Isolation and Characterization of a Human Dual Specificity Protein-Tyrosine Phosphatase Gene*

(Received for publication, August 17, 1993, and in revised form, September 17, 1993)

Seung P. Kwak, David J. Hakes, Karen J. Martell, and Jack E. Dixon†
From the Department of Biological Chemistry, The University of Michigan, Ann Arbor, Michigan 48109-0606

Vaccinia phosphatase VH-1 and its mammalian counterparts, including protein-tyrosine phosphatases (PTPase) CL100 and VHR, constitute a novel subfamily of protein-tyrosine phosphatases that exhibits dual substrate specificity for phosphotyrosine- and phosphoserine/threonine-containing substrates. The expression of human VH-1-like PTPase CL100 is rapidly inducible by mitogen stimulation and oxidative stress, suggesting that the gene is transcriptionally regulated. In order to study the mechanism underlying this transcriptional regulation, we isolated the first human gene of this subfamily, the CL100 gene, and characterized its promoter. The gene consists of four exons intervened by three short introns 400-500 base pairs in length. Analysis of the protein sequence encoded by each exon revealed that there is a second region of similarity between CL100 protein and cdc25 in addition to the PTPase catalytic domain. Promoter analysis of the CL100 gene indicates that an 800-base pair region flanking the transcriptional initiation site is sufficient to confer a transcriptional response to serum and 12-O-tetradecanoylphorbol-13-acetate stimulation. The CL100 gene is expressed in numerous tissues, including nonmitotic cells in the brain. Within the brain, CL100 mRNA is localized in discrete neuronal populations, suggesting that this PTPase is likely to play a role in neurotransmission as well as in mitotic signaling. Finally, although extracellular signal-regulated kinase has recently been shown to act as substrate for CL100 in vitro, we find no clear correspondence between the expression of extracellular signal-regulated kinase and CL100 mRNA in the brain. The potential significance of a second cdc25 homology domain of CL100 is discussed.

The vaccinia virus H1 (VH-1) gene encodes a protein phosphatase that exhibits dual specificity for phosphoserine- and phosphotyrosine-containing substrates (1). Other viruses of the pox family also express this class of phosphatase (2), all of which contain the common peptide motif, HCXAGXXR, within the catalytic domain. This motif is considered the signature sequence for protein-tyrosine phosphatases (PTPases),1 a family of enzymes that exhibit an absolute substrate requirement for phosphotyrosines. Thus, the VH-1-like phosphatases constitute an emerging subfamily of PTPases that possess dual catalytic functions.

Three human PTPases of the VH-1 subfamily have been isolated to date (3-5). The protein structure of two of these VH-1-like PTPases are similar. Proteins encoded by PAC-1 and CL100 mRNAs (pPAC-1 and pCL100, respectively) are similar in size, possess the catalytic domain at the C terminus, and have an extended N terminus. The third human VH-1-like PTPase, pVHR, is a smaller protein that lacks the long N terminus and resembles the viral VH-1 PTPase in structure (3). Additional characteristics shared between PAC-1 and CL100 are the rapid transcriptional response to mitogen stimulation and rapid degradation of their respective mRNAs (5-8). PAC-1 becomes translocated into the nucleus after stimulation (5), whereas the intracellular localization of pCL100 is presently unknown. These observations suggest that the activity of some members of this PTPase subfamily is regulated at the transcriptional level and additionally by intracellular translocation.

An interesting feature of this novel subfamily of PTPases is the ability to dephosphorylate phosphotyrosine and phosphoserine/threonine residues. Both VH-1 and pVHR PTPases exhibit this dual specificity toward artificial substrates such as casein and myelin basic protein (1-3), whereas pCL100 catalyzes phosphotyrosine and phosphothreonine residues of extracellular signal-regulated kinase (ERK1) in vitro (9). The only other eukaryotic PTPase shown to possess dual substrate specificity is cdc25. Protein cdc25a is involved in the cell cycle and activates the cdc2-cyclin B complex by dephosphorylating residues Thr14 and Tyr15 of cdc2 (10-14). However, the sequence homology between cdc25s and the VH-1-like PTPase is low within the catalytic domain, suggesting that they constitute two separate subclasses of dual specificity phosphatases (15, 16).

The existence of a novel class of PTPase whose expression appears to be induced by a variety of mitogenic signals allows one to speculate on the roles VH-1-like PTPases play during cell division. It is suggested from the transient and immediate early time course of mRNA expression that trans-acting and cis-acting elements provide exquisite control over the transcriptional activity of these PTPase genes. However, the gene structure of this relatively novel class of PTPases is not known. We therefore characterized the first gene of this subfamily of PTPases, the CL100 gene, and partially defined the promoter region necessary for expression of this gene in HeLa cells. We subsequently determined the tissue distribution of CL100 mRNA and investigated the potential relationship between O-tetradecanoylphorbol-13-acetate; PCR, polymerase chain reaction; CRE, cAMP-responsive element; CREB, cAMP-responsive element-binding protein; PDGF, platelet-derived growth factor; EGF, epidermal growth factor; FGF, fibroblast growth factor.

