Crystal Structure of the Repetitive Segments of Spectrin

Y. Yan, E. Winograd, A. Viel, T. Cronin, S. C. Harrison, and D. Branton
Crystal Structure of the Repetitive Segments of Spectrin

Y. Yan, E. Winograd, A. Viel, T. Cronin, S. C. Harrison, D. Branton

The elongated proteins of the spectrin family (cytosphillin, actin, spectrin) contain tandemly repeated segments and form resilient cellular meshworks by cross-linking actin filaments. The structure of one of the repetitive segments of α-spectrin was determined at 1.8 angstrom resolution. A segment consists of a three-helix bundle. A model of the interface between two tandem segments suggests that hydrophobic interactions between segments may constrain intersegment flexibility. The helix side chain interactions explain how mutations that are known to produce hemolytic anemias disrupt spectrin associations that sustain the integrity of the erythrocyte membrane.

Spectrin, actin, and dystrophin are actin cross-linking proteins. Their associations with actin generate meshworks that support the plasma membrane and sustain interactions between cellular structures responsible for cell motility and shape (1, 2). Although all spectrins have a similar domain organization, important functional distinctions among these proteins appear to arise from the differing number of tandemly repeated segments that each contains. Because the repeated segments account for most of the length of these proteins, the structure of these segments must in large part determine the flexibility, and hence the functional and mechanochromical properties, of the meshworks they form (3).

Sequence analysis of spectrin and measurements of contour length have suggested that the repeated segments each consist of 100 to 120 amino acid motifs that fold into a three-helix bundle (4, 5). Our recent demonstration that single repeats of spectrin can fold into stable conformations similar to those in the native protein (6) presented a fresh opportunity to examine segment structure. Polyproteins representing single repeats exist in solution as a mixture of monomers and homodimers (7). We have been able to crystallize homodimers of the 14th segment of Drosophila α-spectrin (Fig. 1) and determine its structure.

The details of the structure determination are presented in Table 1. A 3 Å resolution, multiple isomorphous replacement (MIR) map based on two heavy-atom derivatives and anomalous dispersion showed that the α14 fragment had crystallized as a dimer, with the two polyproteins

![Figure 1](100 amino acids in segment 14 of Drosophila α-spectrin (23). The bracke).

**Table 1.** Data collection, structure determination, and refinement (25, 26).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Native III</th>
<th>Native II</th>
<th>Native I*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH3HgCl</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>K2O3S4</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>1.8</td>
<td>2.0</td>
<td>3</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>68201</td>
<td>46546</td>
<td>19853</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>20541</td>
<td>13437</td>
<td>4928</td>
</tr>
<tr>
<td>Complete reflections</td>
<td>4928</td>
<td>4459</td>
<td>4779</td>
</tr>
<tr>
<td>Completion (%)</td>
<td>95</td>
<td>86</td>
<td>96</td>
</tr>
<tr>
<td>Complete reflections</td>
<td>3</td>
<td>2.9</td>
<td>3</td>
</tr>
<tr>
<td>Number of crystals</td>
<td>85</td>
<td>90</td>
<td>0.046</td>
</tr>
<tr>
<td>Number of sites</td>
<td>0.069</td>
<td>0.065</td>
<td>0.053</td>
</tr>
<tr>
<td>Root mean square</td>
<td>0.060</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td>Root mean square</td>
<td>2(4)</td>
<td>1</td>
<td>0.10</td>
</tr>
<tr>
<td>Root mean square</td>
<td>0.00</td>
<td>1.75</td>
<td>2.29</td>
</tr>
</tbody>
</table>

*Mean figure of merit for Native I = 0.50, \( J_{R_{\text{merge}}} = \sum_{i} j_{i} / (\sum_{i} j_{i}) \), where \( j_{i} \) is intensity. \( J_{R_{\text{merge}}} = \sum_{i} j_{i} / \sum_{i} j_{i} \), where \( f_{\text{calc}} \) and \( f_{\text{obs}} \) are the derivative and native structure factor amplitudes, respectively. \( R_{\text{merge}} = \sum |f_{\text{calc}} - \sum f_{\text{obs}}| / \sum f_{\text{calc}} \), where \( f_{\text{calc}} \) is the heavy-atom structure factor amplitude. | Root mean square |

Y. Yan, E. Winograd, A. Viel, T. Cronin, D. Branton, Departments of Biochemistry and Molecular Biology and Cellular and Developmental Biology, Harvard University, Cambridge, MA 02138.

