*Proc. Natl. Acad. Sci. USA*
Vol. 89, pp. 2980–2984, April 1992
Biochemistry

# Cloning and expression of a cytosolic megakaryocyte protein-tyrosine-phosphatase with sequence homology to retinaldehyde-binding protein and yeast SEC14p

(platelets/signal transduction)

MINXIANG GU, ILKA WARSHAWSKY, AND PHILIP W. MAJERUS*

Division of Hematology-Oncology, Washington University School of Medicine, 660 South Euclid, Box 8125, St. Louis, MO 63110

*Contributed by Philip W. Majerus, January 2, 1992*

**ABSTRACT** **Protein tyrosine phosphorylation is important in the regulation of cell growth, the cell cycle, and malignant transformation. We have cloned a cDNA that encodes a cytosolic protein-tyrosine-phosphatase (PTPase), MEG2, from MEG-01 cell and human umbilical vein endothelial cell cDNA libraries. The 4-kilobase cDNA sequence of PTPase MEG2 corresponds in length to the mRNA transcript detected by Northern blotting. The predicted open reading frame encodes a 68-kDa protein composed of 593 amino acids and has no apparent signal or transmembrane sequences, suggesting that it is a cytosolic protein. The C-terminal region has a PTPase catalytic domain that has 30–40% amino acid identity to other known PTPases. The N-terminal region has 254 amino acids that are 28% identical to cellular retinaldehyde-binding protein and 24% identical to yeast SEC14p, a protein that has phosphatidylinositol transfer activity and is required for protein secretion through the Golgi complex in yeast. Recombinant PTPase MEG2 expressed in *Escherichia coli* possesses PTPase activity. PTPase MEG2 mRNA was detected in 12 cell lines tested, which suggests that this phosphatase is widely expressed. The structure of PTPase MEG2 implies that a tyrosine phosphatase could participate in the transfer of hydrophobic ligands or in functions of the Golgi apparatus.**

Protein tyrosine phosphorylation plays a role in cell growth, differentiation, and oncogenesis (1, 2). It is also involved in other cell functions such as neuronal activation (3), platelet activation (4–6), and insulin signaling (7). Protein tyrosine phosphorylation is balanced by the antagonistic effects of protein-tyrosine kinases (PTKases) and protein-tyrosine-phosphatases (PTPases). PTPases are a unique class of protein phosphatases, the physiological functions of which are beginning to be elucidated (8, 9). For example, the PTPase CD45 is essential for T-cell activation (10, 11). Phosphotyrosyl turnover is required for insulin-mediated signal transduction in 3T3-L1 cells (12). Another PTPase, cdc25, controls the cell cycle by dephosphorylating and thereby activating maturation-promoting factor (13). PTPases may be anti-oncogenic, since NIH 3T3 cells overexpressing PTPase 1B are resistant to transformation by the *neu* oncogene (14). Density-dependent growth arrest of fibroblasts is associated with an increase in membrane tyrosine phosphatase activity (15) and baby hamster kidney cells overexpressing a C-terminally truncated T-cell PTPase grow at a reduced rate (16).

The known PTPases can be divided into two classes: transmembrane PTPases and cytosolic PTPases. The former are divided into five subgroups based on their extracellular structure and intracellular phosphatase domain(s) (17). There

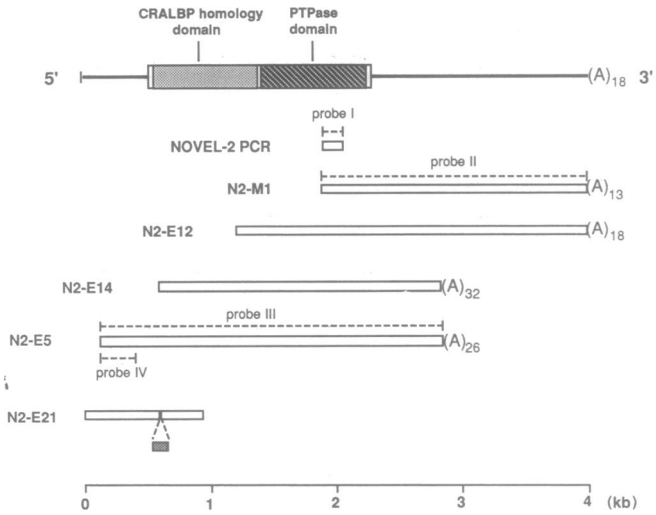


FIG. 1. Predicted protein structure of PTPase MEG2, its relationship to the PCR product and the five cDNA isolates, and the probes designed for screening. Thin segments indicate noncoding sequences and thick segment indicates the open reading frame encoded by PTPase MEG2 cDNA. Relative positions of the CRALBP homology domain and the PTPase domain are indicated. Open bars show the portion of the sequence contained in each of the cDNA isolates, and dotted lines indicate the probes. The interruption in clone N2-E21 and the stippled box below it indicate the insertion position and the inserted sequence. Clone N2-M1 is from the MEG-01 cDNA library, and clones N2-E5, N2-E12, N2-E14, and N2-E21 are from the HUVEC cDNA library. kb, Kilobases.

are two types of cytosolic PTPases wherein the phosphatase domain is found either at the N terminus or at the C terminus. The phosphatase domains of PTPase 1B and T-cell PTPase are located at the N terminus (18, 19). The C termini of these phosphatases determine their cellular localization and their activity (16, 19). The other known cytosolic phosphatases have C-terminal phosphatase domains with N-terminal sequences that have been found to, or predicted to, interact with specific proteins. PTPase MEG and PTPH1 have homology to the cytoskeleton-associated proteins band 4.1, ezrin, and talin (20, 21). PTP1C has two adjacent copies of the Src homology region 2, which forms high-affinity complexes with activated growth factor receptors and other phosphotyrosine-containing proteins (ref. 22; R. J. Mathews, D. B. Bowne, E. Flores, and M. L. Thomas, personal communication).