* This work was supported in part by funding from the Walter Cancer Institute, Indianapolis, IN (to S. P. K. and J. E. D.), Grant MO1 RR00042 awarded to GCRC, University of Michigan, and Grant 18024 from the National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 16 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed.

1 The abbreviations used are: PTPase, protein-tyrosine phosphatase; ERK, extracellular signal-regulated kinase; nt, nucleotide(s); TPA, 12-

Vol. 269, No. 5, Issue of February 4, pp. 3596-3604, 1994
Printed in U.S.A.
Characterization of PTPase CL100 Gene

CL100 and its potential substrate, ERK1, by ascertaining their anatomical localization within the rat brain.

MATERIAL AND METHODS

Genomic Library Screening—Specific primers 24 nt in length coding for the N terminus peptide sequence PVEILPFL (5′GGTGGAGAGCCCTTTUC3′) and complementary to the C-terminal peptide sequence PYSVPVHS (5′GGTGGAGAGGACGCTG3′) of pCL100 were used to amplify a 560-bp CL100 cDNA fragment from a human placental cDNA library (Stratagene, La Jolla, CA). The amplified fragment containing the catalytic region of pCL100 was subsequently labeled with random primers in the presence of [32P]dCTP to be used as a probe to screen the genomic library. The genomic library was constructed from human leukocyte DNA cloned into EMBl3 (Promega, Madison, WI). Palo Alto, CA). Filters were prehybridized at 42 °C in 5× SSPE (750 mm NaCl, 5.5 mm EDTA, 48.7 mm NaPO4), 30% formamide, 5 × Denhardt's, 100 μg/ml sheared sperm DNA, and 0.1% SDS for 2 h, then further hybridized overnight after addition of the labeled fragment (2 × 106 cpm) into the prehybridization solution. Filters were washed in 1 × SSC at 45 °C, 55 °C an finally at 60 °C for 1 h, then the filters were exposed on film (Kodak X-Omat AR) for 2 days with intensifying screens. Positive plaques were identified, plaque-purified, and the inserts were excised, melanolized, and sequenced using Sequence V2.0 (Stratagene).

PTPase Sequences and Alignments—Sequences encoded by exon 1 (amino acids 1–122), exon 2 (amino acids 123–237), exon 3 (amino acids 172–244), and exon 4 (amino acids 245–367) were used in a search of the SwissPro protein data base using the FASTA command in GCG (version 7.0, Genetics Computer Group, Madison, WI). Primer Extension Studies—Two oligonucleotides 17 nt in length complementory to nucleotide positions 210–227 (primer 210) and 234–251 (primer 234) on the cDNA were used as primers in a reverse transcription reaction (see Fig. 2). Primers (100 ng) were labeled with 50 μCi of [γ-32P]ATP (5000 Ci/mmol) and T4 Polynucleotide kinase (Life Technologies, Inc.). Poly(A)-purified RNA (1 μg) was extracted from HeLa cells as described (18) and annealed with a labeled primer (10 ng) at 75 °C for 3 min. The reaction was cooled slowly to room temperature, then incubated at 37 °C in the presence of 200 units of avian myeloblastosis virus reverse transcriptase and 20 units of RNasin (Promega, Madison, WI) for 30 min. The reaction was organically extracted, precipitated with 0.3 M NaAc (pH 5.0), and 2.5 volumes of EtOH, resuspended in 100 μl of TE (pH 8.0), and 5 μl of EtOH, resuspended 50% formamide (5 μl), then electrophoresed on 8% acrylamide–urea gel in parallel with a sequencing reaction of the CL100 gene primed by primer 210. The gel was dried and exposed to film for 2 days with intensifying screens.

HeLa Cell Transfection—Human cervical carcinoma (HeLa) cells were seeded on a 35-mm culture plate and grown in Dulbecco's modified Eagle's medium (Life Technologies, Inc.) with 10% fetal calf serum. Once the cells were grown to 70% confluence, plasmid DNA (5 μg) was transfected using Lipofectin according to the manufacturer's instructions (Life Technologies, Inc.). After 6 h of transfection, the cells were washed in Dulbecco's phosphate-buffered saline, then allowed to recover in serum-free Dulbecco's modified Eagle's medium. Two days following transfection, the cells were either stimulated with 20% fetal calf serum or with specific growth factors. PDGF (AB heterodimeric subunit), EGF, and bovine FGF (Life Technologies, Inc.) were used at final concentrations of 50 ng/ml, 1 μg/ml, and 100 ng/ml, respectively. TPA (Calbiochem) was used at a final concentration of 30 nM. All transfections were conducted on either duplicate or triplicate plates in order to minimize experimental variation.

