Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 5867–5871, July 1991
Biochemistry

Identification, cloning, and expression of a cytosolic megakaryocyte protein-tyrosine-phosphatase with sequence homology to cytoskeletal protein 4.1

(cell signaling/protein phosphorylation)

MINXIANG GU, JOHN D. YORK, ILKA WARSHAWSKY, AND PHILIP W. MAJERUS*

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

Contributed by Philip W. Majerus, April 5, 1991

ABSTRACT **We have isolated a cDNA encoding a third type of protein-tyrosine-phosphatase. We screened human megakaryoblastic cell line (MEG-01) and umbilical vein endothelial cell cDNA libraries to obtain a 3.7-kilobase cDNA designated PTPase MEG. Northern blot analysis of MEG-01 RNA detected a 3.7-kilobase transcript, suggesting that a full-length cDNA has been identified. PTPase MEG cDNA contains an open reading frame of 926 amino acids. The cDNA has a G+C-rich 5′ untranslated region of 771 nucleotides that has the potential to form stable stem–loop structures and has two upstream ATG codons. The predicted protein (M_r = 105,910) has no apparent membrane-spanning region and contains a single protein-tyrosine-phosphatase domain (amino acids 659–909) that is 35–40% identical to previously described tyrosine-phosphatase domains. The recombinant phosphatase domain possesses protein-tyrosine-phosphatase activity when expressed in *Escherichia coli*. The amino-terminal region (amino acids 31–367) is 45% identical to the amino terminus of human erythrocyte protein 4.1, a cytoskeletal protein. The identification of a protein-tyrosine-phosphatase that is related to cytoskeletal proteins implies that cell signaling activities reside not only in transmembrane receptors but in cytoskeletal elements as well.**

Protein tyrosine phosphorylation is regulated by protein-tyrosine kinase (PTKase) and protein-tyrosine-phosphatase (PTPase) activities. Since PTKases are implicated in cell growth and transformation (1, 2), PTPases must have related functions (3–5). Nonproliferating cells such as blood platelets and neurons contain high levels of protein tyrosine phosphate (6, 7), suggesting that tyrosine phosphorylation may serve roles other than promoting cell growth. For example, thrombin, collagen, and ionophore A23187 induce protein tyrosine phosphorylation in platelets (8–10); addition of vanadate and molybdate to electropermeabilized platelets increases tyrosine phosphorylation and stimulates secretion. These effects may result from inhibition of PTPases (11).

Two types of PTPases have been defined: transmembrane and intracellular. The transmembrane PTPases include leukocyte common antigen (LCA, CD45; ref. 12), LCA-related molecule (LAR; ref. 13), *Drosophila* PTPase (DPTP; ref. 14), several tyrosine phosphatases from human placenta (HPTP; ref. 15), three tyrosine phosphatases from human brain (RPTP; ref. 16), and LCA-related phosphatase (LRP; ref. 17). LAR may function in cell adhesion or in cell–cell interactions, since its extracellular domain is structurally similar to the neural adhesion molecules N-CAM and fasciculin II (13). T-cell activation and proliferation require CD45 expression (18, 19). CD45 also activates a PTKase, pp56^{lck}, by dephosphorylating a putative negative regulatory site (20). Intracellular PTPases include PTPase 1B from human placenta (21, 22), a related rat brain PTPase (23), and T-cell PTPase (24). The biological functions of soluble PTPases are not known, although overexpression of T-cell PTPase in BHK cells reduces cellular protein phosphotyrosine (25).

We find that platelets contain very high levels of PTPase activity. To characterize these enzymes we used a human megakaryoblastic cell line (MEG-01) cDNA as a template for polymerase chain reaction (PCR)-mediated amplification of cDNA sequences between conserved heptapeptide sequences located in the catalytic domains of several PTPases. In this way we identified several known PTPases and two novel cDNA sequences (26). Here we report the sequence of a novel cDNA† that was used to screen MEG-01 cell and human umbilical vein endothelial cell (HUVEC) cDNA libraries. The isolated full-length cDNA clone encodes a third type of PTPase, which contains a region that is similar to erythrocyte protein 4.1.

MATERIALS AND METHODS

Human megakaryoblastic cell line MEG-01 was a gift from Hidehiko Saito (Nagoya University, Nagoya, Japan). The HUVEC cDNA library was kindly provided by Evan Sadler (Washington University). Epidermal growth factor (EGF) receptor was provided by Linda Pike (Washington University). The λZAP cDNA synthesis kit was purchased from Stratagene. Sequenase, *Taq* polymerase, and 7-deaza-2′-deoxyguanosine 5′-triphosphate were from United States Biochemical. ^{32}P-labeled deoxyribonucleotides were from ICN. Deoxyadenosine 5′-[α-[^{35}S]thio]triphosphate was from Amersham. Oligo(dT)-cellulose was from Pharmacia. Restriction enzymes, phage T4 DNA ligase, and Klenow fragment of DNA polymerase I were from Bethesda Research Laboratories or New England Biolabs. Oligonucleotides were synthesized on an Applied Biosystems 380B DNA synthesizer in the Washington University Protein Chemistry Facility.

