Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 1730–1734, March 1990
Cell Biology

cDNA sequence for human erythrocyte ankyrin

(skeletal protein/erythrocyte membrane)

STEPHEN LAMBERT*, HUILAN YU†, JOSEF T. PRCHAL†, JACK LAWLER‡, PAUL RUFF*, DAVID SPEICHER§, MEI CHI CHEUNG¶, YUET WAI KAN¶, AND JIRI PALEK*‖

*Department of Biomedical Research, Saint Elizabeth's Hospital of Boston, Tufts University Medical School, Boston, MA 02135; †Division of Hematology/Oncology, University of Alabama at Birmingham, Birmingham, AL 35294; ¶Department of Medicine, Cancer Research Institute and Howard Hughes Medical Institute, University of California, San Francisco, CA 94143; §The Wistar Institute, Philadelphia, PA 19104; ‡Vascular Research Division, Department of Pathology, Brigham and Women's Hospital, and ‡Harvard Medical School, Boston, MA 02115

Contributed by Yuet Wai Kan, August 28, 1989

ABSTRACT **The cDNA for human erythrocyte ankyrin has been isolated from a series of overlapping clones obtained from a reticulocyte cDNA library. The composite cDNA sequence has a large open reading frame of 5636 base pairs (bp) with the complete coding sequence for a polypeptide of 1879 amino acids with a predicted molecular mass of 206 kDa. The derived amino acid sequence contained 194 residues that were identical to those obtained by direct amino acid sequencing of 11 ankyrin proteolytic peptides. The primary sequence contained 23 highly homologous repeat units of 33 amino acids within the 90-kDa band 3 binding domain. Two cDNA clones showed evidence of apparent mRNA processing, resulting in the deletions of 486 bp and 135 bp, respectively. The 486-bp deletion resulted in the removal of a 16-kDa highly acidic peptide, and the smaller deletion had the effect of altering the COOH terminus of the molecule. Radiolabeled ankyrin cDNAs recognized two erythroid message sizes by RNA blot analysis, one of which was predominantly associated with early erythroid cell types. An ankyrin message was also observed in RNA from the human cerebellum by the same method. The ankyrin gene is assigned to chromosome 8 using genomic DNA from a panel of sorted human chromosomes.**

Erythrocyte ankyrin (also known as band 2.1) is a large peripheral protein (M_r = 210,000) that interacts with the integral membrane protein band 3 and the β subunit of erythrocyte spectrin to link the supporting erythrocyte membrane skeleton with the lipid bilayer (see refs. 1 and 2 for review). In membrane preparations, ankyrin is observed as the major species of a series of immunologically (3, 4) and structurally (5, 6) similar molecules, bands 2.2 (M_r = 186,000) and 2.3 (M_r = 160,000). Although the origin of these minor ankyrin-like polypeptides remains unclear, calcium-mediated proteolysis of ankyrin has been suggested (7, 8). Recent studies (9), however, have demonstrated that band 2.2 is not produced by calpain cleavage and that it is an activated form of ankyrin that may even possess its own distinct class of membrane binding sites on kidney microsomes (10). Ankyrin-like molecules have also been described in various nonerythroid tissues such as brain (11), kidney (12), and striated muscle (13).

Low-resolution mapping of ankyrin has allowed the elucidation of two major structural and functional domains of a bipolar nature (14–16). The band 3 binding region of ankyrin has been mapped to a basic 90-kDa domain, whereas the spectrin binding site has been mapped to a 32-kDa tryptic peptide within a neutral phosphorylated domain of 72 kDa.

Recently, ankyrin has been associated with the hemolytic anemia, hereditary spherocytosis (HS), in a mouse model system (17) and in two kindred with an atypical variant of the disorder associated with a combined spectrin and ankyrin deficiency (18).

