Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 1069–1073, February 1991
Neurobiology

Precerebellin is a cerebellum-specific protein with similarity to the globular domain of complement C1q B chain

(cerebellin/cDNA/development/distribution/mRNA)

Y. URADE*, J. OBERDICK, R. MOLINAR-RODE, AND J. I. MORGAN†

Department of Neurosciences, Roche Institute of Molecular Biology, Roche Research Center, Nutley, NJ 07110

Communicated by John J. Burns, November 2, 1990

ABSTRACT **The cerebellum contains a hexadecapeptide, termed cerebellin, that is conserved in sequence from human to chicken. Three independent, overlapping cDNA clones have been isolated from a human cerebellum cDNA library that encode the cerebellin sequence. The longest clone codes for a protein of 193 amino acids that we term precerebellin. This protein has a significant similarity (31.3% identity, 52.2% similarity) to the globular (non-collagen-like) region of the B chain of human complement component C1q. The region of relatedness extends over ≈145 amino acids located in the carboxyl terminus of both proteins. Unlike C1q B chain, no collagen-like motifs are present in the amino-terminal regions of precerebellin. The amino terminus of precerebellin contains three possible N-linked glycosylation sites. Although hydrophobic amino acids are clustered at the amino terminus, they do not conform to the classical signal-peptide motif, and no other obvious membrane-spanning domains are predicted from the cDNA sequence. The cDNA predicts that the cerebellin peptide is flanked by Val-Arg and Glu-Pro residues. Therefore, cerebellin is not liberated from precerebellin by the classical dibasic amino acid proteolytic-cleavage mechanism seen in many neuropeptide precursors. In Northern (RNA) blots, precerebellin transcripts, with four distinct sizes (1.8, 2.3, 2.7, and 3.0 kilobases), are abundant in cerebellum. These transcripts are present at either very low or undetectable levels in other brain areas and extraneural structures. A similar pattern of cerebellin precursor transcripts are seen in rat, mouse, and human cerebellum. Furthermore, a partial genomic fragment from mouse shows the same bands in Northern blots as the human cDNA clone. During rat development, precerebellin transcripts mirror the level of cerebellin peptide. Low levels of precerebellin mRNA are seen at birth. Levels increase modestly from postpartum day 1 to 8, then increase more dramatically between day 5 and 15, and eventually reach peak values between day 21 and 56. Because cerebellin-like immunoreactivity is associated with Purkinje cell postsynaptic structures, these data raise interesting possibilities concerning the function of the cerebellin precursor in synaptic physiology.**

A basic goal of developmental neurobiology is to identify the cellular and molecular genetic mechanisms that regulate assembly of the nervous system. The rodent cerebellum and, in particular, cerebellar Purkinje cells represent a tractable system with which to investigate biological processes and genetic regulatory factors critical in neurogenesis (1–5). Previous studies have been aimed at identifying genes that show specific or preferential expression in cerebellar Purkinje neurons (6–11). Isolation and characterization of these genes will provide an understanding of the transcriptional strategies used in the central nervous system to generate and maintain diversity (5).

The Purkinje neuron contains a hexadecapeptide, termed cerebellin (6–8), that is enriched in the postsynaptic spine (12). The peptide is developmentally regulated, and its levels reflect synapse formation during the immediate postpartum period and subsequent synapse loss during remodeling (7, 8, 12). Furthermore, perturbation of Purkinje cell synaptogenesis, as occurs in several murine mutants, such as staggerer and weaver, diminishes cerebellin levels, suggesting a link between cerebellin biosynthesis and synaptic activity (7, 13). However, the function of cerebellin is unknown, and few clues are provided by its structure other than a similarity to a sequence encoded within the polyimmunoglobulin transport protein (7).

Biochemical studies have identified several truncated peptides derived from the cerebellin sequence (6, 7, 13). These peptides, like cerebellin, are present in the cerebella of vertebrates from human to chicken, where their sequences are completely conserved (6, 7, 14; Y.U., J. L. Hempstead, and J.I.M., unpublished work), suggesting a role for these small molecules and, presumably, the precursor from which they are derived, in Purkinje cell physiology. Cerebellin-like immunoreactivity has been found in extracerebellar sites (15, 16), although the major sites are cerebellar Purkinje cells and cartwheel neurons of the dorsal cochlear nucleus (15). Because the latter cells are believed to be related to Purkinje cells (15, 17, 18), this finding would suggest that the function of cerebellin and its related peptides and putative precursor is of particular significance for the physiology of this neuronal lineage.

