Phosphacan, a chondroitin sulfate proteoglycan of brain that interacts with neurons and neural cell-adhesion molecules. is an extracellular variant of a receptor-type protein tyrosine phosphatase

PATRICE MAUREL*, UWE RAUCH*, MANUELA FLAD[†], RENÉE K. MARGOLIS[†], AND RICHARD U. MARGOLIS^{*}

*Department of Pharmacology, New York University Medical Center, New York, NY 10016; and [†]Department of Pharmacology, State University of New York, Health Science Center, Brooklyn, NY 11203

Communicated by Philip Siekevitz, December 16, 1993

We have identified cDNA clones encoding a **ABSTRACT** chondroitin sulfate proteoglycan of rat brain (previously designated 3F8 and now named phosphacan) that binds to neurons and neural cell-adhesion molecules. A sequence of 1616 amino acids deduced from a 4.8-kb open reading frame contains the N-terminal amino acid sequence of the 3F8 core glycoprotein as well as four internal CNBr, tryptic, and endoproteinase Lys-C peptide sequences from the proteoglycan. The deduced amino acid sequence, beginning with a 24-amino acid signal peptide, reveals an N-terminal domain of 255 amino acids homologous to carbonic anhydrases. The entire amino acid sequence deduced from our cDNA clones corresponds to the extracellular portion of a human receptor-type protein tyrosine phosphatase $(RPTP\zeta/\beta)$ with which it has 76% identity, and the proteoglycan may represent an mRNA splicing variant of the larger transmembrane protein. RNA analysis demonstrated that a probe to the N-terminal carbonic anhydrase domain of the proteoglycan hybridizes with rat brain mRNA of 9.5, 8.4, and 6.4 kb, whereas probes to the phosphatase domains hybridize with only the 9.5-kb message and with the 6.4-kb message (which corresponds to a previously identified variant of the transmembrane protein in which half of the extracellular domain is deleted). The 30 N-terminal amino acids of the 3H1 chondroitin/keratan sulfate proteoglycan of brain are identical to those of the 3F8 proteoglycan, and six internal tryptic peptide sequences also matched those found in sequenced peptides of the 3F8 proteoglycan and/or amino acid sequences deduced from the cDNA clones. We therefore conclude that the 3H1 chondroitin/keratan sulfate proteoglycan and the 3F8 chondroitin sulfate proteoglycan represent glycosylation and possible extracellular splicing variants of a receptor-type protein tyrosine phosphatase. These proteoglycans may modulate cell interactions and other developmental processes in nervous tissue through heterophilic binding to cell-surface and extracellular matrix molecules, and by competition with ligands of the transmembrane phosphatase.

We have previously described a chondroitin sulfate proteoglycan isolated from a phosphate-buffered saline extract of rat brain by immunoaffinity chromatography with the 3F8 monoclonal antibody and which is developmentally regulated with respect to its sulfation, carbohydrate composition and oligosaccharide structure, and immunocytochemical localization (1). A chondroitin/keratan sulfate proteoglycan (designated 3H1) was also isolated from rat brain by using monoclonal antibodies to the keratan sulfate chains (1). 3F8 and neurocan, another chondroitin sulfate proteoglycan of brain, for which the primary structure has been described (2), interact with neurons and the neural cell-adhesion molecules (CAMs), Ng-CAM and N-CAM (3). From radioligandbinding studies it was found that the brain proteoglycans bind with high affinity ($K_d \approx 0.5$ nM) to Ng-CAM and N-CAM but not to other cell-surface and extracellular matrix proteins such as laminin, fibronectin, several collagens, epidermal growth factor and fibroblast growth factor receptors, or the myelin-associated glycoprotein (4). The 3F8 proteoglycan and neurocan inhibited neurite outgrowth and binding of neurons to Ng-CAM when mixtures of these proteins were adsorbed to polystyrene dishes, and direct binding of neurons to the proteoglycan core glycoproteins was demonstrated with an assay in which cell-substrate contact was initiated by centrifugation (3, 4). Recent studies have also shown that embryonic chicken brain neurons bind the 3H1 proteoglycan and contain cell-surface keratan sulfate chains (M. Flad, R.U.M., and R.K.M., unpublished results). These results indicate that brain proteoglycans can bind to neurons and that Ng-CAM and N-CAM may be heterophilic ligands for neurocan and the 3F8 and 3H1 proteoglycans.

The primary structure of the 3F8 proteoglycan has now been determined by cDNA cloning, and we have identified this major chondroitin sulfate proteoglycan of brain as a possible mRNA splice variant of a receptor-type protein tyrosine phosphatase (PTP) named PTP ζ (5) and RPTP β (6) $(RPTP\zeta/\beta)$. The tyrosine phosphatases act in concert with tyrosine kinases to regulate the phosphorylation state of proteins, and their structural features and potential physiological roles in signal transduction and cell-cycle regulation have recently been reviewed (7-9). With probes based on conserved sequences in their phosphatase domains, over 30 PTPs have now been cloned, but a major question concerning this family of key regulatory enzymes is the identification of their ligands and substrates. Our findings indicate that neural CAMs and the extracellular matrix protein tenascin may serve as ligands for RPTP ζ/β . Amino acid sequencing of the 3H1 chondroitin/keratan sulfate proteoglycan of brain demonstrated that it is a glycosylation variant of the 3F8 proteoglycan. We have named the 3F8/3H1 proteoglycan phosphacan[‡] and suggest that it may modulate cell interactions and other developmental processes in nervous tissue.

MATERIALS AND METHODS

Preparation of Peptides and Amino Acid Sequence Analysis. Proteoglycans were isolated from a phosphate-buffered saline extract of rat brain by immunoaffinity chromatography using the 3F8 or 3H1 monoclonal antibodies (1). Chondroitinase and protease digestions, CNBr treatment, electrophoresis, and transfer to ProBlott membranes for N-terminal amino

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.

