Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 7242–7246, August 1991
Neurobiology

Molecular characterization of a protein-tyrosine-phosphatase enriched in striatum

(basal ganglia/molecular cloning)

PAUL J. LOMBROSO*†, GEOFFREY MURDOCH†‡, AND MICHAEL LERNER†

*Child Study Center, †Section of Molecular Neurobiology and Howard Hughes Medical Institute, Yale University School of Medicine, 330 Cedar Street, New Haven, CT 06510

Communicated by Charles F. Stevens, April 25, 1991

ABSTRACT **A cDNA clone encoding a neural-specific putative protein-tyrosine-phosphatase (protein-tyrosine-phosphate phosphohydrolase, EC 3.1.3.48) has been isolated from a rat striatal cDNA library. The deduced amino acid sequence predicts a protein of ≈369 amino acids with a strong homology to other members of the family of protein-tyrosine-phosphatases. *In vitro* translation produces a protein with an apparent molecular mass of 46 kDa. A potential attachment mechanism to the cytoplasmic membrane is suggested by a myristoylation amino acid-consensus sequence at the N terminus of the protein. RNA analyses of various regions of rat brain reveal a 3-kilobase (kb) and a 4.4-kb mRNA. The 3-kb mRNA is highly enriched within the striatum relative to other brain areas and has been termed a "*st*riatum *e*nriched *p*hosphatase" (STEP). In contrast, the 4.4-kb message is most abundant in the cerebral cortex and rare in the striatum. These two messages appear to be alternatively processed RNA transcripts of a single gene.**

It is well established that modulation of tyrosine phosphorylation is one mechanism for controlling cellular proliferation and differentiation (1). The addition and removal of phosphate groups on tyrosine residues is thought to be tightly controlled by the opposing actions of protein-tyrosine kinases and protein-tyrosine-phosphatases (protein-tyrosine-phosphate phosphohydrolase, EC 3.1.3.48; PTPases) (for recent review, see ref. 2). Until recently, research has concentrated on the role of protein-tyrosine kinases in cellular control mechanisms.

Originally, phosphatases were viewed as a small group of enzymes with a rather broad substrate specificity. The recent isolation of a number of PTPases suggests that this may not be correct. PTPases were first identified when the human placental PTPase 1B was isolated, purified, and shown to be homologous to a repeated sequence in the cytoplasmic domain of the leukocyte cell-surface protein CD45 (also known as leukocyte common antigen, or LCA) (3). Although the function of CD45 was not known, the fact that CD45 and PTPase 1B were homologous suggested that in lymphocytes, CD45 might regulate the state of phosphorylation of a group of substrate proteins. Subsequent experiments confirmed that PTPases dephosphorylate tyrosine residues, with little or no effect on serine/threonine residues (4, 5).

Since the initial report on PTPase 1B, a number of PTPases have been cloned and their amino acid sequences determined. Two basic variants have emerged. Those in the first class have extracellular domains linked to cytoplasmic, catalytic domains through single transmembrane regions. These PTPases may interact with soluble factors, surface molecules on neighboring cells, or extracellular matrix molecules. Several members of this first class of PTPases have been isolated (6–11). PTPases in the second class are located intracellularly and lack a transmembrane or extracellular domain. Several have been isolated and characterized as well (5, 12, 13).

Northern analyses of different tissues indicate a wide variability in the regional expression of PTPases. For example, one transmembrane PTPase has been found in all tissues and cell lines tested, suggesting a ubiquitous role (11). Meanwhile, others have been identified as present in a limited number of tissues and cell lines. Finally, another PTPase (CD45) is expressed only in cells of hemapoietic lineage (14). However, none of the PTPases identified to date have been shown to be nervous system specific.

Here, we report the isolation and characterization of a cDNA clone that encodes for a putative PTPase found predominantly in the brain.§ Within the brain, this PTPase shows further tissue specificity by appearing to be a "striatum enriched phosphatase" and has been designated STEP. This PTPase contains regions homologous to ones found in previously characterized PTPases and appears to be a distinct member of the cytoplasmic class of PTPases.

MATERIALS AND METHODS

RNA Isolation. Total RNA from bovine putamen was isolated by using a modification (15) of the guanidinium isothiocyanate procedure of Glisin *et al.* (16). For library construction, poly(A)$^+$ RNA was twice selected by affinity chromatography using oligo(dT)-cellulose (17). For Northern analyses, poly(A)$^+$ RNA from rat organs and dissected brain regions was selected once using oligo(dT)-cellulose (Fast Track, Invitrogen).

Subtractive Putamen cDNA Library. First-strand cDNA was synthesized from 5 μg of bovine putamen poly(A)$^+$ RNA using mouse Moloney leukemia virus reverse transcriptase (Boehringer Mannheim) (18) and was hybridized in the presence of 40 μg of poly(A)$^+$ RNA isolated from bovine cerebellum in 0.125% SDS/2 mM EDTA/0.3 M phosphate buffer in a final volume of 20 μl (19). The sample was covered with paraffin oil, denatured at 90°C for 30 sec, and hybridized at 65°C for 24 hr. The sample was passed over a hydroxylapatite column (HAP), and single-stranded cDNA was eluted with 0.12 M phosphate buffer/0.1% SDS, and a second cycle of subtractive hybridization was performed. Second-strand cDNA was synthesized by using DNA polymerase I (Klenow fragment) (20, 21), followed by S1 nuclease digestion (21) and C-tailing (22). The plasmid pUC-9 (Stratagene) was linearized at its *Pst* I site and G-tailed to allow hybridizing to C-tailed cDNA.