To further the study of these postsynaptic peptides we report the cloning of a series of cDNAs that encode cerebellin sequence and represent the postulated human cerebellin precursor, precerebellin.‡ Details of the structure and developmental expression of the precerebellin transcript are described.

MATERIALS AND METHODS

Cloning of Human Cerebellin Precursor cDNA. A human cerebellum cDNA library constructed in λZAP vector was purchased from Stratagene. The library was screened by using a single 48-mer oligonucleotide complementary to that predicted for cerebellin by using the most frequent codon usages in humans. Sequence of the synthetic oligonucleotide used for screening was 5′-GTGGTTGGTGCTTCTGATGGCGCTGAAGGCCACCTTGGCGCTGCCGCT-3′. The probe was radiolabeled with [γ-^{32}P]ATP (3000 Ci/mmol; 1 Ci = 37 GBq; Amersham) by oligonucleotide kinase (GIBCO–BRL) and used to screen the cDNA library by standard plaque-hybridization techniques (19). After screening ≈1 ×

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.

*Present address: Towa City Corporation-Nishinotohin, Room 706, 440 Konishi-cho, Takoyakushi-sagaru, Nishinotoin, Nakagyo-ku, Kyoto 604, Japan.
†To whom reprint requests should be addressed.
‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. M58583).

10^6 plaques, three independent positive clones, from 1 to 1.5 kilobases (kb) in size, were isolated.

By using one of these cDNAs (clone 2; see Fig. 1) as probe, a genomic clone was isolated from a mouse Charon 35 library (gift of R. Kinloch, Roche Institute). The insert was digested with *Pst* I and subcloned into pGEM-3. A subclone comprising ≈4 kilobase pairs (kbp) of the mouse precerebellin gene was identified by colony hybridization. This clone was used to confirm results obtained with the human cDNA probe.

DNA Sequencing. The inserts from the three positive cDNA clones were excised and ligated into the *Eco*RI cloning site of pBluescript (Stratagene). DNA sequences were determined by the dideoxynucleotide chain-termination procedure (20) employing a T7 DNA polymerase sequencing system (Promega), 2′-deoxy-7-deazaguanosine 5′-triphosphate (in place of dGTP), deoxycytidine 5′-[α-[^{35}S]thio]triphosphate (1000 Ci/mmol) (Amersham), and a series of synthetic oligonucleotide primers. Fig. 1 shows the sequencing strategy and the position of the primers relative to the three clones.

RNA (Northern) Transfer and Hybridization. Rats and mice were sacrificed by decapitation, and tissues were rapidly dissected out and immersed in liquid nitrogen. Human cerebellar tissue was obtained at surgery for removal of a malignant brain tumor. After excision from the patient, a small quantity of cerebellar tissue was dissected away from the tumor and immersed in liquid nitrogen. Total RNA was extracted using the procedure of Auffray and Rougeon (21). Total RNA was electrophoresed on 1% agarose gels and transferred to a nitrocellulose filter as described (22). Membranes were hybridized to the human precerebellin cDNA probe labeled with [α-^{32}P]dCTP (3000 Ci/mmol; Amersham) using a multiprime kit (Amersham), according to the recommendations of the manufacturers. In one experiment a subcloned 4-kbp fragment of the murine precerebellin gene was labeled in a similar manner and used to probe the same blots to ensure authenticity of the bands.

