The Complete Sequence of *Drosophila* Alpha-Spectrin: Conservation of Structural Domains between Alpha-Spectrins and Alpha-Actinin

Ronald R. Dubreuil, Timothy J. Byers, Amy L. Sillman, Dudy Bar-Zvi, Lawrence S. B. Goldstein, and Daniel Branton

Department of Cellular and Developmental Biology, Harvard University, Cambridge, Massachusetts 02138

Abstract. We report the complete sequence of *Drosophila* alpha-spectrin and show that it is similar to vertebrate nonerythroid spectrins. As in vertebrates, the alpha subunit consists of two large domains of repetitive sequence (segments 1–9 and 11–19) separated by a short nonrepetitive sequence (segment 10). The 106-residue repetitive segments are defined by a consensus sequence of 54 residues. Chicken alpha-spectrin (Wasenius, V.-M., M. Saraste, P. Salven, M. Eramaa, L. Holm, V.-P. Lehto. 1989. *J. Cell Biol.* 108:79–93) shares 50 of these consensus positions. Through comparison of spectrin and alpha-actinin sequences, we describe a second lineage of spectrin segments (20 and 21) that differs from the 106-residue segments by an 8-residue insertion and by lack of many of the consensus residues. We present a model of spectrin evolution in which the repetitive lineage of spectrin segments and the nonrepetitive lineage of segments found in spectrin and alpha-actinin arose by separate multiplication events.

THE amino acid sequences of the erythroid alpha- and beta-spectrin subunits include ~18 tandem repeats of 106 amino acid segments that presumably account for the observed shape of the molecule (Speicher and Marchesi, 1984). The amino acid sequence has been confirmed by DNA sequence analysis (Wasenius et al., 1985; Birkenmeier et al., 1985) and sequence data for vertebrate nonerythroid spectrin cDNA has since been obtained. Partial sequences of nonerythroid spectrin from rat brain (Leto et al., 1988), human fibroblasts, frog and chicken brain (McMahon et al., 1987) and, most recently, the full length sequence of chicken brain alpha-spectrin (Wasenius et al., 1989) indicate that these molecules are also composed largely of 106 amino acid repetitive segments.

We recently described the purification, cloning of cDNAs and immunofluorescence localization of spectrin in *Drosophila* (Dubreuil et al., 1987; Byers et al., 1987, 1989; Pesacreta et al., 1989). Here we assemble complete cDNA sequence data for the alpha subunit of *Drosophila* spectrin. Our results, which show that this protein is strikingly similar in invertebrates and vertebrates, imply an important conservation of spectrin function. In addition, other studies have shown that alpha-spectrin–like sequences occur in alpha-actinin and dystrophin (Baron et al., 1987; Koenig et al., 1988; Davison et al., 1989). We have recently identified a region of sequence similarity between *Drosophila* beta-spectrin and alpha-actinin (Byers et al., 1989) that complements the region of similarity between alpha-spectrin and alpha-actinin. These observations are incorporated into a model of the structure and evolution of spectrin.

Materials and Methods

Isolation of cDNA Clones and Sequencing Strategy

Isolation of cDNA clones 9 and 10 was previously described (Byers et al., 1987). Clones 9 and 10 were used to generate sequence data from 6,733 bases of the alpha-spectrin coding sequence (Fig. 1). The remaining 5′ and 3′ sequence was obtained from full-length alpha-spectrin cDNA that was obtained from a size-fractionated 0–4 h *Drosophila* embryo library (Brown and Kafatos, 1988). Briefly, uncut plasmids were fractionated on a 0.8% low-melt agarose gel (NuSieve GTG, FMC Bioproducts, Rockland, ME) that was subsequently cut into 12 slices representing different size classes. Plasmids from the six largest classes were transfected into *Escherichia coli* DH5alpha, plated, and colonies were transferred in duplicate to nylon membranes (GeneScreen; DuPont Co., Wilmington, DE). The filters were probed with a ^{32}P-labeled 255-bp Hind III fragment of cDNA clone 10 (Fig. 1), and eight colonies were identified from sized fractions that included 6–10-kb inserts. Two of these clones (N1 and N8) were found to encode the full-length spectrin protein sequence. Sequence data for the region upstream from clone 10 was obtained using spectrin-specific primers on both clones N1 and N8. N1 and N8 are identical between residues 244 and 868 (Fig. 1), but restriction mapping revealed that these two clones have different 3′ ends. The 3′ termini of clone N1, clone 9, and a third embryonic cDNA (Fig. 1, *SM*) occurred within an A-rich region of the N8 3′ noncoding sequence that probably serves as an alternate priming site for reverse transcription during cDNA synthesis. Thus, the observed heterogeneity between 3′ ends appears to be an artifact of cDNA cloning rather than a difference in spectrin mRNAs.

D. Bar-Zvi's present address is the Department of Biology, Ben-Gurion University, Beer-sheva, Israel.

The *Drosophila* alpha-spectrin DNA sequence that appears in this paper is filed with the GenBank under the accession number M26400.

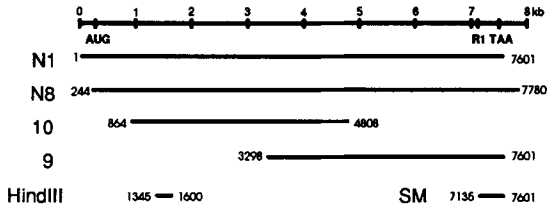

Figure 1. cDNA map of *Drosophila* alpha-spectrin. Two alpha-spectrin cDNAs (clones *9* and *10*) were initially isolated from a *Drosophila* head cDNA library in lambda gt11 (Byers et al., 1987). Clones *N1* and *N8* were selected from a 0–4 h *Drosophila* embryo library using a Hind III fragment of clone *10* as probe. The complete sequence of clones *9* and *10* was obtained by dideoxy, double-strand DNA sequencing of both DNA strands. The cDNA sequence was completed using the full-length clones *N1* and *N8* and spectrin specific primers on both DNA strands. The positions of the start codon (ATG), stop codon (TAA), and the Eco RI site were determined by DNA sequencing. *SM* is a 3′ terminal Eco RI fragment of a third embryonic cDNA clone.