At various times following growth factor stimulation, cells were scraped in Dulbecco's phosphate-buffered saline, pelleted, then washed once in Dulbecco's phosphate-buffered saline before lysis. The luciferase assay was performed in the presence of 1 mM ATP and luciferin (Sigma) as described previously (19). For each cell extract, three aliquots were prepared for TCA from duplicate reactions. A plasmid containing the CMV promoter fused to the β-galactosidase gene was co-transfected in all assays such that the β-galactosidase activity could be used as an internal standard to correct for differences in transfection efficiency. Thus, luciferase activities were expressed as relative light units per mg of protein (RLU/mg). The colorimetric assay for β-galactosidase activity was performed as described (20).

RNA Blot Analysis—Slot blot analysis was performed by denaturing 10 μg of total RNA extracted from HeLa cells, then transferring onto a Nittran membrane (Schleicher & Schuell) via vacuum suction. A human multiple tissue Northern blot was obtained commercially (Clontech). Both blots were hybridized overnight at 45 °C with the PCR-amplified fragment of CL100 in 50% formamide, 4 × SSC, 5 × Denhardt's, 10 μg/ml sheared salmon sperm DNA, and subsequently washed in a final condition of 0.1 × SSC, 0.1% SDS at 60 °C. Filters were exposed on film for 1 h at –70 °C with intensifying screens. Data obtained from Northern analysis were reported elsewhere (21).

In Situ Hybridization—A partial cDNA fragment of rat CL100 was cloned by PCR using a pair of oligonucleotides 22 nt in length that corresponded to the human cDNA sequence at nucleotide position 92–114 (5′GCATCTCTTGGAGAGAACCC3′) and complementary to position 1165–1187 (5′AGCCAGCTGGCCATGAAGCTGA3′). A fragment of expected size (250 bp) was observed from a rat cortex cDNA library (a generous gift from Dr. M. J. Brownstein, National Institutes of Health, Bethesda, MD, by Dr. M. J. Brownstein, National Institutes of Health). After 35 rounds of amplification (94–55–72 °C per cycle). Sequencing the amplified DNA fragment revealed that its proposed sequence matched that of mouse pCL100 (3CH134) perfectly over 84 amino acid residues and differed from that of the human pCL100 by 2 residues. At the nucleic acid level, the rat clone exhibited 93% identity to mouse and human CL100, respectively. Thus, this DNA fragment was considered to be the rat homologue of human CL100 and was subsequently used to synthesize riboprobes for in situ hybridization.

Results

Screening 800,000 plaques of the human genomic library with the amplified cDNA of CL100 at moderate stringency yielded 21 positive signals. Upon further characterization of these clones, the positive plaques were identified as overlapping or sister clones of four PTPase genes. The four genes included our cognate gene CL100, PAC-1, and two novel genes that were most similar to the VH-1 subclass of PTPases (Table 1), suggesting that a large family of VH-1-like PTPases exist. Further details on these genes and their chromosomal localization are reported elsewhere (22).

Characterization of the CL100 Gene—In order to characterize the first mammalian gene in this family of PTPases, one clone with a 16-kb insert containing the entire CL100 gene was analyzed in detail. The CL100 gene consisted of four exons that were separated by relatively short introns of 400–500 bp (Fig. 1A). The gene was approximately 3.8 kb from the putative transcription initiation site to the end of exon 4. Our genomic sequence and characterization of this gene are discussed below (23).

Table 1. Mapping of VH-1-like PTPases on the human genome

| VH-1-like PTPase clones | Number of sequences | Chromosomal assignment: |
|------------------------|-------------------|-------------------------|
| CL100                  | 2                 | 5q35                    |
| PAC-1                  | 5                 | 7p22                    |
| Novel PTPase hVH-3     | 5                 | 10q25                   |
| Novel PTPase hVH-4     | 2                 | 10q11                   |
| Uncharacterized clones | 7                 |                         |

* K. J. Martell, S. P. Kwak, D. J. Hakes, J. E. Dixon, and J. Trent, manuscript in preparation.

** K. J. Martell, S. P. Kwak, D. J. Hakes, J. E. Dixon, and J. Trent, manuscript in preparation.
Characterization of PTPase CL100 Gene

Fig. 1. A, CL100 gene structure. The exons are depicted by rectangular boxes, and the splice sites are indicated by grey lines. Translational start and stop sites are indicated by the arrows. The PTPase domain on exon 4 is colored red. Two segments of the CH2 domain (noncatalytic region homologous to cdc25) are drawn in blue. B, alignment of the CH2 domain of CL100 and the cdc25 family of proteins, including: Drosophila cdc25 homologue (String), human cdc25a (Hcdc25a), Saccharomyces pombe cdc25 (Pcdc25), Caenorhabditis elegans cdc25 (Cecdc25), and mitosis initiation protein of Saccharomyces cereviiae (Mih). The N-terminal region of pCL100 schematized above the sequences was searched against GenBank using the command FASTA, and the aligned regions were displayed by command PILEUP (Genetics Computing Group, Madison, WI). The conserved residues between the noncatalytic cdc25 and CL100 N terminus are shown in blue. Identical residue motifs include DCRXXXXXGHI and LXXGGYXSF. C, alignment of the PTPase domains of cdc25 and CL100. Conserved residues are in red, identical residues within this region are limited to the motif HCXXXXXR.

clone contained an additional 6.5 kb of 5' flanking sequence and 5.7 kb downstream of exon 4. Exon 1 also contained the 5' untranslated sequence (5'-UT) and the putative translational start site. The sequence encoding the active site cysteine for PTPase activity was found on exon 4. This exon also contained 663 nt of 3' UT, the mRNA destabilization signal (23), and the putative polyadenylation signal.