Sequence and Homology Analyses. DNA sequences were assembled using Genetics Computer Groups (GCG) programs

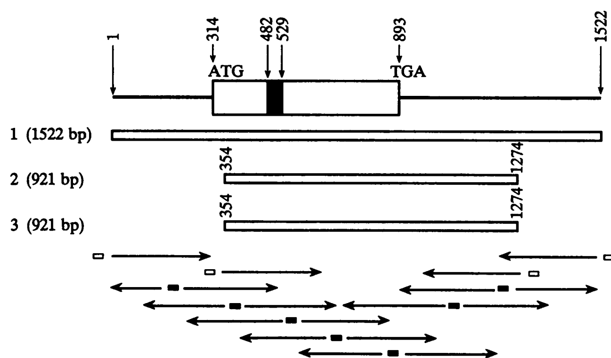

FIG. 1. Structure and sequencing strategy of precerebellin cDNAs. Three clones were isolated that hybridized to a single 48-mer oligonucleotide complementary to the cerebellin sequence. The longest clone, (clone 1) contained a 1522-base-pair (bp) insert that encodes cerebellin hexadecapeptide, indicated by the dark area. Upon sequencing, the two other clones (clones 2 and 3) were both found within the long clone, between residues 354 and 1274 relative to clone 1. Clone 1 has a single open reading frame between residues 314 and 893 that encodes a possible initiator methionine (residue 314), the cerebellin peptide sequence (residues 482–529), and a stop codon (residue 893). Nucleic acid sequence 5′ to the ATG contains stop codons in all three frames and no other candidate initiator sequences, suggesting that the ATG at residue 314 is the translation start site. Sequencing was done as described by using oligonucleotide primers indicated by boxes. Open boxes represent primers sited in the vector; closed boxes are oligonucleotides synthesized from deduced sequences. All three clones were sequenced on both strands, as indicated by arrow direction.

(23). Homology searches were made of the GenBank (release 65.0, European Molecular Biology Laboratory (release 24.0), National Biomedical Research Foundation (release 36.0), and Swiss-Prot (release 14.0) databases using the FASTA program. In addition, the databases were searched by using the program of Nishikawa *et al.* (24) employing the statistical treatment described by Toh *et al.* (25). The latter search and analysis was performed by H. Toh of the Protein Engineering Research Institute (Osaka).

RESULTS

Screening a human cerebellum cDNA library with a single 48-base oligonucleotide antisense to the predicted cerebellin