In this paper we report the isolation of cDNA clones coding for the human ankyrin molecule.** The amino acid sequence derived from these clones shows the presence of a highly conserved repeating structure within the 90-kDa band 3 binding domain of the molecule. Use of these cDNAs as probes reveals two major forms of the ankyrin message, one of which is predominantly associated with earlier stages of erythroid development. Finally, ankyrin is assigned to chromosome 8 by hybridizing one of these cDNAs to genomic DNA from sorted human chromosomes.

MATERIALS AND METHODS

Reagents. Aliquots of a reticulocyte cDNA library cloned in the expression vector λgt11 were a gift from J. Conboy (University of California, San Francisco), and details of its construction have already been published (19). Restriction enzymes were supplied by New England Biolabs, and [^{32}P]dATP and [^{32}P]dCTP were from New England Nuclear. Plasmid vectors pGem-3 and pGem-4 were from Promega Biotec. Sequenase was obtained from United States Biochemical, and all DNA and protein sequence data were analyzed using the "DNAstar" (Madison, WI) computer package.

Screening of Reticulocyte cDNA Expression Library. The reticulocyte cDNA expression library was screened immunologically according to the method of Huynh *et al.* (20) using rabbit (kindly supplied by C. Korsgren, Saint Elizabeth's Hospital, Boston) and goat antibodies raised against electrophoresis-purified ankyrin. Positive clones were identified by peroxidase staining combined with avidin-biotin enhancement (Vector Laboratories). Further cDNA clones were obtained by rescreening the library with radiolabeled cDNA inserts obtained from immunological screening. These inserts were labeled with [^{32}P]dCTP using the primer extension method (21).

DNA Sequence Analysis. cDNA inserts from positive clones were subcloned into the plasmid vectors pGem-3 and pGem-4 and sequenced by the dideoxynucleotide chain-termination method (22) as adapted for double-stranded sequencing using a modified T7 DNA polymerase (Sequenase).

Preparation of RNA. Subjects undergoing phlebotomies for hemochromatosis served as a source of normal reticulocyte RNA. Total RNA was isolated by the method of Goosens and Kan (23) and modified by additional purification of the RNA precipitate using guanidine hydrochloride and phenol (24).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: HS, hereditary spherocytosis.

‖To whom reprint requests should be addressed.

**The sequence reported in this paper has been deposited in the GenBank data base (accession no. M28880).

Normal human bone marrow was prepared from a resected human rib of a patient undergoing exploratory thoracotomy, where 50% of nucleated cells in this preparation were recognizable erythroid precursors. Adult erythroblasts were obtained by direct panning of human bone marrow preparations on Eo-1 antibody-coated plates (25). Human purified fetal erythroblasts were obtained from human fetal liver by a similar method. The human erythroleukemia cell line HEL was used as a source of RNA and maintained as described (25). Total RNA from these tissues and cell lines (including human cerebellum) was prepared by the guanidine hydrochloride/phenol method (24) and poly(A)$^+$ RNA was isolated by oligo(dT)-cellulose chromatography and by one to three passages through oligo(dT)-resin. All human tissues were obtained from the Tissue Procurement Facility of the University of Alabama at Birmingham (UAB) using protocols previously approved by the UAB Human Use Committee.

Northern Blot Hybridization. Poly(A)$^+$ RNAs were fractionated by formaldehyde/1.4% agarose gel electrophoresis and transferred to nitrocellulose filters as described (26). An RNA ladder (Bethesda Research Laboratories) visualized before transfer was used to calculate the relative size of the ankyrin messages. Filters were hybridized with authentic ankyrin [^{32}P]cDNAs and washed under stringent conditions before autoradiography. Relative proportions of different size ankyrin transcripts were determined by scanning the autoradiograph with an LKB laser densitometer.

Chromosomal Assignment of Ankyrin cDNA. Authentic ankyrin [^{32}P]cDNAs were hybridized to a panel of filters containing sorted human chromosomes prepared as described (27). Filters were washed under stringent conditions and positive signals were visualized by autoradiography.