Abbreviations: PTP, protein tyrosine phosphatase; CAM, cell-adhesion molecule; RPTP ζ/β , receptor-type PTP.

[‡]The sequence reported in this paper has been deposited in the GenBank data base (accession no. U04998).

acid microsequencing were done as described (2). N-terminal amino acid sequences of proteins in solution were obtained after drying on Polybrene-coated glass fiber filters. Internal amino acid sequences were also obtained after transfer of the 3F8 proteoglycan core glycoprotein and a CNBr peptide to nitrocellulose followed by in situ digestion with trypsin or endoproteinase Lys-C (10) and by trypsin digestion of the intact 3H1 proteoglycan, in both cases followed by reversephase HPLC fractionation of the resulting peptides before microsequencing of individual peaks.

Generation of Probes by the PCR and Isolation of cDNA Clones. Degenerate oligonucleotide primers based on the N-terminal amino acid sequence of the 3F8 proteoglycan and the sequence of an internal CNBr peptide (Fig. 1) were synthesized with BamHI and HindIII linkers. An oligo(dT)primed 6-week rat brain AZAPII cDNA library (Stratagene) was used as template for amplification with Taq polymerase (Perkin-Elmer/Cetus), using 50 temperature-step cycles of 94°C (0.5 min), 55°C (1.5 min), and 72 °C (5 min). Agaroseethidium bromide gel electrophoresis of the PCR reaction products demonstrated a 780-bp band that was not produced from a control reaction containing only single primers or empty AZAPII vector. The PCR product was subcloned into pGEM3, and antisense RNA transcripts were prepared for screening of the same cDNA library, as described (2).

A probe to the extracellular domain of $RPTP\zeta/\beta$ (5, 6) near the transmembrane region was synthesized by reverse transcription/PCR (2) using 4-day rat brain mRNA and primers designed on the basis of the human sequence (sense, ⁵'- CCCTCGAGGAAGGCATGTATTTGC-3'; antisense, ⁵'- GCCTCTAGATTCACCACCACCCACTAAACC-3'). The 346-bp PCR product was subcloned into the Xba ^I and Sma I sites of pGEM7Zf, linearized with BamHI, and a ³²P-labeled RNA probe was transcribed and used to screen an oligo(dT)/ random-primed AZAPII rat brain stem/spinal cord cDNA library (Stratagene). Positive clones were selected for further study only if they gave no PCR product using primers complementary to a sequence in the first phosphatase domain of the transmembrane protein (see below). These clones were converted to Bluescript plasmids by in vivo excision and used for sequencing.

DNA Sequencing. Subclones for sequencing (using Taq polymerase in conjunction with dye-labeled terminators and the Applied Biosystems model 373A DNA sequencing system) were generated by deletions produced with restriction enzymes, by subcloning of restriction fragments, and by

progressive unidirectional exonuclease III deletions. Sequencing was also done by using synthetic primers corresponding to the ends of previously determined sequences. Both strands of the DNA were sequenced, with sequence alignment and analysis accomplished with the software package from the Genetics Computer Group (Madison, WI). The reading frame was verified by our N-terminal amino acid sequence data for the core glycoprotein and for peptide fragments derived from it.

RNA Blots. As a probe for the N-terminal portion of the 3F8 proteoglycan, the nucleotide sequence from the original 780-bp PCR product used for cDNA library screening was transcribed into digoxigenin-labeled antisense RNA and used for hybridization with Northern blots of mRNA prepared from brain and other tissues (2). Probes for the cytoplasmic domain of the phosphatase were synthesized by PCR using rat brain cDNA as template, and sense and antisense primers (with Xho ^I and HindIII linkers) based on amino acid and nucleotide sequences of the first and second phosphatase domains of human PTP ζ/β (5, 6). The primers for the first phosphatase domain were as follows: 5'-GCGAATTCTC-GAGAARTGYGAYCARTAYTGGCCIGCIGAYG-3' (sense), and 5'-GCTCTAGAAGCTTGTAGTCTGTGCCT-TCCCCACTCAGG-3' (antisense); and for the second phosphatase domain: 5'-GCAATTCTCGAGCCTGTGGAAA-GATCAAGGGTTGGC-3' (sense) and 5'-GCAATTAAGCT-TATGCTCATCATGAACAATCATAGG-3' (antisense). The resulting PCR products were subcloned into pGEM, their identity was confirmed by sequencing, and they were then transcribed into digoxigenin-labeled antisense RNA.

RESULTS

cDNA Clones Corresponding to the N-Terminal Portion of the 3F8 Proteoglycan. Identical N-terminal amino acid sequences were obtained from both the native proteoglycan and the core glycoprotein (resulting from chondroitinase treatment) transferred to a ProBlott membrane after SDS/PAGE (Fig. 1). This same N-terminal amino acid sequence was also obtained from two CNBr peptides (22 and 40 kDa) derived from the core glycoprotein and transferred to ProBlott (data not shown). The N-terminal amino acid sequence of an internal 14-kDa CNBr peptide was obtained in a similar manner, and other internal peptide sequences were obtained from in situ tryptic or endoproteinase Lys-C digestion of CNBr peptides or of the entire core glycoprotein transferred

N-terminus of 3F8 proteoglycan: YYRQQRKLVEEIGWSYTGALNQKNWGKK

N-terminus of 3H1 proteoglycan: YYRQQRKLVEEIGW - YTGALNQKN - GK - YP

3F8 CNBr peptide: DYLQNNFREQQYKFSRQVFSSYTGKEEIHEA (279-309)

3F8 endoprotelnase Lys-C peptide: FAVLYQPLEGNDQTK (346-360)

Tryptic peptides from the 3F8 proteoglycan: FQGWEKPSLENTFIHNTG (82-99) FDADRF--FEE-VK-K--L (158-176)

Tryptic peptides from the 3H1 proteoglycan:

FQGWEK (82-87) TVEINLTNDY (101-110) AIIDGTE (195-201) QVFSSYTGKEEIHEA (295-309) WYDTMI (337-343) YSDQLIVDMPTEDAELDLFPELIGT (400-424)

FIG. 1. N-terminal amino acid sequences of the 3F8 and 3H1 proteoglycan core proteins, and of 10 internal peptides obtained by trypsin, endoproteinase Lys-C, or CNBr treatment. Underlined sections of the peptide sequences indicate regions used for the design of oligonucleotide primers, and the position of the internal peptides in the sequence shown in Fig. 2 is given in parentheses.