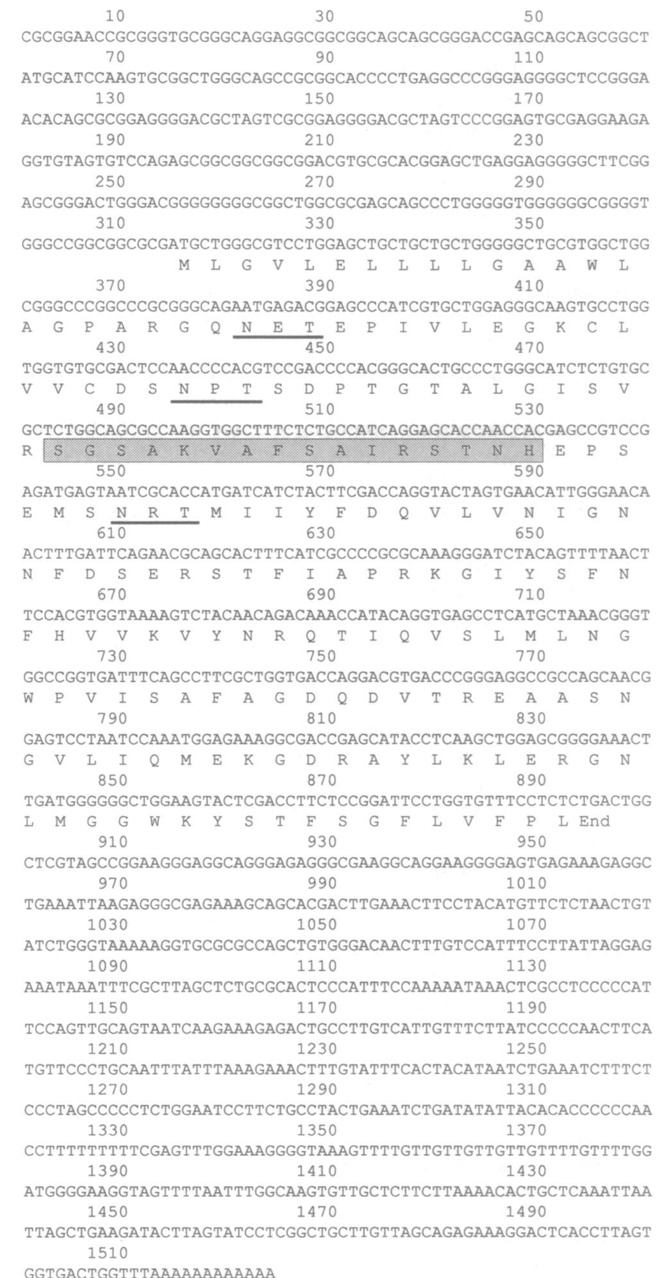

FIG. 2. Sequence of human precerebellin. Sequence was derived as described in Fig. 1 legend, and overlapping sequences were assembled by using the University of Wisconsin Genetics Computer Group programs (23). Complete nucleic acid sequence of clone 1 and amino acid (single-letter code) sequence of the longest open reading frame are shown. Shaded amino acids indicate the cerebellin peptide sequence, whereas underlining indicates possible glycosylation sites.

sequence by using the most frequent codon usage in human yielded three positive clones. All three cDNAs were related (Fig. 1) and upon sequencing were found to contain a nucleotide sequence encoding the 16 amino acids of cerebellin (Figs. 1 and 2). The longest clone (1.5 kb) has a potential initiator methionine and a stop codon in the same reading frame as the cerebellin sequence and encodes a protein of 193 amino acids (Fig. 2). Upstream of the putative initiator methionine, there are stop codons in all reading frames before another ATG is encountered. Neither the nucleotide sequence of the clone nor its predicted amino acid sequence is present in any of the databases searched. Therefore, we identify this cDNA as encoding the cerebellin precursor precerebellin. The homology of the cerebellin hexadecapeptide sequence to the polyimmunoglobulin transporter (7) does not extend further in the precursor, and thus the relationship was fortuitous. However, the extended cerebellin sequence does show a similarity to the complement component C1qB of both human (significance, 9.7 SD) and mouse (significance, 8.8 SD) (Fig. 3). Over a 145-amino acid stretch, precerebellin is 31.3% identical with human C1qB and is 52.2% similar (Fig. 3, calculated according to refs. 24 and 25). Precerebellin is less related to human C1qA, where identity is 20.7%, and similarity is 44.8% (significance, 2.4 SD). Chicken collagen X is intermediate between C1qA and C1qB with 28.6% identity and 50.6% similarity (significance, 5.2 SD).

The predicted precerebellin sequence bears few canonical structural motifs. Three potential N-linked glycosylation sites are present (Fig. 2). A markedly hydrophobic domain is present at the amino terminus, whereas a less hydrophobic region is found at the carboxyl terminus. The former does not conform to the normal signal peptide motif; however, it cannot be excluded formally that this or other hydrophobic sequences could be involved in membrane insertion or association.

Northern blot analysis with total RNAs from human, mouse, and rat cerebellum shows the same transcripts (Fig. 4*B*). These multiple transcripts have approximate sizes of 1.8, 2.3, 2.8, and 3.0 kb in the rat cerebellum and are detected by both the human precerebellin cDNA and mouse precerebellin genomic fragment (Fig. 4*A*). No signal is evident in extraneural tissues, although a weak signal appears in some other areas of brain—notably thalamus (Fig. 5). These data support our earlier biochemical analyses of cerebellin levels in various tissues and brain regions (6, 15).

During development in the rat, precerebellin transcripts are already evident at low levels on embryonic day 20 and postpartum day 1 (Fig. 6). Levels of precerebellin mRNA rise slightly between day 1 and 5 after birth and then much more dramatically by day 15 and 21 postpartum (Fig. 6). Only trace levels of precerebellin mRNA are detected in cerebral cortex, predominantly at early times, during the equivalent time period (Figs. 5 and 6). This pattern of temporal expression of precerebellin mRNA is in excellent agreement with cerebellin peptide levels during the same period of rat development (8).