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; CRALBP, cellular retinaldehyde-binding protein; HUVEC, human umbilical vein endothelial cell; AML, acute myelogenous leukemia; RCM, reduced, carboxamidomethylated, maleylated.

*To whom reprint requests should be addressed.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. M83738).

We previously obtained two PTPase cDNA sequences by using a PCR strategy (20). We used the "novel 1" cDNA sequence to clone and express PTPase MEG. Here we report that using the "novel 2" sequence, we have cloned and expressed another PTPase, MEG2, that shows homology in the N terminus to human cellular retinaldehyde-binding protein (CRALBP) and yeast SEC14p.†

## MATERIALS AND METHODS

Materials were obtained as described (20). Cultured cell lines were from the Washington University Tissue Culture Support Center. The serine/threonine protein phosphatase assay system was from GIBCO/BRL.

**Isolation of PTPase MEG2 cDNA.** A "novel 2" PCR product probe was made as described (20) by using a reduced concentration of dNTPs in the presence of [$\alpha$-$^{32}$P]dCTP and [$\alpha$-$^{32}$P]dATP. The 159-base-pair (bp) radiolabeled probe was used to screen $10^6$ recombinants of the MEG-01 cDNA library ($\lambda$ZAP) (20) at 52°C in 6× standard saline citrate (SSC)/5× Denhardt's solution/0.5% SDS containing salmon sperm DNA at 100 $\mu$g/ml. Subsequent screening of a human umbilical vein endothelial cell (HUVEC) cDNA library (5 × $10^6$ recombinants) (23) was carried out with various random hexamer-primed radiolabeled cDNA probes at 42°C in 50% (vol/vol) formamide/5× SSC/5× Denhardt's solution/0.1% SDS containing salmon sperm DNA at 100 $\mu$g/ml. The candidate clones were subcloned into pBluescript II SK(+) (Stratagene) and sequenced on both strands by the dideoxy chain-termination method. Several persistent band compressions were resolved by using 7-deaza-dGTP and performing electrophoresis in 6% (wt/vol) polyacrylamide gels containing 7M urea and 30% (vol/vol) formamide (24). Primer extension was done as described (25).

**Northern Blot Analysis.** Total RNA was isolated (26) from a variety of cultured cell lines and from leukocytes of a patient with acute myelogenous leukemia (AML). Poly(A)$^+$ RNA was selected with oligo(dT)-cellulose. Five micrograms of each poly(A)$^+$ RNA was fractionated in a 1% agarose/formaldehyde gel and transferred to nitrocellulose. Random hexamer-radiolabeled cDNAs from PTPase MEG2 and human actin were used as probes. Hybridization was performed at 42°C in 50% formamide/5× SSC/5× Denhardt's solution/25 mM potassium phosphate, pH 7.4/10% dextran sulfate containing salmon sperm DNA at 50 $\mu$g/ml. The nitrocellulose was washed twice at 25°C in 2× SSC/0.1% SDS for 15 min and then twice in 0.25× SSC/0.1% SDS at 37°C for 15 min.