.-CTAACCCGGCCGTCACGGCTCCCCGTCTGGTGTCCCCTCTTGCACCTAACTGTCCCC."GCGCTGGGAGGGCCGGGACCGGCGGAGAGCGATCCTCAGAC ¹²³ CCTAGACCCTGGCCAGTCACCGGCGTCCCCTGTCTCGGTGTTCCCACTCTCTGCACCCTAAGCTGTACCCCTGCGGCTGGGAGGGCCCGGGACCCGGGATCGGAATCCTGAACAG
TTCCTCGCGTGCGTTCAGCTACTGTGCGTGTGTCGCCTGGACTGGGCTTATGGATACTACAGACAACAGAAAAACTTGTTGAAGAGATTGGCTGGTCCTATACAGG AATTGGGAAGAAAATCCATATTAATGCCCAAGCGTCTCTATAATATGATAAGACTTAACAATAAA=GA~C~r AGAA~rGAATTTAGGGTOWAAAAC ³⁶⁹⁶ NEW OKEKY P IC ^N ^S ^P ^K ^Q S PI N ^I ^D ^E ^D L TQ V NV ^N ^L ^K ^K ^L K FQ OWU2 ^K ^P ⁸⁸ TCCTTGGAAAACACGTTTATTCACAACACTGGGAAAACAGTGGAAATAAA<u>TCTCACCAATGACTACTATGG</u>AGTGGAGGACTTTCAGAAAAGGTCTTCAAGGCAAACGTATTTCAC 492
S L B M T F I H M T G K T V E I N - L T N D Y Y L S G L S B K V F K A S K M T F H 129 TGGGGGAAATGCAATGTGTCATCGGAAGGATCGGAACACATGTTAGAAGGGCAGAAGTTTCCACTCGAGATGCAAATCTACTGCTTCGATGCGAACCGATTTTCAAGTTTTGAGGAAACAGTT 615
W G K C N V S S E G S E G S T H S L E G Q K F P L E M 2 I T C S A D A D A D A T S T S A D A D K 0 K 0 R L R A L S ^I L F ^E ^I G V E E N L D Y K A ^I ^I D G T E S V S R ^F G K Q A A 211 TTAGATCCGTTCATCTTGCAGAACCTCCTGCCAAACTCCACTGACA.AGTACTACATTTACAACGGATCACTGACATCCCCTCCCTGCACAGACACCGTCGAATGGATTGTTTTTAAGGATACA 861 L D ^P ^F ^I L Q N L L P N S T D K Y Y ^I Y N G S L T S P ^P C T D T V E W ^I V F K D T 252 GTTAGCATCTCTGAAAGCCAGCTGGCTGTATTTTGTGAAGTTCTCACAATGCAACAGTCCGGGTATGTCATGTTGATGGATTACTAA<u>AACAATTTCCGAGAACAACAGTACAA</u>GTTTTCC 984
A G I S E S Q L A V F C E V L T M Q Q S G Y V M L M D **T L <mark>Q M M F R B Q Q T</mark> K B P A G A CAGC** CCTCGGTTTTTACACAAATGTTGAAAGTTGCGTTCGTACAWCATGGAGGAACGACAAACAAGCTGAGTTTAACAATGGTATCGGACTGGGGCCATC ¹²³⁰³ P R V V Y D T M ^I E K F A V L V Q P L ³ 0 N D Q 'F K H E ^F L T D G Y Q D L G A ^I L 375 A TAXISTAATAGTAATAGTAATAGTAATAGTAATAGTAATAGTAATAGTAATAGTAAGGECAATAGTAAGGINGAATAGTAATAGTAAGTAATAGTAAGTAAGTAAGTA
AATAACTTAATAGTAAGTAATAGTAATAGTAATAGTAATAGTAATAGTAAGTAAGTAAGTAAGTAAGTAAGTAATAGTAATAGTAATAGTAAGTAAGTAAGTAAGTAAG
I N N L I P N M S Y V L Q I V A I C S N G L Y G K Y S D Q L I V D M P T E D A E L D 416
CTCTTTCCTGAATTAATTGGAACTGAAGAAATAATCAAGGAGGAAAACTATGGAAAAGGCAATGAAGAAGACACTGGCTTGAATCCCGGTAGAGACAGTGCCACAAACCAAATAAGGAAAAAG 1476 L F P E L I G T E E I I K E E N Y G K G N 'E E D T G L N P G R D S <mark>A T N Q I R K K 457</mark>
GAACCCCAGGTTTCTACCACAACTCACTATAATCACATGGGGACTAAATATAATGAAGCCAAGACTAACCGATCTC<mark>CAACGAGAGGATCTGAATTCTCGGAAAGAGTGATGTTCTCAACAC</mark>A E P Q V S T T T H Y N H M G T K Y N E A K T N R S P T R G S E F S G K S D V L N T 498
TCCCTGAATCCTACTTCCCAACAGGTTGCTGAATTCAATACCAGAAAGAGAAATGTCCTTGCCTTCTCAGATTGGAACTAACCTGCCACACAGTGTGGAAGGCACCTCATCAAC 1722 S L N ^P T ^S Q Q V A E ^F N ^P E R E M S ^L ^P ⁵ Q ^I G T N L ^P ^P H S V E G T S A S L N 539 AGTGCTCTAAACCTCCTGTCTCCCCAGAGAACTGTCGGGATGCAAATCTTAATATGTTTCATACAGATACAAGAGTGTCGCTGCCTCGTGAGAAGAA ¹⁸⁴⁵⁴ S G S K T L L V F P Q M N L S C T A ^E ^S L N M V S ^I T ^E Y K E V S A D L S E E E N 580 TTACTGACTGATTTCAAGCTCGATAGTGGAGCAGATGATTCGTCAGGCTCTAGCCCTGCATCCTCCCCTGTCCCCTTTTCCACAGATAATCTATCATATACGTCTTCTTCAGACACG 1968