DISCUSSION

The peptide cerebellin is derived from a previously undescribed precursor (Fig. 2) by proteolytic processing events that do not involve cleavage at dibasic amino acids. Precerebellin transcripts show the same temporal (Fig. 5) and spatial (Fig. 6) distribution as the cerebellin peptides (6, 7, 15). Thus, transcripts are only abundant in the cerebellum; they are present at low levels at the time of birth; they only accumulate to maximum values by the third week postpartum. Northern blot analyses demonstrate four precerebellin mRNA species, one of which is predominant (Fig. 4*B*). However, no particular variation occurs in the relative distribution of these transcripts, either among different brain regions and tissues or during development. Thus, whatever the significance of these various mRNAs, they all appear to be expressed in the same brain regions—almost exclusively cerebellum—and individual transcripts are not specifically linked to any maturation phase. We do not know yet whether these multiple transcripts are derived from the same gene by alternative splicing or whether they are products of independent genes.

Electron microscopic studies have indicated that a significant proportion of the cerebellin-like immunoreactivity in Purkinje cell spines is associated with postsynaptic densities and other membrane-like structures within the postsynaptic

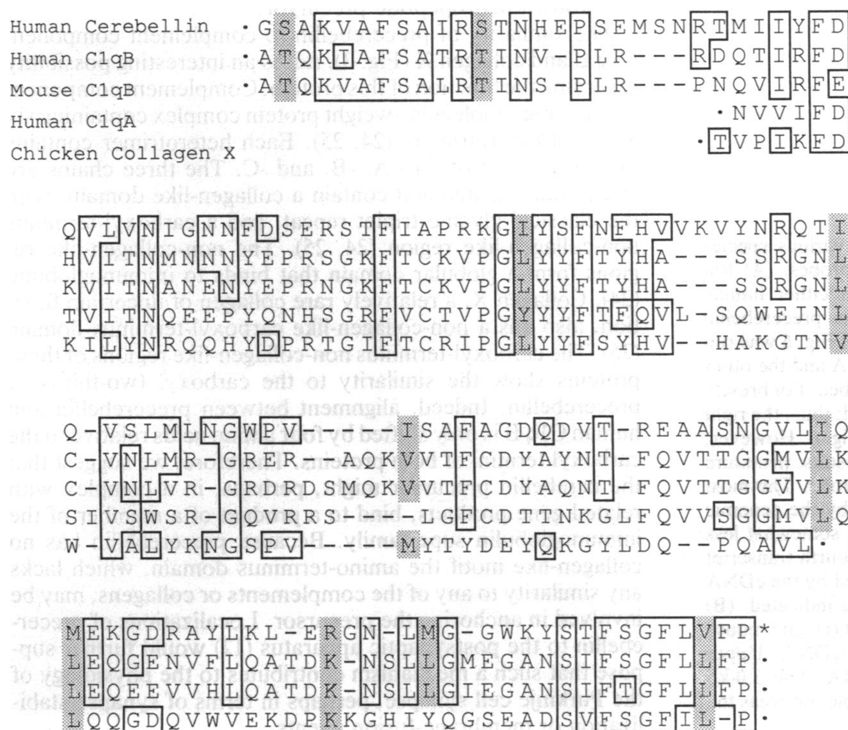

FIG. 3. Sequence comparison of precerebellin to A and B chains of complement C1q and to collagen X. Alignment of amino acid (single-letter code) sequences of human and murine C1q chains, chicken collagen X, and human precerebellin cDNA (alignment was generated by using the program described in ref. 24; further details are given in text). Boxed sequences, identity of amino acids; shaded residues, conservative substitutions; dashes, breaks in sequence introduced by program; dot, sequence continues in protein but lacks sufficient similarity for program comparison (calculated according to ref. 25); asterisk, end of a coding sequence. Sequences shown are resident in the databases and were originally published in refs. 26–28.

apparatus (12). This would suggest that precerebellin might be bound to, or associated with, a membrane. Precerebellin does have a hydrophobic amino terminus that could be an atypical signal sequence. In addition, the presence of potential N-linked glycosylation sites may also indicate that precerebellin is a membrane protein. As cerebellin is selectively localized to dendrites (8, 12), precerebellin could contain motifs that direct it to dendrites. In this context, it might be speculated that the amino-terminus hydrophobic region constituted such a motif.

