Journal of Cellular Biochemistry 30:245–258 (1986)
Membrane Skeletons and Cytoskeletal–Membrane Associations 131–144

The Present Status of Erythrocyte Spectrin Structure: The 106-Residue Repetitive Structure Is a Basic Feature of an Entire Class of Proteins

David W. Speicher

Department of Pathology, Yale University School of Medicine, New Haven, Connecticut 06510

Spectrin, the major component of the erythroid membrane skeleton, is a long, asymmetrical rodlike protein that interacts with several other proteins to form a two-dimensional membrane skeleton. Progress in several laboratories over the past few years including substantial partial peptide and nucleotide sequence determination has greatly enhanced our knowledge of the structrual properties of this large molecule (heterodimer = 465,000 daltons). The alpha and beta subunits are homologous with approximately 30% identity. They are aligned in an antiparallel side-to-side orientation with the amino- and carboxy-termini near opposite physical ends of the molecule. The predominant structural feature elucidated from sequencing this large molecule is the nearly universal occurrence in both subunits of a single type of repetitive structure. The periodicity of this homologous structure is exactly 106 amino acid residues. As many as 36 homologous, but nonidentical, repeats exist and comprise more than 90% of the mass of the heterodimer. Each of these repetitive units is folded into a triple-stranded structure that is highly helical. Peptide maps, antibody crossreactivity, peptide sequence analysis, and more recently nucleic acid sequences have defined several major properties of the erythroid molecule and related proteins in other tissues. Tissue-specific spectrins have the same 106-residue repetitive structure and show sequence homology to erythroid spectrin.

Key words: erythrocyte spectrin, sequence analysis, spectrin structure, spectrin homology, spectrinlike, protein conformation, spectrin evolution

Spectrin was first isolated from low-ionic-strength extracts of erythrocyte membranes nearly two decades ago [1]. Since its discovery, spectrin and other components of the red cell membrane have been extensively studied by numerous investigators. Spectrin is the predominant component of the membrane skeleton, which is a largely two-dimensional protein lattice on the cytoplasmic side of the membrane. This

Received August 13, 1985; revised and accepted October 9, 1985.

© **1986 Alan R. Liss, Inc.**

structure is commonly thought to be primarily responsible for maintaining membrane integrity. Despite spectrin's large size (the heterodimer is approximately 465,000 daltons), it has been extensively studied structurally and several functionally important protein-protein interactions have been defined.

The simplest molecular form of spectrin that can be isolated without denaturants is an $\alpha\beta$ heterodimer indicating that the two subunits are strongly, noncovalently associated. The heterodimer can undergo several additional levels of self-association. Dimers self-associate to tetramers under appropriate conditions [2,3], and this association is in a head-to-head orientation. Tetramers are approximately twice the length of the individual dimers [4]. More recently, higher-order oligomers of spectrin have been identified and characterized in vitro [5], although some investigators question the significance of these oligomers in vivo.

Spectrin interacts directly with at least several other proteins of the red cell membrane including ankyrin [6–8] and actin [9–12]. Protein 4.1 modulates spectrin-actin interaction [13–17] and these three proteins form a high-affinity ternary complex that requires both subunits of spectrin [18]. Protein 4.1 binds to the spectrin dimer at the end opposite the self-association site [19]. In addition to these better-characterized interactions, other proteins play an important role in the red cell membrane skeleton, and some may interact directly or indirectly with spectrin. Numerous recent reviews cover functionally related interactions of red cell membrane proteins and their roles in hemolytic disorders [20–27].

Relatively recently, spectrinlike proteins have been identified in other tissues [28–31], and the old notion that most red cell membrane proteins were unique to this highly specialized cell had to be abandoned. It quickly became evident that spectrinlike proteins could be found in almost any type of cell including single cell organisms such as Acanthameoba [32]. It is possible that some identifications of spectrinlike proteins in nonerythroid tissues that rely solely on antibody crossreactivity may ultimately prove to be inaccurate. However, there are sufficient cases where the spectrinlike proteins have been investigated in great deal, including careful peptide mapping and even determination of partial sequence, to establish clearly a class of closely related, but nonidentical, spectrins.

This review will concentrate on the structural properties of human erythrocyte spectrin that have been elucidated over the past several years. This protein has been studied in great detail, thereby providing a basis for comparison to other spectrinlike proteins.

RESULTS AND DISCUSSION

Basic Properties and Domain Structure of Human Erythrocyte Spectrin

Historically, the shape, size, and molecular homogeneity of the spectrin dimer had been the subject of controversy as a result of conflicting early physical and protein chemical experiments [see earlier reviews, ref. 33–35]. These uncertainties have now been largely resolved. Currently, most investigators in the field seem to accept the electron microscopic images that Branton and co-workers obtained after low-angle rotary shadowing [4,36] as closely representative of gross molecular topography in solution. Using this technique, the spectrin dimer is a long flexible rod of roughly uniform thickness with a unit length of 100 nm per dimer. This rodlike structure is highly flexible throughout its length with tight associations near the ends of the 100

nm rod between two side-to-side strands that are the individual subunits. Other studies suggest that the rotary shadowed images are reasonably close representations of the molecular shape in the membrane as well as in solution [37–39].