Identification of a Second cdc25 Homologous Domain on CL100—Upon characterization of several other genes isolated from the genomic screen, we observed that the intron-exon boundaries were conserved among some genes of this class (data not shown). Thus, we subsequently asked whether the exons of the CL100 gene were segregated such that they encoded separate functional domains of the protein. The amino acid sequences encoded by each of the 4 exons of the CL100 gene were used separately to search a protein data base. Searches with both exon 1 and exon 2 yielded a significant alignment with the cdc25 phosphatase. This was unexpected as neither exon 1 nor exon 2 contained the catalytic residues of CL100. Upon alignment of the protein sequence encoded by exon 1 and exon 2 with various members of the cdc25 family, the catalytic regions of all proteins were preferentially aligned (red boxes in Fig. 1B). On the cdc25 protein, these regions flank the catalytic domain (red box, Fig. 1B). However, there were no residues within the N terminus of pCL100 that resembled the catalytic domain of phosphatases. In fact, the catalytic domain of pCL100 was situated on the C terminus in a manner such that when the entire protein sequence of CL100 was aligned with members of the cdc25 family, the catalytic regions of all proteins were preferentially aligned (red box, Fig. 1C). Thus, it appeared that pCL100 possessed two domains with sequence similarity to cdc25. The first, at the C terminus, contained the active site motif common to all PTPases, and the second domain, which we termed CH2 (for the cdc25 homology domain 2) resided at the N terminus and exhibited two segments of similarity to the region surrounding cdc25 active site. These
two domains were located on different exons, with the CH2 domain residing on exons 1–2 and the catalytic domain on exon 4 of the CLlOO gene (see Fig. 1A).

Transcriptional Start Site of CLlOO Gene—The nucleotide sequence of the CLlOO gene is shown in Fig. 2. The intron splice junctions were sequenced for each exon, and an additional 1 kb of 5'-flanking sequence was characterized. All splice junctions conformed to the donor and acceptor consensus. The ATG consensus that confers mRNA instability is repeated near the translation termination site (Fig. 2). The sequence immediately upstream from the TATA box consensus to position -150 was extremely high in GC content. The transcriptional start site was identified by primer extension analysis using primers complementary to nucleotide positions 210 and 234 of the cDNA (Fig. 3A). A control reaction consisting of RNA harvested from unstimulated HeLa cells low in CLlOO mRNA was run on adjacent lanes. Several minor bands arising from nonspecific priming were observed in the control lane using primer 210. Primer extension of CLlOO from TPA-stimulated cells using primer 210 yielded a unique band terminating near residues T and C, 20–21 nt downstream of the TATA box. Reverse transcription using primer 234 appropriately yielded a band 22–25 nt longer than the first reaction, confirming that transcription of this gene initiates primarily from a single site 19–21 residues downstream of the TATA consensus.

Promoter Analysis—The region upstream of the transcriptional start site was sequenced to identify potential cis elements which may be involved in the regulation of CLlOO gene expression. Analysis of DNA sequences revealed several poten-
characterization of PTPase CL100 Gene

A primer extension study. Two primers 24 nt apart (primers 210 and 234) were used in primer extension reactions and electrophoresed alongside a sequencing reaction generated from primer 210. HeLa cells were either serum-deprived (control) or subsequently treated with TPA (30 nM)/cyclohexamide (10 μg/ml) mixture (TPA) for 3 h to induce CL100 mRNA. The approximate positions of the transcription start site are dotted. A band 23–25 nt longer is synthesized using primer 234, confirming that transcription initiation occurs primarily from this region.

B, the 5′-flanking region of CL100 gene was searched for enhancer elements proximal to the promoter region (Fig. 3B). These included two CREs at positions −163 and −118 bp from the transcription start site (24), three AP2 sites at positions −400, −76, −13 (25), three Sp1 site at positions −219, −123, −104 (26), one AP1 site at position −321 (27), and a NF1-E3 site at −365 (28).

Stimulation with calf serum increased the expression of CL100 mRNA rapidly and transiently in HeLa cells. Consistent with the previous data obtained from the mouse homologue 3CH134 in Swiss 3T3 cells (7, 8), the induction of CL100 mRNA was maximal at 30 min and decreased afterwards such that mRNA levels were back to baseline by 3 h (Fig. 4A). Cyclohexamide produced a prolonged rise in CL100 mRNA content as reported previously for 3CH134 mRNA (7, 8). Thus, human CL100 gene and murine 3CH134 are similarly activated in an immediate-early manner and are independent of novel protein synthesis.