One study has reported cerebellin-like immunoreactivity in extraneural tissues (16). However, this result disagrees with previous biochemical (6) and immunological (7, 12, 15) studies, as well as the Northern blot analysis presented here (Fig. 5). Preliminary evidence suggests that precerebellin is a member of a small gene family (data not shown). Thus, the cerebellin-like immunoreactivity reported in extraneural tissues (16) could be a cross-reaction with the product of a related gene. Cerebellin has also been suggested to be a neuromodulator (16). However, in addition to its postsynaptic localization (12), several aspects of precerebellin structure

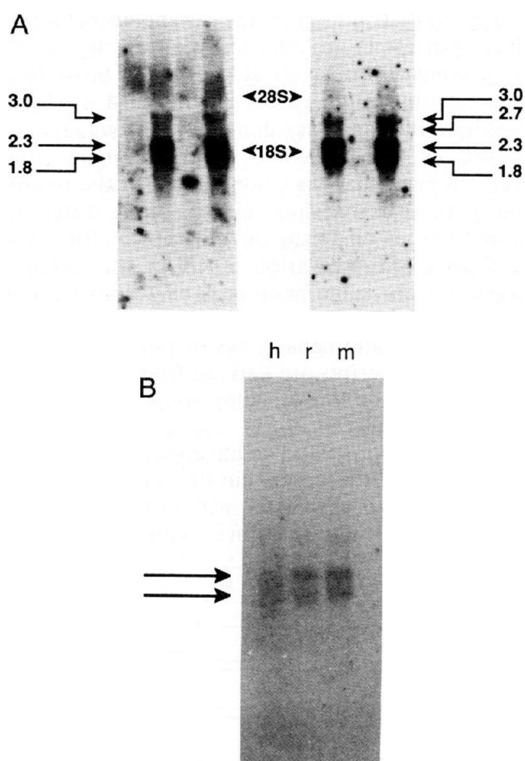

FIG. 4. Precerebellin transcripts in cerebella of various species analyzed with human cDNA and mouse genomic probes. (*A*) Rat cerebellar total RNA (5 μg per lane) probed with either human precerebellin cDNA 2 (three right lanes) or mouse precerebellin genomic fragment (four left lanes). The experiment of Fig. 6 was run in duplicate—one filter was probed with human cDNA and the other with mouse precerebellin genomic fragment, as described. For brevity and clarity, only data for the oldest rats were depicted; thus, the right three lanes are identical with the last three lanes of Fig. 6. However, at all ages genomic and cDNA probes gave the same results. In mature rats four transcripts are apparent with the cDNA probe. The most abundant transcript is $\approx$2.3 kb and is also detected by the genomic probe. Two other transcripts of $\approx$3.0 and 1.8 kb are somewhat less abundant and are also detected by both reagents. A fourth transcript of 2.7 kb is of low abundance and only clearly revealed by the cDNA probe. Positions of 18S and 28S ribosomal RNA are indicated. (B) Total mRNA (5 μg per lane) from adult human (h), rat (r), and mouse (m) cerebella probed with the human precerebellin cDNA. Upper arrow, 2.3-kb transcript; lower arrow, 1.8-kb mRNA. The 3.0-kb message is just visible, particularly in the murine sample, whereas the minor 2.7-kb mRNA is not detected at this exposure.