RESULTS

Isolation and Characterization of Ankyrin cDNA Clones. Fig. 1*A* shows the restriction map of human erythrocyte cDNA as constructed by analysis of overlapping cDNA clones (Fig. 1*B*). Regions of clones that were sequenced on both or either strand are shown by arrows. Clones were verified by comparing the derived protein primary sequence with residues from the direct amino acid sequencing of 11 proteolytic peptides from the ankyrin molecule. The relative positions of these peptides are shown in Fig. 1*C*. The composite nucleotide sequence encoded by these overlapping clones and the predicted amino acid sequence are shown in Fig. 2. Boxed-in residues were identical to those found by direct amino acid sequencing of the peptides mentioned above. The reading frame of this composite sequence starts at nucleotide 88 and remains open until residue 5724, thereby coding for a protein of 1879 amino acids with a molecular mass of 206 kDa compared with the predicted mass for erythrocyte ankyrin as determined by SDS/PAGE of 215 kDa (28). No ambiguities between regions of clone overlap could be detected with the exception of two missing regions of sequence as shown in Fig. 1*B*. These regions contained 486 base pairs (bp) and 135 bp, respectively, and were flanked by residues consistent with consensus sequences for eukaryotic donor/acceptor mRNA splice sites (29). A 265-bp probe (nucleotides 4656–4921) from the larger of the putative splices sites was found to hybridize with both of the erythroid transcripts as well as the cerebellum ankyrin mRNA transcript shown in Fig. 4*A*. The smaller splice site had the effect of creating a different COOH terminus for the molecule as shown in Fig. 2.

Primary and Secondary Structure of Ankyrin. The most striking feature of ankyrin primary structure is the presence of 23 highly homologous 33 amino acid repeats located in the 90-kDa band 3 binding domain. An alignment of these repeats showing areas of homology is shown in Fig. 3. Homologies between repeats ranged from 27% to 45% with one atypical repeat of 29 residues. An analysis of the secondary structure encoded by this primary sequence using the Chou–Fasman algorithm (30) reveals no obvious secondary structure for these repeats, although they do have a higher degree of ordered secondary structure than the rest of the ankyrin molecule. A computer search of the National Biomedical Research Foundation data base (31) reveals a 25.2% homology between the 5th and 15th repeat unit of ankyrin (amino acids 190–610) and amino acids 1840–2260 of the *Drosophila* notch protein (Z score of 16). However the significance of this homology is not clear as this region is outside the epidermal