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: PTPase, protein-tyrosine-phosphatase; STEP, striatum enriched phosphatase.

‡Present address: Division of Neuropathology, Oregon Health Sciences University, Portland, OR 97201.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M65159).

Isolation of Additional STEP cDNAs. A randomly primed striatal λgt11 cDNA library (Clontech) was screened by using randomly primed [^{32}P]dCTP (Amersham) radiolabeled probes (23). The initial probe was a 300-base pair (bp) bovine cDNA clone. Filters were hybridized in buffer containing 50% (vol/vol) formamide/5× Denhardt's solution/5× standard saline phosphate/EDTA (SSPE; 1× SSPE is 0.18 M NaCl/10 mM phosphate, pH 7.4/1 mM EDTA)/1% SDS/boiled salmon sperm (100 μg/ml) at 42°C for 18 hr, followed by washes in 0.5× SSC/0.1% SDS at 65°C. *Eco*RI inserts of purified clones were subcloned into the plasmid vector Bluescript (Stratagene), and DNA was purified by CsCl ultracentrifugation (24).

DNA Sequence Analysis. Nucleotide sequences were obtained using the chain-termination methods of Sanger (25) after either subcloning inserts into M13 bacteriophage or by using cDNA clones in Bluescript, which were progressively digested using Exo III nuclease (Erase-a-Base, Promega). Sequences deposited in GenBank were searched for homology with a sequence analysis software package (University of Wisconsin).

Northern Blot Analyses. Two micrograms of poly(A)$^+$ RNA was electrophoresed on 1% formaldehyde/agarose gels, transferred to nylon membrane and hybridized with randomly primed probe (1×10^6 cpm/ml) under high-stringency conditions in 50% (vol/vol) formamide/5× Denhardt's solution/5× SSPE/1% SDS/boiled salmon sperm (100 μg/ml) at 42°C, and washed at 65°C in 0.1× SSC/0.1% SDS. Probes were either a 2.4-kb cDNA clone (9.6), which consists of ≈600 bp on either side of the open reading frame or a 400-bp *Apa* I–*Pvu* II fragment within the 3′ untranslated region.

***In Vitro* Translation of STEP.** A plasmid containing the complete STEP open reading frame was constructed from overlapping partial clones by splicing at the unique *Bam*HI site (Fig. 5, for restriction map). This sequence was transcribed using T3 RNA polymerase (mRNA capping reaction, Stratagene). As a control experiment, antisense message was transcribed using T7 RNA polymerase. Two and one-half micrograms was added to a rabbit reticulocyte *in vitro* translation mix (Promega) in the presence of [^{35}S]methionine (Amersham). Twenty microliters of this reaction was loaded onto a 12% SDS/polyacrylamide gel and processed by standard procedures (26).

Southern Blot Analysis. Rat genomic DNA was digested with restriction enzymes that either had no restriction sites in the cDNA clone (*Xba* I and *Hin*dIII) or had a single site (*Eco*RI and *Bam*HI). Five micrograms of digested genomic DNA was fractionated on an 0.7% agarose gel and, after brief acid depurination, transferred in 0.4 M NaOH onto Zeta-Probe membrane (Bio-Rad).

RESULTS

Isolation and Characterization of STEP cDNA Clones. A putamen-enriched plasmid cDNA library was constructed from bovine tissue. Over 90% contained inserts with an average size of 300 nucleotides, and 10% of these represented mRNAs that were enriched in the putamen relative to cerebellum (data not shown). One of these clones containing a 300-bp insert was used to screen a rat striatal λgt11 library. The initial screen yielded a 1-kb cDNA clone (9.2) containing portions of an open reading frame (see Fig. 5 for all clones isolated). The 5′ region was used to rescreen λgt11 libraries to obtain overlapping clones encoding a complete open reading frame. The sequence obtained from these overlapping cDNA clones is shown in Fig. 1.