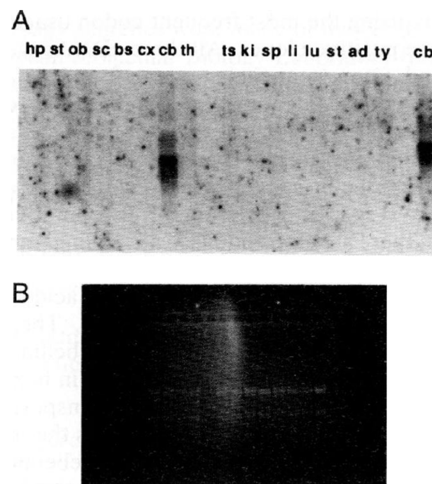

FIG. 5. Tissue distribution of precerebellin mRNA (A). Distribution of precerebellin transcripts in various rat tissues and brain regions. Total RNA (5 μg per lane) from indicated sources was analyzed by Northern transfer and hybridization against the human precerebellin cDNA clone. At left are RNAs from various brain regions, whereas at right are RNAs from several extraneural tissues and organs. Lanes contain RNA from the following sites: hippocampus (hp), striatum (st), olfactory bulbs (ob), brain stem (bs), cerebral cortex (cx), cerebellum (cb), thalamus (th), testis (ts), kidney (ki), spleen (sp), liver (li), lung (lu), stomach (st), adrenal gland (ad), thymus (ty), and cerebellum (cb) (added as positive control). (*B*) Ethidium bromide-stained gel before transfer to nitrocellulose to show that all lanes were loaded with approximately equivalent amounts of total RNA (as judged by intensity of ribosomal RNA bands). An additional side lane was loaded with an independent rat cerebellar RNA sample to show reproducibility and to indicate that no uneven transfer was encountered at the gel edge. Blot has been overexposed intentionally to emphasize the lack of hybridization in extracerebellar tissues. Spotting in the blot is a function of the probe and seems to be a product of its G+C content.

argue against this contention. (*i*) Cerebellin is not bounded by the classical dibasic processing motif present in many (although not all) neuropeptide precursors. (*ii*) Precerebellin appears related to complement and collagens rather than to any known neuropeptide precursor.

The similarity of precerebellin to complement component C1q B and collagen X (Fig. 3), raises an interesting possibility regarding the function of this protein. Complement component C1q is a high-molecular-weight protein complex containing six identical heterotrimers (24, 25). Each heterotrimer contains one chain each of C1q A, -B, and -C. The three chains are structurally related and contain a collagen-like domain, with the classical glycine triplet repeat, and a carboxyl-terminus non-collagen-like region (24, 25). The non-collagen-like regions form a globular domain that binds to immunoglobulin (24). Collagen X, a relatively rare collagen of uncertain function, also has a non-collagen-like carboxyl-terminus domain (25). The carboxyl-terminus non-collagen-like regions of these proteins show the similarity to the carboxyl two-thirds of precerebellin. Indeed, alignment between precerebellin and human C1q B is only shifted by four amino acids relative to the carboxyl termini of both proteins. Therefore, we suggest that the cerebellin precursor might, perhaps, in a complex with related gene products, bind to a product of a member of the immunoglobulin superfamily. Because precerebellin has no collagen-like motif the amino-terminus domain, which lacks any similarity to any of the complements or collagens, may be involved in anchoring the precursor. Localizations of precerebellin to the postsynaptic apparatus (12) would further suppose that such a mechanism contributes to the physiology of the Purkinje cell synapse, perhaps in terms of synapse stabilization or membrane-fusion events.

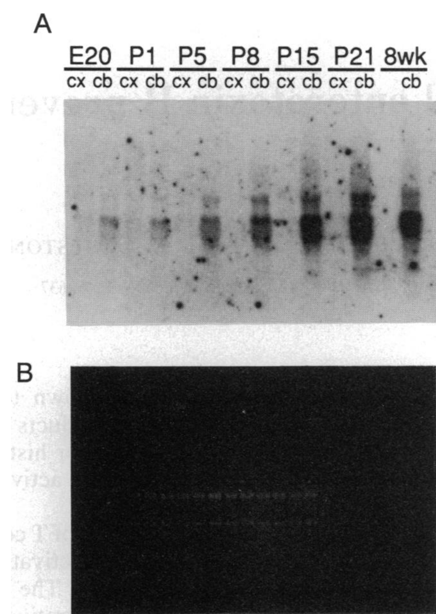

FIG. 6. Developmental expression of precerebellin. Total RNA from cerebellum (cb) or cerebral cortex (cx) was isolated from rats at indicated embryonic (E) or postpartum (P) ages and analyzed for precerebellin transcripts by hybridization to human precerebellin cDNA probe. (*A*) Hybridization of total RNA at 5 μg per lane to cDNA probe. (*B*) Ethidium bromide-stained gel before transfer to indicate equal loading of lanes. An identical blot was probed with murine precerebellin genomic fragment with an essentially identical result (some of these data are in Fig. 4*A*). Blot was intentionally overexposed (as in Fig. 5) to demonstrate the low levels of precerebellin transcripts in the neonate, the absence of transcripts in cortex, and the weaker hybridizing species in cerebellum.