**Expression of PTPase in *Escherichia coli*.** The entire coding region with 92 bp of 3′ untranslated sequence was excised

```
GGCTGAGGCAGCCGCGCAGGTCGCAGGGCCAGCGTCCGCGAGGACGGCCCGGCTGGGCGGCCCCGAGCTCTGTGGCGCTGTGGAGGAGCGGGAGCGCGGC   100
CGAGGAAGCGGGGCGCCGAGGGGGTCGGCGGCCTTCGGGAAATTTCCGCCGACCCTTCGCTCCCGGCTCTAAAAGTTCCTGATTTCCTATTTCCTTTTAA   200
ATCCCGAGTGGCTGTTAGCTCTTCGCCTGCACTTTTTCTTCCCCAGGAGATAAGGGGGAGTGTGAGGAACGGAGCGAATAATATAAAAAAGGATTTCCTC   300
CCGGAAGAGAGCGGCAGTTCGGAGAGATTTTTCTTAAGGAAGCAGAAGCGGCGTTTGCGGCCGCTGCAGGCGCCGGGCCCTGCCGGCCACACTATGCGCG   400
AGCCGGCCCCGGGCTGCTGAGGCGCGGGGACGCGGAAGCGGAGGCCGAGCGCGCCGGGCTCCCGCGCTCGCGAGCGAGTTTTGTCCGCCCGGCGGCGGTG   500
GCGGGGGGATGGAGCCCGCGACCGCGCCCCGGCCCGACATGGCGCCGGAGCTGACCCCGGAGGAGGAGCAGGCTACCAAGCAGTTTCTCGAAGAGATTAA   600
  1          M  E  P  A  T  A  P  R  P  D  M  A  P  E  L  T  P  E  E  E  Q  A  T  K  Q  F  L  E  E  I  N
CAAGTGGACAGTTCAGTACAATGTTTCCCCGCTGTCTTGGAATGTGGCTGTCAAGTTCCTCATGGCAAGGAAGTTTGATGTGCTCCGTGCCATAGAATTG   700
 32   K  W  T  V  Q  Y  N  V  S  P  L  S  W  N  V  A  V  K  F  L  M  A  R  K  F  D  V  L  R  A  I  E  L
TTCCACTCCTACAGAGAAACTCGAAGGAAGGAAGGCATTGTAAAGCTGAAACCTCATGAGGAACCTCTTCGTTCTGAGATCCTCAGTGGAAAATTCACCA   800
 65   F  H  S  Y  R  E  T  R  R  K  E  G  I  V  K  L  K  P  H  E  E  P  L  R  S  E  I  L  S  G  K  F  T
TCTTAAATGTTCGGGACCCAACAGGAGCCTCCATTGCCCTCTTTACTGCCAGGTTGCATCATCCCCACAAGTCAGTCCAACATGTGGTACTTCAGGCTCT   900
 98   I  L  N  V  R  D  P  T  G  A  S  I  A  L  F  T  A  R  L  H  H  P  H  K  S  V  Q  H  V  V  L  Q  A  L
GTTTTACTTGCTAGACAGAGCTGTGGATAGCTTTGAAACTCAGAGGAATGGACTGGTGTTTATCTATGACATGTGTGGTTCTAATTATGCCAACTTTGAG  1000
132   F  Y  L  L  D  R  A  V  D  S  F  E  T  Q  R  N  G  L  V  F  I  Y  D  M  C  G  S  N  Y  A  N  F  E
CTGGATCTTGGCAAGAAAGTCCTAAACCTGCTGAAGGGAGCATTTCCAGCTCGTTTGAAGAAGGTGCTGATTGTGGGGGCACCCATATGGTTCCGAGTGC  1100
165   L  D  L  G  K  K  V  L  N  L  L  K  G  A  F  P  A  R  L  K  K  V  L  I  V  G  A  P  I  W  F  R  V
CCTATTCCATCATCAGTCTCCTCCTGAAGGACAAAGTCCGGGAGAGGATTCAAATATTAAAGACATCTGAGGTCACGCAGCATCTGCCCAGGGAGTGTCT  1200
198   P  Y  S  I  I  S  L  L  L  K  D  K  V  R  E  R  I  Q  I  L  K  T  S  E  V  T  Q  H  L  P  R  E  C  L
TCCAGAAAACCTGGGTGGGTACGTCAAAATTGATCTCGCCACTTGGAATTTCCAGTTCCTACCCCAGGTGAACGGCCACCCAGATCCCTTCGATGAGATC  1300
232   P  E  N  L  G  G  Y  V  K  I  D  L  A  T  W  N  F  Q  F  L  P  Q  V  N  G  H  P  D  P  F  D  E  I
ATCCTGTTCTCCCTCCCTCCTGCCTTAGACTGGGACTCAGTACATGTTCCAGGTCCCCATGCTATGACCATCCAAGAGTTGGTGGACTATGTTAATGCCA  1400
265   I  L  F  S  L  P  P  A  L  D  W  D  S  V  H  V  P  G  P  H  A  M  T  I  Q  E  L  V  D  Y  V  N  A
GGCAAAAGCAAGGAATCTATGAGGAATATGAAGACATTCGTCGTGAGAACCCTGTTGGCACTTTCCACTGTTCCATGTCTCCAGGAAACCTAGAGAAAAA  1500
298   R  Q  K  Q  G  I  Y  E  E  Y  E  D  I  R  R  E  N  P  V  G  T  F  H  C  S  M  S  P  G  N  L  E  K  N
CCGTTATGGGGATGTACCCTGCCTGGACCAAACTAGAGTGAAGCTAACAAAGCGAAGTGGCCATACTCAGACAGATTACATCAATGCCAGTTTCATGGAT  1600
332   R  Y  G  D  V  P  C  L  D  Q  T  R  V  K  L  T  K  R  S  G  H  T  Q  T  D  Y  I  N  A  S  F  M  D
GGCTACAAGCAGAAGAATGCTTACATTGGCACACAAGGTCCTTTGGAGAATACCTATCGTGATTTCTGGCTCATGGTATGGGAGCAAAAAGTCTTGGTGA  1700
365   G  Y  K  Q  K  N  A  Y  I  G  T  Q  G  P  L  E  N  T  Y  R  D  F  W  L  M  V  W  E  Q  K  V  L  V
TTGTCATGACCACCCGCTTTGAGGAAGGCGGCAGGAGAAAGTGTGGCCAGTACTGGCCTTTAGAAAAAGACTCTCGGATCCGATTTGGCTTCCTCACAGT  1800
398   I  V  M  T  T  R  F  E  E  G  G  R  R  K  C  G  Q  Y  W  P  L  E  K  D  S  R  I  R  F  G  F  L  T  V
GACCAATCTAGGCGTGGAGAACATGAATCATTATAAGAAAACAACGCTAGAAATTCACAACACAGAGGAACGGCAGAAACGCCAGGTGACCCACTTCCAG  1900
432   T  N  L  G  V  E  N  M  N  H  Y  K  K  T  T  L  E  I  H  N  T  E  E  R  Q  K  R  Q  V  T  H  F  Q
TTCTTGAGCTGGCCAGACTATGGTGTCCCTTCCTCAGCAGCTTCCCTCATTGACTTCTTGAGAGTGGTCAGAAACCAGCAGAGTCTGGCTGTGAGCAACA  2000
465   F  L  S  W  P  D  Y  G  V  P  S  S  A  A  S  L  I  D  F  L  R  V  V  R  N  Q  Q  S  L  A  V  S  N
TGGGAGCACGCTCCAAAGGGCAGTGCCCTGAGCCACCCATTGTGGTCCATTGCAGTGCAGGCATTGGCAGGACAGGTACCTTCTGCTCACTGGACATCTG  2100
498   M  G  A  R  S  K  G  Q  C  P  E  P  P  I  V  V  H  C  S  A  G  I  G  R  T  G  T  F  C  S  L  D  I  C
CCTGGCACAGCTGGAGGAGCTTGGCACCCTTAATGTGTTCCAGACGGTGTCACGCATGAGGACCCAGAGGGCCTTCAGCATCCAGACCCCTGAGCAGTAC  2200
532   L  A  Q  L  E  E  L  G  T  L  N  V  F  Q  T  V  S  R  M  R  T  Q  R  A  F  S  I  Q  T  P  E  Q  Y
TATTTTTGCTACAAGGCCATCCTGGAGTTCGCAGAGAAGGAGGGCATGGTATCCTCTGGCCAAAACCTGCTGGCCGTGGAGAGTCAGTAACTCTCCTACG  2300
565   Y  F  C  Y  K  A  I  L  E  F  A  E  K  E  G  M  V  S  S  G  Q  N  L  L  A  V  E  S  Q  *
AACCTCCTACCTGTTGGCCAGCCTTCCTTAAACTACCCTGGACACCGCTGAGCCTATAGGTTGCCATCAGTTACGCTGAAGCCATGGATCAACTTCTTTC  2400
TTGTGTCTCCCAGCACACGTGTGCACGCAAGGCAGCCTTTCCCTTTGGCTAGATAAATGTGGTGAAATTGCCACTAGAAAGGGCCAGTAGCAATGTGTTC  2500
TTAATTCCTAGCATCTGCTATCAAACTGTGCCTTATTAAAACCAAATTGTGGCTATTGATTTTTGCTAGGGTCCCCGAGGTAAGTGGGGAAGGAACCTCC  2600
CTCCTCCTTTCTCCCGCCATTCTTTTTCTGAGGTCTCCTGTCTCCCACTCCTTCCCTTTCCTAGGATTTTATGGGGAGGTTTGATGAGCATTTGCCTTTC  2700
TCCCCAGAGAGCTTGACCAAGTTACATATTCTAGAGATCGTCCTGAGTGCCAAGCCAGTTTATAAGTAGGGCGTGTATTCCAGTCCTTTCTGTCATTCTG  2800
TGTGTGTGTGTGTGTGCACTTATGTGTATGGGTCCATATGTACGTGTGTATATAATATATATTGTATGTGATAATATGATCGTATTCTATAAATACATAT  2900
ACACACGATACTCCTTTCCAGAAGTTCCTGGGGCTCCGTGTTGCAGTTCAGGACTCCTTGCTTCTTGGACCCTACTATTTATCCTGGACTAGCTTGGGTT  3000
GGTGATCATGTCTCCTCTTGTCAGGCTACAGAGTGGTGGAAGAGGACACACACACATCTCCTCCATGTTCCTGCCACAGGGCCCCTTCCTTAAGTAATGA  3100
CTTATCTCCTTCCAGTTGCCACCTGTTGGAGCCAAGAATGTATTAGTTTGTGGTGACACTGGGTTAGCTCAGCATGGATTTCCCTTATCCGATGATTTTT  3200
CTACATTTACTTGGCCAATTTGGGGAACAGACCTCCACTGTGATTCCATACTCTCTCTTTGCTTTCACTATCCCAGGGATGGGAGGCGGAGGGTAGAGAT  3300
GAACTCACTGAGCAGCCAGAGCCAAATGCAATCGGTACGAATCTTAGAAGGAAGGAGGGGGAGCCCAGGTATGAGAAAGAAAAAACCACTAAGAAAATAC  3400
CTCCCTGGGAGGATGAGCTGGGGCCCTTTTTCTTTTGCTGGATGGTTCCTTTATGCAGCTTGGCCCTGTCTACCGAGATGCCCATCTCTTCCTGCCTGCT  3500
AGCCTGCTAGACCCTCAAACTGGGTGGGTTCTGTGTCAATAAAAAGCTTCACCCCCTGGCTGAGTGAGGTGGTCCCCTGCAATCACTGTTTGTCCCCTAC  3600
CCACCCAACCTGTCCCTGCCTGCTCCCAGCCCACTCATCCTTATGTGCTAGGGATAAATCAAGAGTCCTCAGCACTCCACATTCCCAAAAAATCCCAGGA  3700
ACTCCTAAACCTTCCCCTGTGACAGAAGATGAGGTTGGCAGCTGATCAGACCTCAATAATTTTATACTGTAATAAGCTCTTTGAATGTGTATATTATTAT  3800
CTTTTACAATCATTTCATTTTGTATTTAATATCAATGGACTGAGTCTGATTCAGAAAATAATTGTAATGTTCATATTCAATATCTTACAGTGATACAGTT  3900
TTGTATTTACATATAAATATACCGACATGATCAAAACCAAAAAAAAAAAAAAAAA                                                 3956
```