The predicted amino acid sequence of the single, 1107-base open reading frame identified the gene product as a potential member of the phosphotyrosine phosphatase gene family. The highly conserved 12-amino acid domain originally identified by Charbonneau *et al.* [His-Cys-Ser-Ala-Gly-(Ile/Val)-Gly-Arg-(Ser/Thr)-Gly-Xaa-(Phe/Tyr)] is highlighted in Fig. 1 (12). A comparison of the predicted amino acid sequence of STEP with the previously described intracellular PTPases, T-cell PTPase and placental PTPase 1B, reveals ≈30% identity. The degree of similarity increased to >50% by allowing conservative substitutions. As illustrated in Fig. 2, most of this similarity falls within a 250-amino acid region characteristic of the catalytic domain of all PTPases. Similar to other intracellular PTPases, this domain is present only once. In contrast, the transmembrane receptor-like PTPases have two phosphatase domains. Although the catalytic domain is lo-

```
   1                G GGGGGATGTTGAGGG ATGAGGGCCTTTCTG GTGGCCTCCGTTACT TCAAGTCTTCATGAC
  62 TGGAACGAAGGGAAG ACTGGAGTTCTGAAC TACATCTCAGGAAT TAATTTCCTCGTATC CTGCCTACCCTAACC
 137 CATAAATAAAAAGAC AGGTATCCTGGGTCT CTCTGAAGACCTCCT GTCTAAGGAGGCCTT TCTGGAAGACTTCCT
 212 TTCTATGTGTACCTG TGGGGTCTTTTGTGG GGAGGGATGTCTCAG TCGCCTGTGATTAGC TTTTCTCCTGAAGAC
 287 TGCTGAGGCCTCCTC TGGGTACACCACCTC TCCCTCCACTGCCCA GTCTGAGAGGCTCAG GAGCTTCCTGGGGGC
 362 TGAAGGTGAGCTCTG CAGGGCAATATAGGG GCAAAGGGGTGCAGC TGTCCACTCAGTCTG CGCCCCCGACTGCCC
 437 CAGCTTGCCTGTGGG GTGCTCTGGCTCAGT GGCCATGGCCACTCC TGGCTGCAGACCACC ACAGACCTCATCTTC
 512 TCCTCGCTCACCGTG TTGAACCACCTGGGA CCTGCAGCCTGGCTA GGTTCTGGGACCTGG GGGACCACCAAGTCTG
 587 CTGTTAGTCTCTCTG ACTGTGAGCCTGGGC ATCATCACCACCCTG GTGTGGCACCTTCTC AAGACACCCCCAGAG
 662 CCACCTGTCCCACTG CCCCCAGAGGACAGA CGACAATCAGTGAGC CGGCAGCCCTCCTTC ACTTACTCAGAGTGG

   1  Met Glu Glu Lys Val Glu Asp Asp Phe Leu Asp Leu Asp Ala Val Pro Glu Thr Pro Val
 737  ATG GAG GAG AAG GTA GAG GAT GAC TTT CTG GAC CTG GAC GCA GTG CCG GAG ACA CCC GTG

  21  Phe Asp Cys Val Met Asp Ile Lys Pro Glu Ala Asp Pro Thr Ser Leu Thr Val Lys Ser
 797  TTT GAC TGT GTG ATG GAC ATC AAG CCT GAG GCT GAT CCC ACC TCG TTG ACT GTC AAG TCC

  41  Met Gly Leu Gln Glu Arg Arg Gly Ser Asn Val Ser Leu Thr Leu Asp Met Cys Thr Pro
 857  ATG GGT CTG CAG GAG AGG AGG GGC TCC AAT GTC TCC TTG ACC CTG GAC ATG TGT ACT CCT

  61  Gly Cys Asn Glu Glu Gly Phe Gly Tyr Leu Val Ser Pro Arg Glu Glu Ser Ala His Glu
 917  GGC TGC AAT GAG GAG GGC TTT GGC TAC CTG GTG TCC CCA CGA GAG GAG TCA GCC CAT GAA

  81  Tyr Leu Leu Ser Ala Ser Arg Val Leu Arg Ala Glu Glu Leu His Glu Lys Ala Leu Asp
 977  TAC CTG CTC AGT GCC TCC CGT GTC CTC CGG GCA GAG GAG CTG CAC GAA AAG GCC CTG GAC

 101  Pro Phe Leu Leu Gln Ala Glu Phe Phe Glu Ile Pro Met Asn Phe Val Asp Pro Lys Glu
1037  CCT TTC TTG CTG CAG GCG GAA TTC TTT GAA ATC CCC ATG AAC TTT GTG GAT CCA AAA GAA

 121  Tyr Asp Ile Pro Gly Leu Val Arg Lys Asn Arg Tyr Lys Thr Ile Leu Pro Asn Pro His
1097  TAT GAC ATC CCA GGG CTG GTA CGG AAG AAT CGG TAC AAA ACC ATC CTT CCC AAT CCT CAC

 141  Ser Arg Val Arg Leu Thr Ser Pro Asp Pro Glu Asp Pro Leu Ser Ser Tyr Ile Asn Ala