PCR Amplification and Isolation of cDNA. A unidirectional MEG-01 cell cDNA library was prepared using a λZAP cDNA synthesis kit and 5 μg of poly(A)$^+$ RNA from MEG-01 cells. The resultant library contains 3.6 × 10^6 independent recombinants. Degenerate oligonucleotides A_1, A_2, and B were designed from sequences encoding amino acids conserved among human PTPases (17). Sense oligonucleotides A_1 and A_2 were based on sequences WPDHGVP [5′-TGGCC(A or T)GA(C or T)CA(C or T)GGAGTCCCT-3′] and WPDFGVP [5′-TGGCC(A or T)GA(C or T)TT(C or

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: PTKase, protein-tyrosine kinase; PTPase, protein-tyrosine-phosphatase; LCA, leukocyte common antigen (CD45); LAR, LCA-related molecule; HUVEC, human umbilical vein endothelial cell; EGF, epidermal growth factor; RCM, reduced, carboxamidomethylated, and maleylated.

*To whom reprint requests should be addressed.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. M68941).

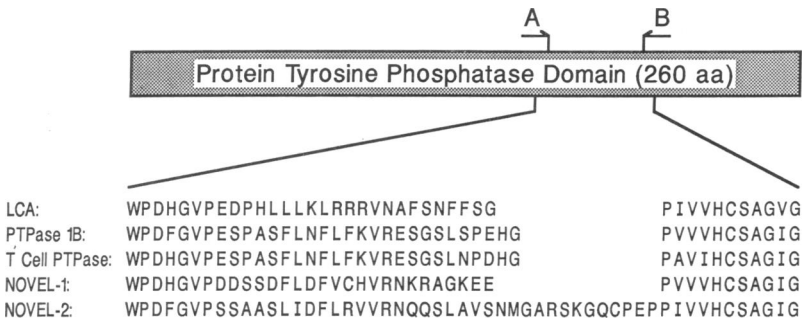

FIG. 1. Diagram of a PTPase domain [260 amino acids (aa)] showing the positions of the oligonucleotides used to prime PCR. The novel sequences of two PCR products from the MEG-01 cDNA library and comparisons with the analogous regions of human PTPase 1B, T-cell PTPase, and LCA are shown.

T)GGAGTCCCT-3']. Antisense oligonucleotide B was based on the sequence HCSAGIG [5'-(C or G)CC(A or G)AT-GCCTGCACT(A or G)CAGTG-3']. A template DNA from the MEG-01 cDNA library (50 ng) was amplified using *Taq* polymerase in the presence of 300 ng of each oligonucleotide as described (27). The PCR cycle, repeated 30 times, consisted of denaturation at 94°C for 1 min, annealing at 50°C for 1 min, and extension at 72°C for 1 min. The PCR products were cloned into the *Sma* I site of pBluescript SK(+) plasmid (Stratagene) (28) and sequenced.

The MEG-01 cDNA library (2×10^6 recombinants) was screened with radiolabeled Novel-1 PCR product, and subsequently it and the HUVEC cDNA library (29) were screened with various radiolabeled partial cDNA clones as described (28). Nucleotide sequence was determined on both strands by the dideoxy chain-termination method. Northern blot analysis was carried out on MEG-01 RNA (28).

Expression of Active PTPase in *Escherichia coli*. The 3' portion of PTPase MEG cDNA that encoded the PTPase domain (bases 1264–2890) was excised from the PTPase MEG-14 clone. The resulting 1.6-kilobase (kb) cDNA fragment was expressed in *E. coli* as described (30). Lysozyme was reduced, carboxamidomethylated, and maleylated (RCM lysozyme) according to Tonks *et al.* (21). RCM lysozyme (2 mg/ml) was then labeled on tyrosine by using [γ-^{32}P]ATP (1 mCi = 37 MBq) and EGF receptor kinase as described (21) except that the reaction mixture did not contain unlabeled ATP. PTPase activity was measured in 50 μl of 10 mM imidazole/HCl (pH 7.2) containing 0.1% (wt/vol) bovine serum albumin and 0.1% (vol/vol) 2-mercaptoethanol. After 10 min at 30°C, 17 μl of 2.5% bovine serum albumin was added and protein was precipitated with trichloroacetic acid. Radioactivity in the supernatant was measured with a liquid scintillation counter.