DNA Sequencing

DNA inserts were subcloned into the Bluescript vector (Stratagene, La Jolla, CA) using either XL-1 Blue (Stratagene) or 71-18 as host. Nested deletions of clones 9 and 10 and the 3′ Eco RI fragment of clone N8 were generated by Exo III digestion as described by Henikoff (1984). A deletion series was generated from each end of each clone to allow sequencing of both DNA strands. The sequence of clones N8 and N1 in the region upstream of clone 10 and downstream of clone 9 were obtained using spectrin-specific synthetic primers. DNA to be used for double strand sequencing was prepared by the boiling method (Holmes and Quigley, 1981) as described by Stratagene. Sequencing reactions were carried out using the dideoxy chain termination method (Sanger et al., 1977) with a Sequenase kit from United States Biochemical Corp. (Cleveland, OH) according to their instructions using alpha [^{35}S]dATP (New England Nuclear Research, Boston, MA). The priming reaction was modified as follows: ~2 μg of purified double stranded DNA was mixed with 2 pmol of primer, denatured at elevated pH (Chen and Seeburg, 1985), precipitated with ethanol, resuspended in Sequenase buffer, and incubated at 37°C for 15 min.

DNA sequence was assembled and analyzed using the Bionet national computer facility and the University of Wisconsin Genetics Computer Group (UWGCG) sequence analysis package (Devereux et al., 1984). Dot matrix comparisons were done with the UWGCG programs "compare" and "Dotplot." A comparison window of 100 residues was used with a varied stringency as stated in the figure legends. All of the comparisons were based on amino acid identity. Stringency refers to the number of identical matches required to generate a signal. The program placed a dot at the position of each amino acid identity within each window of comparison. Structural predictions using the Garnier method (Garnier et al., 1978) were done with the program "Pepplot." The molecular weight of alpha-spectrin was predicted using the Bionet program Amino Acid Composition.

Oligonucleotide Synthesis

Sequencing reactions with clones 9 and 10 were done with either the M13 primer or T3 primer. Sequencing reactions upstream and downstream of the clone 9 and 10 boundaries were done with spectrin-specific oligonucleotide primers on both DNA strands. Primers were synthesized on a DNA synthesizer (Cyclone; Milligen/Biosearch Division, Burlington, MA) using nucleotide phosphoramidites from Fisher Scientific Co. (Pittsburgh, PA) or Cruachem, Inc. (Bend, OR).

Results

The Predicted Amino Acid Sequence of Drosophila Alpha-Spectrin

The composite length of clones N1 and N8 (Fig. 1) is consistent with the major 8-kb mRNA species detected in Northern blots of S3 cell RNA and fly head RNA (not shown). The cDNA sequence of *Drosophila* alpha-spectrin contains one long open reading frame of 7,245 bases that encodes a polypeptide 2,415 residues long. The predicted molecular mass of *Drosophila* alpha-spectrin is 278,364, which is somewhat larger than estimates based on mobility in SDS gels (234-kD

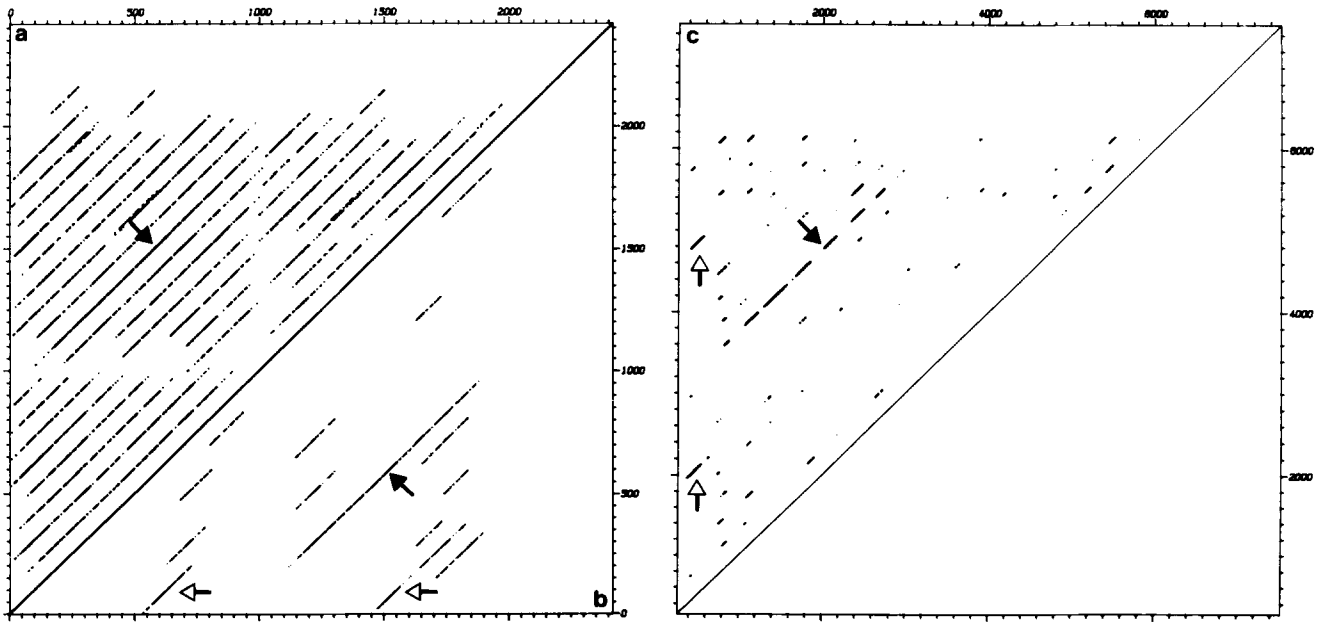

Figure 2. Identification of repetitive domains within alpha-spectrin. *a* and *b*, The full-length sequence of alpha-spectrin was compared with itself in windows of 100 residues at low stringency (*a*, 20 identical matches/window) or high stringency (*b*, 34 matches/window). A dot marks the position of each amino acid identity within the region of comparison. *c*, The coding region of the alpha-spectrin DNA sequence was also compared with itself using a window of 100 residues and a stringency of 56 nucleotide matches. The solid arrow marks a line of duplication between segments 2–8 and 11–17. The open arrow marks the duplication line of segment one.