1157  AGC AGG GTA CGT CTG ACG TCA CCA GAC CCT GAA GAT CCT CTG AGT TCC TAC ATC AAT GCC

 161  Asn Tyr Ile Arg Gly Tyr Asn Gly Glu Glu Lys Val Tyr Ile Ala Thr Gln Gly Pro Ile
1217  AAC TAT ATC CGG GGC TAC AAT GGG GAG GAG AAG GTG TAC ATC GCC ACC CAG GGA CCC ATC

 181  Val Ser Thr Val Val Asp Phe Trp Arg Met Val Trp Gln Glu Arg Thr Pro Ile Ile Val
1277  GTC AGC ACT GTG GTC GAC TTC TGG CGC ATG GTG TGG CAG GAG CGC ACA CCC ATC ATC GTC

 201  Met Ile Thr Asn Ile Glu Glu Met Asn Glu Lys Cys Thr Glu Tyr Trp Pro Glu Glu Gln
1337  ATG ATC ACC AAC ATC GAG GAG ATG AAC GAG AAG TGC ACG GAA TAT TGG CCA GAA GAG CAG

 221  Val Val His Asp Gly Val Glu Ile Thr Val Gln Lys Val Ile His Thr Glu Asp Tyr Arg
1397  GTC GTC CAC GAT GGC GTA GAG ATC ACT GTG CAG AAA GTC ATC CAC ACA GAG GAC TAC CGG

 241  Leu Arg Leu Ile Ser Leu Arg Arg Gly Thr Glu Glu Arg Gly Leu Lys His Tyr Trp Phe
1457  CTG CGA CTT ATC TCC CTC AGG AGA GGG ACT GAA GAG AGA GGC TTG AAG CAT TAC TGG TTC

 261  Thr Ser Trp Pro Asp Gln Lys Thr Pro Asp Arg Ala Pro Pro Leu Leu His Leu Val Arg
1517  ACA TCC TGG CCT GAC CAG AAG ACC CCC GAC CGG GCA CCC CCA CTC CTG CAC CTG GTA CGG

 281  Glu Val Glu Glu Ala Ala Gln Gln Glu Gly Pro His Cys Ser Pro Ile Ile Val His Cys
1577  GAG GTG GAG GAG GCA GCC CAG CAA GAG GGA CCC CAC TGT TCC CCG ATC ATT GTT CAC TGC

 301  Ser Ala Gly Ile Gly Arg Thr Gly Cys Phe Ile Ala Thr Ser Ile Cys Cys Gln Gln Leu
1637  AGT GCA GGG ATT GGC AGA ACC GGG TGC TTC ATC GCC ACC AGC ATC TGC TGC CAG CAG CTG

 321  Arg Arg Glu Gly Val Val Asp Ile Leu Lys Thr Thr Cys Gln Leu Arg Gln Asp Arg Gly
1697  CGA CGT GAG GGC GTG GTA GAC ATC CTA AAG ACC ACG TGC CAG CTC CGT CAG GAC AGG GGC

 341  Gly Met Ile Gln Thr Cys Glu Gln Tyr Gln Phe Val His His Ala Met Ser Leu Tyr Glu
1757  GGC ATG ATC CAA ACA TGC GAA CAG TAT CAG TTT GTG CAC CAC GCC ATG AGC CTC TAT GAG

 361  Lys Gln Leu Ser Leu Gln Ser Ser Glu ***
1817  AAG CAG CTA TCC CTC CAG TCC TCA GAG TGA CCA CAC TCTTCCTCTGAGATC CACTGGGCACTGCCG

1883  AGTCTGAGGCTGGGG CTTCACCCCAACTAC CCAAGGTCTTACCTT GGGTCCTGGGCCTGT TCCCTGGCTCTCCCC
1958  TTCAGTCTGCCCCCA TCTGTTGTCCTGGCC TGGTCCCTACCCCAA GCCTAGCTCTTCTAC TGTACATACTGCAGA
2033  GTTCGGGAGGGCAGG GGAGGAATGTGCCAG GCCAGGCCTGGGGCC CTGGGCCTGACCCAC ACTATATAGACCTGG
2108  GGCCTCAGTTTTAAC CATGATTCCACAGCT ACTTGATCCAAAATG TTTTCATGCCACACC CCAAGTGTCCTACCA
2183  CAACATGTTCTGTCT GTCCATCCATCACAG CTATCTGTACCAAAC ACTGTGTCCCCTCAG CCAGGAAAGGAGTAC
2258  TGATCTGAGCTCTCA CCCCTCAGCCCTCAC CCACATCGCCATTCC CAGACACCCACCCTG ATCCCCAGGCAGAGA
2333  TGTCTACCAATCTCT GGCCTCACAGGATGT GGATGGTGGAGACCA GGCAGAGGAGAATGT GGTGGTTGACCGAGA
2408  TCCCAGCTGAGAAGG TTCCCAGGCAAGGGT GGGCATCTCTGGCCT ATGAGACACTGGAGA GAAAGAGGACAGATA
2483  TAGTGGGAAACCCTA AGGATCTGAGCCCTG AGGGGGCCCTCCAAG AGTAGGGTGGGGACA GCTCCAGACCTGCAA
2558  GGCATCTTTCTGTGT CACTATGGGAAACCC CAAATGTCTCCTCGC TACATGTTGCTCCGT GTGTTCCCTTGTCCA
2633  TGTTTGTGTTGAGAG CCCATGAGCAGGGCA TGCATGACTCTTTGG CAACATGTGTTATCT TGGAGCCACGTGTTT
2708  TTATCGCTGACTTTA AATATTTATTCCACG GCATACAGAGACATT TGGTGTCTTTTTATA ATCTGCTCGTGGTCA
2783  TTGAATAGAGCAATA AACAGAGCATTTTGA GCA(60)
```