Human red cell spectrin is comprised of a single pair of nonidentical polypeptide chains. Peptide maps of the subunits have established that the smaller subunit is distinctly different from the larger subunit [40,41]. More recently, peptide sequence analysis has shown that the two subunits are related, but the degree of homology is quite low [42]. The amino-terminals of both subunits are blocked [40,41], a circumstance that probably contributed to misleading early reports that spectrin was composed of many different polypeptides with multiple termini. At least four covalent phosphates are located near the carboxyl-terminal end of the beta subunit in a protease-sensitive region of approximately 10,000 daltons [43]. While the precise size of each subunit will not be established until the complete sequence is known, commonly used molecular weights are 240,000 daltons for the alpha subunit and 225,000 daltons for the beta subunit, yielding a heterodimer weight of 465,000 daltons.

This large molecule can be subdivided into more manageable pieces that have been termed chemical domains [41]. Under appropriate conditions, mild tryptic proteolysis produces a complex, reproducible pattern of intermediate-sized peptides that can be resolved on two-dimensional polyacrylamide gels. The ordering of nine unique domain peptides representing most of both subunits and precursor-product relationships of nonunique peptides have been established using two-dimensional mapping on cellulose sheets [41,46,47]. A domain map of both subunits has been constructed [47, Fig. 1] that has been used for both structural and functional studies. This method of subdividing the molecule has also proven useful as a convenient diagnostic tool for preliminary localization of molecular defects in hemolytic disorders [25,26,48–56].

While mild proteolytic treatment to produce domain peptides is useful, there are practical limitations to this method. The mapping method, based on iodination of tyrosines, produces simple enough maps for analysis of large polypeptide chains such as spectrin; however, apparently unique peptides may actually contain overlapping unlabeled regions. Also, at least several spectrin peptides migrate anomalously on polyacrylamide gels yielding spurious molecular weights [44,45]. A more serious limitation of the domain map is the loss of regions between the identified domain peptides during mild proteolytic cleavage. Regions between domains that are now known to be lost (proteolysed to small peptides) include both blocked amino-terminals [42], the phosphorylated carboxyl-terminal 10,000 daltons of the beta subunit [43,47], and regions between at least some of the domain peptides [42, Fig. 1]. Actual loss of interdomain regions may be responsible for loss of some protein-binding sites, ie, the ankyrin binding site [58]. As indicated below, a more precise model of spectrin substructure is based on peptide and nucleic acid sequence data. Although the model of spectrin structure is being further refined, division of the spectrin dimer into intermediate-sized domain peptides still remains a useful experimental tool in some cases.

Spectrin Structure Is Highly Repetitive

Probably ever since its discovery, some investigators have thought that spectrin might contain repeating sequences. Repeat unit sizes of approximately 8,000 daltons [59] and 50,000 dalton [46] were hypothesized by different investigators based

primarily upon sizes of peptide fragments produced by mild proteolysis. These conclusions were quite tentative since peptide maps of fragments from mild tryptic digestion of spectrin [41,47] failed to detect conserved or repetitive elements. Also, a number of monoclonal antibodies were specific for individual domains [60].

Despite these negative results, the idea that spectrin should contain repetitive sequences was quite reasonable. Current notions of molecular evolution [61–65] indicate that protein sizes increase by two mechanisms, either through fusion of two dissimilar genes to produce a larger hybrid product, or by duplication and fusion of a single gene to produce a larger product with internal repetitive sequences. Both methods of protein enlargement are known to occur, sometimes simultaneously in the same molecule, and the sizes of genetic elements used as basic building blocks vary substantially.

Our sequence determination of the 80,000 dalton α-I doman provided the first solid evidence that spectrin contained repetitive structure [44,45]. A single type of repetitive structure was identified with a periodicity of 106 amino acids. We subsequently completed the sequence of the α-II domain and found that this repetitive structure extended through most of the amino-terminal half of the alpha subunit [42,57]. The low degree of homology between repeats indicates that these duplications occurred a long time ago on an evolutionary timescale or that the protein has a rapid mutation acceptance rate.

Computer-assisted sequence comparisons provide the best methods for examining distant homologies, and most programs provide statistical comparisons indicating significance levels. Table I summarizes results from searches of the α-I and α-II domains using the computer program RELATE [70,71], which searches a single protein sequence for internal repeat units. Alignments are achieved by sliding segments of the sequence down the remainder of the sequence and scoring each possible alignment. The indicated displacements (number of residues the test segment is

TABLE I. Identification of the 106 Residue Repetitive Structure*

Domain	Sequence displacement	Frequency	Average score
Alpha-I[a]	106	193	80.37
	212	144	84.79
	318	118	81.04
	424	25	71.68
	530	9	84.44
Alpha-II[b]	106	150	56.99
	212	41	51.20
	189	12	35.58

*Sequences were searched using the computer program RELATE [70] with the following parameters: search sequence length—50 residues, random runs—50, scoring matrix—mutation data, 250 PAMs [71].
[a]Displacements of the top 489 scoring segment matches are shown. The statistical summary of the top 249 scores with a mean score of 90.9 is 25 S.D. units. A total of 149,331 real comparisons were performed in this analysis.
[b]Displacements of the top 203 scoring segment matches are shown. The statistical summary of the top 146 scores with a mean score of 60 is 12 S.D. units. A total of 58,311 real comparisons were performed in this analysis.

