

Cloning and Characterization of 4.1G (EPB41L2), a New Member of the Skeletal Protein 4.1 (EPB41) Gene Family

M. Parra, P. Gascard, L. D. Walensky, S. H. Snyder, N. Mohandas, and J. G. Conboy¹

Life Sciences Division, Lawrence Berkeley National Laboratory, University of California, Berkeley, California 94720; and Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

Received September 30, 1997; accepted January 28, 1998

The prototypical erythrocyte membrane skeletal protein 4.1 (HGMW-approved symbol EPB41), here designated 4.1R, is encoded by a large, complexly spliced gene located on human chromosome 1p32–p33. In this paper we report evidence for a second 4.1 gene, 4.1G (HGMW-approved symbol EPB41L2), which maps to human chromosome 6q23 and is widely expressed among human tissues. The complete nucleotide sequence of 4.1G cDNA predicts a 113-kDa protein that exhibits three regions of high homology to 4.1R, including the membrane binding domain, the spectrin-actin binding domain, and the C-terminal domain. Interspersed among the shared domains are unique sequences that may define functional differences between 4.1R and 4.1G. Specific isoforms of 4.1R and 4.1G exhibit differential subcellular localizations. These results expand the 4.1 gene superfamily and demonstrate that the diverse cellular complement of 4.1 isoforms results from both alternative splicing and expression of distinct genes. © 1998 Academic Press

INTRODUCTION

Protein 4.1 is a cytoskeletal protein whose structure and function have been well characterized in the context of the erythroid plasma membrane. As an essential component of the spectrin-based membrane skeleton, 4.1 plays a critical role in determining the morphology and biophysical properties of red cells. Extensive biochemical studies have defined several key protein–protein interactions that mediate these cellular effects. An internal spectrin–actin binding domain is a key element in the formation of the junctional complexes that laterally crosslink spectrin heterodimers in the plane of the skeleton. In addition, an N-terminal membrane binding domain specifies vertical interactions between 4.1 and overlying integral membrane proteins

such as band 3 and glycophorin C. These latter interactions play a critical role in anchoring the membrane skeleton to the plasma membrane.

Our understanding of the erythroid membrane skeleton serves as a useful paradigm for exploring potentially analogous structures in nonerythroid cells both at the plasma membrane and in various subcellular compartments (Beck *et al.*, 1994, 1997; Devarajan *et al.*, 1996). Many of the prototypical erythroid skeletal proteins and/or close homologues encoded by evolutionarily related genes are expressed in a variety of cell types. For example, small gene families encoding erythroid and nonerythroid homologues of band 3 (anion exchanger), ankyrin (Devarajan *et al.*, 1996; Kordeli *et al.*, 1995; Lambert and Bennett, 1993; Otto *et al.*, 1991; Peters *et al.*, 1995), and spectrin (Chang *et al.*, 1993; Hu *et al.*, 1995; Winkelmann *et al.*, 1990) have been characterized. The cognate protein isoforms typically retain regions of high evolutionary conservation which presumably encode critical “core” functional domains and regions of lesser homology likely to specify novel subcellular localization and/or protein interaction sites.

A detailed characterization of nonerythroid 4.1 proteins has not been reported at the molecular level. An intriguing collection of experiments revealing a remarkable diversity in molecular weight and subcellular compartmentalization of 4.1-immunoreactive proteins in nucleated cells suggests that the story is likely to be quite complex. Western blots have revealed that 4.1 polypeptides range in size from 30 to 210 kDa among various nonerythroid cells (Anderson *et al.*, 1988). Furthermore, immunofluorescence studies of nucleated cells have demonstrated colocalization of 4.1 protein(s) with stress fibers (Cohen *et al.*, 1982), intranuclear compartments (Correas, 1991; De Carcer *et al.*, 1995; Krauss *et al.*, 1997b; Lallena and Correas, 1997), centrosomes (Krauss *et al.*, 1997a), and neuronal axons and dendrites (Krebs *et al.*, 1987b; Zimmer *et al.*, 1992).

The genetic origin of these putative 4.1 homologues is unknown. Until now, a single 4.1 gene² encoding the

Sequence data from this article have been deposited with the GenBank Library under Accession No. AF027299.

¹ To whom correspondence should be addressed at the Life Sciences Division, MS 74-174, Lawrence Berkeley National Laboratory, Berkeley, CA 94720. Telephone: (510) 486-6973. Fax: (510) 486-6746.

² The HGMW-approved symbols for the genes described in this paper are EPB41 (4.1R) and EPB41L2 (4.1G).

prototypical red cell protein 4.1R has been cloned and mapped to chromosome 1. The salient features of this gene include its large size (Baklouti *et al.*, 1997; Huang *et al.*, 1993), its demonstrated expression in selected nonerythroid cell types, and its extraordinarily complex expression pattern resulting from extensive alternative pre-mRNA splicing (Conboy *et al.*, 1988, 1991; Tang *et al.*, 1990). Importantly, regulation of splicing in tissue- and differentiation-specific patterns gives this gene the potential to encode a characteristic and distinct complement of 4.1 polypeptides in different cell types. Whether alternative splicing alone can generate the observed 4.1 protein diversity, or whether this isoform diversity might also result from expression of multiple 4.1R-related genes, has not been explored systematically.

As one strategy for exploring the genetic origin of nonerythroid 4.1 proteins, we have performed extensive computer searches of available cDNA databases. This approach led to the identification in the EST database of novel 4.1-like cDNAs potentially encoded by distinct gene(s). Using this information as a starting point, we have carried out extensive characterization of a unique gene product, which we now designate 4.1G. By criteria of nucleotide sequence, mRNA size, mRNA tissue distribution, and chromosomal localization, 4.1G and the prototypical 4.1R are encoded by two distinct genes. Initial protein expression studies suggest that a major 113-kDa isoform of 4.1G localizes differently from the familiar 80-kDa 4.1R in nucleated cells. Furthermore, on the basis of its extensive homology to the membrane binding, spectrin-actin binding, and COOH-terminal domains of 4.1R, 4.1G represents a closer homologue to protein 4.1 than the previously described 4.1 superfamily members ezrin, radixin, and moesin.