FIG. 1. Complete nucleotide and deduced amino acid sequence of STEP. Three potential initiating codons are underlined. Polyadenylylation signal is indicated by an open box. The 12-amino acid consensus sequence [His-Cys-Ser-Ala-Gly-(Ile/Val)-Gly-Arg-(Ser/Thr)-Gly-Xaa-(Phe/Tyr)] found in all tyrosine phosphatases isolated to date is indicated by the filled-in box. *, Stop codon.

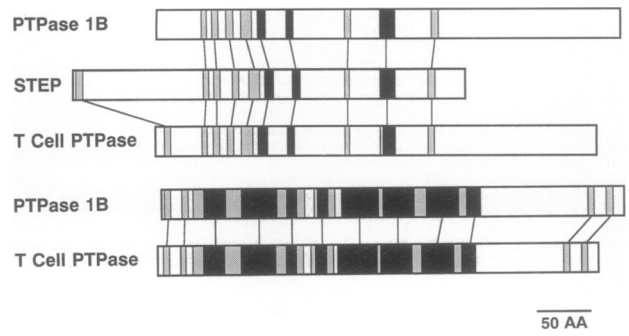

FIG. 2. Sequence similarities of intracellular PTPases. The amino acid sequences of STEP, PTPase 1B, and T-cell PTPase were compared by pairwise Pustell matrix analysis (IBI/MacVector) using an 8-amino acid window. Regions with a similarity score of >90% are shaded black; regions with a similarity score of >50% are stippled.

cated near the C terminus of STEP, in the other cytosolic PTPases described to date it is closer to the N terminus.

The overlapping striatal cDNA clones account for 2816 bases of coding and noncoding sequence including the poly(A) addition signal (see Fig. 1). A putative polyadenylylation signal, AATAAA, was found at nucleotide 2794 and was followed after 17 additional nucleotides by a poly(A)$^+$ tail of 60 adenosine residues.

The predicted size of the STEP protein is between 38 and 42 kDa. There are three methionines at the 5′ end of the open reading frame (underlined in Fig. 1). The first of these potential initiator codons predicts a 369-amino acid protein with an approximate molecular mass of 42 kDa. The nucleotide sequence surrounding the third methionine, however, gives a better match with the Kozak sequence for eukaryotic initiation sites (RCCAUGG) (27). This third potential initiator methionine predicts a 329-amino acid protein of 38 kDa. To gain insight into which of these methionines could act as the initiation codon, *in vitro* RNA transcripts were translated by using a rabbit reticulocyte translation system. A nearly full-length fusion cDNA under the control of a T3 polymerase promoter was constructed (clone 9.6). Fig. 3 shows that the resultant RNA transcripts direct the synthesis of a single prominent protein with an apparent molecular mass of 46 kDa. A few minor smaller proteins were seen with prolonged exposure. Although these results are most consistent with the use of the first potential initiator methionine, PAGE may not be sensitive enough to distinguish between the three sites accurately. Generation of antisera will be required to determine the actual size of STEP protein *in vivo* (see *Discussion*).

Expression of STEP mRNA. The initial nucleic acid-subtraction strategy used to isolate STEP was designed to identify gene products enriched in striatum. Fig. 4*a* demonstrates that STEP is selectively expressed in the nervous system. Whole-brain RNA contains a predominant 3-kb mRNA species and a second 4.4-kb species that is more distinct on longer gels when probed with the 400-bp 3′ fragment (Fig. 4*A*, *Right* and see *Discussion*.) With prolonged exposures (2 weeks), these mRNAs were barely detectable in

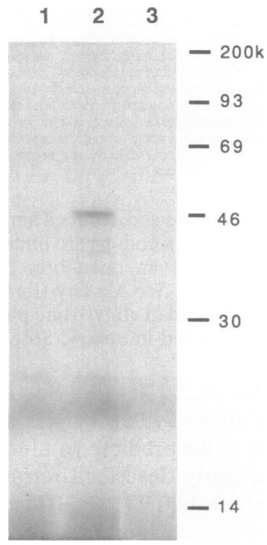

FIG. 3. *In vitro* translation of STEP RNA transcripts. *In vitro* RNA transcripts were translated with a rabbit reticulocyte lysate system in the presence of [^{35}S]methionine and analyzed by 12% SDS/PAGE, as described. Lanes: 1, no added RNA; 2, 2.5 μg of clone 9.6 sense RNA; 3, 2.5 μg of clone 9.6 antisense RNA. Autoradiogram was exposed overnight.