FIG. 2. Nucleotide and predicted amino acid sequence of PTPase MEG2. The nucleotide sequence is numbered at right. The amino acid sequence is shown in single-letter code below the nucleotide sequence and is numbered at left. The first two ATG codons are in boldface type. The sequence of the PCR product is noted by a solid underline. The CRALBP homology region and the PTPase homology region are indicated (⌜⌟). Arrowhead indicates the position where the extra 117-bp sequence is inserted.

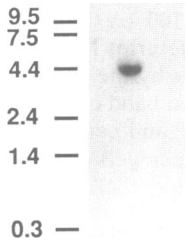


FIG. 3. Detection of PTPase MEG2 mRNA in MEG-01 cell RNA by blot hybridization. Poly(A)$^+$ RNA (10 $\mu$g) from MEG-01 cells was electrophoresed, transferred to nitrocellulose, and hybridized with the cDNA insert of clone N2-M1. The filter was exposed at −70°C for 2 hr with two intensifying screens. The mobility of RNA standards (kb) is shown at left.

from clone N2-E14 and introduced into a bacterial expression vector, pTrp (27). *E. coli* XL1-Blue cells were transformed with this pTrp-PTPase MEG2 plasmid and grown at 37°C for 12 hr. Cells were harvested by centrifugation at 2000 × *g* for 5 min, resuspended in 25 mM Tris, pH 7.5/50 mM NaCl/0.25 mM phenylmethylsulfonyl fluoride, and sonicated at 100 W for 30 sec. The supernatant was separated from the pellet by centrifugation at 13,000 × *g* for 10 min. PTPase activity was measured as described (20). Supernatant protein (20 $\mu$g) was tested for serine or threonine phosphatase activity with the protein phosphatase assay system according to the manufacturer's instructions.