MATERIALS AND METHODS

Isolation of 4.1G cDNA. The human 4.1R amino acid sequence was used as a query sequence to screen the EST database, resulting in identification of two groups of novel cDNA sequences with high homology to the membrane binding domain (e.g., clone W26747) and C-terminal domain (e.g., clone W31744) of 4.1R. Sequencing of W31744, purchased from Research Genetics (Huntsville, AL), revealed that it extends through the entire 1.25-kb 3'UTR to the poly(A) tail. This sequence was distinct from 4.1R and confirmed that it was derived from a distinct gene. To clone the intervening region of this novel cDNA, we employed an RT-PCR strategy using primers located within W26747 and W31744, respectively (sense strand, 5'-GGGAGACCAAGGAAGTGC-3'; antisense strand, 5'-GAGCAATAGACCAGAGCTGGTAAG-3'). The antisense primer was used to prime cDNA synthesis using brain RNA as a template. Conditions for PCR were as follows: 35 cycles for 30 s at 94°C, 30 s at 58°C, 3 min at 72°C. The authenticity of this PCR product was later confirmed by sequence analysis of clone AA333370 (purchased from ATCC), which extended from the membrane binding domain to the 3'UTR. To clone the full 5' coding sequence, a modified 5' RACE experiment was performed. DNA from a human heart cDNA library (Clontech HL3026a) was amplified using a gene-specific antisense primer (5'-GCACTTCCTTGGTCTCCC-3') located near the 5' end of EST clone W26747, together with a λ gt10 forward primer. One of the resulting PCR fragments extended the existing 4.1G sequence to an apparent

translation start codon that (a) contained an AUG sequence situated in the context of a good Kozak consensus site and (b) encoded an N-terminal sequence with identity to the first four amino acids of 4.1R. As an additional control to verify the authenticity of this RACE product, we amplified a region of brain 4.1G sequences extending from near the 5' end of the RACE product (sense primer 5'-GATCCCGGG-AATGACTACTGAAGTAGGCT-3'; start codon is underlined and sequence preceding it is added to facilitate cloning) to a position farther downstream in the region encoding the membrane binding domain (antisense primer 5'-TCTGTCTTTGTAATGAGAA-3'). These multiple approaches enabled us to assemble the entire 4.1G cDNA sequence.

RNA blot analysis. 4.1G mRNA size was determined by hybridization to multiple tissue Northern blots (Clontech, Palo Alto, CA). The tissue distribution of 4.1G mRNA expression was examined by hybridization to master RNA blots containing poly(A)⁺ RNA from a number of different tissues (Clontech). Probes used for RNA blots were derived from the N-terminal coding region of 4.1G (nt 1–615), the 3' untranslated region of 4.1G (nt 3364–4139), and a portion of the 3' untranslated sequence of 4.1R (nt 3167–3899).

Chromosomal mapping of 4.1G gene. A bacterial artificial chromosome (BAC) library of human genomic DNA was screened using a PCR strategy with oligonucleotide primers in a unique region of the 4.1G cDNA (sense strand primer, 5'-GCCCAAAAGATACCCGGAGA-3'; antisense strand primer, 5'-CTCTGAACAACAAATAATTGTGGCA-3'). Among several positive 4.1G genomic clones identified in this manner, one having a 137-kb insert was judged likely to contain the entire 4.1G gene by virtue of its ability to be amplified with a primer pair situated near the 5' end of the cDNA (sense strand primer, 5'-GATCCCGGGAATGACTACTGAAGTAGGCT-3'; antisense strand primer, 5'-TTGCTGCTCCTCCTTCTA-3') and another primer set within the 3' untranslated region (sense strand primer, 5'-CTGGCTCAGGCGATCAGGGA-3'; antisense strand primer, 5'-CCTTCCTCAGCCAATC-3') (results not shown). A biotinylated probe was prepared from this 137-kb 4.1G clone by random priming using the BioPrime kit according to manufacturer's instructions (Gibco BRL), hybridized to metaphase chromosomes, and visualized using FITC-avidin. Images were acquired and analyzed using Quips-XL software (Vysis, Downers Grove, IL) as described previously (Piper *et al.*, 1995).

Immunofluorescence of transfected cells. COS cells cultured on coverslips in DMEM were transiently transfected with 2.5 μ g of purified plasmid DNA using LipofectAMINE according to the manufacturer's instructions (Life Technologies, Inc., Gaithersburg, MD). Plasmids contained cDNA encoding either the major erythroid 80-kDa isoform of 4.1R (bearing a C-terminal HA epitope tag) or the full coding region of 4.1G (similarly with an HA epitope tag). Forty-eight hours after transfection, cells were washed twice in PBS, fixed in PBS containing 3% paraformaldehyde for 30 min at room temperature, and incubated in PBS containing 0.5% Triton X-100 for 10 min at room temperature. After extensive washing with PBS, coverslips were blocked with PBS plus 10 mg/ml BSA and 10% (v/v) goat serum, washed again with PBS containing 0.1 mg/ml BSA, and incubated for 60 min at room temperature with primary anti-HA epitope tag (Zymed Laboratories Inc., South San Francisco, CA) antibody in PBS plus 10 mg/ml BSA. Secondary anti-rabbit IgG conjugated to FITC (Molecular Probes, Inc., Eugene, OR) was added for 60 min at room temperature, and samples were prepared for image analysis using a Zeiss Axioplan microscope.

Western blot analysis. COS cells (10⁶) seeded in 100-mm culture dishes were grown for 24 h and transfected as above using 18 μ l of LipofectAMINE and 7.5 μ g of DNA per dish. After 48 h cells were washed in PBS and lysed in 500 μ l of ice-cold RIPA buffer (10 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1% IGEPAL CA-630, 0.1% sodium dodecyl sulfate, 2 mM Pefabloc, 2 mM benzimidazole, 10 μ g/ml aprotinin, 5 μ g/ml leupeptin, 2 μ g pepstatin A) for 30 min on ice, and the supernatant was cleared of particulate materials by centrifugation. One hundred microliters of protein A agarose beads, preincubated overnight with 5 μ g anti-HA epitope tag antibody, was added and allowed to bind for 3 h at 4°C. Beads were spun down and washed