RESULTS

A PCR strategy was devised to identify different types of PTPases represented in the MEG-01 cell cDNA library. Partially degenerate primers based on sequences that are highly conserved among PTPases but that flank a region of variable length and sequence were used to amplify cDNAs for multiple enzymes in a single reaction (Fig. 1). Using primers A_1, A_2, and B, we observed numerous PCR products and 22 were subcloned and sequenced (Table 1). Sixteen of the clones were identified as previously known PTPases including T-cell PTPase, PTPase 1B, and LCA. The remaining 6 encoded two sequences that we dubbed Novel-1 and Novel-2 (Fig. 1). When we used the Novel-1 cDNA (probe I, Fig. 2) to screen a MEG-01 cDNA library, the clone PTPase MEG-1 was identified. This truncated clone encodes an entire PTPase domain, a short 3' noncoding region, and a poly(A) tail (although no polyadenylylation signal is present).

Table 1. Number of independent cDNA subclones

Apparent PCR product size, base pairs	No. of clones sequenced	No. with identical DNA sequence				
		T-cell PTPase	PTPase 1B	LCA	Novel-1	Novel-2
107	7	3	–	2	2	–
117	6	1	–	4	1	–
123	6	4	2	–	–	–
153	3	–	–	–	–	3
	22	8	2	6	3	3

PCR was carried out using primers A_1, A_2, and B with MEG-01 cDNA template. The products were size-fractionated by agarose gel electrophoresis, eluted, subcloned, and sequenced.

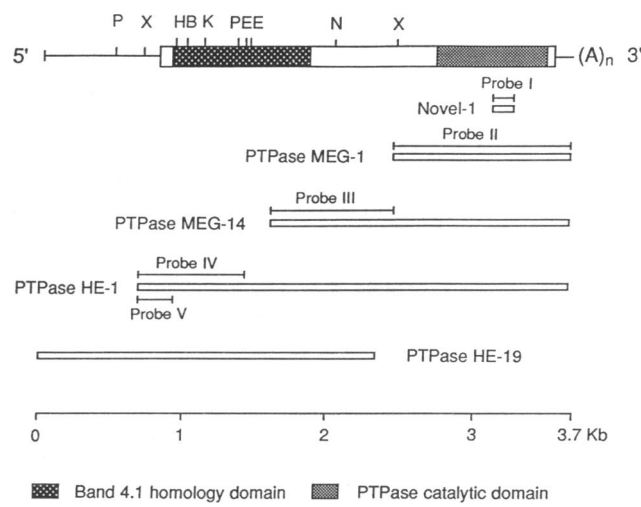

FIG. 2. The relationship of four cDNA isolates, a partial restriction map of the composite full-length PTPase MEG cDNA, and the probes designed for screening. The thin segments indicate noncoding sequences, and the thick segment indicates the open reading frame that encodes PTPase MEG. Selected restriction sites are shown: E, *Eco*RI; B, *Bam*HI; H, *Hin*dIII; K, *Kpn* I; N, *Nde* I; P, *Pst* I; X, *Xho* I. The thin open bars show the portion of the sequence contained in each of the cDNA isolates: PTPase MEG-1 and PTPase MEG-14 (from the MEG-01 library) and PTPase HE-1 and PTPase HE-19 (from the HUVEC library). All clones were sequenced on both strands.

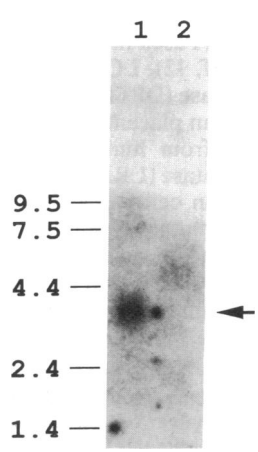

FIG. 3. Detection of PTPase MEG mRNA in MEG-01 cell RNA by blot hybridization. Total (10 μg, lane 2) and poly(A)$^+$ (2 μg, lane 1) RNA samples from MEG-01 cells were electrophoresed in a 2.2 M formaldehyde/1% agarose gel and transferred to a nitrocellulose filter. The cDNA insert of the PTPase MEG-1 clone was used to probe the RNA blot. Mobility of RNA standards (kb) is shown at left.

This clone (probe II) was used to isolate PTPase MEG-14, a 2-kb clone. An 800-base-pair (bp) *Xho* I fragment from the 5′ end of this clone (probe III) was used to screen a HUVEC library, yielding the 3-kb clone PTPase HE-1 (Fig. 2). When we attempted to use probe IV to further screen the HUVEC library, false positives that did not contain a PTPase domain were obtained, so that a combination of probes III and V was used for further screening to isolate PTPase HE-19.