## RESULTS

**Isolation and DNA Sequence Analysis of a PTPase from MEG-01 and HUVEC cDNA Libraries.** The relationship of the major clones obtained from screening MEG-01 and HUVEC cDNA libraries and the various probes used are shown in Fig. 1. Among $10^6$ recombinants from a MEG-01 library screened with the PCR probe and 4 × $10^6$ recombinants from a HUVEC library screened with various partial cDNA probes, 19 positive clones were plaque-purified. The length of the inserts obtained from these clones ranged from 1 to 3 kb. Clones N2-E5 and N2-E14 end at positions 2910 and 2915,

A

```
PTPase MEG2  307  YEDIRRENPVGTFHCSMSPGNLEKNRYGDVPCLDQTRVKLTKRSGHTQTD  356
PTPase MEG   659  FDQLYRKKPGMTMSCAKLPQNISKNRYRDISPYDATRVILKGNE----D  703
PTPase 1B     20  YQDIRHEASDFPCRVAKLPKNKNRNRYRDVSPFDHSRIKLHQED----ND   65
LAR D1      1330  YESIDPG-QQFTWENSNLEVNKPKNRYANVIAYDHSRVILTSIDGVPGSD 1378

PTPase MEG2  357  YINASFMDGYKQKNA----YIGTQGPLENTYRDFWLMVWEQKVLVIVMTT  402
PTPase MEG   704  YINANYINMEIPSSSIINQYIACQGPLPHTCTDFWQMTWEQGSSMVVMLT  753
PTPase 1B     66  YINASLIKMEEAQRS----YILTQGPLPNTCGHFWEMVWEQKSRGVVMLN  111
LAR D1      1379  YINANYIDGYRKQNA----YIATQGPLPETMGDFWRMVWEQRTATVVMMT 1424

PTPase MEG2  403  RFEEGGRRKCGQYWPLEKDSRIRFGF--LTVTNLGVENMNHYKKTTLEIH  450
PTPase MEG   754  TQVERGRVKCHQYWPEPTGSSSYGC---YQVTCHSEEGNTAYIFRKMTLF  800
PTPase 1B    112  RVMEKGSLKCAQYWPQKEEKEMIFEDTNLKLTLISEDIKSYYTVRQLELE  161
LAR D1      1425  RLEEKSRVKCDQYWP-ARGTETCGL---IQVTLLDTVELATYTVRTFALH 1470

PTPase MEG2  451  NTEERQKRQVTHFQFLSWPDYGVPSSAASLIDFLRVVRNQQSLAVSNMGA  500
PTPase MEG   801  NQEKNESRPLTQIQYIAWPDHGVPDDSSDFLDFVCHVRNKRAG-------  843
PTPase 1B    162  NLTTQETREILHFHYTTWPDFGVPESPASFLNFLFKVRESGSL-------  204
LAR D1      1471  KSGSSEKRELRQFQFMAWPDHGVPEYPTPILAFLRRVKACNPL-------1513

PTPase MEG2  501  RSKGQCPDPPIVVHCSAGIGRTGTFCSLDICLAQLEELGTLN-VFQT--V  547
PTPase MEG   844  ------KEEPVVVHCSAGIGRTGVLITMETAMCLIECNQP---VYPLDIV  884
PTPase 1B    205  ----SPEHGPVVVHCSAGIGRSGTFCLADTCLLLMDKRKDPSSVDIKKVL  250
LAR D1      1514  ------DAGPMVVHCSAGVGRTGCFIVIDAMLERMKHEKT---VDIYGHV 1554

PTPase MEG2  548  SRMRTQRAFSIQTPEQYYFCYKAILE                         573
PTPase MEG   885  RTMRDQRAMMIQTPSQYRFVCEAILK                         910
PTPase 1B    251  LEMRKFRMGLIQTADQLRFSYLAVIE                         276
LAR D1      1555  TCMRSQRNYMVQTEDQYVFIHEALLE                        1580
```

B

```
PTPase MEG2    6  APRPDMAPELTPEEEQAT---KQFLEEINKWTVQYNVSPLSWNVAVKFL   51
CRALBP        57  ETREEAVREL-QEMVQAQ---AASGEELAVA-VAERVQEKDSGFFLRFI  100
SEC14p        20  DALPGTPGNLDSAQEKALAELRKLLEDAG------FIERLDDSTLLRFL   62

PTPase MEG2   52  MARKFDVLRAIELFHSYRETRRKE----GIVK-LKPHEEPLRSEILSGK   95
CRALBP       101  RARKFNVGRAYELLRGYVNFRLQY-----PELFDSLSPEAVRCTIEAGY  144
SEC14p        63  RARKFDVQLAKEMFENC-EKWRKDYGTDTILQDFHYDEKPLIAKFYPQY  110

PTPase MEG2   96  FTILNVRDPTGA---SIALFTARLHHPHK-SVQHVVLQALFYLLDRAVD  140
CRALBP       145  PGVLSSRDKYGR---VVMLFNIENWQSQE-ITFDEILQAYCFILEKLLE  189
SEC14p       111  Y---HKTDKDGRPVYFEELGAVNLHEMNKVTSEERMLKNLVWEYESVVQ  156

PTPase MEG2  141  SFETQ---RNGLVFI---YDMCGS-NYANFELDLGKKVLNLLKGA----  178
CRALBP       190  NEETQ---INGFCII ENFKGFTMQ-QAASLRTSDLRKMVDMLQDS----  230
SEC14p       157  -YRLPACSRAAGHLV---ETSCTIMDLKGISISSAYSVMSYVREASYIS  201

PTPase MEG2  179  ---FPARLKKVLIVGAPIWFRVPYSIISLLLKDKVRERIQLLKTS---E  221
CRALBP       231  ---FPARFKAIHFIHQPWYFTTTYNVVKPFLKSKLLERVFVHGDD-LSG  275
SEC14p       202  QNYYPERMGHFYIINAPFGFSTAFRLFKPFLDPVTVSKIFILGSSYQKE  250

PTPase MEG2  222  VTQHLPRECLPENLGGYV-KIDLATWNFQ-FLPQVNGHPDPF         261
CRALBP       276  FYQEIDENILPSDFGGTLPKYDGKAVAEQLFGPQAQAENTAF*        317
SEC14p       251  LLKQIPAENLPVKFGGKS-EVDESKGGL--YLSDIGPWRDPK        289
```

