unfolding limb  $\sim 1.2 M^{-1}$ .

<sup>e</sup> Kinetic experiments not performed.