translocated) show alignments with high homology (high scores). All high-scoring displacements in α-I are 106 or multiples of 106. All possible multiples of 106 give high scores, indicating that the only identifiable repetitive structure has this periodicity and that it extends throughout the α-I domain. This repeating structure is highly significant for both domains as indicated by the high statistical scores of 25 (α-I) and 12 (α-II) normalized standard deviation units (S.D. units) for the listed matches. The α-II domain comparison shows the same repeat unit size, and none of these matches include the carboxyl terminal region of α-II (α-10), which is not homologous to the repeat unit. The occasional alignments with a periodicity of 189 and other periodicities that sometimes are detected do not stand up to more detailed homology comparisons. By contrast, the 106-residue periodicity is always detected with very high statistical significance regardless of the search parameters used.

We have subsequently obtained partial sequences from domain peptides throughout the remaining regions of both subunits [42]. These sequences together with the sequences of α-I and α-II total about 37% of the approximately 4,100 residues in both subunits. All of these additional partial sequences are homologous to the basic 106-residue repeat identified in the α-I and α-II domains. We hypothesized at that point that most of the molecule is comprised of this 106 residue repetitive structure [42]. Recent genetic sequences of nonhuman spectrins indicate that this idea is reasonable.

With a repeat unit size of 106 residues, a maximum of 39 full repeats could occur in the heterodimer ($106 \times 39 = 4{,}134$ residues). However, at least one nonhomologous segment was known to exist in the middle of the alpha subunit that was designated α-10 [42]. We also hypothesized that the carboxyl-terminal alpha subunit segment (α-20) was probably derived from the α-10 segment during the last duplication step of the gene for the alpha subunit [42,57].This idea is supported by recent reports of cDNA sequence of the carboxyl terminal regions of a chicken muscle spectrinlike protein that contained a nonhomologous sequence at the carboxyl terminal-most region [66,67]. The carboxyl-terminal region of the beta subunit has several unique properties (in addition to the phosphorylation sites) that suggest that it may also be nonhomologous.

These results suggest that as many as 36 homologous repeat units may exist in spectrin (18 in each subunit). We had previously identified 22 repeat units (13 for alpha and 9 for beta) and, recently, nucleic acid sequence analysis of cDNA clones from both and human erythroid alpha spectrin by Curtis and co-workers provided two additional repeats for the alpha subunit [68,69]. Another alpha subunit repeat is indicated by the closely related chicken muscle protein (see below), and we have subsequently obtained peptide sequence for an additional beta segment. When the currently available peptide and nucleic acid sequence information is combined, 16 of 18 possible repeats in alpha and 10 of 18 possible repeats in beta have been confirmed. Since beta is related to alpha and evolved from a common gene as discussed below, it is likely that the remaining regions in the beta subunit will have the same homologous structure.

The Subunits are Homologous and Have Evolved From a Common Gene

Early efforts to identify homology between the spectrin subunits using antibody crossreactivity and peptide maps were inconclusive. Peptide maps of intact chains [40,41] showed that the beta subunit was not a processed version of the larger alpha subunit, and most investigators interpreted the minimal overlap of the densely popu-

lated maps of intact subunits as incidental. This interpretation of the subunit maps was confirmed by the striking lack of similarity between peptide maps of domains from the two subunits [41,46,47].

The frequent occurrence of the 106-residue repeat in both subunits suggests that most portions of both subunits must have originated from a single ancestral gene coding for a primordial 12,000-dalton peptide (106 residues). This implies that the subunits diverged from a single gene at some intermediate point in the course of their evolution to the current molecule. While the subunits are clearly related, it is difficult to estimate overall homology between the subunits with precision, despite substantial available sequence information. The extensive repetitive structure throughout both subunits actually means that most sequences are homologous to most other sequences. This makes accurate matching of corresponding regions from the different subunits more difficult, especially using incomplete sequence information. The different sizes of the two subunits further complicate sequence comparisons since it is not obvious whether the first repeat in the beta subunit corresponds with the first, second, or third repeat of the alpha subunit.

Actually, one would hope to resolve all of these questions and deduce the path of evolutionary development from the primordial 12,000-dalton ancestor to the current-sized protein by careful analysis of the known sequences. Alignment of known beta subunit segments against all known alpha subunit sequences should produce the best fits (highest homologies) when the most closely related sequences from the alternate subunit are matched. Unfortunately, this approach does not lead to unambiguous answers.