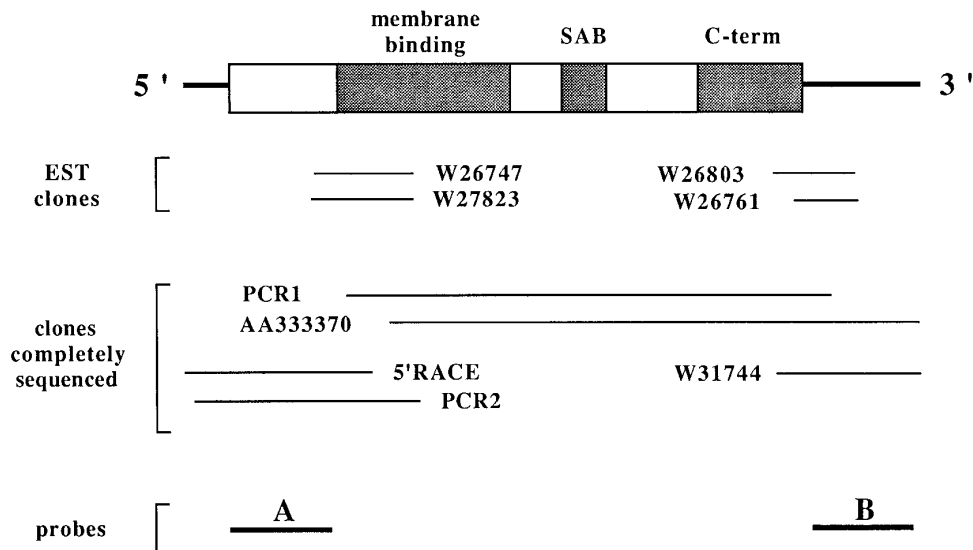


FIG. 1. Strategy for assembly of the 4.1G cDNA sequence. **Top** represents the structure of the 4.1G protein deduced from analysis of the cDNAs shown below. The two pairs of EST clones depicted are representative of a larger number of overlapping cDNAs. The full-length structure of 4.1G was derived by complete sequence analysis of the cDNA clones and PCR fragments illustrated. Shown at the **bottom** are the cDNA probes used for Northern blot analysis.

in RIPA buffer, and associated proteins were denatured by boiling in SDS-PAGE sample buffer and resolved by electrophoresis in a 7.5% polyacrylamide gel. Proteins were transferred to a nitrocellulose membrane using a semidry electroblotter (Integrated Separation Systems, Inc., Natick, MA). After a blocking step (1 h, room temperature) performed with TBS (10 mM Tris-HCl, pH 7.4, 150 mM NaCl) containing 0.1% Tween 20, 4% nonfat dry milk, 1% BSA, and 0.02% sodium azide, blots were probed for 1 h at room temperature with anti-HA antibody at 1 μ g/ml (diluted in the same TBS buffer). Immunoreactive bands were visualized after incubation with the secondary antibody, anti-rabbit IgG coupled to horseradish peroxidase, using the Renaissance chemiluminescence detection kit (NEN Life Science Products, Boston, MA).

RESULTS

Cloning of 4.1G cDNA. The existence of novel 4.1R-related cDNAs was revealed by computer searches of the DNA databases. When the human erythroid 4.1 amino acid sequence was used to query the EST database, two groups of cDNAs with high homology to the 4.1R membrane binding domain (e.g., clone W26747) and C-terminal domain (e.g., clone W31744) were identified. Since the sequences within each group were virtually identical, we hypothesized that they represented fragments of a single novel transcript highly homologous to erythroid 4.1. RT-PCR was therefore employed to demonstrate directly that these two islands of sequence were indeed derived from a larger common transcript and to clone the intervening region (Fig. 1, clone PCR1). By analysis of multiple independent PCR subclones, we deduced the entire amino acid sequence

of this 4.1 homologue except for the N- and C-terminal ends; this structure was later confirmed by sequence analysis of cDNA AA333370 which extended across the same region. The N-terminal region was derived from a 5' RACE product that contained an AUG at nt 45–47, assigned as the start codon by the following criteria: it is the first AUG in the cDNA, it is preceded by an in-frame stop codon, it is situated in a good Kozak consensus sequence for translation initiation, and the first four amino acids of the predicted product are identical to residues 1–4 of the high-molecular-weight isoform of 4.1R. Finally, at the 3' end of the cDNA sequence, complete sequence analysis of EST clone W31744 showed that it encodes the C-terminal coding sequence and entire 3' untranslated region.

The sequence of this novel cDNA, designated 4.1G, exhibits an open reading frame of 1005 amino acids, predicted to encode a protein of \sim 113 kDa (Fig. 2). Alignment of 4.1G sequences with the prototypical 4.1R protein, as illustrated by the dot matrix plot of Fig. 3, revealed an interesting pattern of alternating conserved and unique sequence domains. Within the membrane binding domain, the two proteins are 76% identical over a span of 336 amino acids. The reported binding site for band 3, however, is altered from LEEDY in 4.1R (Jons and Drenckhahn, 1992) to LEKDY in 4.1G. This substitution of a basic residue for an acidic is likely to alter its function dramatically. Another functionally critical domain, which is required for binding to spectrin-actin, was also highly conserved (73% iden-

FIG. 2. Structure of 4.1G cDNA and predicted protein. Below the cDNA sequence is the derived amino acid sequence using the single-letter code. The first ATG, which is preceded almost immediately by an in-frame stop codon, predicts a short N-terminal peptide identical to the start of 135-kDa 4.1R. The putative polyadenylation sequence is underlined. Amino acid sequence is numbered in italics. This sequence has been deposited with GenBank (Accession No. AF027299).

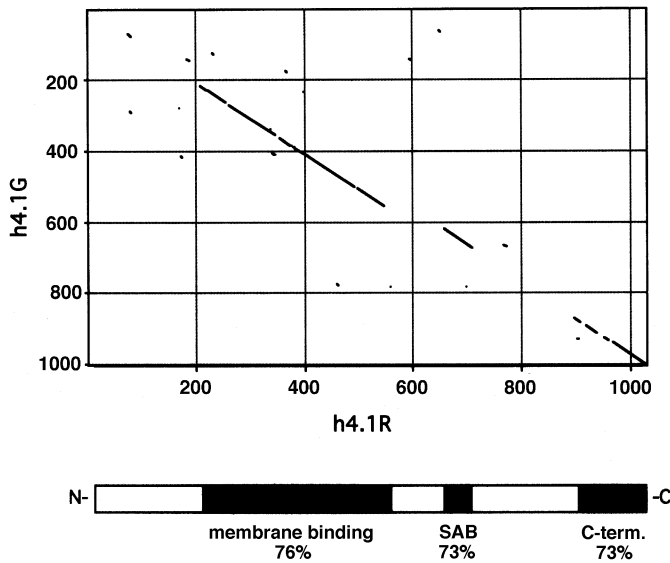


FIG. 3. Sequence comparison of 4.1R and 4.1G polypeptides. 4.1R and 4.1G sequences were aligned using the dot matrix program (Macvector). Each dot represents a minimum identity of 60% over a window of eight residues. The full amino acid sequence of 4.1R, derived from all known constitutive and alternative exons of the gene, was used for this comparison. Below is a domain map of the 4.1R showing the three regions of homology to 4.1G.

tity over 72 amino acids). Interestingly, this region of homology corresponds approximately to the minimal high-affinity spectrin-actin binding domain defined in earlier biochemical studies of 4.1R (Schischmanoff *et al.*, 1995). Finally, the COOH-terminal domain of the two proteins is also highly conserved (73% identity over 135 amino acids). This region has been reported to bind NuMA (Mattagajasingh *et al.*, 1996).