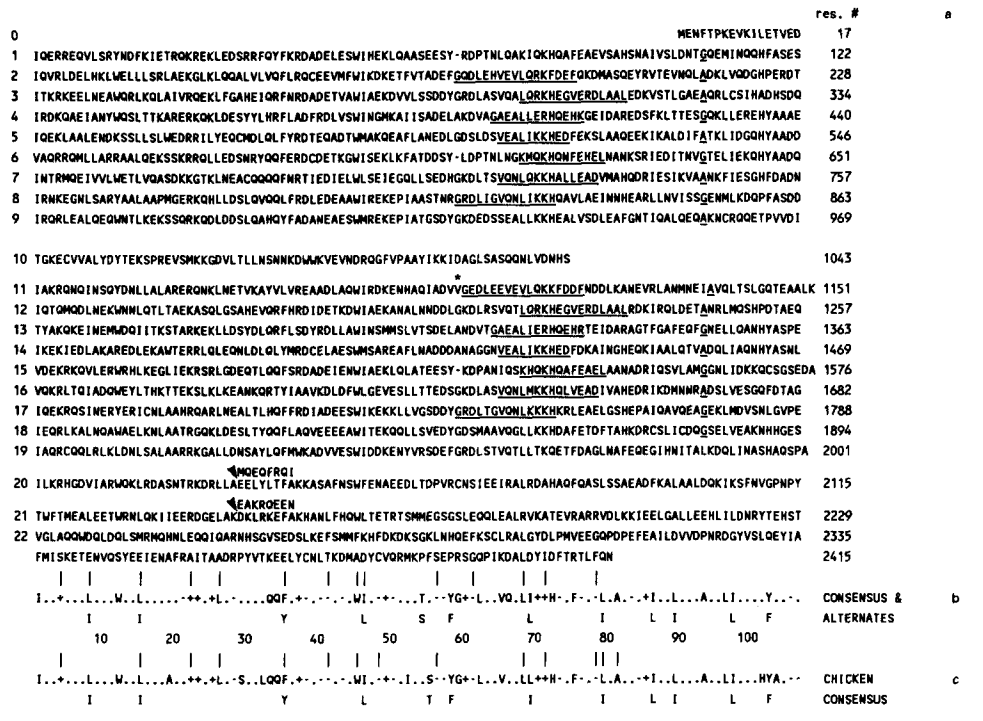

Figure 3. Alignment and consensus of the full-length amino acid sequence of alpha-spectrin. *a*, The alpha-spectrin sequence is presented in a vertical register of 106 amino acid repeats. Repetitive and nonrepetitive segments are numbered consecutively and the residue number at the end of each row is indicated to the right. Single residue gaps (−) must be added in segments 1, 6, and 15 to optimize the alignment of repetitive residues. The asterisk over segment 11 marks the approximate location of a single residue insertion. An 8-residue sequence (*arrowheads*) was removed from segments 20 and 21 to maximize their alignment with the other segments. Some of the sequences that give rise to the line of duplication shown in Fig. 2 are underlined. The positions of segments 0, 10, and 22 do not represent a known relationship to the repetitive segments. *b*, A 54-residue consensus of the repetitive segments was based on the occurrence of a specific residue or conservative substitution (alternates) in at least half of the segments. Conserved residues that occur in at least 15 segments are indicated by a vertical bar. *c*, A 59-residue consensus was determined from the chicken alpha-spectrin sequence (Wasenius et al., 1989) using the same criteria as for the *Drosophila* consensus.

for *Drosophila* alpha-spectrin; Dubreuil et al., 1987). The molecular mass of chicken alpha-spectrin is also larger than estimates based on gel mobility (Wasenius et al., 1989).

Dotplots of the predicted amino acid sequence (Fig. 2, *a* and *b*) demonstrate the presence of two major repetitive domains within alpha-spectrin. The first begins at the amino terminus of the protein and extends to about residue 1,000, thus defining repetitive segments 1–9. A similar pattern con-

Table I. Conservation of Consensus and Nonconsensus Residues between Drosophila and Chicken Alpha-Spectrin

Segment (residues)	At Drosophila consensus positions						
	(A) Consensus residues in fly	(B) Chicken consensus matches	(C) Nonconsensus residues in fly	(D) Chicken nonconsensus matches	(E) Chicken matches at nonconsensus positions	(F) Total nonconsensus matches	(G) Total percent similarity
1 (105)	36	36	18	17	18	35	68
2 (106)	38	37	16	12	22	34	67
3 (106)	44	39	10	4	23	27	62
4 (106)	32	28	22	17	27	44	68
5 (106)	39	37	15	13	33	46	78
6 (105)	36	35	18	11	16	27	60
7 (106)	42	38	12	7	26	33	67
8 (106)	39	37	15	10	23	33	66
9 (106)	37	33	17	13	27	40	69
11 (108)	35	28	19	7	14	21	45
12 (106)	40	34	14	6	26	32	62
13 (106)	29	28	25	18	29	47	71
14 (106)	37	36	17	8	25	33	65
15 (107)	38	38	16	14	27	41	76
16 (106)	39	37	15	5	28	33	66
17 (106)	37	36	17	9	29	38	70
18 (106)	41	35	13	5	26	31	62
19 (107)	35	35	29	12	25	37	68
20 (114)	19						
21 (114)	15						
	Totals:	627	308	188	444	632	

tinues between residues 1,100 and 2,000, defining repetitive segments 11–19. At low stringency (Fig. 2 *a*) nearly all of the repetitive segments show similarity to one another with a periodicity of 106 residues.

The protein sequence can be aligned in a vertical register of repeating segments (Fig. 3 *a*) beginning with residue 18. Most of the segments are 106 residues in length and are numbered using the nomenclature of Speicher and Marchesi (1984) so that the first repetitive segment is numbered "1." The repetitive segments are defined by a consensus of conserved residues at 54 positions within the repeating unit (Fig. 3 *b*). A consensus residue is defined as one that occurs in at least half (9 out of 18) of the repeats. Vertical bars indicate the more stringent consensus positions that are conserved in 15 out of the 18 repeating segments. Only one position is conserved in all of the repetitive segments: the tryptophan at position 45. Five types of conservative substitution have been allowed in the determination of consensus positions: lysine and arginine often appear to be interchangeable, as do glutamic and aspartic acid. These residues are indicated by a "+" or "−", respectively, in the consensus. Likewise, leucine and isoleucine, tyrosine and phenylalanine, and serine and threonine have been grouped together as indicated in Fig. 3 *b*. Segments 1–19 (excluding segment 10) include from 29–44 (average = 37 ± 4, Table I) of these consensus residues.