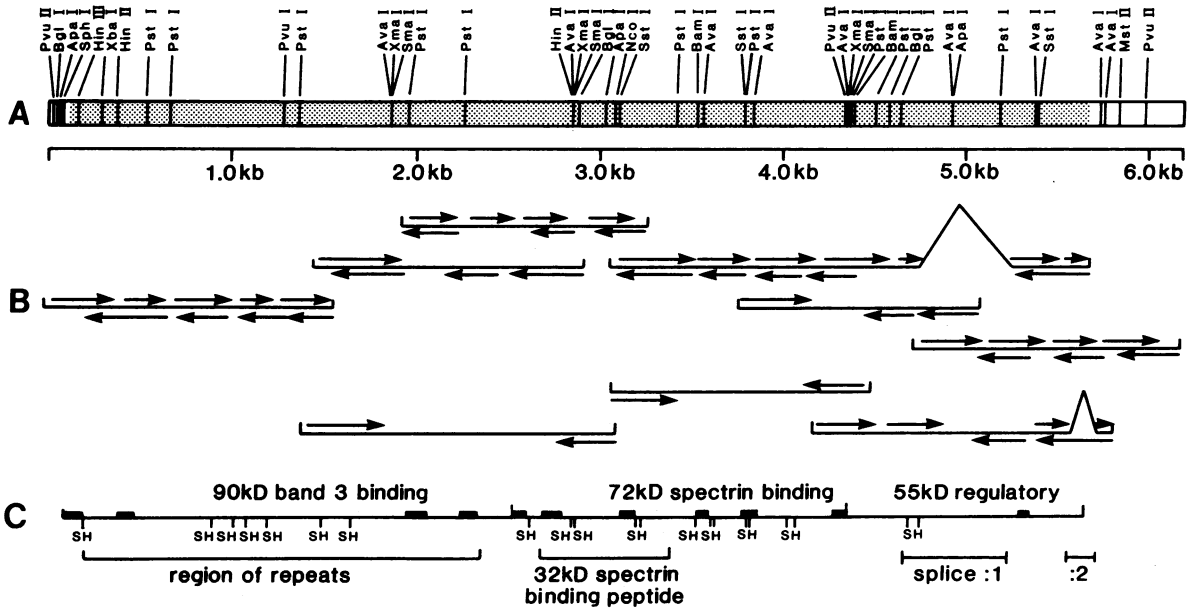

FIG. 1. Organization of the ankyrin molecule as provided by cDNA and amino acid sequence data. (A) Restriction map of the composite cDNA clones. The stippled region (nucleotides 88–5724) represents the coding region of the cDNA. kb, Kilobases. (*B*) Alignment of clones included in the composite cDNA sequence. Arrows represent areas of clones sequenced. ∧ represents regions of cDNA that are missing due to probable alternative mRNA processing in the clones shown. (*C*) Linear map of the ankyrin molecule, showing the major functional and structural domains as described (14–16), indicating the position of 11 peptide sequences used in the confirmation of cDNA clone authenticity.