In addition to the similarities described above, 4.1R and 4.1G also exhibit several distinctive features, including apparent differences in initiation codon utilization as well as interspersed regions of unique sequence. Two translation initiation sites existing within 4.1R mRNA isoforms can encode two size classes of polypeptide of ~135 kDa (from upstream AUG-1) and ~80 kDa (from downstream AUG-2). In contrast, only one apparent initiation site has been detected thus far in 4.1G cDNAs. This site corresponds in location to the upstream AUG-1 site in 4.1R and promotes expression of an N-terminal domain similar in size to the N-terminal extension of high-molecular-weight 4.1R isoforms. However, aside from two or three short conserved peptides, the N-terminal domains of these two proteins exhibit little sequence resemblance. 4.1G does not appear to have a counterpart of the 4.1R downstream AUG and therefore is not predicted to encode a direct homologue of the ~80-kDa red cell 4.1R isoform.

A second region of nonhomology between 4.1R and 4.1G is located between the membrane and the spectrin-actin interaction sites and corresponds to the region encoded by exon 13 of the 4.1R gene. It has previously been noted that this region of ~63 amino acids is the least conserved domain among 4.1R proteins

from mammalian (human, mouse, dog) and amphibian (frog) species (Winardi *et al.*, 1995). Together these data suggest that this domain represents a spacer region whose overall length, but not primary sequence, is important for 4.1 protein function.

Finally, both 4.1R and 4.1G genes contain alternative exons capable of encoding unique sequence motifs in the region between the spectrin-actin binding domain and the COOH-terminal domain. In 4.1R, such splicing events are manifested in muscle by inclusion of exon 17A (Baklouti *et al.*, 1997; Schischmanoff *et al.*, 1997) and in certain epithelial tissues by inclusion of exon 17B (Schischmanoff *et al.*, 1997). 4.1G isoforms also contain unique sequence inserted at the position corresponding to the boundary of predicted exons 17 and 18; however, this domain is not homologous to exons 17A/17B of 4.1R. Preliminary experiments suggest that this region of 4.1G is also subject to alternative splicing, but the precise exon structure of this part of the gene remains to be determined (data not shown).

Characterization of 4.1G mRNA. Northern blot analysis was performed with cDNA probes representing unique sequences in the 5' coding region and 3'UTR that exhibit little or no homology to 4.1R mRNA. Figure 4 shows that 4.1G mRNA is widely expressed among human tissues. Most tissues exhibited a prominent ~5-kb mRNA that was detected with both 5' (Fig. 4A) and 3' (Fig. 4B) probes, supporting the notion that we have assembled a bona fide full-length 4.1G cDNA sequence. In brain mRNA, in contrast, both

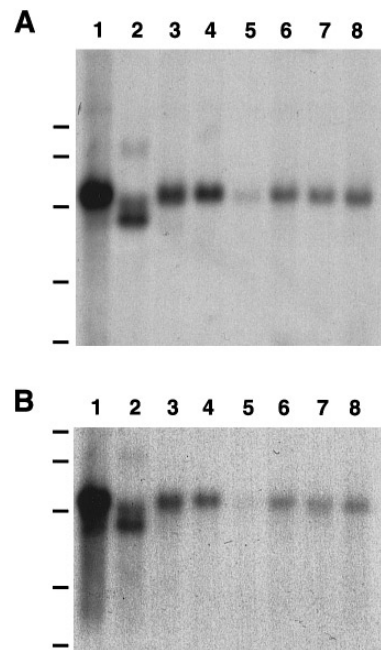


FIG. 4. Northern blot analysis of 4.1G mRNA. Human mRNA was hybridized under stringent conditions with cDNA probes derived from 5' coding sequences (A, probe A in Fig. 1) and the 3'UTR (B, probe B in Fig. 1). RNA size standards of 9.5, 7.5, 4.4, 2.4, and 1.35 kb are shown at the left. Sources of RNAs: lane 1, heart; lane 2, brain; lane 3, placenta; lane 4, lung; lane 5, liver; lane 6, skeletal muscle; lane 7, kidney; and lane 8, pancreas.