The degree of similarity between repetitive segments provides information on their origin. Most of the similarity points in low-stringency dotplots (Fig. 2 *a*) result from consensus residues that are shared between segments. Higher stringency dotplots (Fig. 2 *b*) score only those segments with a high degree of similarity at both consensus and nonconsensus positions. At higher stringency, a sequence that extends from segments 2–8 appears to be similar to a sequence that extends from segments 11–17 (Fig. 2 *b*, *solid arrow*). The line of duplication (which also occurs in chicken alpha-spectrin; Wasenius et al., 1989) indicates that the two sequences probably arose by duplication of a common ancestor. Some of the nonconsensus residues that give rise to the line of duplication are underlined in Fig. 3 *a*. Further evidence for a duplication event comes from dotplots of the DNA sequence (Fig. 2 *c*) that show a line of duplication between segments 2–8 and 11–17. The sequence corresponding to segment 1 also shows similarity to segments 6 and 15 in protein and DNA dotplots (Fig. 2, *open arrows*).

Segments 10 and 20–22 do not appear to be repetitive either by dotplot analysis or by consensus residue content. Segments 20 and 21 share relatively few consensus residues with segments 1–19 (Table I) or with each other. As a result, most of the repetitive segments do not score matches with segments 20 and 21 in low stringency dotplots (Fig. 2 *a*). In addition, eight residues must be deleted to maximize their alignment relative to the consensus (Fig. 3 *a*, *arrowheads*). Although they were probably derived from the same ancestor as the repeating segments, segments 20 and 21 appear to represent a different lineage of spectrin segments whose sequences have not been subject to the same constraints as segments 1–19. Segments 10 and 22 are not similar in sequence to any of the other spectrin segments. Thus we consider segments 10 and 20–22 to be nonrepetitive when compared to segments 1–9 and 11–19.

Repetitive Aspects of Spectrin Structure

The current methods of structural prediction are of limited accuracy when applied to a single sequence, but the average predicted structure from 18 segments will reveal repetitive sequence patterns that are likely to be related to spectrin structure. The amino acid sequence from each repetitive segment of *Drosophila* spectrin was divided into blocks of 5 residues and scored for the presence or absence of predicted alpha helical structure or turn sequences. The results for each block of residues obtained from the Garnier prediction (using no decision constant) were tabulated in the histograms shown in Fig. 4. With the exception of residues near positions 60 and 100, the repetitive segments were predicted to be largely alpha helical. There was a high probability of turn structure around positions 59 (a conserved glycine) and 100 where many segments contain a glycine or proline. The Garnier method with no decision constant also predicted that segment 10 is largely nonhelical with a high probability of turns throughout. Conversely, segments 20–22 are predicted to be largely alpha helical with no turn sequences in segment 21 (not shown).

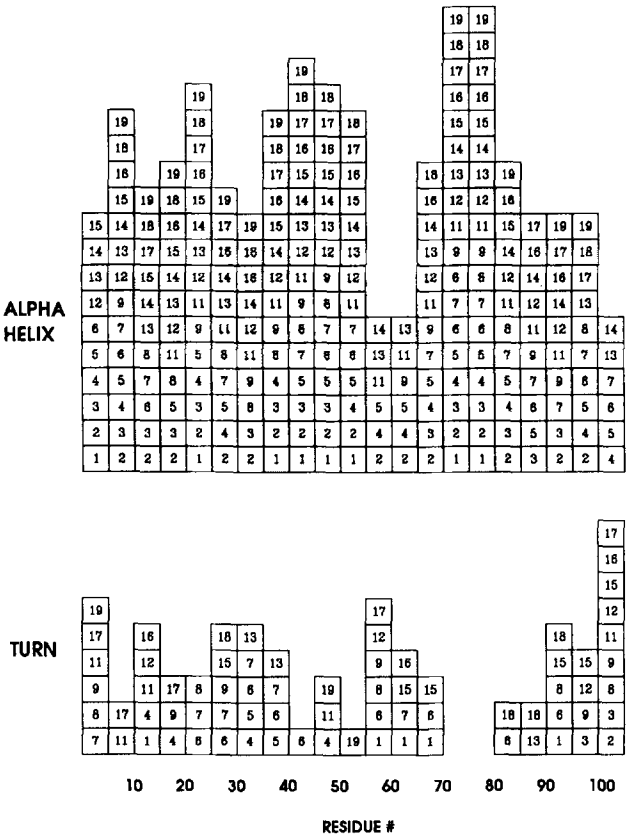

Figure 4. Predicted structural consensus of the repetitive segments of alpha-spectrin. Each repetitive segment from Fig. 3 was divided into sequential 5 residue blocks and scored for the presence of either predicted alpha helical sequence (*upper*) or turn sequences (*lower*). Predictions were based on the method of Garnier et al. (1975) using no decision constant. One or more turn residues in a block resulted in a turn score for the entire block. A block was only scored alpha helical if there were three or more helical residues in a row. Blocks that did not meet either criterion were not included in the histograms. The segment number (as in Fig. 3) is indicated in each block.

Comparison of the Drosophila Spectrin Sequence to Vertebrate Spectrins

A comparison of partial spectrin sequence data from a number of sources indicated that while nonerythroid spectrins are ~90% identical to one another, human erythroid spectrin and human nonerythroid spectrin are <60% identical (McMahon et al., 1987; Leto et al., 1988). Based on full-length sequence comparisons, we found that the *Drosophila* alpha-spectrin sequence is 63% identical to chicken brain alpha-spectrin (Wasenius et al., 1989) and 46% identical to human erythroid alpha-spectrin (P. Curtis and B. G. Forget, Yale University School of Medicine, personal communication). In both comparisons gaps were considered as nonidentical positions.

The primary structures of vertebrate and invertebrate alpha-spectrin are remarkably similar: both molecules consist of two large repetitive domains, a nonrepetitive segment 10 and a large carboxy terminal nonrepetitive sequence (segments 20–22). The position of the carboxy terminus of fly spectrin is identical to that reported for chicken alpha-spectrin (Wasenius et al., 1989), and both of these sequences extend considerably further than the reported carboxy terminal sequence of rat brain alpha-spectrin and human brain alpha-spectrin (Leto et al., 1988). In segment 10, 56 out of 74 residues are identical between *Drosophila* and chicken alpha-spectrin. These residues also correspond to a region of similarity between v-src, v-crk and phospholipase C (Lehto et al., 1988). A major difference between fly and chicken alpha-spectrin is a 21-residue extension of chicken segment 10 that is not found in the fly. Another important difference between the fly and chicken sequences lies in segment 11: the chicken sequence has a 35-residue extension that is not present in *Drosophila* spectrin. The latter region of vertebrate nonerythroid alpha-spectrins includes a substrate site for calcium-activated protease and a binding site for calmodulin (Harris et al., 1988). While *Drosophila* spectrin also binds calmodulin in blot overlay experiments (Dubreuil et al., 1987), it lacks the high-affinity binding site present in segment 11 of the vertebrate protein and instead appears to bind calmodulin at a site within segment 15 (Dubreuil, R. R., manuscript in preparation). The extensions of segments 10 and 11 in vertebrate nonerythroid spectrins are probably specializations that are important in the regulation and function of vertebrate spectrin but that are not important in the general function of spectrin as a component of the membrane skeleton in invertebrates.