The structure of the composite 3.7-kb cDNA is outlined in Fig. 2. Its length corresponds to the size of the mRNA transcript detected by Northern blot hybridization of MEG-01 poly(A)$^+$ RNA with probes II and III (Fig. 3). The full 3661-bp nucleotide sequence and predicted amino acid sequence of PTPase MEG are shown in Fig. 4. Base pairs −771 to 845 are from HUVEC cDNA clones while base pairs 846 to 2890 are from MEG-01 clones.

The cDNA predicts a protein of 926 amino acids with M_r 105,910. Amino acids 31–367 of PTPase MEG are 45% identical to the amino-terminal region of human erythrocyte cytoskeletal protein 4.1 (Fig. 5). The protein is also related to

```
      CCTGCGTGTCCCTCTGCGCTCCGACTGGTGCGACTTCTCCCTGCGCTAGCGAGGCAGGGTTTTGGCCTCGCCTCTCGCGAGATCGCCTCCTGTTGCTGC   -673
      CGCCGCCGCTCCTGGCCACTGACTGGCGGCGCCTGCGCAGCCGCCATGTTCGGTTGCTATGCTGCGGCCTAGGAGAGGGGGTGTGCTTGAGGGAGGAGG   -574
      AAGAGATAGAGGAGGAGGAGGGGGAGGAAGAGGAGGTGGAGAAGGAGGGGGGTGACTGAGCTCCTCTTGCACTCTCACACACAAACGCTGCCCAGGATT   -475
      ACCCGCCAGCTCACGCCGCGCAGTGCGCTTTTCCGCTCCTCGCGCCCCACCACCAACATTGTTCTCTCAGGACTCCTGGGTCCCAGGGGTCGGAATTGG   -376
      GCCTGAGCGGGAGAGGAAAGAGACTTGGCTTTGGCCGCGGGGTCGGAGGATTGGGGCCAGGCCCCCTCCCCCACGCACTTTTGGGGGTGTGGATTATCT   -277
      CATCCCTGCAGGGAGGTAGGAGAGGTCGCCGGCTGCCCGCCTCCCTGCCACCTCCCCAGCGGCGCCGGCCCGCGGCTGCCCAGCAGCATGAGGTGGTGC   -178
      TGGCGGCTCCGGGTCGTGGCGCGACCGCTGCGGCGGCGGCTGCTCGGGGGGCGCTGAGGTAGCCCCCCGGAGCGGCACGGAGGACGCGCTTCTCCTCTG   -79
      CGCGCCGGGGCCTCGAGGCTTTTTTTCTCCAGCCGAGAGGACGCGGCTGTGATATACGAAGACTTTGTGTGGACAGTAATGACCTCACGTTTCCGATTG   21
  1                                                                                   M  T  S  R  F  R  L
      CCTGCTGGCAGAACCTACAATGTACGAGCATCAGAGTTGGCCCGAGACAGACAGCATACTGAAGTGGTTTGCAACATCCTTCTTCTGGATAACACTGTA   120
  8    P  A  G  R  T  Y  N  V  R  A  S  E  L  A  R  D  R  Q  H  T  E  V  V  C  N  I  L  L  L  D  N  T  V
      CAAGCTTTCAAAGTCAATAAACATGATCAGGGGCAAGTCTTGTTGGATGTCGTCTTCAAGCATCTAGATTTGACTGAGCAGGACTATTTTGGTTTACAG   219
 41    Q  A  F  K  V  N  K  H  D  Q  G  Q  V  L  L  D  V  V  F  K  H  L  D  L  T  E  Q  D  Y  F  G  L  Q
      TTGGCTGATGATTCCACAGATAACCCAAGGTGGCTGGATCCAAACAAACCAATAAGGAAGCAGCTAAAGAGAGGATCTCCTTACAGTTTGAACTTTAGA   318
 74    L  A  D  D  S  T  D  N  P  R  W  L  D  P  N  K  P  I  R  K  Q  L  K  R  G  S  P  Y  S  L  N  F  R
      GTCAAATTTTTTGTAAGTGACCCCAACAAGTTACAAGAAGAATATACAAGGTACCAGTATTTTTTGCAAATTAAACAAGACATTCTTACTGGAAGATTA   417
107    V  K  F  F  V  S  D  P  N  K  L  Q  E  E  Y  T  R  Y  Q  Y  F  L  Q  I  K  Q  D  I  L  T  G  R  L
      CCCTGTCCTTCTAATACTGCTGCCCTTTTAGCTTCATTTGCTGTTCAGTCTGAACTTGGAGACTACGATCAGTCAGAGAACTTGTCAGGCTACCTCTCA   516