A summary of this type of comparison using the computer program ALIGN [70,71] is shown in Table II. There are several properties of the repeat unit that must be considered in evaluating these comparisons. All portions of the 106-residue repeat are not equally well conserved [see ref. 42], and most available beta subunit sequences are less than full repeats. This results in significant variation in homology comparisons unrelated to evolutionary distance when using less than full repeats. Optimal

TABLE II. Optimal Alignments of Beta Subunit Sequences With the Alpha Subunit*

Beat repeat segment	Most homologous alpha repeat	Identity (%)	Real score	Alignment score (SD units)
1	1	20	59	7.6
	3	18	56	5.5
2,3[a]	5,6	28	149	14.0
	7,8	31	153	17.8
	3	32	108	13.7
	7	28	116	13.3
8	7	26	80	7.5
11,12	7,8	32	147	17.0
17,18	4,5	30	98	10.8

*Entire known sequences from the beta subunit [42] were compared with all alpha subunit sequences using the computer program ALIGN [70]. Parameters: random runs—50, scoring matrix—mutational data, 250 PAMs [71], gap penalty—40. Two alignments to individual beta sequences are reported for cases where differences between comparisons (real score) were not significant. Real scores are partially dependent on sequence length and should only be compared for alignments of a single beta sequence. Identity and alignment score are independent of sequence length.

[a]As indicated in reference 42, this sequence could correspond to either repeats 2,3 or 3,4 based on peptide mapping data.

alignments of sequences from poorly conserved regions of the repeat unit would understate the overall homology between subunits, while alignments involving well-conserved regions would have the opposite effect. It is also known that selected regions of both subunits have developed new functions during the evolutionary development of this molecule. New or modified functions might be expected to place new constraints on selected sequences resulting in anomalously low homologies when such regions are compared to corresponding regions in the alternate subunit. This may explain the unusually low homology of the β-1 repeat (see Table II) that is near the actin-protein 4.1 binding site.

Based on the partial sequence information currently available, a reasonable estimate of subunit homology is approximately 30% identity. This diversity between the erythrocyte spectrin subunits can be contrasted with other proteins with homologous, nonidentical subunits. An interesting example in the same cell is hemoglobin; the alpha and beta subunits are 43% identical in humans, indicating the spectrin subunits have diverged more extensively.

Properties of the 106-Residue Repeat Unit

The nearly universal presence of the 106-residue repetitive structure in the spectrin heterodimer makes this unit the most important substructural feature of the entire molecule. It is worth examining the properties of this basic unit in further detail. Essentially, the repetitive structure makes a very large, complex protein much simpler; a detailed understanding of this elemental structural unit should enhance our understanding of general features of the entire molecule. This is especially true since related proteins, as well as internal repetitive sequences, have highly conserved conformations [61,62,72,73]. Polypeptide chain folding is actually better conserved than sequence homology. Proteins that do not show any detectable homology but have very similar conformation from X-ray structural analysis are thought to be distantly related through a common ancestor, and even proteins that have evolved new functions retain their basic conformation. Based upon these well-established principles, each homologous spectrin repeat unit should have highly conserved conformation. Even functional sites of spectrin that are homologous to the 106-residue repeat should not have a grossly different conformation; although, the unique amino acid sequence that comprises a specific binding site imparts unique and important characteristics to that specific region.

Several properties of the 106-residue repeat may be indicators of important structural features. The most unusal and least understood property is the precise conservation of the length of the repeat unit [42,45]. As genes coding for homologous proteins or internal repeat units mutate, insertions and deletions occur with appreciable frequency. One comparison of gaps relative to amino acid replacements showed one gap event on average per 42 amino acid substitutions in 14 protein families [74]. In contrast, we previously showed that almost 1,500 residues of spectrin sequences could be aligned in the 106-residue repeat and only a single one-residue gap was required [42]. This remarkable absence of gaps is especially unusual when evolutionary distances between these sequences are considered; some spectrin repeats are less than 20% identical, yet all except the first alpha repeat have exactly 106 residues.

The precise 106-residue repeat is not currently understood, but we can assume that the exact size of the repeat unit is critical. The importance of the exact repeat unit size is further indicated by the observation by Birkenmeier and co-workers that

chicken smooth-muscle spectrin has the same 106-residue repeat and its size is exactly preserved [66,67].

The exact preservation of the 106-residue repeat aids in the evaluation of spectrin sequences. Since gaps are apparently not allowed, very high gap penalties are justified. The use of high gap penalties allows statistically significant alignments of even the most distantly related spectrin repeats. We were also previously able to place many partial sequences within the overall framework of the spectrin subunits by using the approximate size of the domain peptides and the precise positions of the determined sequences within the repeat unit [42] (Fig. 1).

Several characteristics of the repeat units are illustrated in Table III, which shows the degree of identity for all possible pairwise comparisons between repeat units from the amino-terminal half of the alpha subunit. The repeat units show very low homology, indicating that they have diverged extensively since the time of gene duplication. While all comparisons involving homologous segments are statistically significant, the homology is so low that it is difficult to deduce a detailed scheme for the evolutionary intermediates between the basic ancestral 12,000-dalton precursor and the present molecule using evolutionary distance comparisons. Also, sequence segments covering only partial repeats are misleading in these comparisons since all regions of the repeat unit are not equally well preserved. This differential conservation of sequence was previously illustrated by identification of generally conserved residues in available sequences; most conserved residues were in the first half of the repeat (see [42] and α-6 in Table III). A more extensive evaluation of evolutionary