The occurrence of a consensus sequence in *Drosophila* alpha-spectrin repeats suggests that certain residues are important to the presumed repetitive structure of the molecule. If so, we expected that chicken alpha-spectrin would have a similar set of consensus residues. 50 of the 54 consensus positions found in fly alpha-spectrin are also found in the chicken sequence (Fig. 3 *c*). Two of the fly consensus positions are shifted by one residue in the chicken sequence and two of the fly positions do not appear to be conserved in chicken. In addition, there are seven consensus positions in which a residue is conserved in at least 9 out of 18 chicken repetitive segments, but in fewer than 9 out of 18 fly segments.

Nonconsensus residues in alpha-spectrin appear to be conserved between *Drosophila* and chicken at least as often as consensus residues. In comparing the individual fly and chicken segments (Table I), we find that most of the consensus positions conserved in any fly segment (column *A*) are also present in the corresponding chicken segment (column *B*). For example, segment 1 of *Drosophila* includes 36 of the consensus residues and all of these are present in chicken segment 1. Where a fly residue in an individual segment does not match the consensus of other fly segments (column *C*), there is often a match between the fly residue and the chicken residue in the corresponding chicken segment (column *D*). In segment 1 from *Drosophila*, there are 18 nonconsensus residues at consensus positions and 17 of these residues are identical in chicken segment 1. In addition, there are a large number of nonconsensus positions that are conserved between a given fly segment and the corresponding chicken segment (column *E*). In segment 1, there are 18 identical nonconsensus positions between fly and chicken. The sum of the nonconsensus matches at nonconsensus positions (column *E*) and the nonconsensus matches at consensus positions (column *D*) is the total number of nonconsensus matches (column *F*). A comparison of the total nonconsensus matches (in *F*) to the number of consensus matches (in *B*) shows that consensus and nonconsensus positions have been equally conserved between vertebrate and invertebrate spectrins. This segment by segment comparison also shows that a given segment in *Drosophila* is more similar to the corresponding segment in chicken alpha-spectrin than it is to other segments in *Drosophila*.

Sequence Comparison of Drosophila Alpha-Spectrin and Alpha-Actinin

Previous studies showed that part of the sequence of alpha-spectrin was similar to part of the sequence of alpha-actinin (Wasenius et al., 1989; Baron et al., 1987; Noegel et al., 1987). To identify the precise boundaries of similarity, alpha-spectrin and alpha-actinin were further compared by sequence alignment (Fig. 5). Because the first 356 residues of alpha-actinin did not show significant similarity to alpha-spectrin in dotplots (not shown) the alignment begins at residue 357 of chicken alpha-actinin and residue 1895 of alpha-spectrin (segment 19). The region of alpha-actinin between residues 357 and 509 is compared to the alpha-spectrin consensus, since it is actually more similar to the consensus than to spectrin segment *19*. From residue 510 to the carboxy terminus of alpha-actinin (residues 2,040 to the carboxy terminus of alpha-spectrin: segments 20–22), the comparison is to the actual alpha-spectrin sequence, since the remainder of alpha-spectrin and alpha-actinin are more similar to each other than to the repeat consensus. Alpha-spectrin was aligned with alpha-actinin from *Drosophila*, chicken, and *Dictyostelium* (E. Fyrberg, Johns Hopkins University, personal communication; Baron et al., 1987; Noegel et al., 1987), and the collated results were used to identify the conserved residues of all four sequences (Fig. 5, upper case, *Conserved*). Additional conserved residues could be identified if *Dictyostelium* alpha-actinin was omitted from the comparison (lower case, *Conserved*). Without *Dictyostelium* there were 134 matches (25%), and with *Dictyostelium* there were 70 matches (13%) out of 546 positions. Segment 3 of alpha-actinin, which most closely resembles the spectrin repeats, and seg-