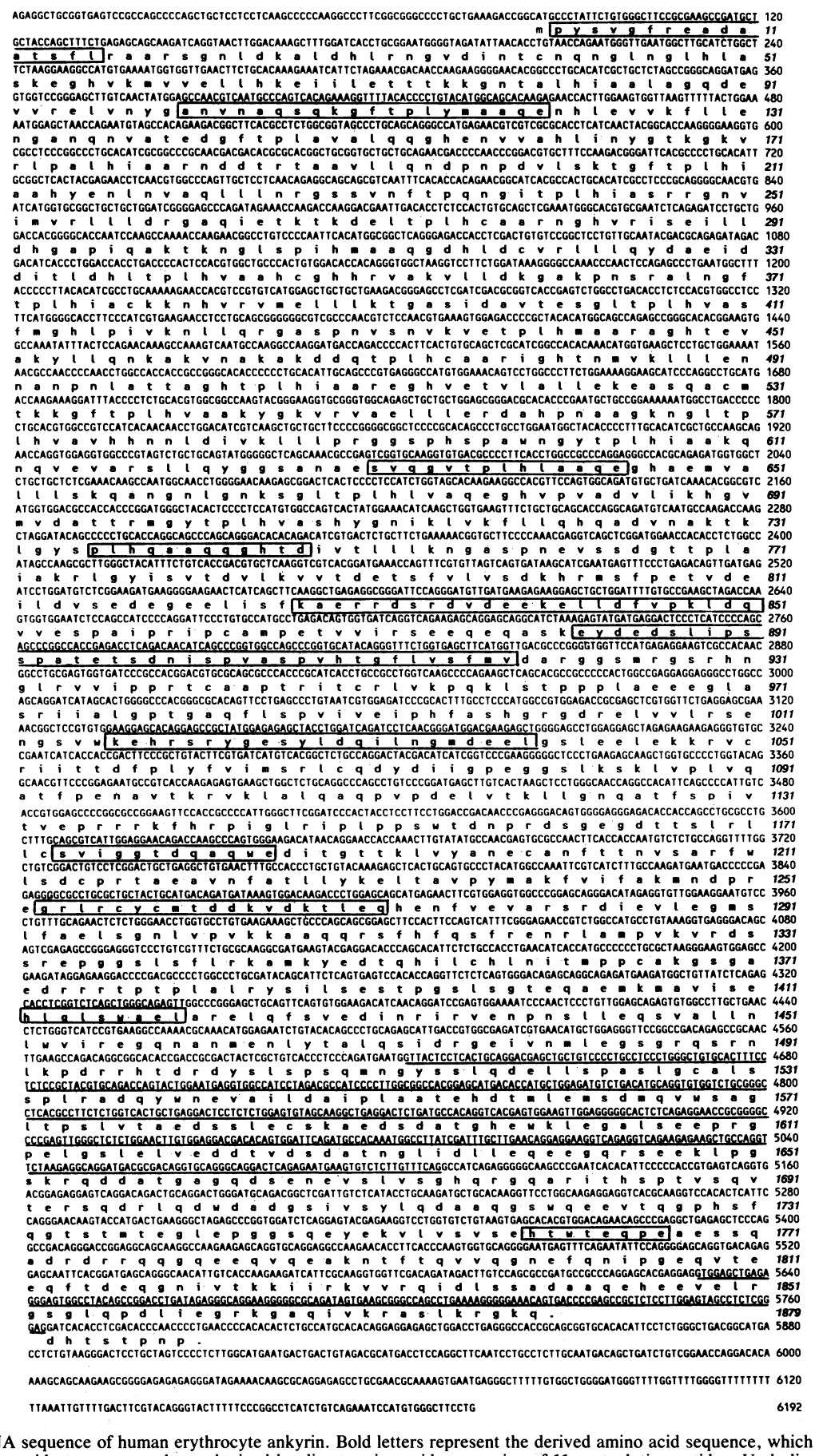

FIG. 2. cDNA sequence of human erythrocyte ankyrin. Bold letters represent the derived amino acid sequence, which is numbered with italics. Boxed-in residues represent those obtained by direct amino acid sequencing of 11 proteolytic peptides. Underlined areas of cDNA sequence represent those missing in the positions shown in Fig. 1*B*.

FIG. 3. Alignment of the 23 repeat units within the 90-kDa band 3 binding domain of ankyrin. Shaded regions represent identical residues between repeats. A consensus sequence for 18 of the 33 residues is shown.

growth factor-like repeat sequences of *Drosophila* notch and does not offer any clues as to its functional role in this protein (32).

Sequence from the large splice site encodes an acidic (pI = 3.42) peptide of 16 kDa that is downstream from the calculated COOH terminus of the 72-kDa spectrin binding domain and is in the same relatively hydrophilic domain of ankyrin proposed by Hall and Bennett (9) to be associated with the production of protein 2.2 by either a proteolytic or mRNA splicing event. On the contrary, the smaller splice encodes a highly alkaline peptide (pI = 11.4).

RNA Blot Analysis of Erythroid and Nonerythroid Ankyrin mRNA. Fig. 4*A* shows a Northern blot analysis of poly(A)$^+$ RNA from human reticulocytes and human cerebellum using a probe that spans the spectrin binding domain of ankyrin. Two sizes for the human erythrocyte ankyrin mRNA of

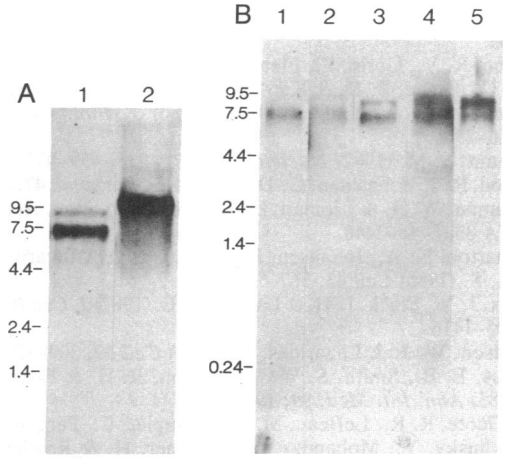

FIG. 4. Northern blot analysis of poly(A)$^+$ RNA hybridized to an ankyrin [^{32}P]cDNA. (*A*) Lane 1, 40 μg of poly(A)$^+$ RNA from normal human reticulocytes; lane 2, 40 μg of poly(A)$^+$ RNA from a normal human cerebellum. (*B*) Lane 1, 25 μg of poly(A)$^+$ RNA from normal human reticulocytes; lane 2, 60 μg of total RNA from human bone marrow; lane 3, 24 μg of total RNA from human adult erythroblasts; lane 4, 40 μg of total RNA from human fetal erythroblasts; lane 5, 40 μg of poly(A)$^+$ RNA from HEL cells. Sizes are given in kb.