FIG. 4. Alignment of PTPase MEG2 with homologous regions of related proteins. (*A*) C-terminal region of PTPase MEG2 is compared with the PTPase catalytic domains of PTPase MEG (20), PTPase 1B, and the first phosphatase domain of CD45 (29). (*B*) N-terminal region of PTPase MEG2 is compared with the homology region of human CRALBP (30) and yeast SEC14p (31). Gaps (-) were introduced to optimize alignments.

respectively, and both have poly(A)$^+$ tails longer than the oligo(dT) used to prime the cDNA library. Primer extension using an oligonucleotide from the 5′ end of clone N2-E5 as the primer and MEG-01 cell poly(A)$^+$ RNA as the template suggested that the full-length clone contained an additional 120 bp of 5′ sequence (data not shown). A 0.3-kb fragment from the 5′ end of clone N2-E5 was excised, labeled by random hexamer priming, and used to screen $10^6$ recombinants from the HUVEC library. The clone obtained, N2-E21, has overlapping sequences at the 3′ end and contains an additional 100 bp of sequence at the 5′ end. Moreover, it contains 117 bp of additional sequence that is an in-frame insertion between positions 572 and 573. This sequence (5′-TCTTACATGTCCTGTGCTAAATGTCATCTGTCTC-CCTGGAAGTTACCATACGTCAAGGACCAAATGATT-TATGATGTTAAATTGAGGTCTTCTACAGAGACAAA-GAAGGAAGGAAAG-3′) has no termination codons and lacks the typical GT intron splicing signal at the 5′ end (45).

The nucleotide sequence of the composite PTPase MEG2 cDNA (Fig. 2) is 3956 bp long and corresponds to the size of the mRNA transcript detected by Northern blot analysis of MEG-01 cell poly(A)$^+$ RNA (Fig. 3). It has 508 bp of 5′ untranslated sequence that is 68% G+C, and it contains an out-of-frame upstream ATG followed by an 8-amino acid open reading frame. The open reading frame is 1779 bp and is followed by a 3′ untranslated region of 1669 bp. The second ATG, at position 509, has a strong consensus sequence for translation initiation (GGGATGG) (28) with upstream termination codons in all frames. No consensus polyadenylylation signals (AATAAA) was found in the 3′ untranslated region.

The open reading frame of PTPase MEG2 encodes 593 amino acids and predicts a protein product with a molecular weight of 68,019 (Fig. 2). There is no leader sequence or strong hydrophobic stretch of amino acids, suggesting that the protein is cytosolic. A conserved tyrosine phosphatase catalytic domain at amino acids 307–573 is 30–40% identical to known phosphatase domains (Fig. 4*A*). The N terminus of PTPase MEG2 is related to the human CRALBP (Fig. 4*B*). The relation is strongest between residues 8–261 of PTPase MEG2, wherein 70 of 254 amino acids are identical (28%). There is also a relation to the yeast protein SEC14p in this region wherein 60 of 254 amino acids are identical (24%) when the two proteins are compared by the ALIGN program (32). The alignment scores are 7 and 11 standard deviations for yeast SEC14p and CRALBP, respectively; a score of 7 corresponds to a *P* value of $10^{-10}$ and that of 11 is $>10^{-30}$. Yeast SEC14p is homologous to CRALBP (33).

**Bacterial Expression and PTPase Activity.** To confirm that MEG2 encoded a PTPase, the MEG2 cDNA was cloned into the bacterial expression vector pTrp. Homogenates of *E. coli* transformed with pTrp-PTPase MEG2 contain PTPase activity as determined with [$^{32}$P]phosphotyrosine-containing reduced, carboxamidomethylated, maleylated (RCM) lysozyme as a substrate (Fig. 5). The phosphatase activity curve was linear for 15 min and was proportional to protein concentration. This activity was activated by EDTA and inhibited by $Zn^{2+}$, $Ca^{2+}$, and vanadate (data not shown). We did not detect a Coomassie blue-staining protein band upon SDS/PAGE of protein from *E. coli* expressing pTrp-PTPase MEG2. The recombinant protein has no serine or threonine phosphatase activity for the substrate glycogen phosphorylase, which had been phosphorylated by phosphorylase kinase (data not shown).

**PTPase MEG2 mRNA Expression.** PTPase MEG2 mRNA was detected in all cells tested, including hematopoetic cells (MEG-01, HL-60, HEL, leukocytes from a patient with AML), epithelial cells (Hep G2, HeLa, Caco-2, A431), fibroblastic cells (Hs-27, A172), and choriocarcinoma cells (BeWo and JAR) (Fig. 6). JAR cells and the leukocytes of the AML patient clearly contained MEG2 mRNA upon longer exposure than shown in Fig. 6. A172 and BeWo cells showed the highest levels of the MEG2 mRNA after normalization for the amount of β-actin mRNA.