Since these observations indicate that a region on the CL100 gene confers responsiveness to mitogen stimulation, we transfected HeLa cells transiently with gene fusion constructs to partially characterize this segment of the CL100 promoter. Three fusion constructs were made with varying amounts of the CL100 5′-flanking sequence and luciferase reporter gene (Fig. 4B). The 1.6-kb XbaI fragment spanning −1500 nt to +100 nt of exon 1 responded to serum stimulation in HeLa cells (Fig. 4C). A shorter 0.8-kb construct (−703 XhoI site to +100 of exon 1) exhibited a similar response to serum, although both

Fig. 3. A, primer extension study. Two primers 24 nt apart (primers 210 and 234) were used in primer extension reactions and electrophoresed alongside a sequencing reaction generated from primer 210. HeLa cells were either serum-deprived (control) or subsequently treated with TPA (30 nM)/cyclohexamide (10 μg/ml) mixture (TPA) for 3 h to induce CL100 mRNA. The approximate positions of the transcription start site are dotted. A band 23–25 nt longer is synthesized using primer 234, confirming that transcription initiation occurs primarily from this region.

Fig. 4. A, CL100 mRNA levels increase in response to serum stimulation in HeLa cells. Cells were serum-deprived for 48 h, then stimulated with 20% fetal calf serum for the indicated lengths of time (open circles). Co-incubation with cyclohexamide (10 μg/ml) potentiates the rise and stabilizes mRNA content during serum stimulation (closed circles). Total RNA harvested from HeLa cells were analyzed by a Northern blot. CL100 mRNA signal on film was digitized and quantitated by optical densitometry. B, fusion gene constructs were made in plasmid pLUC PS. The CL100 fragments containing 5′-flanking DNA and exon 1 were subcloned in either direction and transfected into HeLa cells to test for promoter activity. C, luciferase activity arising from these three fusion gene constructs was assayed at the indicated times after stimulation with 20% fetal calf serum (n = 2–3, mean ± S.E.) and normalized to β-galactosidase activity (A420) as described under "Materials and Methods." All groups were serum-deprived for 48 h prior to stimulation.

Fig. 3. A, primer extension study. Two primers 24 nt apart (primers 210 and 234) were used in primer extension reactions and electrophoresed alongside a sequencing reaction generated from primer 210. HeLa cells were either serum-deprived (control) or subsequently treated with TPA (30 nM)/cyclohexamide (10 μg/ml) mixture (TPA) for 3 h to induce CL100 mRNA. The approximate positions of the transcription start site are dotted. A band 23–25 nt longer is synthesized using primer 234, confirming that transcription initiation occurs primarily from this region. B, the 5′-flanking region of CL100 gene was searched for enhancer elements proximal to the promoter region (Fig. 3B). These included two CREs at positions −163 and −118 bp from the transcription start site (24), three AP2 sites at positions −400, −76, −13 (25), three Sp1 site at positions −219, −123, −104 (26), one AP1 site at position −321 (27), and a NF1-E3 site at −365 (28).

Stimulation with calf serum increased the expression of CL100 mRNA rapidly and transiently in HeLa cells. Consistent with the previous data obtained from the mouse homologue 3CH134 in Swiss 3T3 cells (7, 8), the induction of CL100 mRNA was maximal at 30 min and decreased afterwards such that mRNA levels were back to baseline by 3 h (Fig. 4A). Cyclohexamide produced a prolonged rise in CL100 mRNA content as reported previously for 3CH134 mRNA (7, 8). Thus, human CL100 gene and murine 3CH134 are similarly activated in an immediate-early manner and are independent of novel protein synthesis.

Since these observations indicate that a region on the CL100 gene confers responsiveness to mitogen stimulation, we trans-
Characterization of PTPase CL100 Gene

Transcription of CL100 and PAC-1 genes appear to be induced rapidly during cell stimulation. Since this feature may be commonly exhibited by the mammalian VH-1-like subfamily of PTPases, characterization of a VH-1-like PTPase gene and its promoter activity may help us understand how this class of

found in heart and skeletal muscle. Lower levels of gene expression was observed in brain and kidney upon longer exposure (data not shown).