140    P  C  P  S  N  T  A  A  L  L  A  S  F  A  V  Q  S  E  L  G  D  Y  D  Q  S  E  N  L  S  G  Y  L  S
      GATTATTCTTTCATTCCTAATCAACCTCAAGATTTTGAAAAAGAAATTGCAAAATTACATCAGCAACACATAGGCTTATCTCCTGCAGAAGCAGAATTT   615
173    D  Y  S  F  I  P  N  Q  P  Q  D  F  E  K  E  I  A  K  L  H  Q  Q  H  I  G  L  S  P  A  E  A  E  F
      AATTACCTAAACACAGCACGTACCTTAGAACTCTATGGAGTTGAATTCCACTATGCAAGGGATCAGAGTAACAATGAAATTATGATTGGAGTGATGTCA   714
206    N  Y  L  N  T  A  R  T  L  E  L  Y  G  V  E  F  H  Y  A  R  D  Q  S  N  N  E  I  M  I  G  V  M  S
      GGAGGAATTCTGATTTATAAGAACAGGGTACGAATGAATACCTTTCCATGGTTGAAGATTGTAAAAATTTCTTTTAAGTGCAAACAGTTTTTTATTCAA   813
239    G  G  I  L  I  Y  K  N  R  V  R  M  N  T  F  P  W  L  K  I  V  K  I  S  F  K  C  K  Q  F  F  I  Q
      CTTAGAAAAGAATTGCATGAATCTAGAGAAACATTATTGGGATTTAATATGGTGAATTACAGAGCATGTAAAAATTTGTGGAAAGCATGTGTAGAACAT   912
273    L  R  K  E  L  H  E  S  R  E  T  L  L  G  F  N  M  V  N  Y  R  A  C  K  N  L  W  K  A  C  V  E  H
      CACACATTCTTCCGTTTGGACAGACCACTTCCACCTCAAAAGAATTTTTTTGCACATTATTTTACATTAGGTTCAAAATTCCGGTACTGTGGGAGAACT   1011
305    H  T  F  F  R  L  D  R  P  L  P  P  Q  K  N  F  F  A  H  Y  F  T  L  G  S  K  F  R  Y  C  G  R  T
      GAAGTCCAATCAGTTCAGTATGGCAAAGAAAAGGCAAATAAAGACAGGGTATTTGCAAGATCCCCAAGTAAGCCCTTGGCACGGAAATTAATGGATTGG   1110
338    E  V  Q  S  V  Q  Y  G  K  E  K  A  N  K  D  R  V  F  A  R  S  P  S  K  P  L  A  R  K  L  M  D  W
      GAAGTAGTAAGCAGAAATTCAATATCTGATGACAGGTTAGAAACACAAAGTCTTCCATCACGATCTCCACCGGGAACTCCTAATCATCGAAATTCTACA   1209
371    E  V  V  S  R  N  S  I  S  D  D  R  L  E  T  Q  S  L  P  S  R  S  P  P  G  T  P  N  H  R  N  S  T
      TTCACGCAGGAAGGAACCCGGTTACGACCATCTTCAGTTGGTCATTTGGTAGACCATATGGTTCATACTTCCCCAAGCGAAGTGTTTGTAAATCAGAGA   1308
404    F  T  Q  E  G  T  R  L  R  P  S  S  V  G  H  L  V  D  H  M  V  H  T  S  P  S  E  V  F  V  N  Q  R
      TCTCCGTCATCAACACAAGCTAATAGCATTGTTCTGGAATCATCACCATCACAAGAGACCCCTGGAGATGGGAAGCCTCCAGCTTTACCACCCAAACAG   1407
437    S  P  S  S  T  Q  A  N  S  I  V  L  E  S  S  P  S  Q  E  T  P  G  D  G  K  P  P  A  L  P  P  K  Q
      TCAAAGAAAAACAGTTGGAACCAAATTCATTATTCACATTCGCAACAAGATCTAGAAAGTCATATTAATGAAACATTTGATATTCCATCTTCTCCTGAA   1506
470    S  K  K  N  S  W  N  Q  I  H  Y  S  H  S  Q  Q  D  L  E  S  H  I  N  E  T  F  D  I  P  S  S  P  E
      AAACCCACTCCTAATGGTGGTATTCCACATGATAATCTTGTCCTAATCAGAATGAAACCTGATGAAAATGGGAGGTTTGGATTCAATGTAAAGGGAGGA   1605
503    K  P  T  P  N  G  G  I  P  H  D  N  L  V  L  I  R  M  K  P  D  E  N  G  R  F  G  F  N  V  K  G  G
      TATGATCAGAAGATGCCTGTGATTGTGTCTCGAGTAGCACCAGGAACACCTGCTGACCTCTGTGTCCCTAGACTGAATGAAGGGGACCAAGTTGTACTG   1704