L L T D F K L D S G A D D S S G S S P A S S T V P F S T D N L S H G Y T S S S D T 621 CCCGGGCGTCACTATGTGTCTTAGCCAGATCTCGAG~kATCTCTGAGGTTCGCTCCTCAGTTCAAAGATCACAAAGATCCTCTCTGAAGGAGCTGTGT ²⁰⁹¹⁹ P E A V T Y D V L R P E S T R N A L E D S A P S G S E E S L K D P S L B G S V W F 662
CCTGGATCCACAGACCTAACAACAGTCTGAGACTGGATCTGGGAGAGGGGGCTTTCTCCAAGTTAACTCCACGGACTTCCAAGTTGATGAATCGAGACAACTGAGACATTTTCTCCA 2214 P G S T D L T T Q S E T G S G R ^E G F L Q V N S T D F Q V D E S R ^E T T E T F S P 703 GATGCACCGGTCCGGGGCCTTGGTCCAGAATGGAATGCACATATTCACCTTGCCACCCCCGATGAATAACACACAGCTTCACTCGTCTCCAACCATTG ²³³⁷³ 0 A T A S R C P ⁵ V T 0 M ^E H P H Y S T F A Y ^P P T E V T S H A F T P ⁵ 5 R P L D 744 TTGGTCCCCTAGAACACCTCATTCCAGCAACCAACAGTAACAAGGTGGACACTCTCAACTTCTACATAGTAAGTTTTCTCTGTCACCC''r'iTTGCTG ²⁴⁶⁰⁶ L A ^P T ^S N ^I L H S Q T T Q P V Y N G E T ^P L Q ^P 5 Y S S E V F P L V T P L L L D 785 AATCAACCCC~kAACTACCCTCTWTCAAGAGTGTTCGGCTTGATGCACGCTGTATCCCAGTGTGGTTGTCATTGATCCACCTGCTTCTATGTGATCAC ²⁵⁸³⁸ N Q T L N T T ^P A A S S S D S A L H A T ^P V ^F ^P 5 V G V S F D S ^I L S S Y D D A P 826 AN COLLINITITIC CONTROL AND STRIP OF A ASSESSED SALL HAN THE VEH SALL CONTRACT CONSTRUCT DISTURBED SALL HAND T
CTGCTCCCATTTTCCTCTTCCTTCCTTCCTTCAGTAGTCACTTGTTCACCATCCGCATACGGTTTCTCAAACCCTTCCGCAAGTTACTTCAGCCTGCTGCATAGCCTCTTT L L P F S S A S F S S D L F H H L H T V S Q T L P Q V T S A A E R D E L S L H A 5 867 CTGCGGTGCTGGGGTATTTCTGTAGACCCGCCTGTTCGTATCTGTGTATGTACATAGGCACTTTCTGCT~rrCGACCATTGAATTGGAGTGGTCrG ²⁸²⁹² L L V A C C 0 L L L ^E P ⁵ L V Q Y ⁵ 0 V M S H Q V T ^I H A A S D T L E F C S E S A 908 GTCCTTATAAACGCTATGTTTTCAATCGATCACCAGCGTGAGTCGTATWTGCATrrCGCAGGCCTGAACT~rrAGCCATGAGGCTCCACCCGTGTCA ²⁹⁵²⁵ V L Y K T S M V S Q ^I ^E S ^P ⁵ ⁵ 0 V V M H A Y S ^S C ^P ^E T S Y A ^I E C S H H V L T 949 GTTTTTCCGTTCGCATACCGTGCTGATCTGCGGTTAGCGATCGGGTCCTACTTTCAACCTGCCAATATACTGCTGGTCCCATTATTATCCACTGAT ³⁰⁷⁵⁷ V S S S S A ^I P V H D S V G V A 0 Q C S L L ^I N P ⁵ H ^I S L P E S S F ^I T P T A 5 990 TTACGCAGTTCCCCTCCCTTCTGTGATGGGGTGGCTGGGCCTCTCGATATGAATGCTTTACTGAACAGTGGGTGAAACTTTAAATGTTTCCCTGTT ³¹⁹⁸⁹ L L Q L P ^P A L S C 0 C E W 5 G A 5 ⁵ 0 5 E L L L P D T D C L R T L N M S S P V 5 1031 GTAGCTGATTTTACATACACGACATCTGTGTGTGCGATGATATTAAGCCGCTTTCTAAAGGTGAAATGATGTATGAAATGAGCGAAAATGTCTTTTCAGTGACATGGCATAC
VADFTYTTSVSGDDIKPIKPLSVSGDDIKPLSKGBMMYGNBTLBLKMBTELKMSSFSDMAY V A D F T Y T T S V S G D D I K P L S K G E M M Y G N E T E L K M S S F S D M A Y 1072
CCTTCTAAAAGCACAGTCGTGCCAAAGATGTCTGATATTGTAAATAAGTGGAGTGAATCTTTAAAAGAAACCTCTGTTTCCGTATCTAGACATAGACGCGTGTTTACAGAGTCTCTTGTTTAT 3444 ^P ⁵ K ^S T V V P K M ⁵ 0 ^I V N K W ^S ^E ^S L K ^E T ^S V S V S ^S ^I N S V ^F T E S L V Y 1113 CCCAAACTAGGTTTTATCAGAGATAGTGAGTCCAAGATATCTCCCGTTAACCTCACAACACATCCAAGATCTGTGCACTGGCTAAACCGGCTTACA ³⁵⁶⁷⁶ P I T K V F D Q E I S R V P E I I F P V K P T H T A S Q A S G D T W L K P G L S T 1154
AACTCAGAGCCTGCGCTCTCTGACACTGCTTCTAGTGAAGTATCACACCCCTTCAACAGCCCCTTGCTCTATGAGGCCGCATCTCCTTTTAATACGGAAGCATTGCTGCAACCTTTCCG 3690 N S E P A L S D T A S S E V S H P S T Q P L L Y E A A S P F N T E A L L Q P S F P 1195