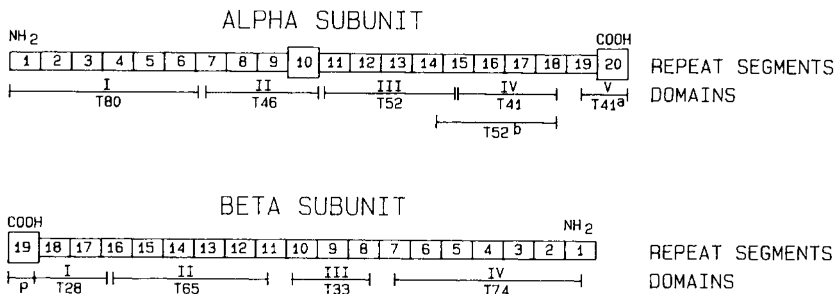

Fig. 1. A highly schematic model of human erythrocyte spectrin structure illustrating the relationship between domains and repeat segments (repeat units). The subunits occur in an antiparallel side-to-side orientation as illustrated diagramatically. The self-association site is on the left and the actin/protein 4.1 site is on the right. Repeat segments represent the underlying evolutionary structure and have been numbered starting from the amino-terminal of each subunit. Most units are homologous (rectangles) and are exactly 106 residues in length; nonhomologous segments are represented as squares and may have somewhat different lengths. Domains (Roman numerals) were derived before amino-terminal orientation of the subunits was established, and were numbered from the self-association site. The peptides from mild tryptic treatment that represent the domains are labeled by domain number and by molecular weight in thousands, ie, the α-I domain is represented by T80 (an 80,000-dalton tryptic peptide).

[a]This peptide representing the α-V domain migrates on polyacrylamide gels as 41,000 daltons, but it is probably somewhat smaller as illustrated (see text).

[b]This peptide was originally defined as the α-IV domain [47], but it probably overlaps the α-III domain [42] as illustrated. A smaller peptide T41 is now defined as the largest unique tryptic fragment representing the α-IV domain as shown.

TABLE III. The Degree of Identity Between Alpha Subunit Repeats†

Repeat segment	Repeat segment 1	2	3	4	5	6	7	8	9	10	11
1	*	19	27	21	17	37	21	26	22	13	22
2		*	23	17	26	27	28	26	23	16	27
3			*	21	23	27	28	29	33	13	19
4				*	21	25	13	25	21	10	15
5					*	22	23	25	22	10	17
6						*	14	37	25	14	27
7							*	21	17	10	10
8								*	27	17	24
9									*	10	15
10										*	13

†Sequences as reported in reference 42 (except repeat 10, unpublished results) were used for this comparison. The sequence for most of these repeats has been completely or nearly completely determined with the exception of repeat 6. Only the amino-terminal half of repeat 6, which is more strongly conserved than the carboxyl-terminal half of the repeat, is known; this leads to anomalously high values for comparisons involving this segment. Segment 10 is not homologous to the other segments and has been included for comparative purposes.

development of this complex molecule requires more complete sequence information, especially from the beta subunit.

The Importance of Domains and Repeat Units

The term domain has traditionally been used to describe two entirely different submolecular features of proteins. The term has been applied to experimentally generated fragments of a protein (generally using mild proteolysis). These divisions of proteins are often useful for structural studies and sometimes segregate functions into smaller peptide units. A well-known example would be the myosin domains (globular head group, light meromyosin, heavy meromyosin). In other cases, the term domain has been applied to an internal repetitive segment indicative of gene duplication, such as the 110-residue periodicity of immunoglobulins.

The study of spectrin is complicated by the existence of both types of structure, and each type might appropriately be called domains. The potential confusion is obvious, and a comprehensive standard is needed. Domains were initially defined for the spectrin molecule as mild proteolytic fragments [41,46,47] and these fragments or domains continue to be used by a number of groups, especially for the investigation of hemolytic disorders. The importance of establishing a uniform system for identifying these fragments is further accentuated by the slightly different (but insignificant) molecular weights that various investigators in the field have assigned to apparently identical tryptic fragments.

It seems appropriate to continue to describe the peptides produced by mild tryptic treatment using the domain nomenclature previously applied and to develop a new nomenclature for the more detailed and highly repetitive structure indicated by sequence information. We propose adopting the convention of referring to the more detailed structure as repeat segments or repeat units as illustrated in Figure 1. To achieve as simple a system as possible, both homologous and nonhomologous segments are included in this scheme (it is not known whether the several nonhomologous segments were originally derived from a different gene, or if they selectively mutated to a distinct structure). Also we have adopted the convention of using Roman numerals

for domains (α-I,II,III,IV,V and β-I,II,III,IV) and arabic numbers for repeat segments (α-1,2,...,20 and β-1,2,...,19) to differentiate further between the two types of structure. Using this scheme, most of the 20 alpha subunit repeat segments are homologous and exactly 106 residues in length. The α-10 and α-20 segments are not homologous to the other 18 repeats and their precise lengths have not yet been determined, although their sizes are close to the homologous segment size.