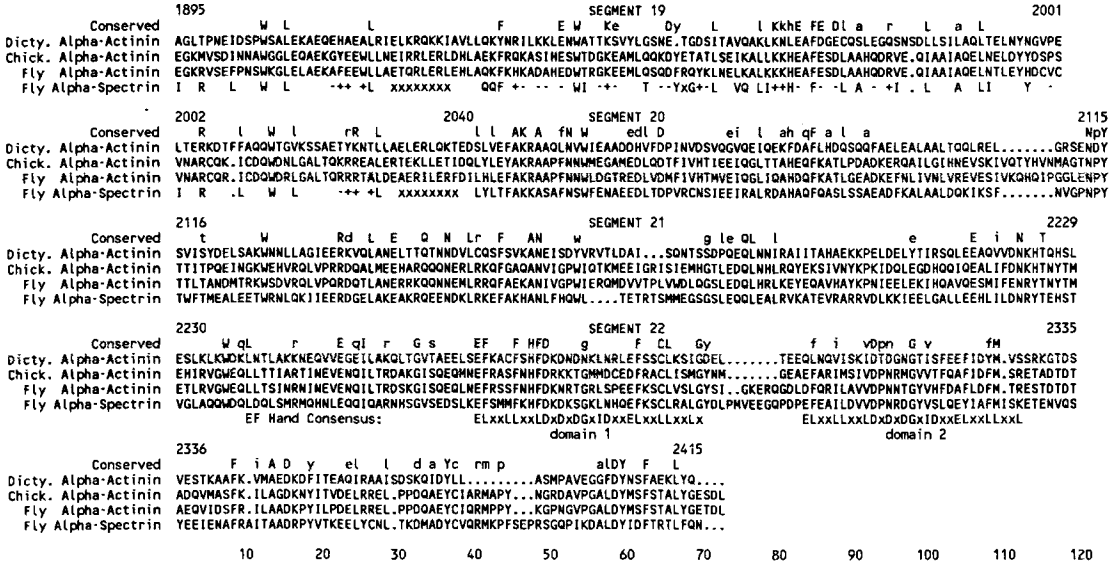

Figure 5. Sequence similarity between alpha-spectrin and alpha-actinin. *Drosophila* alpha-spectrin was aligned in pairwise combinations with *Dictyostelium* alpha-actinin (Noegel et al., 1987), chicken smooth muscle alpha-actinin (Baron et al., 1987) and *Drosophila* alpha-actinin (E. Fyrberg, Johns Hopkins University, personal communication) and assembled into a register of all four sequences with appropriate gaps. The region of comparison in alpha-spectrin extends from the beginning of segment 19 (residue 1,895) through the end of the molecule. The corresponding region in alpha-actinin extends from residue 357 of chicken alpha-actinin through the end of the molecule. Since the region of alpha-actinin between residues 357 and 509 is actually more similar to the alpha-spectrin consensus than to segment 19, the consensus sequence is included in the lineup through residue 2,040. Beyond residue 2,040, the alpha-spectrin sequence is more like alpha-actinin than the consensus. Residues shared by all four sequences (*upper case*), and residues shared by the sequences excluding *Dictyostelium* alpha-actinin (*lower case*) are shown on the *Conserved* line. Two 8-residue gaps (*x*) are necessary to align the spectrin consensus with alpha-actinin. In segment 22, the EF hand consensus sequence (Kretzinger, 1980) is shown below the alpha-spectrin sequence.

ment 4 of alpha-actinin include 8-residue insertions (*x*) relative to the alpha-spectrin consensus.

The region of similarity between alpha-spectrin segment 22 and alpha-actinin includes two EF hand domains that are thought to confer calcium sensitivity to the interaction between *Dictyostelium* alpha-actinin and actin (Noegel et al., 1987; Fig. 5). Like vertebrate nonerythroid spectrin (Wasenius et al., 1989), *Drosophila* alpha-spectrin includes EF hand sequences that match the consensus sequence described by Kretzinger (1980) for calcium binding domains (not shown).

A Structural Comparison of Drosophila Spectrin and Alpha-Actinin

We have incorporated features of the *Drosophila* alpha-spectrin sequence into a schematic model that delineates important structural boundaries within the molecule (Fig. 6).

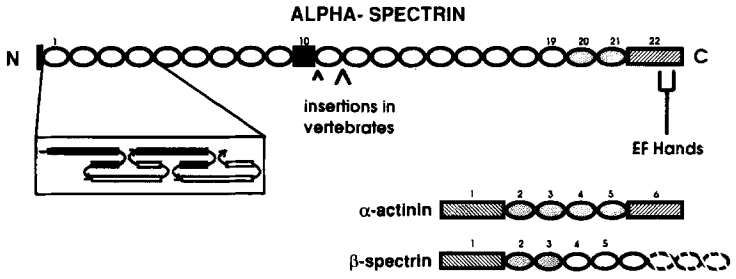

Figure 6. A structural model of alpha-spectrin. The repetitive segments of alpha-spectrin are presented as uniform ellipses and nonrepetitive sequences are presented as rectangles. Shaded ellipses represent alpha-actinin–like segments, which are generally longer than alpha-spectrin repetitive segments because of an 8-residue insert and are only marginally similar to the alpha-spectrin repeat consensus. All sequence lengths are represented uniformly. Vertebrate nonerythroid spectrins include a 21- and 35-residue extension (*arrowheads*) relative to the *Drosophila* spectrin sequence. All sequences are presented with the amino terminus (*N*) to the left. Segments 19–22 of alpha-spectrin resemble segments 4–6 of alpha-actinin. Segments 1–3 of alpha-actinin resemble the first three segments of *Drosophila* beta-spectrin (Byers et al., 1989). Segment 19 of alpha-spectrin and segment 3 of alpha-actinin and beta-spectrin are all similar, but the greatest resemblance is between the latter two sequences, which contain an 8-residue insert relative to the alpha-spectrin repeat consensus. Segments 4 and 5 of beta-spectrin are more closely related to the alpha-spectrin consensus than to any alpha-actinin sequence (Byers et al., 1989). *Inset*, structural predictions suggest that *Drosophila* alpha-spectrin (like human erythroid spectrin; Speicher and Marchesi, 1984) is arranged in a series of triple barrel alpha-helices connected by short turn regions. The phase of the structural repeat is somewhat staggered relative to the sequence repeat so that the amino terminal end of segment 1 projects out of the triple barrel series. Boundaries between sequence repeats are indicated by arrowheads, and alternate repeats are shaded.

The model also facilitates a comparison of the relevant domains of alpha-actinin and beta-spectrin (Byers et al., 1989). Baron et al., (1987) suggest that alpha-actinin is composed of an amino terminal segment that binds actin, four spectrin repeat-like segments, and a carboxy terminal segment with two EF hands (numbered *1–6*, respectively, in Fig. 6). Segments 3–6 of alpha-actinin resemble the segments 19–22 of alpha-spectrin (Fig. 6). The native alpha-actinin molecule is an antiparallel homodimer, and Noegel et al. (1987) have suggested that EF hand sequences near the carboxy terminus of the polypeptide interact with the putative actin binding amino-terminal sequence to regulate the actin-cross-linking activity of the native molecule. The native spectrin molecule is a tetramer composed of two antiparallel heterodimers joined head to head so that each end of the native molecule includes the carboxy terminus of the alpha subunit and the amino terminus of the beta subunit. We recently discovered that the amino terminus of *Drosophila* beta-spectrin is closely related to the amino terminus of alpha-actinin (Byers et al., 1989). These carboxy terminal and amino terminal sequence similarities between alpha-actinin and spectrin (Fig. 6) clarify the relationship of the two molecules. Both are actin cross-linkers with functionally similar ends. But, because the actin binding domains of alpha-actinin cannot be farther apart than 40 nm (the length of the alpha-actinin dimer), whereas the actin binding domains of spectrin may be as far apart as 180 nm (the contour length of the spectrin tetramer), these two molecules are likely to have different functional roles in the cytoskeleton.

An 8-residue insert distinguishes the segments of alpha-actinin and their homologues in spectrin (which are ~114 residues long; Fig. 6, *shaded segments*) from the repetitive 106-residue segments. Segments 1–3 of alpha-actinin are most closely related to beta-spectrin (Byers et al., 1989), whereas segments 4–6 of alpha-actinin are related to alpha-spectrin. If all of the segments of alpha-actinin correspond to the unique 114-residue-long segments of alpha- and beta-spectrin, what then was the origin of the 106-residue repeat?