GCTTCTGATGTTGACACCTTGCTTAAAACTGCCCTTCCCTCTGGGCCTCGTGATCCAGTGCTGAACCCCCATGGTTGAGCAAAGTAGCTCTTCCCTTATCTCTCCCTGGCATCAGAG 3813 A ⁵ 0 V 0 T L L K T A L ^P 5 C ^P R 0 P V L T E T P M V E Q S S S S V S L ^P L A S E 1236 TCTGTTCAGCAAAGCCGCTCACTTACACTGACCATTCTAATAGTCCCTTTGATTGCACCCCTTCCTTCACGCTAAAGTTCTCATCGAGGAGAATTT ³⁹³⁶³ S A S S K S T L H F T S V P V L N N S P ⁵ 0 V H ^P T S L Q R L T V P H S R E E Y F 1277 GAACAGGTTGCTTAGAGAAAATCCCAGCAGTCCGCCGCCTTCACACCATACGATTTTCCAACTGCCATCGGACTTAGCAGGCTACCTCCA~a~aA ~ ⁴⁰⁵⁹⁵ E Q C L L K S K S ^P Q Q V L ^P 5 L H S H D E ^F F Q T A H L D ^I S Q A Y P P K C R H 1318 GCATTGCTCTCCATTTATCATCAAGAACACAAATACCTTAAAACCGCTGTGATTCGAGGCATTTCATCACCTGAATTTCATTATGATAGGCCTTATG ⁴¹⁸²⁸ A F A T ^P ^I L S ^I N E P Q N T L ^I N R L V Y S E 0 ^I F M H P E I S ^I T 0 K A L T C 1359 CTACAACGCCGTTCTATGTCTTAAGCTCTGCCATCTGTCCATAGGAGTGGCCCTTTCATGCAACGTTTTCCCACAAGATATTCGTGACACACCAGT ⁴³⁰⁵⁰ L ^P T T V ⁵ 0 V L ^I A T ⁰ H ^S V ^P L G S C ^P ^I S M T T V S ^P N R 0 0 ⁵ V T T T K L 1400 CTTTTCTCTAAGTATTTAACGTATGTGCAATTGAGCGATTGTAGAGTG GAGTAGCAGTACTATTAGCTGTG ^T ⁴⁴²⁸ L L ^P ⁵ K A T S K P T H S A R ⁵ 0 A 0 L V C G C E 0 G 0 0 Y 0 0 0 0 Y 0 0 ^I 0 5 0 1441 CGCTTCCCTAAAAAGGTATTCATTTCACCTTAGAAATCCAGGAAAGTAAGAATACTCGACCCCAGAAACAGTTTGGGATAGAGGACCAATTCACTT ⁴⁵⁵¹⁵ R F P V N K C M S C S P Y R E S Q E K V M N D S D T Q E S S L V D Q S D P I S H L 1482
CTCTCTGAGAATACCGAAGAAAAATGGAGGCACGGGTGTAACTAGGGTGGATAAAAGTCCTGATAAGTCACCACCAAGTATGCTACCCCAGAAGCACAATGATGGAAGAGGGCCCT 4674 L S E N T E E E N G G T G V T R V D K S P D K S P P P S M L P Q K H N D G R E D R 1523
GACATTCAGATGGGTAGTGCTGTCCTTCCTCACACCCCAGGATCT<mark>AAAGCATGGGCAGTTTTGACAAGTGAAGAGAGTGGGTCAGGGCAAGGCACCTCAGATAGCCTTAATGATAATGA</mark>G 1797 DIQ MGSA VLPHTPGSKA WAVLTSDEESGSGSGQGTSDSLNDNE1564
ACTTCCACAGATTTCAGTTTCCCAGATGTTAATGAAAAGGATGCTGGTGCTGGAAGCAGATGACACAGGCATAGTCCGGGATCTCCACGGTCCTCCACACCATCTGTTACTAGTGGG4920
TSTDFSFDFSFDVNECAGATGTTAATGAAAAAGGATGCTGGTGCTGGAA CCACAGCAATCTCACTTCTATCAAACGTCATACATTATGCTCTATAATAGGTTTACTATGTTTCAATGATATCAAATATTTGGTATTATTTTATAT ⁵¹⁶⁶⁶ TCAAATGCTTGTTGCGTTATAATCCTATCCATTGGTGTGTATAACATGTTGCAAATAGCCTTAGATATACAAGCATGTGTGTGTAAGCAGTTAATG ⁵²⁸⁹⁸ CATACTGTATCTGAACAGTTGACACAACTTATTGATTTAAATTCAGAATAGCTAAAGCCCCACCACAATGGGGACCTTCCATGGCACACAGTTGGGAG ⁵⁴¹²¹ TCTCAAAGACTAAGGCTAAATTTATAGTAAGAAGTTTTATAACATATACGCTTTAAATATCTTGGCTTTTTTTGTGCCTAAGAGGCATATAACCCC ⁵⁵³⁵³ AATAATTCACTTGTGTTGTACAGATTCAACTACTATCGCATGATAGCAGAGATCATAGTATGTAGACAGATGAAAGAAAAGCCACAGCTTTCTGAG ⁵⁶⁵⁸⁵ ATGCTTTCTTGTAATTATTTATGAAATGGTTATTGCCAATGTTCTGGAGAGACGCAACTAAATAAGATTGTGTTTAATGCAATGTAAAAATTCGAA ⁵⁷⁸¹⁸ 'ATTAAGTGTTTAAAAAAAAAACTAGGTGATTCCATTGAATTCACGGCAAGAATTTTATAATTTGCTTTGATTAAGTGATGAGGTTTTTTTATAGCACAAGGAAAGTATTTAAATCTTAG 5904
AGTCAAAAGATAAGATAATGTATAATGTGACATTGGGTAATAATTGCACAGAGCTTAATATTAAAAATCCACAGTATAGAGCCAATCTCATTAAAT TTAGTTTATTAATGTAAGATAAGTCCCCACAATCAAACCAGCATAGGATAAACTCCACTTAGGAATAAACAGTAAGAAGGAGTGGATAAAAGAATA ⁶⁵¹⁹¹ GAAACAATTTTACTGTTTTAAGCTTATCGTAAACTATTCAAGTCGTCTATCAATTTTGATATTCCTATCTGAACTATCATTATGTACTCTAATTGAG ⁶⁶⁴²⁴ AAACCACAGATGTTCCAAATAATTAAAAGGGA~kTATAATGAAGTGACAAGACGTAAATACTGCTTGGTCACTAAAAGTCATCCAAATAACCTGGAA ⁶⁷⁶⁵⁶ TATTATTAGTGTCTCTGTCTTTAGAACATGCTTACAT o 2 3 4 5 6 7 ^a ¹ 1kh $= c12$ \bullet c21 pBSP3 $pBSP4 =$