S. C. Harrison, Departments of Biochemistry and Molecular Biology and Cellular and Developmental Biology and Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138.

*Current address: Departamento de Quimica, Universidad Nacional de Colombia, Bogota, Colombia.
helix. The remaining residues tend to be more polar.

The antiparallel A and B helices in spectrin are in a register such that positions d of one helix are at the same level as positions a of the other (Fig. 4B). Thus, the d side chains project into a cavity on the surface of the opposite helix, packing laterally against a side chains and bounded axially by d side chains of successive heptads. This ridge-to-ridge packing arrangement is similar to that found in coiled coil proteins (13), except that helix A does not coil around helix B. In contrast, helix C is axially displaced relative to helices B and A so that each turn of helix C runs roughly midway between two adjacent turns in helices B and A (Fig. 4B), and the side chain packing between helices C and B and between C and A conforms to the ridges into grooves arrangement common in globular proteins (14) (Fig. 5). The ridges into grooves packing of hydrophobic residues imposes different constraints on the angle between helices than does the ridge-to-ridge packing seen between helices A and B. Consequently, the B and C helices and the A and C helices pack together at angles of 10° and 15°, respectively, instead of the usual coiled-coil angle of 20° (13, 15).

In addition to the hydrophobic packing at the interior of the three-helix bundle, electrostatic interactions also appear to stabilize the spectrin repeat structure (Fig. 4A). Between helices A and C, interchain salt bridges are often formed by charged residues at positions e and g, as in conventional coiled-coil structures, and additional residues at position c in helix C are also involved in some ionic interactions. Salt bridges between the antiparallel A and C helices also involve charged residues at positions c.

Comparisons of the sequence of segment α14 with those of the other repetitive segments of α- and β-spectrin chains show that the key interactions present in our structure are broadly conserved. We therefore believe that all spectrin repeat structures are likely to have the same three-helix packing. In particular, there is conservation of hydrophobic residues at those positions in all three helices that form the critical hydrophobic interface at the inside of the helix bundle. Moreover, the conserved charged residues often form interchain salt bridges. Among the most conserved of all residues are those that are important for maintaining contacts between the straight helix A and the coiled helices B and C. In particular, the residues at A7 (usually Tyr or Phe).
helps space helix A away from B and C at one end, whereas at the other end the conserved, bulky residue at B11 (often Phe) packs against A25 and A24, the latter exceptionally conserved as a hydrophobic residue at a g position. Between these two, near the middle of the helix bundle, the invariant Trp A17 is involved in the way helix A crosses over from B to C, as is the nearly invariant Trp at C15. Indeed, these two tryptophans lie opposite each other, with His B18 between them. Finally, near the AB loop, the interaxial distance between helices A and B is unusually large (14A), and rather open salt bridges between the d position Asp A28, g position Lys B7, and d position Lys C25 replace the usual hydrophobic contacts.