The correlation between major tryptic fragments, domains, and repeat segments is summarized in Figure 1. This model is updated from that shown previously in several important respects. The previous model assumed that small regions at the termini of the alpha subunit may have been added on to the ends of the molecule to provide protein binding sites; these functional sites were thought to be unrelated to the repeat unit and were previously indicated as separate and additional to the 20 segments. Our original phasing of the repeat unit started at residue 18 of the α-I domain sequence [44,45]. It now appears likely that most functional sites including the oligomer binding site at the end of the α-I domain [58,44] are included in the repeat units. A new phasing of the repeat unit starting at the amino-terminal end of the intact alpha subunit (about 10 residues before the α-I sequence—the blocked amino-terminal is lost from α-I) has been deduced [75]. This new phasing of the repeat unit with exactly nine 106 residue repeats before the anomolous α-10 segment is incorporated into the model in Figure 1. It is likely that this phasing represents the phasing of the corresponding genetic segments that have duplicated to produce the repetitive features we observe in the protein.

As indicated by the model, the α-I domain is represented by a tryptic 80,000-dalton peptide (T80) and this domain includes most of the first six repeat segments of the alpha subunit (α-1 through α-6). The α-II domain is represented by a tryptic 46,000 dalton peptide (T46) and this domain includes most of the repeat segments α-7 through α-10. The α-III domain starts near the beginning of α-11 and would contain the next four segments.

Some ambiguity remains in correlating domain peptide sequences with specific repeat segments in the remainder of the molecule. This problem is illustrated by the positioning of the α-V domain that is represented by a tryptic 41,000 peptide in the original domain map [47]. Based on its size, we previously placed the amino-terminal sequence from this domain peptide in α-18. Recently, Curtis apparently did not overlap this sequence with a cDNA clone that extended into the α-18 segment [69]. Based on this observation and new suggestive evidence from our laboratory that the α-V peptide migrates anomolously on gels, it is likely that there is an additional repeat unit between the α-IV and α-V domains and the T41 sequence is probably located in α-19 rather than α-18.

Spectrin Is Comprised of Short Triple Helical Segments

Spectrin is generally regarded as a long flexible rod, and correlating this with possible conformations should enhance our understanding of its properties. The predication of spectrin conformation based on the α-I domain sequence [45] using the method of Chou and Fasman [78,79] indicated a general occurrence of intermediate-sized helicies. At that time, we proposed a simplified model of spectrin structure comprised of a number of linearly arranged triple helical structures with each triple helix connected by a flexible nonhelical region. Precise correlation of this general structure with the 106 amino acid repeat and precise lengths of helicies were not

made. More recently, we refined this model by examining the phasing of the conformational unit with the 106-residue repeat and concluded that there were two reverse turns (a conformation of four contiguous residues that changes the direction of propagation of the polypeptide chain by approximately 180°) per 106-residue repeat [42]. The implications of this observation include the idea that the repeat unit (the 106-residue repetitive structure) and the conformational repeat (the triple helical units) are the same size.

Most recently, we have further refined this model using the fact that homologous structures have similar conformations. A simple computer method that we termed homology reinforcement conformational analysis was used to refine conformational predictions and obtain a detailed model of spectrin conformation [75]. One surprising feature of this newest model is the hypothesis that the repeat unit and the conformational unit are out of phase by approximately half of a 106-residue unit; this difference may have played an important role in the function of primitive spectrin ancestors. Another feature of the current model is that the three helices that comprise a conformational unit are quite different in length and the triple-stranded structure may include small amounts of beta strands and coil regions.

The feasibility of these predicted models of spectrin conformation can be reinforced by independently considering known physical and chemical properties of the molecule. The molecule is highly helical [about 75% helix—ref. 76,77]. with amino- and carboxyl-termini near opposite physical ends of the molecule [41,43,44,47,58]. The molecule must start at one end and fold back and forth upon itself an even number of times producing an odd number of paired strands with a net unidirectional propagation of the polypeptide chain. Considering the high helical content, the total length of the secondary structure must be on the order of 300 nm (2,000 residues × .15 nm propagation/helical residue) and the molecule would have to fold back upon itself two times forming a triple strand in order to fit the secondary structure into a 100-nm rod. A triple-stranded largely helical, structure is therefore a feasible general conformational structure. Extended helices are fairly rigid and the molecule appears highly flexible throughout its length, suggesting that helical segments must be relatively short. A model of spectrin structure comprised of short triple-stranded (largely helical) segments as described above is therefore strongly supported by independently derived properties of the molecule.

The Relationship Between Erythrocyte Spectrin and Other Spectrinlike Proteins

Initial identifications of spectrinlike proteins were based on antibody crossreactivity and/or similar size [28–31]. Subsequently, at least several of these tissue-specific spectrinlike proteins have been extensively studied [80–83] structurally and functionally in multiple species including avian and mammalian tissues. A minimum of three fairly well characterized proteins are generally regarded as major members of the spectrin family. In addition to erythroid spectrin, they are fodrin or brain spectrin [29,31] and TW260/240 from brush border [30,82]. These proteins share at least some morphologic, functional, and structural properties that have been summarized previously [83]. Despite this detailed attention by many laboratories, several important questions remain unresolved, including the total number of unique members of the spectrin family, tissue distribution of specific spectrin types, and the degree of structural and functional similarity between some family members. Since erythrocyte

spectrin has been most extensively studied, it has often served as a prototype of this family both in the design and interpretation of experimental results. While our knowledge of erythrocyte spectrin has certainly accelerated study of related proteins, it appears that erythroid spectrin, especially in mammals, is the most unique or specialized member of this group as suggested by Glenney and Glenney [83].