Since the transcriptional activity of CL100 gene appears to be modulated by mitotic factors, we asked whether the expression of CL100 was restricted to dividing cells or whether it was expressed in terminally differentiated neurons as well. We initially confirmed the expression in the CNS by cloning a rat homologue of CL100 from an adult rat brain library using primers based on human cDNA sequence (see "Materials and Methods" for further detail). Subsequent in situ hybridization analysis using the amplified rat CL100 cDNA on rat brain sections revealed that mature neurons express CL100 mRNA in discrete regions (Fig. 7). High levels of expression was found in the cingulate gyrus within the retrosplenial cortex, ventral, and medial divisions of the anterior thalamus and the medial geniculate nucleus (Fig. 7, A-C). Parietal and temporal cortex also expressed moderate levels of this transcript. Nuclei associated with the limbic circuit, including the anteroventral and the anteromedial thalamic nuclei, the cingulate gyrus, and the hippocampus were intensely labeled. Analysis at higher resolution was performed from emulsion-dipped slides to provide evidence that the neurons in the suprachiasmatic nucleus, the primary circadian rhythm generator in most mammals, and the medial aspect of the cingulate gyrus expressed CL100 mRNA (Fig. 7, D and E). A section was counterstained with cresyl violet to identify silver grains on the lightly stained neurons (black arrows) but absent on smaller and darker glial cells (Fig. 7F).

pCL100 has recently been shown to dephosphorylate ERK1 in vitro, raising the possibility that ERK1 may be the endogenous substrate for pCL100. We asked whether these two proteins are associated in vivo by determining the distribution pattern of CL100 and ERK1 mRNAs within the brain. Hybridization of adjacent sections with CL100 and ERK1 antisense riboprobes revealed that ERK1 mRNA was expressed in all areas of the brain (Fig. 8A). We found no exclusive correspondence between CL100 and ERK1 distribution in the rat brain, although CL100 mRNA was expressed in a subset of areas that were positive for ERK1.

FIG. 6. Human multiple tissue northern blot. The blot (Clontech, CA) containing 2 μg of polyA mRNA isolated from indicated tissues was screened with the amplified CL100 cDNA fragment. Molecular weight markers are indicated to the left.

DISCUSSION

The distribution pattern of CL100 mRNA in human tissues was determined by screening a Northern blot with the human CL100 cDNA fragment. The expression of CL100 mRNA was widespread among peripheral tissues (Fig. 6). High levels of mRNA were detected in lung, liver, placenta, and pancreas, whereas moderate levels were
Characterization of PTPase CL100 Gene

**Fig. 7. In situ hybridization analysis of CL100 mRNA in the rat brain.** Rat CL100 cDNA cloned by PCR was used to synthesize a $^{35}$S-labeled riboprobe. A-C, coronal sections of rostral, middle, and caudal rat brain in low magnification. Analysis of emulsion-dipped sections at higher magnification reveals mRNA signal in the suprachiasmatic nucleus (D) and from single cells in the cingulate cortex (E). It is apparent after counter-staining with cresyl violet that grains surround the lightly stained neurons (arrows in F) but not the darker glial cells. Nuclei abbreviations: cg, cingulate cortex; par, parietal cortex; pir, piriform cortex; cl, claustrum, cpu, caudate putamen; te, temporal cortex; avT, anteroventral thalamus; amT, antero medial thalamus; rsg, retrosplenial granular cortex; MGd, dorsal medial geniculate, MGv, ventral medial geniculate; dg, dentate gyrus; ca1 and ca2, hippocampal subfields; Sch, suprachiasmatic nucleus; OX, optic chiasm; 3v, third ventricle.

**Fig. 8. Comparison of adjacent rat brain sections processed with ERK1 (A) and CL100 riboprobes (B).** ERK1 signal is present throughout the entire section and strongest in the hippocampus and the piriform cortex. CL100 signal is high in nonoverlapping regions, with clear absence of signal in other areas.

PTPases are regulated. To this end, we isolated the first human PTPase gene of this subfamily, the CL100 gene. The catalytic domain of pCL100 was present on the C terminus of the protein and encoded by the fourth exon. The PTPase motif encoded within this region aligned to the catalytic domains of other PTPases, including cdc25.

Interestingly, we observe a second region of sequence similarity to cdc25, the CH2 domain, that is encoded by exons 1 and 2 of the CL100 gene. This domain aligns well with the regions surrounding the catalytic domain of cdc25. We suspect that this N-terminal alignment was not detected previously, since the computer program preferentially aligned the PTPase motif of CL100 with the PTPase domain of cdc25. The alignment of the CH2 domain in pCL100 to cdc25 is interrupted into two segments by the existence of the cdc25 active site. The first segment is encoded on exon 1 of the CL100 gene, whereas the
second segment spans the intron splice junction of exons 1 and 2. Thus, the amino acid residues Lys and Glu of the conserved LKGGY motif are separated on different exons. It is conceivable that these 2 residues are conserved by virtue of the intron splicing mechanism that requires a specific donor and acceptor sequences on exons 1 and 2, respectively.

The CH2-like sequence is also present at the N terminus of PAC-1 cDNA, suggesting that the CH2 domain is shared by several members of this PTPase subfamily. It is noteworthy that pVHR, a short VH-1-like protein isolated from human fibroblast, lacks the CH2 domain. Perhaps this clone defines a subclass of VH-1-like PTPases that lacks the N-terminal cdc25 homologous domain and may therefore possess functions that are distinct from CH2-containing PTPases.