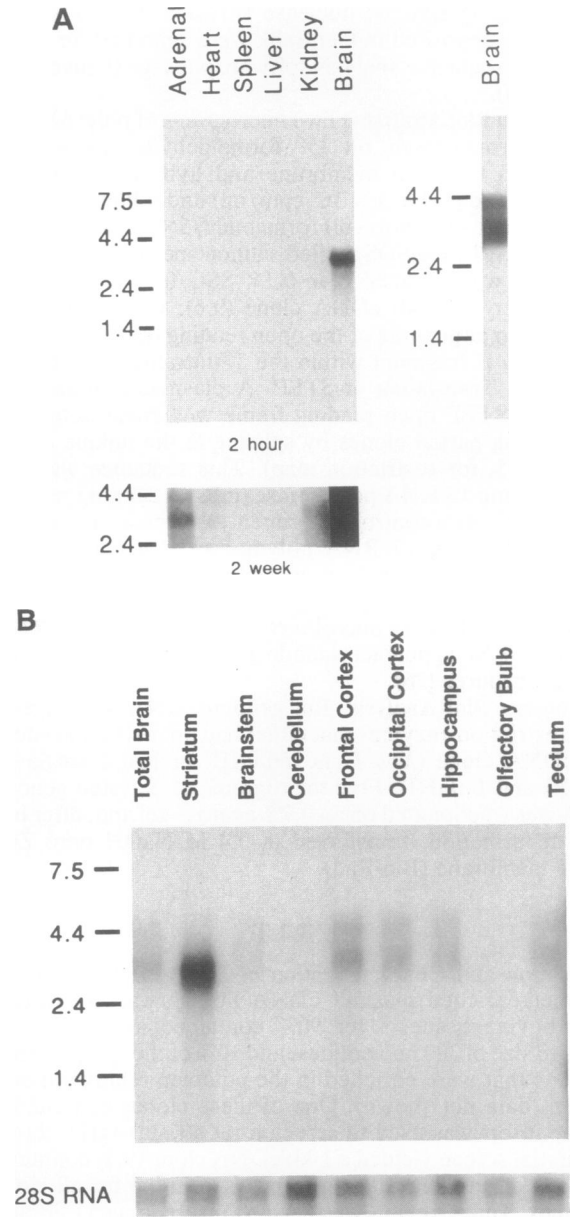

FIG. 4. Analysis of STEP mRNA abundance. (*A*) Northern blot of poly(A) RNA isolated from various rat organs probed with a 2.4-kb fragment spanning the entire open reading frame (clone 9.6). Autoradiograms were exposed as indicated. The blot at right was probed with a 400-bp 3′ fragment and exposed for 24 hr. (*B*) Northern blot of poly(A) RNA isolated from various rat brain regions and probed with 400 bp of 3′ fragment. Autoradiogram was exposed for 24 hr.

adrenal, kidney, and heart. The ratio of band intensities demonstrates that STEP mRNA is much more abundant in brain relative to peripheral tissues.

The distribution of STEP mRNA within the central nervous system is shown in Fig. 4*b*. The 3-kb species is highly expressed in the striatum. In contrast, the 4.4-kb species appears to be expressed selectively outside the striatum. The 3-kb mRNA is also present in other brain regions but at lower levels of expression. As expected from the initial subtraction strategy, neither RNA message is present in the cerebellum.

A Southern blot of rat genomic DNA is shown in Fig. 5. Multiple genomic fragments are present with two restriction enzymes (*Hin*dIII and *Xba* I), which do not cut the STEP cDNA, and with *Bam*HI, which has a single internal site (see Fig. 5 for a restriction map). The STEP gene thus consists of several exons distributed over at least 15 kb of genomic DNA.

DISCUSSION

The identification of STEP as a PTPase is based on its high homology to the core amino acid sequence found in other PTPases isolated to date, but specific phosphatase activity has not yet been demonstrated. Although a number of other PTPases are expressed both in peripheral tissues and in brain, STEP is the only one to be highly expressed in the nervous system. Northern analyses reveal that the intracellular phosphatases PTPase 1 and T-cell PTPase are expressed in brain. PTPase 1 mRNA shows a higher level of expression in lymphoid tissue (spleen) relative to brain and is present in liver and kidney (5). T-cell PTPase is more abundant in spleen, thymus, and placenta relative to brain (13). Several PTPases with transmembrane and extracellular domains have also been isolated from brain cDNA libraries, although these proteins are expressed by a variety of tissues (9, 10). In addition, biochemical fractionation reveals seven distinct types of PTPase activity, although none have been shown to be brain specific (28).

Our results demonstrate that STEP expression is restricted primarily to the nervous system. The small amounts of transcripts present in peripheral tissues may actually be localized to postganglionic autonomic neurons present in peripheral tissues, and immunocytochemical staining will be used in future studies to clarify this issue. Northern analysis of brain poly(A)$^+$ RNA reveals two distinct signals at ≈3 kb and 4.4 kb. The smaller band represents the mRNA that encodes STEP. This mRNA is present in the striatum and, in lesser amounts, in the cortex and midbrain regions. In contrast, the 4.4-kb message is present in all brain areas in which the lower band is found except for the striatum, where it is not visible. Neither band is present in the cerebellum; this observation is not surprising because cerebellar mRNA was used as driver in the original subtractive hybridization procedures. The absence of STEP within the cerebellum also suggests that the cerebellum may express different or unique phosphatase(s). This hypothesis is further supported by the findings that the rat brain contains high levels of tyrosine-specific protein kinase activity in a regional distribution. Interestingly, the cerebellum demonstrated one of the highest level of protein-tyrosine kinase activity (29), further suggesting the possibility of cerebellum-specific PTPases.