FIG. 2. (Legend appears at the bottom of the opposite page.)

to nitrocellulose, followed by reverse-phase HPLC of the digestion products (Fig. 1).

For cloning the 3F8 proteoglycan, we initially used mixed oligonucleotide-primed PCR amplification of cDNA, based on amino acid sequences present in the proteoglycan, for synthesis of an unambiguous nucleic acid probe. The N-terminal sequences of the core glycoprotein and of the 14-kDa CNBr peptide were used for the design of PCR primers: sense, 5'-CCGCGGATCCAARYTIGTIGARGARATIG-GITGG-3'; antisense, 5'-GCTTAAGCTTRTAYTGYTGYT-CICKRAARTTRTTYTG-3'. This primer combination resulted in the enzymatic amplification of a 780-bp product with ^a rat brain cDNA library as template, and dideoxynucleotide chain-termination sequencing demonstrated that both the ³' and 5'-ends contained nucleotides encoding the respective proteoglycan amino acid sequences adjacent to those used in the primer design.

The cDNA library was divided into aliquots, and lysates from each of these were tested by PCR using new primers based on exact sequences of the original PCR product: sense, 5'-TCTCACCAATGACTACTATCTC-3'; antisense, ⁵'- AGTACTTGTCAGTGGAGTTTGG-3'. The nine positive eluates that yielded the expected 370-bp PCR product were each grown on one plate and screened with an antisense RNA transcript of the original PCR product. Supernatants of positive plaques were checked by PCR, and five positive clones from a second screening were converted into Bluescript plasmids by in vivo excision. Two of these $(c12 \text{ and } c21)$ were used for sequencing.

A sequence of ⁴⁶⁹ amino acids deduced from ^a 1.4-kb open reading frame contains the N-terminal amino acid sequence of the 3F8 core glycoprotein as well as all of the CNBr, tryptic, and endoproteinase Lys-C peptide sequences from the proteoglycan (Fig. 2). The deduced amino acid sequence, beginning with a 24-amino acid signal peptide, reveals an N-terminal domain of 255 amino acids (aa 38-292) having significant identity to carbonic anhydrases and 93% identity to ^a carbonic anhydrase domain in the N terminus of human PTP ζ/β (5, 6). The 30 N-terminal amino acids of the 3H1 chondroitin/keratan sulfate proteoglycan were found to be identical to those of the 3F8 proteoglycan, and six internal tryptic peptide sequences also matched those found in sequenced peptides of the 3F8 proteoglycan and/or amino acid sequences deduced from the cDNA clones (Figs. ¹ and 2).

RNA Blot Analyses. An RNA probe corresponding to the cDNA sequence of the extracellular N-terminal carbonic anhydrase domain hybridizes with messages of 9.5, 8.4, and 6.4 kb from both 4-day and adult rat brain mRNA, and a faint band above 9.5 kb was sometimes seen in early postnatal brain (Fig. 3). A 160-kDa probable splice variant of RPTP β has been identified (6), in which nt 2393-4951 (Fig. 2) are deleted. It is likely that this splice variant corresponds to the 6.4-kb message seen on Northern blots. The 8.4-kb message was not detectable using probes corresponding to either the first (data not shown) or second (Fig. 3) cytoplasmic phosphatase domains, suggesting that this mRNA may code for ^a soluble protein containing only the extracellular portion of the phosphatase. No message was detected in liver, kidney, muscle, lung, or rat PC-12 pheochromocytoma cells, in agreement with the reported limitation to nervous tissue of $RPTP\beta$ expression (6) and our finding that the proteoglycan is synthesized by glia (4).

Identification of an Extraceliular Variant of the Phosphatase. Because the results of RNA analyses suggested that the 3F8 proteoglycan isolated in a phosphate-buffered saline extract of brain might represent an extracellular splice variant of the full-length transmembrane RPTP, we used reverse transcription/PCR to synthesize a probe to a sequence ≈ 0.8 kb upstream of the human RPTP ζ/β transmembrane region (corresponding to nt 4050-4380 in Fig. 2). This PCR product was subcloned into pGEM and used as template to transcribe an RNA probe for library screening. As expected, on RNA blots it hybridized with only the 9.5- and 8.4-kb messages (data not shown).