The functional significance of the CH2 domain of pCL100 is presently unclear. The region of homology corresponds to an amino acid stretch flanking the active site of cdc25, a cell cycle protein that dephosphorylates cdc2 at juxtaposed residues Thr14 and Tyr15 (13, 29, 30). It is conceivable that the CH2 domain of PTPase CL100 functions to specify substrates for catalysis. Although cdc25 is a bona fide dual specificity phosphatase, the CH2 domain is not likely to confer dual substrate specificity on pCL100, since the vacuina phosphatase VH-1 lacks this domain but still retains its ability to catalyze phosphotyrosines and phosphoserines. Perhaps the cdc25 homology domain acts to refine the substrate recognition process. The PTPases that possess this region may therefore be more "selective" for their substrates, whereas those that lack it may exhibit activity across a broader spectrum of substrates. This may perhaps explain why pCL100 and its murine homologue exhibit low catalytic activity against artificial substrates (4, 31). In contrast, pCL100 readily exhibits tyrosine phosphatase activity on extended cyclic nucleotide-regulated kinase (ERK) (31) and dual phosphate activity on ERK when phosphorylated by ERK kinase (MEK) (9).

Alternatively, this region of pCL100 may be analogous to the noncatalytic src homology domains (SH2 and SH3) in that they are involved in the recruitment of other proteins (32, 33). Association with other proteins may permit nuclear translocation of this class of PTPases or even regulate its half-life. Several PTPases are known to possess a second domain that specifies nuclear or cytoskeletal localization (34, 35). Whether the cdc25 homology region of VH-1-like PTPases serves as a separate functional domain remains to be determined in future studies.

We determined that 800 bp of 5'-flanking sequence contain the elements necessary for response to serum stimulation. The transcriptional response from the 1.6-kb construct was consistently lower during serum and TPA stimulation. It is therefore conceivable that a negative regulatory element resides between -1600 and -801. During the initial characterization of mouse CL100 (3CH134), Charles et al. (6) observed that cyclohexamide exaggerates and prolongs the increase of this transcript during serum stimulation, which is consistent with the CL100 gene, suggesting that many like other immediate early transscripts, activation of CL100 gene is relatively independent of protein synthesis. Although our transfection studies suggest that 800 bp flanking the promoter encodes most of the elements necessary for response to mitogens, we do not find an serum response element-like consensus on this gene within -1.5 kb of the promoter. Thus, it is likely that the response to serum stimulation is mediated by a mechanism that involves a transcriptional activator other than the serum response factor or ternary complex factor.

The two CREs situated upstream of the promoter are of interest in this context since the CREB/ATF family of DNA-binding proteins are known to be regulated primarily by a phosphorylation event (see Ref. 36 for review). Consistent with this notion, PAC 1 expression has been shown to be increased upon stimulation with Tax (37), a protein encoded by the X region of HTLV-1 gene that can transactivate through a CRE site (38). Further work is required to determine the involvement of one or both CREs in eliciting transcriptional response to mitogens.

The levels of CL100 mRNA were high in several human tissues, although expression was widespread. Since CL100 expression in Swiss 3T3 cells is inducible by a variety of mitogens (6), it appears likely that this gene product is involved in mitotic signaling. However, we suspect that CL100 is important to transduction under a variety of other conditions. We provide evidence here that terminally differentiated neurons in the brain express CL100 mRNA. The discrete localization of this transcript within the rat brain suggests that only a subpopulation of neurons utilizes pCL100-mediated signaling. Since no obvious neurotransmitter is common to all the regions expressing CL100 mRNA, it is likely that several transmitters/factors can elicit CL100 expression in a region-specific manner.

Recently, reports have indicated that ERK1 may be the endogenous substrate for CL100 (9, 31). ERK1 is abundantly expressed in the brain, multiphosphorylated on serine, tyrosine, and on threonine residues (39, 40), and reported to be dephosphorylated by pCL100 in vitro (9, 31). We examined the relationship between ERK1 and CL100 by in situ hybridization and found that the distribution of ERK1 mRNA in the brain correlates poorly with CL100. ERK1 mRNA expression is widespread in the brain whereas CL100 mRNA is localized in a subset of these areas. The distribution of ERK1 and ERK2 assessed by a Northern blot analysis is consistent with this notion (39), suggesting that ERKs cannot be paired exclusively with CL100. Perhaps other VH-1-like PTPases exist and catalyze ERK in specific areas of the brain or conversely an unobserved substrate for CL100 exists in the brain.

Acknowledgment-We thank Dr. Carolyn Minth for careful critique of the manuscript.