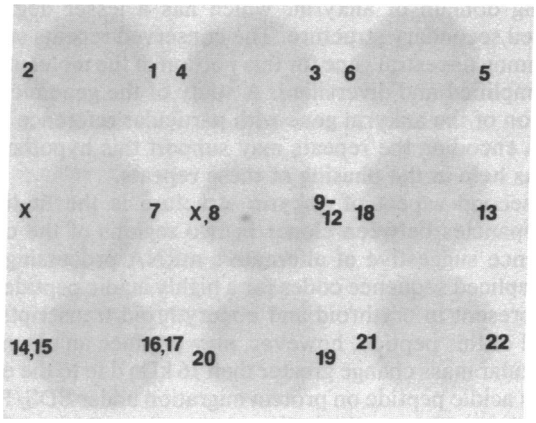

FIG. 5. Autoradiography of a panel of nitrocellulose filters containing cell sorter-separated human chromosomal DNA hybridized to an ankyrin [^{32}P]cDNA.

approximately 7.2 kb and 9.0 kb were observed in reticulocytes, whereas a major message of 9.5 kb cross-hybridized with the erythrocyte ankyrin probe in human cerebellum. Fig. 4*B* shows RNA blot analysis of poly(A)$^+$ mRNA from erythroid cells at various stages of development. Quantitation of the ratio between the 7.2- and 9.0-kb message in these cells by densitometry of the autoradiograph reveals a ratio of 0.7 for HEL cells. These cells have been committed to a pathway of erythroid differentiation by culturing them in the presence of δ-aminolevulinic acid (δ-ALA) for 3 days (25). This same ratio was 0.5 for cells grown in the absence of δ-ALA (not shown) and increased to 1.8 in bone marrow, 3.9 and 3.7 for adult and fetal erythroblasts, respectively, and 6.15 for reticulocytes (using four separate reticulocyte sample preparations).

Chromosomal Assignment of the Human Erythrocyte Ankyrin Gene. Fig. 5 shows a panel of nitrocellulose filters containing human DNA from sorted human chromosomes that has been hybridized to an ankyrin [^{32}P]cDNA. A signal in the region of chromosome X and chromosome 8 is visualized by autoradiography, and no signal is seen in the region of chromosome X alone, indicating that the human erythrocyte ankyrin gene may be assigned to chromosome 8 by this method.

DISCUSSION

The choice of a reticulocyte cDNA library, previously used in the isolation of protein 4.1 cDNA (19), as a potential source of ankyrin clones was suggested by synthetic data in the mammalian (M. Hanspal, J. H. Hanspal, J.T.P., S.L., and J.P., unpublished observations) and avian (33) systems. This work suggested that ankyrin synthesis, like that of protein 4.1, continues until late in erythroid development, including the reticulocyte. The use of such a library has allowed us to isolate a series of overlapping cDNA clones that contain the coding sequence for human erythrocyte ankyrin.

The derived primary structure for ankyrin indicates the presence of a highly conserved repeat structure within the 90-kDa band 3 binding domain. The arrangement of these repeats within the 90-kDa domain is intriguing as secondary structure predictions using the Chou–Fasman (30) algorithm do not indicate a linear extension of the repeat motif, suggestive of a possible globular packing arrangement. This arrangement would agree with observations made by Hall and Bennett (9) suggesting that ankyrin is a globular molecule, with a protruding tail or domain (based on electron microscopy studies and calculations of frictional ratios). The tail in this model may well be represented by the spectrin

binding domain of ankyrin, which has a lesser degree of ordered secondary structure. The conserved repeats suggest a common ancestral gene for this portion of the molecule that has amplified and diversified. A study of the genomic organization of the ankyrin gene with particular reference to the exons encoding the repeats may support this hypothesis as well as help in the phasing of these repeats.