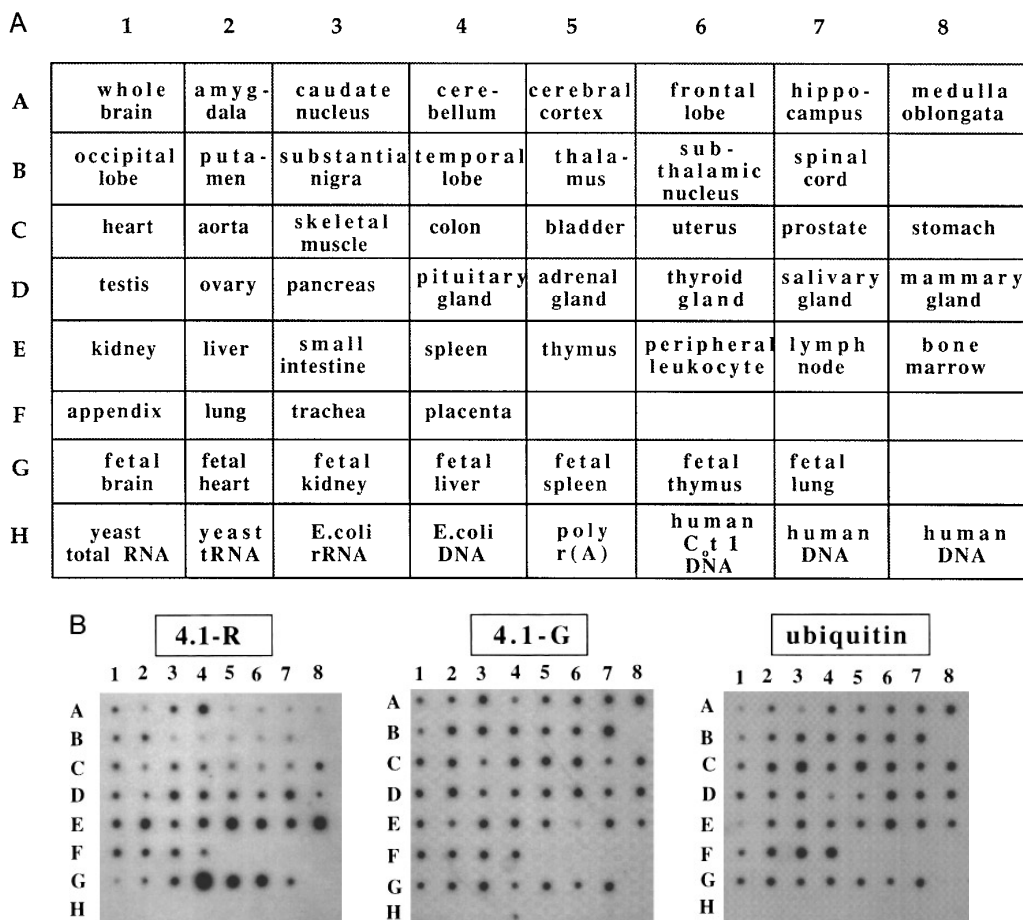


FIG. 5. Expression pattern of 4.1 genes in various human tissues. (A) Grid showing the identification of RNA sources on the RNA master blot (Clontech). (B) Hybridization patterns obtained with probes from 4.1R and 4.1G. Ubiquitin cDNA probe was used as a positive control.

probes hybridized to a distinct set of transcripts of 4.6 and 4.1 kb, as well as to a less abundant 7.3-kb transcript. This observation suggests that alternative splicing and/or differential use of polyadenylation sites may occur in 4.1G transcripts expressed in brain.

To explore the expression patterns of 4.1G mRNA further, and to compare its expression with that of 4.1R mRNA, RNA blot analysis was performed with a larger panel of RNAs extracted from human tissues (Fig. 5). 4.1R mRNA was most abundant in hematopoietic tissues such as bone marrow, fetal liver, and leukocytes and to a lesser extent in kidney, adrenal, and cerebellum. In contrast, 4.1G was generally expressed over a wide range of tissues, but was relatively less abundant in the hematopoietic tissues.

Chromosomal localization. To localize the 4.1G gene to a specific chromosome, we first isolated an ~137-kb BAC clone containing the 4.1G genomic locus (Materials and Methods). Biotin-labeled genomic DNA prepared from this clone was used as a hybridization probe against a metaphase spread of human chromosomes. As shown in Fig. 6, the 4.1G gene locus maps to chromosome 6q22–q23. This location was confirmed by analysis of multiple



FIG. 6. Chromosomal localization of the 4.1G gene. At left is depicted an idiogram of chromosome 6. The center image represents an inverted image of DAPI-stained chromosome 6 to visualize banding patterns. At right is the same chromosome 6 double-labeled with DAPI and FITC-avidin-biotinylated 4.1G probe. The location of 4.1G at 6q22–q23 was observed in at least 10 chromosome spreads; no other site was consistently labeled.

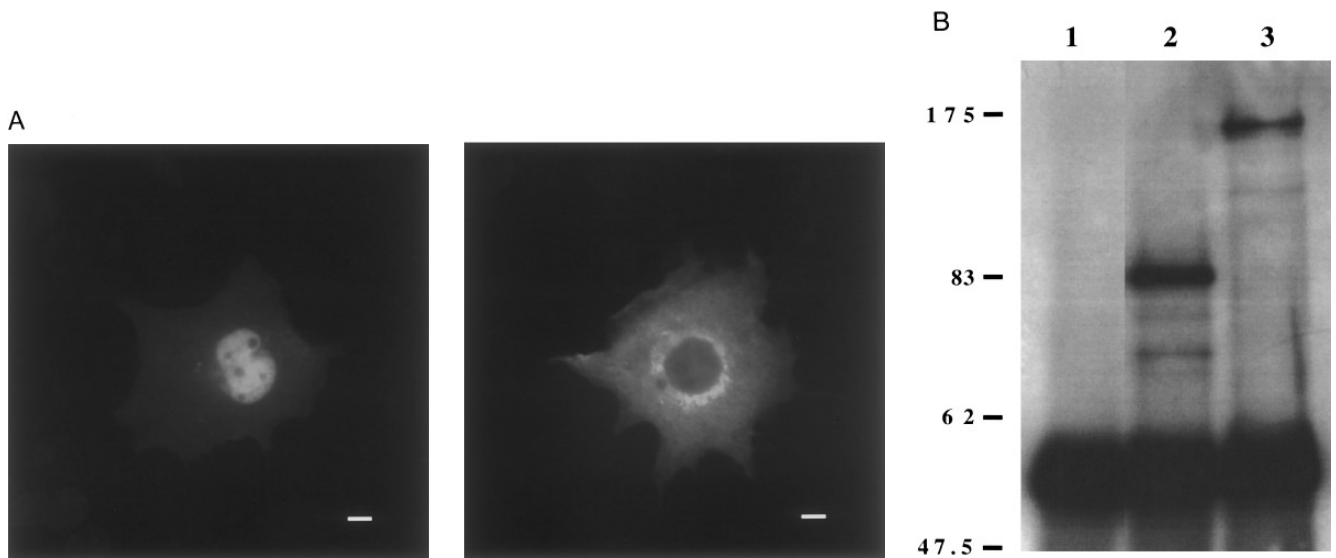


FIG. 7. Intracellular localization of the 4.1G protein. **(A)** COS cells were transfected with plasmids encoding HA-epitope tagged 80-kDa 4.1R (**left**) or the 113-kDa isoform of 4.1G described in Fig. 2 (**right**), then examined by immunofluorescence microscopy using affinity-purified polyclonal anti-HA rabbit IgG and FITC-labeled goat anti-rabbit IgG. Untransfected cells in the same field were not stained with the HA antibody. **(B)** Protein extracts from 4.1R-transfected (lane 2) or 4.1G-transfected (lane 3) cultures were immunoprecipitated using anti-HA antibody, electrophoresed on SDS-polyacrylamide gels, and immunoblotted with the same anti-HA antibody. The control sample in lane 1 was transfected with a 4.1R construct bearing a different epitope tag not recognized by anti-HA antibodies. The strong band at 50–55 kDa represents IgG heavy chains.

metaphase chromosome spreads; no other site was consistently labeled with the 4.1G-specific probe. Since the 4.1R gene was previously assigned to chromosome 1, these two evolutionarily related homologues are not genetically linked in the human genome.