Discussion

Evolution of Alpha-Spectrin

Spectrin and alpha-actinin appear to have evolved from a common ancestral sequence, but the details of their evolution are obscured by the substantial divergence between individual segments in both proteins. Nonetheless, we can deduce some of the evolutionary events that gave rise to the present day proteins through identification of their common properties. First, there are two lineages of spectrin segments that appear to have evolved independently of one another. Both 106-residue and 114-residue segments are found in alpha-spectrin, but only the latter lineage is found in alpha-actinin. The distinguishing features of these two lineages have been maintained through spectrin evolution so that segments 20 and 21 of alpha-spectrin are more like segments 4 and 5 of alpha-actinin, respectively, than like any other segments of alpha-actinin or spectrin. Conversely, the 106-residue segments are characterized by their similarity to one another and by their differences from the 114-residue segments in alpha-actinin and spectrin. The divergence of the two lineages probably occurred once and the multiplication of the segments of each lineage was independent of the other lineage. Second, the 106-residue repetitive segments found in beta-spectrin (Speicher and Marchesi, 1984) and alpha-spectrin probably arose by a single event in a common ancestor, rather than by independent events in two separate genes. Thus, alpha- and beta-spectrin may have evolved from a single ancestral protein that resembled alpha-actinin.

The distribution of sequence patterns in the repetitive regions of alpha-spectrin together with the above conclusions have been incorporated into a model of spectrin evolution (Fig. 7). While steps *5* and *6* are supported by the distribution of unique sequence markers within alpha-spectrin, steps *1–4* are hypothetical in detail but illustrate some of the problems in spectrin evolution. The evolutionary scheme begins with an alpha-actinin-like molecule with multiple 114-residue segments: the duplication of 114-residue segments may have preceded the modification and multiplication of 106-residue segments since the former have diverged from one another more extensively than the latter. The formation of separate alpha- and beta-spectrin subunits probably occurred after duplication of the 106-residue repetitive segment (step *2*). Once complete sequence data are available for beta-spectrin, it should be possible to determine whether the repetitive regions of beta-spectrin arose by the same multiplication steps or steps different from those that generated the repetitive alpha-spectrin segments. Additional duplication events probably account for the expansion of the repetitive domain of alpha-spectrin (steps *3–5*), perhaps through unequal crossing over of ancestral genes. Dotplots reveal the boundaries, as judged by sequence similarity, of two relatively recent duplication events (steps *5* and *6*). Based on amino acid sequence similarity (the sequence constellation around residue 60 and the glycine residue at position 93), segment 1 appears to have arisen by duplication of a sequence like segments 6 and 15 (step *6*). The extent of DNA sequence similarity between segments 1, 6, and 15 suggests that this event occurred at about the same time as the large duplication shown in step *5*. The fact that segment 10 is not repeated in the alpha-spectrin sequence suggests that it was also acquired relatively late in the evolution of alpha-spectrin, after multiplication of the repetitive segments.

Implications for a Spectrin Superfamily of Proteins

The presence of spectrinlike sequences in other proteins has led to the notion that there is a spectrin superfamily whose members include alpha- and beta-spectrins (erythroid and nonerythroid), alpha-actinins (muscle and nonmuscle) and dystrophin (Byers et al., 1989; Davison et al., 1989). In particular, the conservation of segments 1 and 2 between alpha-actinin, beta-spectrin and dystrophin (Byers et al., 1989) suggests that these segments contribute to a common function of all three proteins, probably actin binding activity. However, the relative roles of the two subunits (and their corresponding amino terminal and carboxy terminal sequences) in the actin binding activity of spectrins and alpha-actinins are not yet known.

Spectrin, alpha-actinin, and dystrophin all share repetitive and nonrepetitive sequence elements. If the common function of actin binding activity among these proteins is contained in their homologous nonrepetitive sequences, what then is the function of the repetitive segments? We suggest

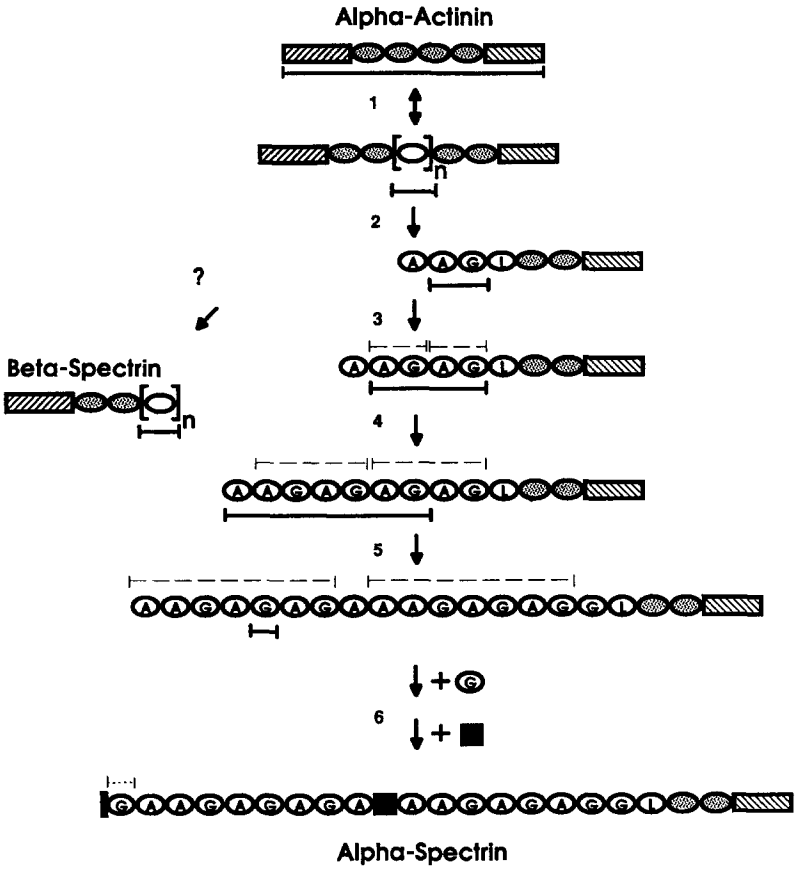

Figure 7. Evolution of alpha-spectrin. Based on features of the alpha-spectrin sequence we propose the following model for the evolution of spectrin: (1) The sequence of alpha-actinin was duplicated at some point in evolution so that similar sequences are now found both in alpha-actinin and spectrin, although it is not clear which came first. Another early event was the divergence of 106-residue segments found in spectrin from the 114-residue segments found in alpha-actinin and spectrin. (2) The 106-residue segment was multiplied at some point before the separation of alpha- and beta-spectrin from a common coding sequence. (*3–6*) The occurrence of alanine and glycine residues at position 93 in alternating repetitive alpha-spectrin segments and the duplication lines found in dotplots were incorporated into a duplication scheme that accounts for their distribution in the present alpha-spectrin molecule; horizontal brackets mark the boundaries of sequence duplications. The single-letter code within segments 1–18 indicates the residue present at position 93 within each segment.

that the repetitive segments serve as molecular spacers that determine both the positioning of the actin binding sites at the ends of the molecule and the functional actin cross-linking distance of these sites. These roles may be clarified if additional members of the spectrin superfamily are found occupying presently unknown niches in the eukaryotic cytoskeleton. The identification of such proteins would also be helpful in the elucidation of duplication steps during evolution of the spectrin superfamily. Finally, the sequences of spectrinlike proteins from *Dictyostelium* (Bennett and Condeelis, 1989) and *Acanthamoeba* (Pollard, 1985) should in the future provide additional insight into the problems of spectrin evolution and function.