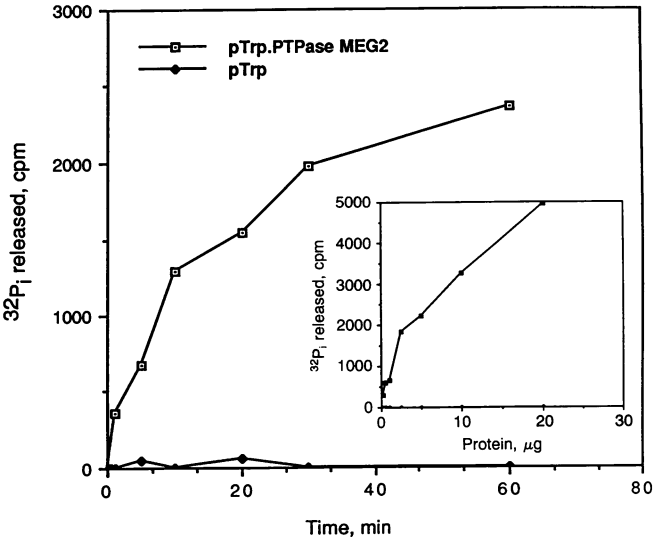


FIG. 5. Activity of PTPase MEG2 expressed in *E. coli*. Time course is shown for dephosphorylation of [$^{32}$P]phosphotyrosine-containing RCM lysozyme (20,000 cpm) incubated with 2 μg of supernatant protein from a homogenate of XL1-Blue cells transformed with either pTrp.PTPase MEG2 or pTrp. (*Inset*) Dephosphorylation of [$^{32}$P]phosphotyrosine-containing RCM lysozyme (20,000 cpm) as a function of the *E. coli* protein concentration. Various amounts of supernatant were incubated with substrate for 10 min.

## DISCUSSION

We have identified a cytosolic PTPase. The results from primer extension and the size of the mRNA transcript detected by Northern blot analysis suggest that the full-length cDNA sequence has been obtained. The nucleotide sequence predicts a 68-kDa protein composed of two domains: a phosphatase catalytic domain at the C terminus and a CRALBP/yeast SEC14p homology domain at the N termi-

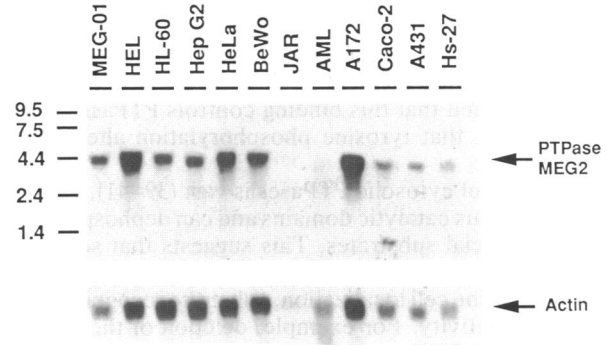


FIG. 6. Detection of PTPase MEG2 mRNA from various cultured human cell lines by Northern blot analysis. Poly(A)$^+$ RNA concentration was measured by $A_{260}$, and each lane was loaded with 5 μg of poly(A)$^+$ RNA. The 3-kb cDNA insert from clone N2-E5 and a 1.6-kb human β-actin cDNA were used to probe the blot separately. The filter hybridized with PTPase MEG2 probe was exposed at −70°C for 4 hr and that with the actin probe for 15 min. RNA sources: MEG-01, megakaryoblastic leukemia; HEL, human erythroleukemia; HL-60, promyelocytic leukemia; Hep G2, liver hepatoma; HeLa, cervical epitheloid carcinoma; BeWo, placental choriocarcinoma; JAR, placental choriocarcinoma; AML, leukocytes from a patient with AML; A172, brain glioblastoma; Caco-2, colon adenocarcinoma; A431, epidermoid carcinoma; Hs-27, foreskin fibroblast. Mobility of RNA standards (kb) is shown at left.

nus. The recombinant protein expressed in *E. coli* possesses PTPase activity. No serine and threonine phosphatase activity was detected by using glycogen phosphorylase *a* as a substrate for the recombinant protein. Northern blot analysis of mRNA from several cell lines indicated that PTPase MEG2 is widely expressed. The distant homology of the PTPase domain to known PTPases, as well as the CRALBP homology, suggests that PTPase MEG2 represents a distinct type of PTPase.

PTPase MEG2 RNA has a long, G+C-rich (68%) 5′ untranslated region that contains an upstream AUG codon. Less than 10% of eukaryotic mRNAs have upstream AUG codons and only 1–2% have 5′ leader sequences longer than 500 nucleotides (28). Most mRNAs that show these two features encode scarce proteins that are critical for cell growth and functions, such as protooncogene products (28, 34).

Clone N2-E21 contains an additional 117 bp of in-frame sequence that inserts 39 amino acids near the N terminus (Fig. 2). This sequence could reflect an alternative exon or a partially spliced intron. Northern blot analysis of 10 $\mu$g of MEG-01 poly(A)$^+$ RNA with the unique insert sequence as a probe detected no mRNA signal, and PCR analysis of MEG-01 cDNA, using primers on either side of the inserted sequence, gave only one PCR product, which corresponded to the sequence without the 39-amino acid insert.

An initial clue to the physiological function of the PTPase MEG2 is provided by the homology of the N terminus to CRALBP and yeast SEC14p. CRALBP is a water-soluble protein found only in the retina and pineal gland. It serves as a carrier protein for 11-*cis*-retinaldehyde or 11-*cis*-retinol and modulates the interaction of these retinoids with visual-cycle enzymes (30, 35, 36). SEC14p is a *Saccharomyces cerevisiae* gene product and is required for protein transport through the Golgi complex. Its function has been determined from study of a temperature-sensitive mutant cell line that lacks SEC14p and from gene disruption experiments (31, 33). SEC14p has also been found to act as a phosphatidylinositol transfer protein, catalyzing the transfer of phosphatidylinositol and phosphatidylcholine between membrane bilayers *in vitro*. Phosphatidylinositol transfer activity may be responsible for the stimulation of protein secretion from the Golgi complex (37, 38). It has been shown that SEC14p and CRALBP are 27% identical (33). PTPase MEG2 is 28% and 24% identical to CRALBP and SEC14p, respectively, in the N-terminal region. Since both CRALBP and SEC14p interact with hydrophobic ligands, we suggest that PTPase MEG2 binds to a lipid moiety and that this binding controls PTPase activity or, conversely, that tyrosine phosphorylation alters ligand binding.