536    Y  D  Q  K  M  P  V  I  V  S  R  V  A  P  G  T  P  A  D  L  C  V  P  R  L  N  E  G  D  Q  V  V  L
      ATCAATGGTCGGGACATTGCAGAACACACTCATGATCAGGTTGTGCTGTTTATTAAAGCTAGTTGTGAGAGACATTCTGGGGAACTCATGCTTCTAGTT   1803
569    I  N  G  R  D  I  A  E  H  T  H  D  Q  V  V  L  F  I  K  A  S  C  E  R  H  S  G  E  L  M  L  L  V
      CGACCTAATGCTGTATATGATGTAGTGGAAGAAAAGCTAGAAAATGAGCCAGATTTCCAGTATATTCCTGAGAAAGCCCCACTAGATAGTGTGCATCAG   1902
602    R  P  N  A  V  Y  D  V  V  E  E  K  L  E  N  E  P  D  F  Q  Y  I  P  E  K  A  P  L  D  S  V  H  Q
      GATGACCATTCCCTGCGGGAGTCAATGATCCAGCTAGCTGAGGGGCTTATCACTGGAACAGTCCTGACACAGTTTGATCAACTGTATCGGAAAAAACCT   2001
635    D  D  H  S  L  R  E  S  M  I  Q  L  A  E  G  L  I  T  G  T  V  L  T  Q  F  D  Q  L  Y  R  K  K  P
      GGAATGACAATGTCCTGTGCCAAATTACCTCAGAATATTTCCAAAAATAGATACAGAGATATTTCGCCTTATGATGCCACACGGGTCATTTTAAAAGGT   2100
668    G  M  T  M  S  C  A  K  L  P  Q  N  I  S  K  N  R  Y  R  D  I  S  P  Y  D  A  T  R  V  I  L  K  G
      AATGAAGACTACATCAATGCGAACTATATAAATATGGAAATTCCTTCTTCCAGCATTATAAATCAGTACATTGCTTGTCAAGGGCCATTACCACACACT   2199
701    N  E  D  Y  I  N  A  N  Y  I  N  M  E  I  P  S  S  S  I  I  N  Q  Y  I  A  C  Q  G  P  L  P  H  T
      TGTACAGATTTTTGGCAGATGACTTGGGAACAAGGCTCCTCTATGGTTGTAATGTTGACCACACAAGTTGAACGTGGCAGAGTTAAATGTCACCAATAT   2298
734    C  T  D  F  W  Q  M  T  W  E  Q  G  S  S  M  V  V  M  L  T  T  Q  V  E  R  G  R  V  K  C  H  Q  Y
      TGGCCAGAACCCACAGGCAGTTCATCTTATGGATGCTACCAAGTTACCTGCCACTCTGAAGAAGGAAACACTGCCTATATCTTCAGGAAGATGACCCTA   2397
767    W  P  E  P  T  G  S  S  S  Y  G  C  Y  Q  V  T  C  H  S  E  E  G  N  T  A  Y  I  F  R  K  M  T  L
      TTTAACCAAGAGAAAAATGAAAGTCGTCCACTCACTCAGATCCAGTACATAGCCTGGCCTGACCATGGAGTCCCTGATGATTCGAGTGACTTTCTAGAT   2496
800    F  N  Q  E  K  N  E  S  R  P  L  T  Q  I  Q  Y  I  A  W  P  D  H  G  V  P  D  D  S  S  D  F  L  D
      TTTGTTTGTCATGTACGAAACAAGAGGGCTGGCAAGGAAGAACCCGTTGTTGTCCATTGCAGTGCTGGAATCGGAAGAACTGGGGTTCTTATTACTATG   2595
833    F  V  C  H  V  R  N  K  R  A  G  K  E  E  P  V  V  V  H  C  S  A  G  I  G  R  T  G  V  L  I  T  M
      GAAACAGCCATGTGTCTCATTGAATGCAATCAGCCAGTTTATCCACTAGATATTGTAAGAACAATGAGAGATCAGCGAGCCATGATGATCCAAACACCT   2694
866    E  T  A  M  C  L  I  E  C  N  Q  P  V  Y  P  L  D  I  V  R  T  M  R  D  Q  R  A  M  M  I  Q  T  P
      AGTCAATACAGATTTGTATGTGAAGCTATTTTGAAAGTTTATGAAGAAGGCTTTGTTAAACCCTTAACAACATCAACAAATAAATAAGAAAGCAAAAAG   2793
899    S  Q  Y  R  F  V  C  E  A  I  L  K  V  Y  E  E  G  F  V  K  P  L  T  T  S  T  N  K  *
      ATCTGGGATATGTGTTGGAAAACTGCTTTCCCTTATGTTCACTGTGCCATAATGCTGCTCGCAGGAAATGGCATTTTACAAAAAAAAAAAAAAAAAA   2890
```