Two cDNA clones (pBSP3 and pBSP4) were identified that gave no PCR product using primers corresponding to sequences in the first phosphatase domain (which should be absent from the putative extracellular splice variant), and these clones were therefore considered good candidates for further study. Restriction analysis and sequencing demonstrated that clone pBSP4 overlapped at its ⁵' end with the N-terminal sequence of the proteoglycan obtained from clones c12 and c21 (Fig. 2). Both clones contained identical 3'-untranslated sequences preceded by a stop codon at the putative 3' splice site of RPTP β (6) and may represent an alternative splicing product corresponding to the 8.4-kb mRNA seen on RNA blots. However, definitive evidence concerning the origin of this extracellular variant must await determination of the intron/exon junctions from genomic cloning studies. The rat amino acid sequence has 76% identity to the human sequence, whereas the ³'- and ⁵' untranslated regions have no identity. RNA analysis using an RNA probe transcribed from ^a subcloned EcoRI fragment from the beginning of the 3'-untranslated region (nt 5034- 5820) showed that this hybridized to the 8.4 kb and, very weakly, to the 9.5-kb messages (Fig. 3). The RNA blot results suggest that the extracellular proteoglycan may be the major product of the phosphatase (or a closely related) gene expressed in rat brain and that the 3'-untranslated regions may not be identical in the three mRNAs. That the extracellular proteoglycan does, in fact, have a different 3'-untranslated region has been demonstrated by sequencing clones for the full-length rat PTP (unpublished results).

DISCUSSION

The studies reported here demonstrate that the 3F8 and 3H1 proteoglycans of brain, which we have named phosphacan, are glycosylation and possible extracellular splice variants of a receptor-type transmembrane PTP designated $PTP\zeta(5)$ and $RPTP\beta$ (6). The proteoglycan contains an N-terminal domain homologous to carbonic anhydrases. At least seven mammalian carbonic anhydrase isozymes are known, and the highest level of identity (32%) of phosphacan is with carbonic anhydrase VI, a protein secreted by salivary glands (11). However, because a catalytically obligatory zinc ion is bound to three essential histidine residues in carbonic anhydrases, whereas in both rat phosphacan and human RPTP ζ/β two of these have been changed to threonine and glutamine (Fig. 2), it is unlikely that this proteoglycan can function as a catalytically active carbonic anhydrase (5). Based on the general definition of fibronectin type III homology domains as ≈ 90 residue modules characterized by conserved tryptophan and tyrosine residues in their N- and C-terminal halves, respectively, but lacking conserved cysteine residues, we can also

FIG. 2 (on opposite page). Nucleotide sequence of phosphacan cDNA and deduced amino acid sequence. The contiguous cDNA sequence determined from overlapping clones is shown, together with the translation ofa 4.8-kb open reading frame. Peptides from which 3F8 proteoglycan amino acid sequence data were obtained are shown in boldface type, and amino acid and nucleotide sequences used for the design of oligonucleotide primers for enzymatic amplification reactions are boxed or underlined. Threonine and glutamine residues, which replace two of the three essential histidines in the zinc-binding site of catalytically active carbonic anhydrases, are indicated in boldface italics. The clones used for sequencing are shown in the lower diagram, with the bold lines representing coding sequence.

FIG. 3. RNA blot of 4 μ g of mRNA from 4-day (lanes A, C, E, and F) and adult rat brain (lanes B and D) electrophoresed on a 1% agarose gel containing 6% formaldehyde. Blots were probed with RNA transcripts corresponding to the N-terminal carbonic anhydrase domain (lanes A, B, and E), to the cytoplasmic second phosphatase domain of the membrane tyrosine phosphatase $RPTP\zeta/\beta$ (lanes C and D), and to an EcoRI fragment from the 3'-untranslated region of the extracellular variant (lane F). The blot was stained with alkaline phosphatase-labeled anti-digoxigenin antibodies (2). Positions of RNA molecular-size markers are indicated at left.

identify such a domain in phosphacan (residues 312-391, including Trp-332 and Tyr-384). The amino acid sequence deduced from our cDNA clones, corresponding to ^a 173-kDa mature protein with a calculated pI of 4.32, has 76% identity to the extracellular portion of the 252-kDa human transmembrane RPTP ζ/β (5, 6), and expression of a full-length human $RPTP\zeta/\beta$ cDNA construct in kidney fibroblasts has demonstrated that the phosphatase is also synthesized in the form of a chondroitin sulfate proteoglycan (G. Barnea, 0. Silvennoinen, J. B. Levy, J. Sap, R.U.M., and J. Schlessinger, unpublished results).

Of the 36 Ser-Gly or Gly-Ser dipeptides in phosphacan, all of the serine residues that could serve as potential chondroitin sulfate-attachment sites are located in the large serineand threonine-rich (\approx 23%) domain outside of the carbonic anhydrase homology region. The potential sites most likely to be used (i.e., those consisting of serine-glycine sequences with a closely preceding or immediately following acidic amino acid) are Ser-595 (DDSSGSS), -1005 (DGEWSGAG), -1549/1551 (DEESGSGQG), and -645 (EDSAPSGSEE). These five sites are sufficient to accommodate the calculated four chondroitin sulfate chains in the 3F8 proteoglycan (1). However, we have no basis on which to identify potential attachment sites for the keratan sulfate chains in the 3H1 proteoglycan, some or all of which are linked through unusual mannosyl-O-serine/threonine linkages, which we have previously described in the chondroitin sulfate proteoglycans of brain (12, 13). The 3F8/3H1 proteoglycans appear to represent glycosylation variants of the same protein, and considered together these proteoglycans account for 15-20% of the total soluble chondroitin sulfate proteoglycan protein in early postnatal and adult rat brain, respectively (1). This result indicates that phosphacan represents a major fraction of brain proteoglycans, like neurocan, which accounts for \approx 20% of the soluble proteoglycan protein and is the only other wellcharacterized chondroitin sulfate proteoglycan present at similarly high levels in brain (1, 2).