Several lines of evidence strongly suggest that the 3-kb and 4.4-kb poly(A)$^+$ species are alternatively processed RNA transcripts derived from a single gene. (*i*) The same pattern of bands was obtained on a Northern (RNA) blot when probes were used that were a nearly full-length cDNA clone or one encompassing a region from the 3′ noncoding region. (*ii*) This observation is also supported by the results shown in Fig. 4, in which the relative abundance of the two species of RNA on Northern blots is approximately equal when probed with cDNA sequence complementary to the 3′ region. In contrast, the 3-kb RNA species appeared more abundant when the blots were probed with the nearly full-length STEP cDNA (Fig. 4*a*). (*iii*) In addition, Southern blots revealed two bands when rat genomic DNA was digested with *Eco*RI and probed with the nearly full-length cDNA clone. This pattern is consistent with the single *Eco*RI restriction site that exists within the open reading frame of the 3-kb message. More detailed direct characterization of the gene structure is necessary to define the precise relationship between the 3-kb and 4.4-kb species and to rule out the alternative possibility that adjacent, duplicated genes encode the two mRNAs.

The striatal STEP transcript generated an *in vitro* translation product with a 46-kDa apparent molecular mass. A 369-amino acid protein is predicted if the first methionine were used as the initiator codon. Kozak consensus rules predict, however, that the third methionine of the open reading frame should be a better initiator codon, and this would produce a protein of 329 amino acids. The resulting polypeptide would have a potential N-terminal myristoylation site similar to those described for other proteins (30). In this situation, the initiating methionine residue is removed, and the α amino group of the N-terminal glycine is modified by the covalent attachment of a myristoyl group. A similar modification has been described for a number of other proteins, including the tyrosine kinases pp60$^{v\text{-}src}$ (31) and the

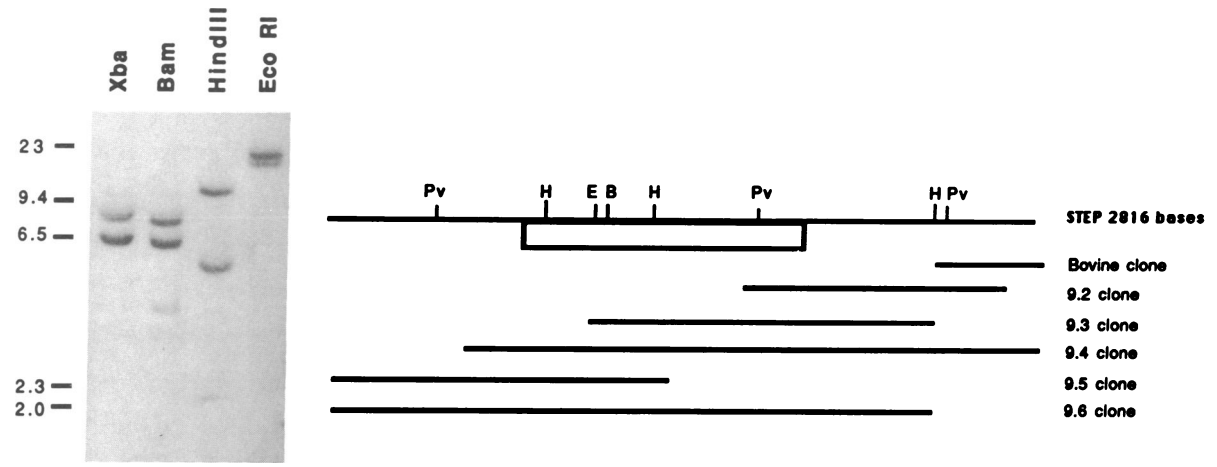

FIG. 5. Restriction enzyme analysis of STEP cDNAs and genomic Southern analysis. (*Left*) Genomic Southern blot of rat genomic DNA digested with the indicated enzymes and probed with the 9.6 clone. E, *Eco*RI; B, *Bam*HI; H, *Hin*fI; Pv, *Pvu* II. (*Right*) Restriction map of striatal STEP mRNA (box indicates location of open reading frame) and distribution of the individual isolated cDNA clones.

serine/threonine protein phosphatase, calcineurin (32). Although myristoylation could be the mechanism for membrane attachment of STEP, this hypothesis will need to be tested by the addition of radiolabeled myristoyl groups in the appropriate translation system. Polyisoprenylation or palmitoylation at the C terminus of the cytoplasmic PTPases has been reported as an alternative mechanism (5). Both mechanisms would allow for the association of these enzymes with their substrates through the preexisting assembly of component parts in the form of complexes. A similar complex has been suggested for the epidermal growth factor receptor, a tyrosine kinase that forms a complex with one of its substrates, phospholipase C-γ (33).

In conclusion, STEP is a protein containing sequence with a high degree of homology to the core phosphatase region found within other PTPases and which shows enrichment within the brain and a high degree of specificity within the striatum. A number of phosphoproteins are known to be enriched within the striatum (34), including the dopamine and cAMP-regulated phosphoprotein DARPP-32 (35). DARPP-32 itself is probably not a substrate for STEP, as it is phosphorylated on serine and threonine residues (34). The identification of STEP will now potentially allow us to determine whether there are other striatal-specific phosphoproteins that might serve as substrate(s) for STEP. It seems likely that future studies will reveal other PTPases with patterns of nervous system expression distinct from STEP. This degree of PTPase heterogeneity in the central nervous system could provide an important mechanism for the regional control of events, such as cell division and differentiation.