The integrity of the erythrocyte membrane and its skeleton depends on the association of spectrin heterodimers into tetramers (3). Tetramers form by associations between the partial repetitive motifs found near the NH2-terminus of the spectrin α chain and the COOH-terminus of the β chain (16, 17). In all spectrins that have been sequenced, the partial motif on the α chain contains only those residues typical of a C helix, whereas the partial motif on the β chain contains residues typical of just the A and B helices. If the mode of association in tetramer formation corresponds to the packing seen in the crystallized fragment (helix C of one chain with helices A and B of the second chain) (Fig. 2A), the observed hydrophobic and ionic interactions should also be those that stabilize the corresponding association between α and β chains in the tetramer. Subgroups of hereditary elliptocytoses are human hemolytic disorders associated with the defective association of erythroid spectrin heterodimers into tetramers (18). Different point mutations have been found to be responsible for the defect. One of these (16), Ala to Pro at position B10 (19) in the incomplete terminal segment of the β chain, would be expected to disrupt the B helix and destabilize the three-helix bundle. Another (20), Arg to Leu, Ser, Cys, or His at position C8 (21) in the partial NH2-terminal segment of the α chains, could disrupt critical interactions with conserved residues at A7 (usually Phe or Tyr) and B29 (usually Ala or Gly) (Fig. 2B). This disruption could occur because the manner in which the charged end of Arg C8 sticks out of the helix bundle to interact with the neighboring e position residue

(Fig. 4A) leaves the base of Arg C8 as a small hydrophobic surface against which the side chains of A7 and B29 pack.

In general, the repetitive segments of spectrin preserve a fixed register between the C helix of one and the A helix of the next, with no prolines or glycines at the segment boundary. It has therefore been proposed that the "two" helices are in fact continuous (Fig. 4B), with the C helix of the first segment and the A helix of the next segment forming a single helix (5, 22). Because the residue positions are well defined along nearly the entire length of the helices, we were able to build a model of two successive repeats (Fig. 6). This model shows that interactions at the segment boundary explain several conserved, but otherwise puzzling, features of the spectrin amino acid sequence. First, the relatively open ends of the three-helix bundle allow the connection between segments to be made without any rearrangement of the AB or BC loops. Second, the α-helical turns that form the CA boundary (the last turns of helix C and the first of A) are involved in multiple interactions with the two turns of the helix at the NH2-terminus of B. Third, because the model predicts that each successive segment is axially rotated 60° in a right-handed sense with respect to the previous segment, it explains why the f position at A2 is frequently occupied by a hydrophobic, rather than a hydrophilic, residue. The 60° rotation places this f position in the A helix of the next successive segment between the a and d (B1 and B4) residues of the preceding segment's B helix (Fig. 6). If this model is correct, the packing of this hydrophobic residue from A2 of one segment against the hydrophobic residues at B1, B4, and B5 of the previous segment should confine the segment-to-segment flexibility within a limited number of planes.

REFERENCES AND NOTES

7. Unpublished observations based on chromatography, cross-linking, and equilibrium sedimentation. Interconversion of the monomer and dimer was slow, probably because the helix-helix contacts must break to convert in either direction.
8. The first (e) end of a helix in a three-helix bundle is labeled A, and the succeeding helices are labeled B and C (this differs from the notation in previous work, in which the "A" referred to our
helix B). The first residue in each helix is numbered 1; thus, A1 refers to the first residue in helix A. Amino acids are represented by their single-letter, uppercase abbreviations, and McCallum (12) positions are in lowercase bold. When the crystalline dimer is referred to, the helices from one monomer are labeled with plain letters and those from the other are labeled with a prime.


10. Unlike a number of the other coiled-coil repetitive motifs, segment 14 of the Drosophila α chain contains no Pro in the region between helix B and C, so that a continuous BC helix is possible. Furthermore, we note that the BC loop is extensive in length from segment 1 to segment 2, in contrast to segment 9, it contains an entire Sro homology 3 domain.


19. Corresponding to Asp B10 in Drosophila α14.


23. Single-letter abbreviations for the amino acids residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; T, Tyr; and Y, Trp.