Partial sequence of several nonerythroid spectrinlike proteins indicates that these proteins are indeed sequence related [66,67, and our unpublished results] and have probably evolved from a common ancestor. The most striking feature of these sequences is that the 106 amino acid repeat which is precisely conserved throughout erythroid spectrin is also precisely preserved even in the distantly related chicken smooth-muscle protein [67]. Since most of the erythroid spectrin molecule is comprised of these 106-residue repeats, it is not surprising that related proteins have this basic building block as a major element of their structure as well. Certainly, sequence homology will be the final determinant for inclusion of members into the spectrin-related protein family, and will be more exacting than antibody crossreactivity or peptide maps. The presence of many 106-residue repetitive units in a high-molecular-weight actin-binding protein appears to be a reasonable, exacting, and sufficient criterion for ultimately discerning which proteins should be included in this protein family.

ACKNOWLEDGMENTS

This work was supported by grants GM21714 and AM27932 from the National Institutes of Health. The author thanks G. Davis for critical review of this manuscript and K. Speicher for computer analyses. The author also thanks Drs. C. Birkenmeier and P. Curtis for sharing unpublished sequence results.

REFERENCES

1. Marchesi VT, Steers EJ: Science 159:203, 1968.
2. Ungewickell E, Gratzer W: Eur J Biochem 88:379, 1978.
3. Ralston GB, Dunbar J, White M: Biochim Biophys Acta 491:345, 1977.
4. Shotton D, Burke B, Branton D: Biochim Biophys Acta 536:313, 1978.
5. Morrow JS, Marchesi VT: J Cell Biol 88:463, 1981.
6. Bennett V. Stenbuck PJ: Biol Chem 254:2533, 1979.
7. Luna EJ, Kidd GH, Branton D: J Biol Chem 254:2625, 1979
8. Yu J, Goodman SR: Proc Natl Acad Sci USA 76:2340, 1979.
9. Tyler J, Anderson JM, Branton D: J Cell 85:489, 1980.
10. Cohen CM, Tyler JM, Branton D: Cell 21:875, 1980.
11. Brenner SL, Korn ED: J Biol Chem 255:1670, 1980.
12. Fowler VM, Luna EJ, Hargreaves WR, Taylor DL, Branton D: J Cell Biol 88:388, 1981.
13. Cohen CM, Foley SF: J Cell Biol 86:694, 1980.
14. Wolfe LC, Lux SE, Chanian V: J Cell Biol 87:203d, 1980.
15. Fowler V, Taylor DL: J Cell Biol 85:361, 1980.
16. Ungewickell E, Bennet PM, Calvert R, Ohanian V, Gratzer WB: Nature 280:811, 1979.
17. Cohen CM, Korsgren C: Biochem Biophys Res Commun 97:1429, 1980.
18. Calvert R, Bennett P, Gratzer, W: Eur J Biochem 107:355, 1980.
19. Tyler JM, Hargreaves WR, Branton D: Proc Natl Acad Sci USA 76:5192, 1979.
20. Lux SE, Wolfe LC, Pease B, Tomaselli MB, John KM, Bernstein SE: Prog Clin Biol Res 45:159, 1981.
21. Haest CWM: Biochim Biophys Acta 694:331, 1982.