Intracellular localization of 4.1G. COS cells were transiently transfected with plasmids encoding either a prototypical ~80-kDa erythroid 4.1R protein or the full-length 4.1G protein described in Fig. 2. In each case the 4.1 proteins were linked to an HA-epitope tag to facilitate specific detection/localization of the transfected products. Immunofluorescence microscopy performed with affinity purified anti-HA plus FITC-labeled secondary antibody revealed a distinct localization of these 4.1 homologues. The 4.1R isoform was compartmentalized mainly in a nuclear fraction as reported earlier (Krauss *et al.*, 1997b) (Fig. 7A, left). In contrast, 4.1G transfected in parallel into identical cultures exhibited a diffuse cytoplasmic staining as well as a more intense perinuclear localization (Fig. 7A, right). To demonstrate that transfected cells were expressing the expected full-length 4.1 proteins, immunoprecipitation of protein extracts was performed (Fig. 7B). The 4.1R-transfected culture expressed the expected 80-kDa polypeptide (lane 2), while the 4.1G-transfected cells expressed a single protein of ~160-kDa. This apparent molecular weight of 4.1G is larger than the calculated size of 113-kDa, a phenomenon also observed in the high-molecular-weight isoforms of 4.1R.

DISCUSSION

In this study we identified and characterized the complete cDNA encoding 4.1G, a close homologue of

the erythroid membrane skeletal protein 4.1R. 4.1G mRNA encodes a protein that exhibits substantial sequence homology to the major protein interaction domains of 4.1R, but is more widely expressed among human tissues. The discovery of a novel 4.1G gene is important in several respects. First, the existence of multiple protein 4.1 genes, coupled with complex alternative splicing of the cognate protein 4.1 pre-mRNAs, may provide a genetic explanation for the heterogeneity of 4.1-immunoreactive proteins previously observed in nonerythroid cells. Second, new 4.1 isoforms may potentially have novel intracellular functions by participating in assembly of internal skeletal elements not previously appreciated (see also below). Finally, the

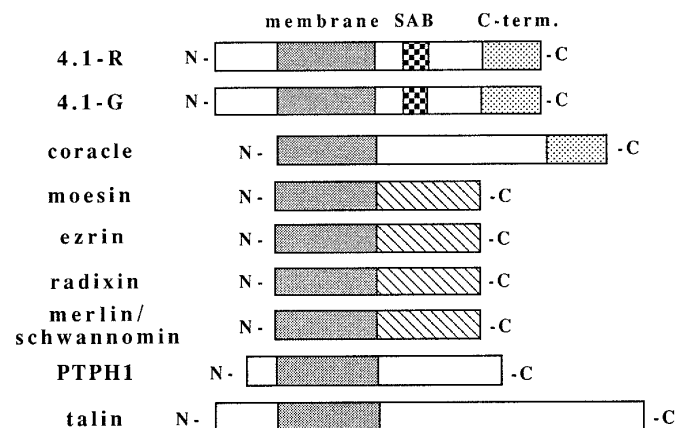


FIG. 8. 4.1 superfamily members. Domain structure of proteins with significant homology to the prototypical protein 4.1R. ■, membrane binding domain; ▨, spectrin-actin binding domain; ▩, C-terminal homology domain. A domain shared among ERM proteins that is not found in 4.1 is represented by ▨.

size and scope of the protein 4.1 superfamily is now expanded to include very close homologues of the prototypical erythroid 4.1 (Fig. 8). Other members of the family include several proteins with homology limited to the membrane binding domain of 4.1R, including ERM proteins such as ezrin, radixin, and moesin, as well as talin and several protein phosphatases. The *Drosophila* protein coracle has domains similar to both the N-terminal membrane binding domain and the C-terminal domain of 4.1R, but no apparent spectrin-actin binding domain. 4.1G is the first new member of this family to include domains homologous to the spectrin-actin binding motif of 4.1R.

The apparent overlap in tissue distribution between 4.1R and 4.1G indicates that these polypeptides may be coexpressed in select nucleated cells. What might be the underlying cell biological rationale for synthesis of multiple 4.1 homologues? The intriguing domain organization of these two proteins, with highly homologous protein interaction domains interspersed with unique regions of undefined function, suggests that 4.1R and 4.1G proteins may play fundamentally analogous roles but at different intracellular sites. Initial localization studies indicate that major isoforms of 4.1R and 4.1G exhibit different subcellular compartmentalization (Fig. 7); however, it should be noted that this pattern reflects the distribution of two specific isoforms, and the composite localization pattern of all 4.1R and 4.1G isoforms may be more complex. Precedence for differential compartmentalization of close skeletal protein homologues has been reported earlier in the case of spectrin and ankyrin gene products expressed in neuronal cells. Brain spectrin (240/235) is found predominantly in neuronal axons, whereas the distinct brain spectrin (240/235E) is compartmentalized mainly in cell bodies and dendrites (Riederer *et al.*, 1987, 1988). Similarly, specific ankyrin isoforms have been localized to neuronal cell bodies and dendrites, but not axons (Kordeli and Bennett, 1991), while others are found in unmyelinated axons (Chan *et al.*, 1993). Finally, isoforms of a 4.1-immunoreactive protein named amelin may be located at different intracellular sites (Krebs *et al.*, 1987a; Zimmer *et al.*, 1992). Such mechanisms may help explain the diverse immunofluorescence data showing that 4.1 protein(s) localize to peripheral membrane sites (Chasis *et al.*, 1993; Lue *et al.*, 1994), stress fibers (Cohen *et al.*, 1982), nuclei (Correas, 1991; De Carcer *et al.*, 1995; Krauss *et al.*, 1997b), and centrosomes (Krauss *et al.*, 1997a).