A second aspect of ankyrin structure is the finding of discrepancies between clones in two regions of the coding sequence suggestive of alternative mRNA processing. The large spliced sequence codes for a highly acidic peptide of 16 kDa present in erythroid and nonerythroid transcripts. Removal of this peptide, however, may produce an anomalous molecular mass change greater than 16 kDa due to the effects of this acidic peptide on protein migration under SDS/PAGE conditions. The smaller splice observed would introduce a different COOH terminus into the ankyrin molecule. Alternative mRNA processing at the 3′ end of the ankyrin message suggests a possible mechanism for the generation of the minor ankyrin-like polypeptides (bands 2.2 and 2.3) observed in erythrocyte membrane preparations.

A similar mechanism of alternative mRNA processing of a tissue-specific nature might account for the larger than reticulocyte mRNA transcript observed in RNA from the human cerebellum hybridized with an erythroid ankyrin cDNA. Ankyrin-like molecules, diverse from their erythroid counterparts, have been observed in a range of tissues. Such a mechanism might explain their origin. A second possible explanation is that the erythroid cDNA is hybridizing with the product of an alternative ankyrin gene even under high stringency conditions. Both mechanisms might also be used to explain the two sizes (9.0 kb and 7.2 kb) of erythroid ankyrin message observed in developing erythrocytes. The larger transcript would seem to be associated with early erythroid precursors and is largely replaced by the smaller message as the cells approach maturation. Nelson and Lazarides (34) may have described a similar phenomenon in the development of the neuronal membrane skeleton in the avian system. Two isoforms of ankyrin (called goblin in this system) are observed in these cells, one of which is predominantly expressed early in the cell's differentiation concomitant with the synthesis of a brain-type spectrin. Synthesis of both of these molecules decreases as the cells mature to be replaced by increased synthesis of the second goblin isoform and an erythroid-like spectrin.

Finally, the assignation of the erythroid ankyrin gene to chromosome 8 indicates that it may have a direct role to play in HS. Bass *et al.* (35) and Chilcote *et al.* (36) have described HS associated with interstitial deletions of chromosome 8 or its translocation to chromosomes 3 and 12, leading them to suggest the presence of a HS loci on chromosome 8. Recently Lux *et al.* (37) have described decreased ankyrin gene dosage associated with an interstitial deletion of chromosome 8, resulting in decreased membrane ankyrin and spectrin levels, and hence leading to HS. A large body of HS patients associated with a membrane spectrin deficiency has also been described (38). As with the atypical cases described earlier (18), this deficiency may be secondary to subtle abnormalities or deficiencies of the major membrane spectrin binding site—i.e., ankyrin.

Note Added in Proof. A similar sequence for human erythrocyte ankyrin has been found by Lux *et al.* (39).

We acknowledge Catherine Korsgren for helpful discussion and assistance and Joan Joos for the artwork. We also acknowledge Lucille Paul's typing assistance. This work was supported by National Institutes of Health Awards HL 37462-01 and HL 27215.