REFERENCES
1. Guan, K. L., Broyles, S. S., and Dixon, J. E. (1991) Nature 350, 359-362
2. Hakes, D. J., Martell, K. J., Zhao, W.-G., Maseung, R. F., Esposito, J. J., and Dixon, J. E. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 4017-4021
3. Ishibashi, T., Tottori, D., Ohan, A., Miki, T., and Arosenios, S. A. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 12170-12174
4. Keynes, S. M., and Emrke, E. A. (1992) Nature 350, 644-647
5. Roban, P. J., Davis, R., Moskaluk, M. K., Kearns, M., Krutzsch, H., Siienbenist, U., and Kelly, K. (1993) Science 256, 1763-1766
6. Charles, C. H., Ahler, A. S., and Lau, L. F. (1992) Oncogene 7, 187-190
7. Leu, L. F., and Nathans, D. (1992) EMBO J. 11, 3146-3151
8. Lau, L. F., and Nathans, D. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 1182-1186
9. Zheng, C. F., and Guan, K. L. (1993) J. Biol. Chem. 268, 16116-16119
10. Russell, P., and Nurse, P. (1996) Cell 85, 145-152
11. Sadhu, K. S., Reed, S. I., Richardson, H., and Russell, P. (1996) Proc. Natl. Acad. Sci. U. S. A. 87, 5139-5143
12. Girard, F., Strausfeld, U., Cavadore, J. C., Russell, F., Fernandez, A., and Lamb, N. J. (1992) J. Cell Biol. 118, 755-764
13. Gautier, J., Solomon, M. J., Boorer, N. R., Bazan, J. F., and Kirshner, M. W. (1991) Cell 71, 197-211
14. Kumagai, A., and Dunaigy, W. G. (1991) Cell 64, 903-914
15. Guan, K. L., Hakes, D. J., Wang, Y., Park, H.-D., Cooper, T. G., and Dixon, J. E. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 12175-12179
16. Lee, M. S., Ong, S. X., Xu, M., Parker, L. F., and Nathans, D. (1992) EMBO J. 11, 3146-3151
17. Pearson, W. R., and Lipman, D. J. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 2444-2448
18. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., and Struhl, K. (1987) Current Protocols in Molecular Biology, J. Wiley & Sons, New York
19. Brasier, A. R., Tate, J., and Habener, J. F. (1989) BioTechniques 7, 1116-1122
20. Maniatis, T., Fratich, E. F., and Sambrook, J. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
21. Marguardt, B., and Stabel, S. (1992) Gene (Amst.) 120, 297-299
22. Watanabe, S., Patel, P., Burke, S., Herman, J., Schaffer, M., and Kwak, S. P. (1988) in Society for Neuroscience Short Course 1 (syllabus) (Sundermam, E., ed.) pp. 4-29, Society for Neuroscience, Washington, D. C.
23. Shaw, G., and Kamel, R. (1986) Cell 48, 659-677
24. Hyman, S. E., Comb, M., Lin, Y. S., Pearlberg, J., Green, M. R., and Goodman, H. M. (1998) Mol. Biol. Cell 8, 4225-5233
Characterization of PTPase CL100 Gene

25. Jones, K. A., and Tjian, R. (1985) Nature 317, 179-182
26. Mitchell, P. J., Wang, C., and Tjian, R. (1987) Nature 321, 529-536
27. Angeli, P., Imagawa, M., Chiu, B, Stein, B., Imbra, R. J., Rahmadosf, H. J., Jonat, C., Herrlich, P., and Karin, M. (1987) Cell 48, 729-739
28. Garcia, J., Wu, F., and Gaynor, R. (1987) Nucleic Acids Res. 15, 8367-8385
29. Strausfeld, U., Lehre, J. C., Pachter, D., Cavare, J. C., Picard, A., Sadhu, K., Russell, P., and Dorsey, M. (1991) Nature 351, 242-245
30. Duro, W. G., and Kuret, A. (1991) Cell 87, 189-196
31. Charles, C. H., Sun, H., Lau, I. F., and Tenks, N. K. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 5220-5226
32. Moran, M. F., Koch, C. A., Anderson, D., Ellis, C., England, L., Martin, G. S., and Pawson, T. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 8622-8626
33. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C., and Pawson, T. (1991) Science 252, 666-674
34. McLaughlin, S., and Dixon, J. R. (1993) J. Biol. Chem. 268, 6839-6842
35. Frangioni, J. V., Reahl, P. H., Shifrin, V., Jost, C. A., and Neel, B. G. (1992) Cell 68, 545-560
36. deGroot, R. F., and Sassone-Corsi, P. (1993) Mol. Endocrinol. 7, 145-153
37. Kelly, K., Davis, P., Misuya, H., Irving, R., Wright, J., Grassman, R., Plekenstein, B., Wano, Y., Greene, W., and Siebenlist, U. (1992) Oncogene 7, 1463-1470
38. Xu, Y. L., Adya, N., Stowers, E., Gao, Q. S., and Giam, C. Z. (1993) J. Biol. Chem. 268, 20235-20242
39. Boulton, T. G., Nye, S. H., Robbins, D. J., Ip, N. Y., Radziejewaski, E., Morgenbesser, S. D., DaPinho, A., Panayotatos, N., Cobb, M. H., and Vangons, G. D. (1991) Cell 65, 663-675
40. Zheng, C. F., and Guan, K.-L. (1993) J. Biol. Chem. 268, 11435-11439