22. Goodman SR, Shiffer K: Am J Physiol 244:C121, 1983.
23. Marchesi VT: Blood 61:1, 1983.
24. Cohen CM: Semin Hematol 20:141, 1983.
25. Palek J, Lux SE: Semin Hematol 20:189, 1983.
26. Knowles W, Marchesi SL, Marchesi VT: Semin Hematol 20:159, 1983.
27. Goodman SR, Shiffer K, Coleman DB, Whitfield CF: Prog Clin Biol Res 165:415, 1984.
28. Goodman SR, Zagon IS, Kulikowski RR: Proc Natl Acad Sci USA 78:7570, 1981.
29. Bennett V, Davis J, Fowler WE: Nature 299:126, 1982.
30. Glenney JR, Glenney P, Osborn M, Weber K: Cell 28:843, 1982.
31. Levine J, Willard M: J Cell Biol 90:631, 1981.
32. Pollard TD: J Cell Biol 99:1970, 1984.
33. Marchesi VT: J Membr Biol 51:101, 1979.
34. Lux SE: Semin Hematol 16:22, 1979.
35. Branton D, Cohen CM, Tyler J: Cell 24:24, 1981.
36. Shotten D, Burke B, Branton D: J Mol Biol 181:303, 1979.
37. Tyler JM, Reinhardt BN, Branton D: J Biol Chem 255:7034, 1980.
38. Liu SC: Cell 37:587, 1984.
39. Lux S, Shohet SB: Hosp Pract Oct: 77, 1984.
40. Luma EJ, Kidd GH, Branton D: J Biol Chem 254:2526, 1979.
41. Speicher DW, Morrow JS, Knowles WJ, Marchesi VT: Proc Natl Acad Sci USA 77:5673, 1980.
42. Speicher DW, Marchesi VT: Nature 311:177, 1984.
43. Harris HW, Lux SE: J Biol Chem 255:11512, 1980.
44. Speicher DW, Davis G, Yurchenco PD, Marchesi VT: J Biol Chem 258:14931, 1983.
45. Speicher DW, Davis G, Marchesi VT: J Biol Chem 258:14938, 1983.
46. Speicher DW, Marchesi VT: J Cell Biochem 18:479, 1982.
47. Speicher DW, Morrow JS, Knowles WJ, Marchesi VT: J Biol Chem 257:9093, 1982.
48. Liu SC, Lawler J, Prchal J, Palek J: Blood (abstract), 58:45a, 1981.
49. Coetzer T, Zail S: J Clin Invest 67:1241, 1981.
50. Lawler J, Liu SC, Palek J, Prchal J: J Clin Invest 70:1019, 1982.
51. Palek J, Lux SE: Semin Hematol 20:189, 1983.
52. Knowles WJ, Morrow JS, Speicher DW, Zarkowsky HS, Mohandas N, Mentzer WC, Shohet SB, Marchesi VT: J Clin Invest 71:1867, 1983.
53. Knowles WJ, Bologna ML, Chasis JA, Marchesi SL, Marchesi VT: J Clin Invest 73:973, 1984.
54. Lawler J, Liu SC, Palek J, Prchal J: J Clin Invest 73:1688, 1984.
55. Zail SS, Coetzer TL: J Clin Invest 74:753, 1984.
56. Becker PS, Lux SE: Clin Haematol 14:15, 1985.
57. Speicher DW, Davis G, Lanzetti TJ, Knowles WJ: Manuscript submitted.
58. Morrow JS, Speicher DW, Knowles WJ, Hsu CJ, Marchesi VT: Proc Natl Acad Sci USA 77:6592, 1980.
59. Anderson JM: J Biol Chem 254:939, 1979.
60. Yurchenco PD, Speicher DW, Morrow JS, Knowles WJ, Marchesi VT: J Biol Chem 257:9102, 1982.
61. Doolittle RF: Science 214:149, 1981.
62. Keim P, Heinrikson RL, Fitch WM: J Mol Biol 151:179, 1981.
63. Li WH: In Nei M, Koehn RK, (eds): "Evolution of Genes and Proteins." Sunderland, Mass: Sinauer Associates. 1983, pp 14–37.
64. Pink JRL: In Gutfreund H (ed): "Biochemical Evolution." Cambridge University Press: Cambridge, England, 1981, pp 231–260.
65. Barker WC, Ketcham LK, Dayhoff MO: In Dayhoff MO (ed):"Atlas of Protein Sequence and Structure," Vol 5(suppl 3). Silver Spring, Maryland: National Biomedical Research Foundation, 1978.
66. Birkenmeier CS, Bodine DM, Repasky EA, Helfman DM, Hughes SH, Barker JE: Proc Natl Acad Sci USA, 82:5671, 1985.
67. Birkenmeier CS, Barker JE, Bodine DM: Submitted.
68. Cioe L, Curtis P: Proc Natl Acad Sci USA 82:1367, 1985.
69. Curtis PJ, Palumbo A, Ming J, Fraser P, Cioe L, Meo P, Shane S, Rovera G: Gene 36:357, 1985.
70. Dayhoff MO: In Dayhoff MO (ed): "Atlas of Protein Sequence and Structure," Vol 5(suppl 3). Silver Spring, Maryland: National Biomedical Research Foundation, 1978, pp 1–8.

71. Schwartz RM: In Dayhoff MO (ed): "Atlas of Protein Sequence and Structure," Vol 5(suppl 3). Silver Spring, Maryland: National Biomedical Research Foundation, 1978, pp 353–358.
72. Schulz GE, Schirmner RH: In Cantor CR (ed): "Principles of Protein Structure." New York: Springer-Verlag., 1979, pp 166–205.
73. Arnheim N: In Nei M, Koehn RK (eds): "Evolution of Genes and Proteins." Sunderland, Mass: Sinauer Associates, 1983, pp 38–61.
74. Dayhoff MO, Barker WC: In Dayoff MO (ed): "Atlas of Protein Sequence and Structure," Vol 5. Silver Spring, Maryland: National Biomedical Research Foundation, 1972, pp 41–45.
75. Speicher DW: Manuscript submitted.
76. Calvert R, Ungewickell E, Gratzer W: Eur J Biochem 107:363, 1980.
77. Gratzer WB: Biochem J 198:1, 1981.
78. Chou PY, Fasman GD: Adv Enzymol 47:45, 1978.
79. Chou PY, Fasman GD: Annu Rev Biochem 47:251, 1978.
80. Repasky EA, Granger BL, Lazarides E: Cell 29:821, 1982.
81. Burridge K, Kelly T, Mangeat P: J Cell Biol 95:478, 1982.
82. Glenney JR, Glenney P, Weber K: Proc Natl Acad Sci USA 79:4002, 1982.
83. Glenney JR, Glenney P: Cell Motil 3:671, 1983.