1. Bennett, V. (1985) *Annu. Rev. Biochem.* **54,** 273–304.
2. Marchesi, V. T. (1985) *Annu. Rev. Cell Biol.* **1,** 531–561.
3. Bennett, V. & Stenbuck, P. J. (1979) *J. Biol. Chem.* **254,** 2533–2541.
4. Bennett, V. & Stenbuck, P. J. (1979) *Nature (London)* **280,** 468–473.
5. Luna, E. J., Kidd, G. H. & Branton, D. (1979) *J. Biol. Chem.* **254,** 2526–2532.
6. Yu, J. & Goodman, S. R. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 2340–2344.
7. Siegel, D. L., Goodman, S. R. & Branton, D. (1980) *Biochim. Biophys. Acta* **598,** 517–527.
8. Grasso, M., Morelli, A. & DeFlora, A. (1986) *Biochem. Biophys. Res. Commun.* **138,** 87–94.
9. Hall, T. G. & Bennett, V. (1987) *J. Biol. Chem.* **262,** 10537–10545.
10. Davis, J., Davis, L. & Bennett, V. (1989) *J. Biol. Chem.* **264,** 6417–6426.
11. Davis, J. Q. & Bennett, V. (1984) *J. Biol. Chem.* **259,** 13550–13559.
12. Drenckhahn, D., Shluter, K., Allen, D. & Bennett, V. (1985) *Science* **230,** 1287–1289.
13. Nelson, W. J. & Lazarides, E. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 3292–3296.
14. Wallin, R., Culp, E. N., Coleman, D. B. & Goodman, S. R. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 4095–4099.
15. Weaver, D. C. & Marchesi, V. T. (1984) *J. Biol. Chem.* **259,** 6165–6169.
16. Weaver, D. C., Pasternack, G. R. & Marchesi, V. T. (1984) *J. Biol. Chem.* **259,** 6170–6175.
17. Bodine, D. M., Birkenmieier, C. S. & Barker, J. E. (1984) *Cell* **37,** 721–729.
18. Coetzer, T. L., Lawler, J., Liu, S.-C., Prchal, J. T., Gualtieri, R. J., Brenin, M. C., Dacie, V. J. & Palek, J. (1988) *N. Engl. J. Med.* **318,** 230–234.
19. Conboy, J., Kan, Y. W., Shohet, S. B. & Mohandas, N. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9512–9516.
20. Huynh, T. V., Young, R. A. & Davis, R. W. (1985) in *DNA Cloning: A Practical Approach*, ed. Glover, D. M. (IRL, Oxford), Vol. 1, pp. 49–78.
21. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
22. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
23. Goosens, M. & Kan, Y. W. (1981) *Methods Enzymol.* **76,** 805–817.
24. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159.
25. Papayannopoulou, T. H., Nakamoto, B., Kurachi, S. & Nelson, R. (1987) *Blood* **70,** 1764–1772.
26. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5201–5205.
27. Lebo, R. V., Gorin, F., Fletterick, R. J., Kao, F. T., Cheung, M. C., Bruce, B. D. & Kan, Y. W. (1984) *Science* **225,** 57–59.
28. Bennett, V. & Stenbuck, P. J. (1980) *J. Biol. Chem.* **255,** 2540–2548.
29. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472.
30. Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol.* **47,** 45–148.
31. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2444–2448.
32. Wharton, K. A., Johansen, K. M., Xu, T. & Artavaris-Tsakonas, S. (1985) *Cell* **43,** 567–581.
33. Cox, J. V., Stack, J. H. & Lazarides, E. (1987) *J. Cell Biol.* **105,** 1405–1416.
34. Nelson, W. J. & Lazarides, E. (1984) *Cell* **39,** 309–320.
35. Bass, E. B., Smith, S. W., Stevenson, R. E. & Rose, W. F. (1983) *Ann. Int. Med.* **99,** 192–193.
36. Chilcote, R. R., LeBeau, M. M., Dampier, C., Pergament, E., Verlinsky, Y., Mohandas, N., Frischer, H. & Rowley, J. D. (1987) *Blood* **69,** 156–159.
37. Lux, S., John, K., Shaley, O., Forget, B., Chilcote, R., Marchesi, S., McIntosh, S., Harris, P., Watkins, P. & Bennett, V. (1988) *Blood* **72,** Suppl. 1, 46a (abstr.).
38. Agre, P., Casella, J. F., Zinkham, W. H., McMillan, C. & Bennett, V. (1985) *Nature (London)* **314,** 380–383.
39. Lux, S. E., John, K. M. & Bennett, V., *Nature (London)*, in press.