The domain structure of 4.1G bears striking resemblance to predicted high-molecular-weight isoforms of 4.1R that contain both (1) an N-terminal extension upstream of the conserved membrane binding domain and (2) a unique region inserted between the SAB and the C-terminal domains (Schischmanoff *et al.*, 1997). We therefore hypothesize that the primordial 4.1 gene encoded a high-molecular-weight protein and that the familiar mammalian erythroid 80-kDa isoform represents a more recent adaptation. By introduction of an

internal translation initiation site (AUG-2), and by utilization of alternative splicing to regulation expression of the upstream AUG-1 and the domain downstream of SAB, a smaller 80-kDa isoform specifically evolved to function in the erythroid membrane skeleton. Interestingly, sequence analysis of the spectrin gene family has supported a related hypothesis, namely that the erythroid spectrin genes have evolved recently and diverged rapidly from the nonerythroid spectrin counterparts (Muse *et al.*, 1997).

Finally, it is interesting to speculate that the gene families encoding membrane skeletal proteins may be coevolving via gene duplication and divergence to generate distinct skeletal structures in various eukaryotic cell types. In this view, the prototypical erythroid proteins 4.1R, ankyrin_R, and β -spectrin would represent components of a coevolving, specialized membrane skeleton adapted to mechanically support the red cell membrane. More widely expressed isoforms of protein 4.1 (this paper), ankyrin [ankyrin_C/ank3 (Devarajan *et al.*, 1996; Kordeli *et al.*, 1995; Peters *et al.*, 1995), and spectrin (β _C-spectrin/ β -fodrin (Chang *et al.*, 1993; Hu *et al.*, 1992)] are encoded by a distinct set of genes; these might represent the primordial skeletal components with a generalized function. A third set of homologues with strong expression in brain includes brain ankyrin [ankyrin_B/ank2 (Chan *et al.*, 1993; Kunitomo *et al.*, 1991; Otto *et al.*, 1991)], brain isoforms of protein 4.1 (e.g., Parra *et al.*, 1997; Walensky *et al.*, 1997), and probably other as yet uncharacterized skeletal proteins. These homologues may assemble into a specialized neuronal skeleton and/or they may functionally interact with components of the "constitutive" widely expressed skeletal elements. Much future work on the structure and function of nonerythroid membrane skeleton proteins will be required to explore these issues in depth.

ACKNOWLEDGMENTS

The authors thank Jan-Fang Cheng, Steve Lowry, Thomas Duell, and Damir Sudar for their generous help with isolation and chromosomal mapping of the 4.1G genomic clones.

REFERENCES

- Anderson, R. A., Correas, I., Mazzucco, C., Castle, J. D., and Marchesi, V. T. (1988). Tissue-specific analogues of erythrocyte protein 4.1 retain functional domains. *J. Cell. Biochem.* **37**: 269–284.
- Baklouti, F., Huang, S.-C., Vulliamy, T. J., Delaunay, J., and Benz, E. J., Jr. (1997). Organization of the human protein 4.1 genomic locus: New insights into the tissue-specific alternative splicing of the pre-mRNA. *Genomics* **39**: 289–302.
- Beck, K. A., Buchanan, J. A., Malhotra, V., and Nelson, W. J. (1994). Golgi spectrin: Identification of an erythroid β -spectrin homolog associated with the Golgi complex. *J. Cell Biol.* **127**: 707–723.
- Beck, K. A., Buchanan, J. A., and Nelson, W. J. (1997). Golgi membrane skeleton: Identification, localization and oligomerization of a 195 kDa ankyrin isoform associated with the Golgi complex. *J. Cell Sci.* **110**: 1239–1249.
- Chan, W., Kordeli, E., and Bennett, V. (1993). 440-kD ankyrinB: Structure of the major developmentally regulated domain and selective localization in unmyelinated axons. *J. Cell Biol.* **123**: 1463–1473.