RPTPs containing immunoglobulin-like and fibronectin type III-like repeats have been shown to mediate cell-cell adhesion by homophilic binding of their extracellular domains (14-16). The ability of phosphacan to bind to N-CAM, Ng-CAM, and tenascin (3, 17) provides evidence that RPTPs may also mediate cell interactions through heterophilic mechanisms. The pattern of expression of three RPTPs in the developing Drosophila central nervous system suggests that they may be involved in axonal pathfinding and stabilization (18, 19). In situ hybridization histochemistry demonstrated a high level of expression of phosphacan in the ventricular zone and other areas of cell proliferation in the embryonic rat central nervous system and that in postnatal cerebellum it is synthesized by glial cells such as the Golgi epithelial cells, from which the Bergmann glia fibers project into the molecular layer where they guide granule cell migration (refs. 4, 17, 20; and M. Engel, M.F., P.M., R.U.M., and R.K.M., unpublished results). Immunocytochemical studies of cerebellum had previously shown that staining of the 3F8 and 3H1 proteoglycans was most intense in the molecular layer (1), and considered together these data indicate that they may play important roles in nervous tissue histogenesis. The effects of developmentally regulated glycosylation (1) on the ability of phosphacan to bind to neurons, neural CAMs, tenascin, and other potential cell-surface or extracellular matrix ligands for membrane tyrosine phosphatases require further investigation. It is possible that developmental changes in glycosylation and in the localization and levels of the extracellular proteoglycan may also serve to modulate the activity of the transmembrane phosphatase.

We thank Drs. Julie Rushbrook and William Lane for sharing their expertise in protein sequencing, Margit Engel for skillful technical assistance, and Dr. Martin Grumet, Dr. Jan Sap, and Gilad Barnea for useful discussions. This work was supported by Grants NS-09348, NS-13876, and MH-00129 from the National Institutes of Health.

- 1. Rauch, U., Gao, P., Janetzko, A., Flaccus, A., Hilgenberg, L. Tekotte, H., Margolis, R. K. & Margolis, R. U. (1991) J. Biol. Chem. 266, 14785-14801.
- 2. Rauch, U., Karthikeyan, L., Maurel, P., Margolis, R. U. & Margolis, R. K. (1992) J. Biol. Chem. 267, 19536-19547.
- 3. Grumet, M., Flaccus, A. & Margolis, R. U. (1993) J. Cell Biol. 120, 815-824.
- 4. Milev, P., Friedlander, D., Grumet, M., Margolis, R. K. & Margolis, R. U. (1993) Glycobiology 3, 535.
- 5. Krueger, N. X. & Saito, H. (1992) Proc. Natl. Acad. Sci. USA 89, 7417-7421.
- 6. Levy, J. B., Canoll, P. D., Silvennoinen, O., Barnea, G., Morse, B., Honneger, A. M., Huang, J.-T., Cannizzaro, L. A., Park, S.-H., Druck, T., Huebner, K., Sap, J., Ehrlich, M., Musacchio, J. M. & Schlessinger, J. (1993) J. Biol. Chem. 268, 10573-10581.
- 7. Charbonneau, H. & Tonks, N. K. (1992) Annu. Rev. Cell Biol. 8, 463-493.
- 8. Brautigan, D. L. (1992) Biochim. Biophys. Acta 1114, 63–77.
9. Walton, K. M. & Dixon, J. E. (1993) Annu. Rev. Biochem
- 9. Walton, K. M. & Dixon, J. E. (1993) Annu. Rev. Biochem. 62, 101-120.
- 10. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E. & Kent, S. B. (1987) Proc. Natl. Acad. Sci. USA 84, 6970-6974.
- 11. Aldred, P., Fu, P., Barrett, G., Penschow, J. D., Wright, R. D., Coghlan, J. P. & Fernley, R. T. (1991) Biochemistry 38, 569-575.
- 12. Krusius, T., Finne, J., Margolis, R. K. & Margolis, R. U. (1986)J. Biol. Chem. 261, 8237-8242.
- 13. Krusius, T., Reinhold, V. N., Margolis, R. K. & Margolis, R. U. (1987) Biochem. J. 245, 229-234.
- 14. Gebbink, M. F. B. G., Zondag, G. C. M., Wubbolts, R. W., Beijersbergen, R. L., van Etten, I. & Moolenaar, W. H. (1993) J. Biol. Chem. 268, 16101-16104.
- 15. Brady-Kalnay, S. M., Flint, A. J. & Tonks, N. K. (1993) J. Cell Biol. 122, 961-972.
- Sap, J., Jiang, Y.-P., Friedlander, D., Grumet, M. & Schlessinger, J. (1994) Mol. Cell. Biol. 14, 1-9.
- 17. Milev, P., Friedlander, D., Grumet, M., Margolis, R. K. & Margolis, R. U. (1993) J. Neurochem. 61, (Suppl.), S110C.
- 18. Tian, S.-S., Tsoulfas, P. & Zinn, K. (1991) Cell 67, 675–685.
19. Yang, X., Seow, K. T., Bahri, S. M., Oon, S. H. & Chia, W. (
- 19. Yang, X., Seow, K. T., Bahri, S. M., Oon, S. H. & Chia, W. (1991) Cell 67, 661-673.
- 20. Canoil, P. D., Barnea, G., Levy, J. B., Sap, J., Ehrlich, M., Silvennoinen, O., Schlessinger, J. & Musacchio, J. M. (1993) Dev. Brain Res. 75, 293-298.