FIG. 4. Nucleotide and deduced amino acid sequence of PTPase MEG. Amino acid sequence is shown in single-letter code below the nucleotide sequence. The major open reading frame begins at position 1. The first three in-frame ATG codons are boldface and the two short upstream open reading frames are underscored with broken lines. A sequence predicted to form a stem–loop structure is shown with two arrows. The sequence of the PCR product is underscored with a solid line. Numbers to the left refer to amino acid position, and those to the right indicate nucleotide position.

ezrin (Fig. 5), a PTKase substrate (32, 33). Amino acids 659–909 of PTPase MEG are 35–40% identical to the phosphatase domains of PTPase 1B, T-cell PTPase, LCA, and LAR. The region between these two domains (amino acids 368–657) does not match any sequences in the GenBank/EMBL data base (January 1991).

The 5′ noncoding region of PTPase MEG cDNA was 771 bp and contains two short open reading frames predicting peptides of 8 and 22 amino acids (Fig. 4). The G+C content of this region is 65% and the region −200 to −1 contains >80% G+C. Several stable stem–loop structures have the potential to form in this region; most notable are 14-bp inverted repeats at positions −139 to −126 and −116 to −102 (Fig. 4). These can form stable stem–loops with a calculated Gibbs energy of $\Delta G^{\circ}_{37} = -26.3$ kcal/mol (−0.7 kcal/mol per base) as analyzed using the FOLD program (34, 35). We searched GenBank and EMBL nucleic acid data bases with this sequence and found a 32-bp segment from the 5′ untranslated region of the human protooncogene *ABL* (36) that was 75% identical (data not shown). This region of *ABL* may form a stem–loop structure similar to that in PTPase MEG, and 6 of 7 bases in the loop of the structures are identical.

Expression of PTPase MEG in *E. coli*. To determine that the protein encoded by PTPase MEG has PTPase activity, a 1.6-kb segment from the 3′ end of the cDNA that encoded the entire PTPase domain was subcloned into the bacterial expression vector pTrp (30). A prominent 59-kDa protein was detected by SDS/PAGE of total, supernatant, and particulate fractions of extracts of *E. coli* transformed with pTrp.PTPase MEG (data not shown). About 90% of the recombinant protein was in the particulate fraction and was presumed to be in inclusion bodies. Extracts expressing pTrp.PTPase MEG contained PTPase activity as measured with [^{32}P]phosphotyrosine-containing RCM lysozyme as substrate; this activity was inhibited by vanadate. Fig. 6 shows the time and protein concentration dependence of the activity. Most of the phosphatase activity was due to recombinant enzyme in the supernatant fraction, since this fraction had twice the activity of the particulate fraction (data not shown).

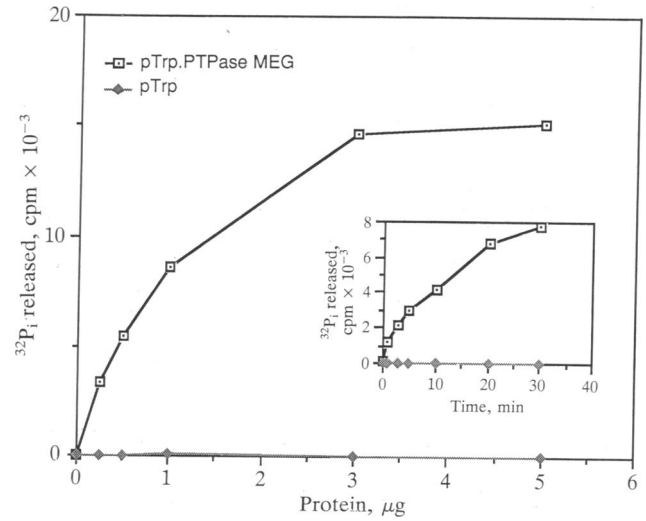

FIG. 6. Dephosphorylation of [^{32}P]phosphotyrosine-containing RCM lysozyme as a function of *E. coli* protein concentration. Various amounts of total homogenate from *E. coli* transformed with pTrp or pTrp.PTPase MEG were incubated with 25,000 cpm of the RCM lysozyme substrate for 10 min. (*Inset*) Time course for dephosphorylation of RCM lysozyme by the recombinant PTPase MEG; 0.5 μg of total homogenate protein was incubated with 25,000 cpm of the RCM lysozyme substrate for various times.