It is now possible to analyze specific gene regulation in cerebellar Purkinje cells by using transgenic approaches (5). Availability of the precerebellin cDNA, as well as a mouse genomic clone, will now permit isolation and expression of the promoter for this gene in mice and comparison of this promoter with other promoters that direct expression in Purkinje cells (5).

1. Ito, M. (1984) *The Cerebellum and Neural Control* (Raven, New York).
2. Miale, I. L. & Sidman, R. L. (1961) *Exp. Neurol.* **4,** 277–296.
3. Altman, J. (1972) *J. Comp. Neurol.* **145,** 465–514.
4. Sidman, R. L. (1983) in *Genetics of Neurological and Psychiatric Disorders,* eds. Kety S. S., Rowland, L. P., Sidman, R. L. & Matthysse, S. W. (Raven, New York), pp. 19–46.
5. Oberdick, J., Smeyne, R. J., Mann, J. R., Zackson, S. & Morgan, J. I. (1990) *Science* **248,** 223–226.
6. Slemmon, J. R., Blacher, R., Danho, W., Hempstead, J. L. & Morgan, J. I. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 6866–6870.
7. Morgan, J. I., Slemmon, J. R., Danho, W., Hempstead, J., Berrebi, A. S. & Mugnaini, E. (1988) *Synapse* **2,** 117–124.
8. Slemmon, J. R., Danho, W., Hempstead, J. L. & Morgan, J. I. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 7145–7148.
9. Ziai, M. R., Pan, Y.-C. E., Hulmes, J. D., Sangameswaran, L. & Morgan, J. I. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8420–8423.
10. Ziai, M. R., Sangameswaran, L., Hempstead, J. L., Danho, W. & Morgan, J. I. (1988) *J. Neurochem.* **51,** 1771–1776.
11. Sangameswaran, L., Hempstead, J. L. & Morgan, J. I. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5651–5655.
12. Mugaini, E., Dahl, A.-L. & Morgan, J. I. (1988) *Synapse* **2,** 125–138.
13. Slemmon, J. R., Goldowitz, D., Blacher, R. & Morgan, J. I. (1988) *J. Neurosci.* **8,** 4603–4611.
14. Yiangou, Y., Burnet, P., Nikou, G., Chrysanthou, B. J. & Bloom, S. R. (1989) *J. Neurochem.* **53,** 886–889.
15. Mugnaini, E. & Morgan, J. I. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 8692–8696.
16. Burnet, P. W. J., Bretherton-Watt, D., Ghatei, M. A. & Bloom, S. R. (1988) *Neuroscience* **25,** 605–612.
17. Osen, K. K. & Mugnaini, E. (1981) in *Neuronal Mechanisms of Hearing,* eds. Syka, J. & Aitken, L. (Plenum, New York), pp. 119–125.
18. Eccles, J. C., Ito, M. & Szentagothai, J. (1967) *The Cerebellum as a Neuronal Machine* (Springer, Berlin).
19. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
20. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
21. Auffray, C. & Rougeon, F. (1980) *Eur. J. Biochem.* **107,** 303–307.
22. Morgan, J. I., Cohen, D. R., Hempstead, J. L. & Curran, T. (1987) *Science* **237,** 192–197.
23. Staden, R. (1979) *Nucleic Acids Res.* **6,** 2601–2610.
24. Nishikawa, K., Nakashima, H., Kanohisa, M. & Ooi, T. (1987) *Protein Sequences Data Anal.* **1,** 107–116.
25. Toh, H., Hayashida, H. & Miyata, T. (1983) *Nature (London)* **305,** 827–829.
26. Reid, K. B. M. (1985) *Biochem. J.* **231,** 729–735.
27. Reid, K. B. M., Gagnon, J. & Frampton, J. (1983) *Biochem. J.* **203,** 559–569.
28. Ninomiya, Y., Gordon, M., van der Rest, M., Schmid, T., Linsenmayer, T. & Olsen, B. R. (1986) *J. Biol. Chem.* **261,** 5041–5050.