25. Crystals (in space group P2₁2₁2) were grown at room temperature with vapor diffusion against a well solution containing 10% polyethylene glycol (PEG) 400, 1 M ammonium acetate, 50 mM MES (4-morpholineethanesulfonic acid), pH 6.0. The starting mixture contained a 1:1 mixture of protein (20 mg/ml) and well solution. The protein used for crystallization (originally called B14) was cleaved from a fusion protein generated in E. coli as described (5). The mercury derivative was obtained by soaking crystals for 2 days in stabilization buffer (14% PEG 400, 1.2 M ammonium acetate, pH 6.0) saturated in CH₃HgCl, followed by a 1-hour backsoak. To prepare the K₃[OsCl₆] derivative, the ammonium acetate in the stabilization buffer was exchanged for 1:2 M sodium acetate, and the crystals were soaked for 1 day in stabilization buffer containing 1 mM K₃[OsCl₆]. Data were collected with a Siemens area detector on an Elliot GX-13 rotating anode generator (Avonics, Borehamwood, United Kingdom). CuKα radiation was provided by a Frank’s double-focusing mirror assembly [S. C. Harrison, J. Appl. Crystallogr. 1, 84 (1968)]. Data collection was controlled as described by M. Blum, P. Metcalf, S. C. Harrison, and D. C. Willey [J. Appl. Crystallogr. 20, 235 (1987)]. The data were processed by the XDS package of W. Kabsch [ibid. 21, 916 (1988)]. Anomalous dispersion measurements were included in calculations of the phasing for the CH₃HgCl derivative. The Hg atom positions for the CH₃HgCl derivative were determined from a difference Patterson function. There were two sites in the asymmetric unit, and each site splits into two substates, 3 Å apart. The K₃[OsCl₆] derivative was solved by a difference Fourier synthesis phased on the CH₃HgCl derivative. There is one site per asymmetric unit. Heavy-atom parameters were refined, and initial phases were calculated with the program HEAVY (T. C. Terwilliger and D. Eisenberg, Acta Crystallogr. Sect. A 40, 813 (1983)).

26. Low-resolution data (3 Å) were collected at room temperature with unit cell a = 47.91 Å, b = 48.70 Å, c = 105.29 Å for native I and two derivatives. High-resolution data (2 and 1.8 Å) were collected at cryogenic temperature (–160°C) (native II and III: before data collection crystals were soaked in 25% PEG 400 and 1.2 M ammonium acetate, pH 6.0; for one day) with unit cell a = 48.60 Å, b = 47.22 Å, c = 104.43 Å. The model built into the map was partially refined with XPLOR [A. T. Brucoleri, XPLOR Manual, Version 3.1 (Yale University, New Haven, CT, 1992)]. Acta Crystallogr. Sect. A 46, 665 (1990)] against the native I (room temperature data set). We then used the model to determine phases for the native II data set, by refining six rigid segments in the resolution range of 12 to 3 Å (R = ΣFobs - Fcalc)/ΣFcalc = 0.36, where Fobs and Fcalc are the observed and calculated structure-factor amplitudes, respectively. A randomly selected 10% subset of the data was set aside for use in "free R factor" calculations [A. T. Brucoleri, Nature 355, 472 (1992)]. The results of the rigid-body search were manually adjusted in 2(Fo - Fe) and F, and the resolution was extended to 2 Å in several stages.}

Cycles of simulated annealing refinement against 6 to 2 Å resolution data, with manual adjustment and the addition of ordered water molecules, gave a R factor of 0.21 (Rfree = 0.34). A still better low-temperature data set (Native III) extending to 1.8 Å resolution, then became available. Simulated annealing was used to refine these data, first at 2 Å and then at 1.8 Å resolution (R = 0.23, Rfree = 0.33) in XPLOR; refinement with TWIN [D. E. Trottin, D. F. TenEyck, B. W. Matthews, Acta Crystallogr. Sect. A 43, 489 (1987)] yielded an R factor of 0.20 (Rfree = 0.304) at 1.8 Å, F > 2σ (91% complete), with excellent geometry (ρcalc = 0.017 Å, ρmin = 2.7 Å). The final model had 107 ordered residues in each monomer (including one from the NH2-terminal cloning artifact) and a total of 156 ordered water molecules in the asymmetric unit. The coordinates of the refined model have been deposited in Brookhaven Protein Data Bank.

27. We thank W. Stafford for equilibrium sedimentation analyses. Supported by National Institutes of Health grants CA 13202 (S.C.H.) and HL 17411 (D.R.).

30 July 1993; accepted 19 October 1993