We thank Saumyendar Sarkar for his assistance with some experiments, Janice Naegele and Randy Blakely for assistance with rat brain dissections, Susan Staurovsky for technical assistance, and Susan Amara and Janice Naegele for comments on the manuscript. This research was supported, in part, by an Alcohol, Drug Abuse, and Mental Health Administration Scientist Development Award for Clinicians MH00856-02 (P.J.L.) and the Howard Hughes Medical Institute (M.L.).

1. Hunter, T. & Cooper, J. A. (1985) *Annu. Rev. Biochem.* **54**, 897–930.
2. Alexander, D. (1990) *New Biologist* **2**, 1049–1062.
3. Charbonneau, H., Tonks, N. K., Walsh, K. A. & Fischer, E. H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7182–7186.
4. Tonks, N. K., Charbonneau, H., Diltz, C. D., Fischer, E. H. & Walsh, K. A. (1988) *Biochemistry* **27**, 8695–8701.
5. Guan, K., Haun, R. S., Watson, S. J., Geahlen, R. L. & Dixon, J. E. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1501–1505.
6. Streuli, M., Krueger, N. X., Hall, L. R., Schlossman, S. F. & Saito, H. (1988) *J. Exp. Med.* **168**, 1553–1562.
7. Streuli, M., Krueger, N. X., Tsai, A. Y. M. & Saito, H. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8698–8702.
8. Matthews, R. J., Cahir, E. D. & Thomas, M. L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4444–4448.
9. Sap, J., D'Eustachio, P., Givol, D. & Schlessinger, J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6112–6116.
10. Kaplan, R., Morse, B., Huebner, K., Croce, C., Howk, R., Ravera, M., Ricca, G., Jaye, M. & Schlessinger, J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 7000–7004.
11. Jirik, F. R., Janzen, N. M., Melhado, I. G. & Harder, K. W. (1990) *FEBS Lett.* **273**, 239–242.
12. Charbonneau, H., Tonks, N. K., Kumar, S., Diltz, C., Harrylock, M., Cool, D., Krebs, E. G., Fischer, E. H. & Walsh, K. A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5252–5256.
13. Cool, D. E., Tonks, N. K., Charbonneau, H., Walsh, K. A., Fischer, E. H. & Krebs, E. G. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5257–5261.
14. Barclay, A. N., Jackson, D. I., Willis, A. C. & Williams, A. F. (1987) *EMBO J.* **6**, 1259–1264.
15. MacDonald, R. J., Swift, G. H., Przybyla, A. E. & Chirgwin, J. M. (1987) *Methods Enzymol.* **152**, 219–227.
16. Glisin, V., Crkvenjakov, R. & Byus, C. (1974) *Biochemistry* **13**, 2633–2637.
17. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
18. Krug, M. S. & Berger, S. L. (1987) *Methods Enzymol.* **152**, 316–325.
19. Alt, F., Kellem, R., Bertino, J. & Schimke, R. (1978) *J. Biol. Chem.* **252**, 1357–1370.
20. Maniatis, T., Fritisch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
21. Gubler, U. (1987) *Methods Enzymol.* **152**, 325–329.
22. Eschenfeldt, W. H., Puskas, R. S. & Berger, S. L. (1987) *Methods Enzymol.* **152**, 337–342.
23. Feinberg, A. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267.
24. Miller, H. (1987) *Methods Enzymol.* **152**, 145–170.
25. Sanger, F. & Coulson, A. (1975) *J. Mol. Biol.* **94**, 444–448.
26. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
27. Kozak, M. (1986) *Cell* **44**, 283–292.
28. Jones, S. W., Erikson, R. L., Ingebritsen, V. M. & Ingebritsen, T. S. (1989) *J. Biol. Chem.* **264**, 7747–7753.
29. Hirano, A., Greengard, P. & Huganir, R. L. (1988) *J. Neurochem.* **50**, 1447–1455.
30. Towler, D. A., Gordon, J. I., Adams, S. P. & Glaser, L. (1988) *Annu. Rev. Biochem.* **57**, 69–99.
31. Garber, E. A., Krueger, J. G., Hanafusa, H. & Goldberg, A. R. (1983) *Nature (London)* **302**, 161–163.
32. Aitken, A., Cohen, P., Santikarn, S., Williams, D. H., Calder, A. G., Smith, A. & Klee, C. B. (1982) *FEBS Lett.* **150**, 314–318.
33. Margolis, B., Bellot, F., Honegger, A. M., Ullrich, A., Schlessinger, J. & Zilberstein, A. (1990) *Mol. Cell. Biol.* **10**, 435–441.
34. Hemmings, H. C., Nairn, A. C., McGuinness, T. L., Huganir, R. L. & Greengard, P. (1989) *FASEB J.* **3**, 1583–1592.
35. Hemmings, H. C., Jr., Nairn, A. C., Aswad, D. W. & Greengard, P. (1984) *J. Neurosci.* **4**, 99–110.