We thank Maureen Dwyer and Eric Brandin for their expert technical assistance in parts of this project. We also thank Drs. Les Kotula, Peter Curtis, Bernard Forget, and Eric Fyrberg for sharing their sequence data before publication. We thank Ray Erikson and Dave Alcorta for use of and assistance with their oligonucleotide synthesizer, and finally we thank Dr. Stephen Harrison for stimulating discussions on the structure of spectrin.

This work was supported by National Institutes of Health grant No. GM39686. R. R. Dubreuil was a postdoctoral fellow of the American Cancer Society.

Received for publication 8 May 1989 and in revised form 30 June 1989.

References

Baron, M. D., M. D. Davison, P. Jones, and D. R. Critchley. 1987. The sequence of chick alpha-actinin reveals homologies to spectrin and calmodulin. *J. Biol. Chem.* 262:17623–17629.

Bennett, H., and J. Condeelis. 1988. Isolation of an immunoreactive analogue of brain fodrin that is associated with the cell cortex of *Dictyostelium* amoebae. *Cell Motil. Cytoskeleton.* 11:303–317.

Birkenmeier, C. S., D. M. Bodine, E. A. Repasky, D. M. Helfman, S. H. Hughes, and J. S. Barker. 1985. Remarkable homology among the internal repeats of erythroid and nonerythroid spectrin. *Proc. Natl. Acad. Sci. USA.* 82:5671–5675.

Brown, N. H., and F. C. Kafatos. 1988. Functional cDNA libraries from *Drosophila* embryos. *J. Mol. Biol.* 203:425–437.

Byers, T. J., R. R. Dubreuil, D. Branton, D. P. Kiehart, and L. S. B. Goldstein. 1987. *Drosophila* spectrin II. Conserved features of the alpha subunit are revealed by analysis of cDNA clones and fusion proteins. *J. Cell Biol.* 105:2103–2110.

Byers, T. J., A. Husain-Chishti, R. R. Dubreuil, D. Branton, and L. S. B. Goldstein. 1989. Sequence similarity of the amino-terminal domain of *Drosophila* beta-spectrin to alpha-actinin and dystrophin. *J. Cell Biol.* In press.

Chen, E. Y., and P. H. Seeburg. 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA (NY).* 4:165–170.

Davidson, M. D., M. D. Baron, D. R. Critchley, and J. C. Wootton. 1989. Structural analysis of homologous repeated domains in alpha-actinin and spectrin. *Int. J. Biol. Macromol.* 11:81–90.

Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387–395.

Dubreuil, R., T. J. Byers, D. Branton, L. S. B. Goldstein, and D. P. Kiehart. 1987. *Drosophila* spectrin I. Characterization of the purified protein. *J. Cell Biol.* 105:2095–2102.

Garnier, J., D. J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* 120:97–120.

Harris, A. S., D. E. Croall, and J. S. Morrow. 1988. The calmodulin-binding site in alpha-fodrin is near the calcium-dependent protease-I cleavage site. *J. Biol. Chem.* 263:15754–15761.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene (Amst.).* 28:351–359.

Holmes, D. S., and M. Quigley. 1981. A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* 114:193–197.

Koenig, M., A. P. Monaco, and L. M. Kunkel. 1988. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell.* 53:219–218.

Kretzinger, R. H. 1980. Crystallographic studies of calmodulin and homologs. *Ann. NY Acad. Sci.* 356:14–19.

Lehto, V. P., V. M. Wasenius, P. Salven, and M. Saraste. 1988. Transforming and membrane proteins. *Nature (Lond.)*. 334:388.

Leto, T. L., D. Fortugno-Erikson, D. Barton, T. L. Yang-Feng, U. Francke, A. S. Harris, J. S. Morrow, V. T. Marchesi, and E. J. Benz, Jr. 1988. Comparison of nonerythroid alpha-spectrin genes reveals strict homology among diverse species. *Mol. Cell Biol.* 8:1–9.

McMahon, A. P., D. H. Giebelhaus, J. E. Champion, J. A. Bailes, S. Lacey, B. Carritt, S. K. Henchman, and R. T. Moon. 1987. cDNA cloning, sequencing and chromosome mapping of a non-erythroid spectrin, human alpha-fodrin. *Differentiation*. 34:68–78.

Noegel, A., W. Witke, and M. Schleicher. 1987. Calcium sensitive non-muscle alpha actinin contains EF hand structures and highly conserved regions. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 221:391–396.

Pesacreta, T. C., T. J. Byers, R. R. Dubreuil, D. P. Keihart, and D. Branton. 1989. *Drosophila* spectrin: the membrane skeleton during embryogenesis. *J. Cell Biol.* 108:1697–1709.

Pollard, T. D. 1984. Purification of a high molecular weight actin filament gelation protein from *Acanthamoeba* that shares antigenic determinants with vertebrate spectrins. *J. Cell Biol.* 99:1970–1980.

Sanger, F., S. Miklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*. 74:5463–5467.

Speicher, D. W., and V. T. Marchesi. 1984. Erythrocyte spectrin is comprised of many homologous triple helical segments. *Nature (Lond.)*. 311:177–180.

Wasenius, V.-M., M. Saraste, J. Knowles, I. Virtanen, and V.-P. Lehto. 1985. Sequencing of the chicken non-erythroid spectrin cDNA reveals an internal repetitive structure homologous to the human erythrocyte spectrin. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:1425–1430.

Wasenius, V.-M., M. Saraste, P. Salven, M. Eramaa, L. Holm, and V.-P. Lehto. 1989. Primary structure of the brain alpha-spectrin. *J. Cell Biol.* 108:79–93.