- Chang, J. G., Scarpa, A., Eddy, R. L., Byers, M. G., Harris, A. S., *et al.* (1993). Cloning of a portion of the chromosomal gene and cDNA for human beta-fodrin, the nonerythroid form of beta-spectrin. *Genomics* **17**: 287–293.
- Chasis, J. A., Coulombel, L., Conboy, J., McGee, S., Andrews, K., *et al.* (1993). Differentiation-associated switches in protein 4.1 expression: Synthesis of multiple structural isoforms during normal human hematopoiesis. *J. Clin. Invest.* **91**: 329–338.
- Cohen, C. M., Foley, S. F., and Korsgren, C. (1982). A protein immunologically related to erythrocyte band 4.1 is found on stress fibres of non-erythroid cells. *Nature* **299**: 648–650.
- Conboy, J. G., Chan, J., Chasis, J. A., Kan, Y. W., and Mohandas, N. (1991). Tissue- and development-specific alternative RNA splicing regulates expression of multiple isoforms of erythroid membrane protein 4.1. *J. Biol. Chem.* **266**: 8273–8280.
- Conboy, J. G., Chan, J., Mohandas, N., and Kan, Y. W. (1988). Multiple protein 4.1 isoforms produced by alternative splicing in human erythroid cells. *Proc. Natl. Acad. Sci. USA* **85**: 9062–9065.
- Correas, I. (1991). Characterization of isoforms of protein 4.1 present in the nucleus. *Biochem. J.* **279**: 581–585.
- De Carcer, G., Lallena, M. J., and Correas, I. (1995). Protein 4.1 is a component of the nuclear matrix of mammalian cells. *Biochem. J.* **312**: 871–877.
- Devarajan, P., Stabach, P. R., Mann, A. S., Ardito, T., Kashgarian, M., *et al.* (1996). Identification of a small cytoplasmic ankyrin (AnkG119) in the kidney and muscle that binds beta I sigma spectrin and associates with the Golgi apparatus. *J. Cell Biol.* **133**: 819–830.
- Hu, R. J., Moorthy, S., and Bennett, V. (1995). Expression of functional domains of beta G-spectrin disrupts epithelial morphology in cultured cells. *J. Cell Biol.* **128**: 1069–1080.
- Hu, R. J., Watanabe, M., and Bennett, V. (1992). Characterization of human brain cDNA encoding the general isoform of beta-spectrin. *J. Biol. Chem.* **267**: 18715–18722.
- Huang, J.-P., Tang, C.-J. C., Kou, G.-H., Marchesi, V. T., Benz, E. J., *et al.* (1993). Genomic structure of the locus encoding protein 4.1. *J. Biol. Chem.* **268**: 3758–3766.
- Jons, T., and Drenckhahn, D. (1992). Identification of the binding interface involved in linkage of cytoskeletal protein 4.1 to the erythrocyte anion exchanger. *EMBO J.* **11**: 2863–2867.
- Kordeli, E., and Bennett, V. (1991). Distinct ankyrin isoforms at neuron cell bodies and nodes of Ranvier resolved using erythrocyte ankyrin-deficient mice. *J. Cell Biol.* **114**: 1243–1259.
- Kordeli, E., Lambert, S., and Bennett, V. (1995). AnkyrinG. A new ankyrin gene with neural-specific isoforms localized at the axonal initial segment and node of Ranvier. *J. Biol. Chem.* **270**: 2352–9.
- Krauss, S. W., Chasis, J. A., Rogers, C., Mohandas, N., Krockmalnic, G., *et al.* (1997a). Structural protein 4.1 is located in mammalian centrosomes. *Proc. Natl. Acad. Sci. USA* **94**: 7297–7302.
- Krauss, S. W., Larabell, C. A., Lockett, S., Gascard, P., Penman, S., *et al.* (1997b). Structural protein 4.1 in the nucleus of human cells: Dynamic rearrangements during cell division. *J. Cell Biol.* **137**: 275–289.
- Krebs, K. E., Zagon, I. S., and Goodman, S. R. (1987a). Amelin and synapsin I are 4.1 related spectrin binding proteins in brain. *Brain Res. Bull.* **18**: 793–798.
- Krebs, K. E., Zagon, I. S., and Goodman, S. R. (1987b). Amelin: A 4.1-related spectrin-binding protein found in neuronal cell bodies and dendrites. *J. Neurosci.* **7**: 3907–3914.
- Kunimoto, M., Otto, E., and Bennett, V. (1991). A new 440-kD isoform is the major ankyrin in neonatal rat brain. *J. Cell Biol.* **115**: 1319–1331.
- Lallena, M. J., and Correas, I. (1997). Transcription-dependent redistribution of nuclear protein 4.1 to SC35-enriched nuclear domains. *J. Cell Sci.* **110**: 239–247.
- Lambert, S., and Bennett, V. (1993). From anemia to cerebellar dysfunction: A review of the ankyrin gene family. *Eur. J. Biochem.* **211**: 1–6.
- Lue, R. A., Marfatia, S. M., Branton, D., and Chishti, A. H. (1994). Cloning and characterization of hdlg: The human homologue of the *Drosophila* discs large tumor suppressor binds to protein 4.1. *Proc. Natl. Acad. Sci. USA* **91**: 9818–9822.
- Mattagajasingh, S. N., Huang, S. C., and Benz, E. J., Jr. (1996). Direct evidence for a nuclear localization and function of protein 4.1 in the nucleus: In vivo association with mitotic apparatus proteins. *Blood* **88**(Suppl. 1): 276a.
- Muse, S. V., Clark, A. G., and Thomas, G. H. (1997). Comparisons of the nucleotide substitution process among repetitive segments of the α - and β -spectrin genes. *J. Mol. Evol.* **44**: 492–500.
- Otto, E., Kunimoto, M., McLaughlin, T., and Bennett, V. (1991). Isolation and characterization of cDNAs encoding human brain ankyrins reveal a family of alternatively spliced genes. *J. Cell Biol.* **114**: 241–253.
- Parra, M., Gascard, P., Walensky, L., Snyder, S., Mohandas, N., *et al.* (1997). A family of distinct protein 4.1 genes is expressed in human tissues. *Mol. Biol. Cell* **8**(Suppl.): 177a.
- Peters, L. L., John, K. M., Lu, F. M., Eicher, E. M., Higgins, A., *et al.* (1995). Ank3 (epithelial ankyrin), a widely distributed new member of the ankyrin gene family and the major ankyrin in kidney, is expressed in alternatively spliced forms, including forms that lack the repeat domain. *J. Cell Biol.* **130**: 313–330.
- Piper, J., Rutovitz, D., Sudar, D., Kallioniemi, A., Kallioniemi, O.-P., *et al.* (1995). Computer image analysis of comparative genomic hybridization. *Cytometry* **19**: 10–26.
- Riederer, B. M., Lopresti, L. L., Krebs, K. E., Zagon, I. S., and Goodman, S. R. (1988). Brain spectrin(240/235) and brain spectrin(240/235E): Conservation of structure and location within mammalian neural tissue. *Brain Res. Bull.* **21**: 607–616.
- Riederer, B. M., Zagon, I. S., and Goodman, S. R. (1987). Brain spectrin(240/235) and brain spectrin(240/235E): Differential expression during mouse brain development. *J. Neurosci.* **7**: 864–874.
- Schischmanoff, P. O., Winardi, R., Discher, D. D., Parra, M., Bicknese, S. E., *et al.* (1995). Defining of the minimal domain of protein 4.1 involved in spectrin-actin binding. *J. Biol. Chem.* **270**: 21243–21250.
- Schischmanoff, P. O., Yaswen, P., Parra, M. K., Lee, G., Chasis, J. A., *et al.* (1997). Cell-shape dependent regulation of protein 4.1 alternative splicing in mammary epithelial cells. *J. Biol. Chem.* **272**: 10254–10259.
- Tang, T. K., Leto, T., Marchesi, V. T., and Benz, E. J. (1990). Heterogeneity of mRNA and protein products arising from the protein 4.1 gene in erythroid and nonerythroid tissues. *J. Cell Biol.* **110**: 617–624.
- Walensky, L. D., Blackshaw, S. S., Conboy, J. G., Mohandas, N., and Snyder, S. H. (1997). Molecular cloning of a novel neuron-specific homologue of the erythrocyte membrane skeletal protein 4.1. *Soc. Neurosci. Abstr.* **23**: 1674.
- Winardi, R., Discher, D., Kelley, C., Zon, L., Mays, K., *et al.* (1995). Evolutionarily conserved alternative pre-mRNA splicing regulates structure and function of the spectrin-actin binding domain of erythroid protein 4.1. *Blood* **86**: 4315–4322.
- Winkelmann, J. C., Chang, J.-G., Tse, W. T., Scarpa, A. L., Marchesi, V. T., *et al.* (1990). Full-length sequence of the cDNA for human erythroid β -spectrin. *J. Biol. Chem.* **265**: 11827–11832.
- Zimmer, W. E., Zagon, I. S., Casoria, L. A., and Goodman, S. R. (1992). Identification of an amelin isoform located in axons. *Brain Res.* **582**: 94–100.