The number of cytosolic PTPases is vast (39–41). PTPases have homologous catalytic domains and can dephosphorylate the same artificial substrates. This suggests that sequences outside of the catalytic domains play a role in regulating PTPases to define cell localization, substrate recognition, and phosphatase activity. For example, deletion of the C terminus from T-cell PTPase activates phosphatase activity and changes its localization and behavior in cells (16). PTPase MEG (20), PTPH1 (21), PTP1C (22), STEP (42), YOP2B (43), VVH1 (44), and cdc25 (12) have long sequences N-terminal to the catalytic domains. The homology of the N-terminal region of PTPase MEG and PTPH1 to protein 4.1 and that of PTP1C to the Src homology 2 domains provide clues to their functions. The finding that PTPase MEG2 has homology to CRALBP and yeast SEC14p provides another example and suggests that the N-terminal region of PTPases is important in defining their function.

We thank John York, A. B. Jefferson, and Theo Ross for helpful suggestions. This research was supported by Grant HL14147 (Specialized Center for Research in Thrombosis), Grant HL16634, and Training Grant HL07088 from the National Institutes of Health.

1. Yarden, Y. (1988) *Annu. Rev. Biochem.* **57**, 443–478.
2. Cantley, L. C., Auger, K. R., Carpenter, C., Duckworth, B., Graziani, S., Kapeller, R. & Soltoff, S. (1991) *Cell* **64**, 281–302.
3. Bothwell, M. (1991) *Cell* **65**, 915–918.
4. Ferrell, J. E., Jr., & Martin, S. (1988) *Mol. Cell. Biol.* **8**, 3603–3610.
5. Golden, A. & Brugge, J. S. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 901–905.
6. Lerea, K. M., Tonks, N. K., Krebs, E. G., Fischer, E. H. & Glomset, J. A. (1989) *Biochemistry* **28**, 9286–9292.
7. Rosen, O. (1987) *Science* **237**, 1452–1458.
8. Banks, P. (1990) *J. NIH Res.* **2**, 62–66.
9. Fischer, E. H., Charbonneau, H. & Tonks, N. K. (1991) *Science* **253**, 401–406.
10. Pingel, J. T. & Thomas, M. L. (1989) *Cell* **58**, 1055–1065.
11. Koretzky, G. A., Picus, J., Thomas, M. L. & Weiss, A. (1990) *Nature (London)* **346**, 66–68.
12. Liao, K., Hoffman, R. D. & Lane, M. D. (1991) *J. Biol. Chem.* **266**, 6544–6551.
13. Gautier, J., Solomon, M. J., Booher, R. N., Bazan, J. F. & Kirschner, M. W. (1991) *Cell* **67**, 197–211.
14. Brown-Shimer, S., Johnson, K. A., Hill, D. E. & Bruskin, A. M. (1992) *Cancer Res.* **52**, 478–482.
15. Pallen, C. J. & Tong, P. H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 6996–7000.
16. Cool, D. E., Tonks, N. K., Charbonneau, H., Fischer, E. H. & Krebs, E. G. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 7280–7284.
17. Saito, H. & Streuli, M. (1991) *Cell Growth Differ.* **2**, 59–65.
18. Chernoff, J., Schievella, A. R., Jost, C. A., Erikson, R. L. & Neel, B. G. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2735–2739.
19. 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.
20. Gu, M., York, J. D., Warshawsky, I. & Majerus, P. W. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 5867–5871.
21. Yang, Q. & Tonks, N. K. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 5949–5953.
22. Shen, S. H., Bastien, L., Posner, B. I. & Chretien, P. (1991) *Nature (London)* **352**, 736–739.
23. Ye, R. D., Wun, T. C. & Sadler, J. E. (1987) *J. Biol. Chem.* **262**, 3718–3725.
24. Martin, R. (1987) *Focus* **9**, 8–10.
25. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
26. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
27. York, J. D. & Majerus, P. W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9548–9552.
28. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8148.
29. Matthews, R. J., Cahir, E. D. & Thomas, M. L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4444–4448.
30. Crabb, J. W., Goldflam, S., Harris, S. E. & Saari, J. C. (1988) *J. Biol. Chem.* **263**, 18688–18692.
31. Bankaitis, V. A., Malehorn, D. E., Emr, S. D. & Greene, R. (1989) *J. Cell Biol.* **108**, 1271–1281.
32. Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91**, 524–545.
33. Salama, S. R., Cleves, A. E., Malehorn, D. E., Whitters, E. A. & Bankaitis, V. A. (1990) *J. Bacteriol.* **172**, 4510–4521.
34. Kozak, M. (1988) *Mol. Cell. Biol.* **8**, 2737–2744.
35. Wolf, G. (1991) *Nutr. Rev.* **49**, 1–12.
36. Crabb, J. W., Gaur, V. P., Garwin, G. G., Marx, S. V., Chapline, C., Johnson, C. M. & Saari, J. C. (1991) *J. Biol. Chem.* **266**, 16674–16683.
37. Aitken, J. F., van Heusden, P. H., Temkin, M. & Dowhan, W. (1990) *J. Biol. Chem.* **265**, 4711–4717.
38. Bankaitis, V. A., Aitken, J. R., Cleves, A. E. & Dowhan, W. (1990) *Nature (London)* **347**, 561–562.
39. Matthews, R. J., Flores, E. & Thomas, M. L. (1991) *Immunogenetics* **33**, 33–41.
40. Krueger, N. X., Streuli, M. & Saito, H. (1990) *EMBO J.* **9**, 3241–3252.
41. Yi, T., Cleveland, J. L. & Ihle, J. N. (1991) *Blood* **78**, 2222–2228.
42. Lombroso, P. J., Murdoch, G. & Lerner, M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7242–7246.
43. Guan, K. & Dixon, J. E. (1990) *Science* **249**, 553–556.
44. Guan, K., Broyles, S. S. & Dixon, J. E. (1991) *Nature (London)* **350**, 359–362.
45. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55**, 1119–1150.