DISCUSSION

We have isolated a cDNA from human MEG-01 cell line and HUVEC cDNA libraries that encodes a third type of PTPase. This protein, PTPase MEG, has a predicted $M_r \approx 106{,}000$ and can be divided into two functional domains, (*i*) a protein 4.1 homology domain and (*ii*) a tyrosine-phosphatase domain, that are separated by 292 residues of unknown function. A hydropathy plot of the deduced amino acid sequence showed no hydrophobic regions that resemble a signal peptide or transmembrane domain, and thus PTPase MEG is likely to be a cytoplasmic protein. The tyrosine-phosphatase domain expressed in *E. coli* has activity, confirming that this cDNA encodes a PTPase. The other intracellular PTPases that have been described all contain two functional domains: the amino-terminal phosphatase catalytic domain and the putative carboxyl-terminal regulatory domain. The phosphatase domains of these proteins are 72–97% identical to each other (21–23). The phosphatase domain of PTPase MEG is only

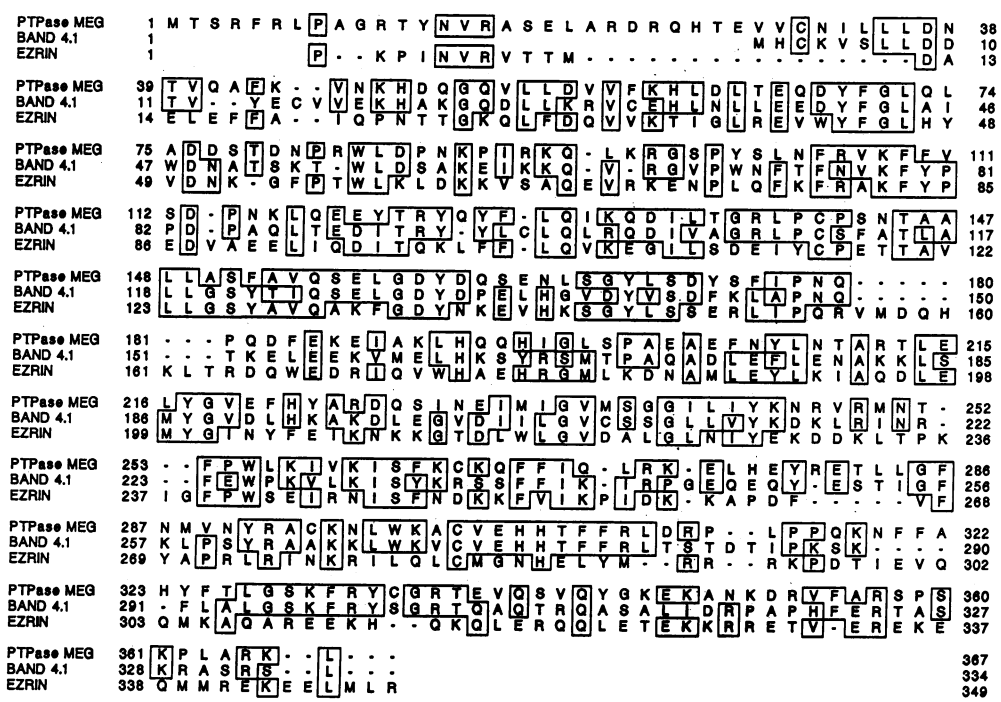

FIG. 5. Protein sequence relationship of the amino terminus of PTPase MEG, human erythrocyte protein 4.1 (31) and human ezrin (32). Identical residues are enclosed in boxes. Gaps have been introduced to optimize alignments.

38% identical to the other cellular PTPases in the catalytic domain. In addition, it has only 17 amino acids carboxyl to the phosphatase domain. Thus, the low degree of identity in the catalytic domain and the lack of a putative carboxyl-terminal regulatory domain indicate that PTPase MEG is only distantly related to the other intracellular PTPases. In addition, the phosphatase domain of PTPase MEG is only 35–40% identical to the phosphatase domains of the transmembrane-type PTPases.

The amino-terminal similarity of PTPase MEG to protein 4.1 suggests possible functions for the phosphatase. Protein 4.1 has been identified not only in erythrocytes but also in tissues such as brain (37), fibroblasts (38), blood platelets (39), and endothelial cells (40). It serves as a crosslinking molecule with binding sites in the amino-terminal region that participate in the inositolphospholipid-dependent binding of protein 4.1 to glycophorin, a transmembrane protein (41, 42). Since this is the region of similarity between PTPase MEG and protein 4.1, PTPase MEG may also associate with a glycophorin-like transmembrane molecule. This suggests that PTPase MEG has a cytoskeletal association with an additional PTPase activity that may function in cytoskeletal rearrangement or organization. The central part of PTPase MEG is unrelated to any previously sequenced protein and may interact with some other molecule to control the activity of the phosphatase domain.

PTPase MEG is related to other cytoskeletal elements, including ezrin (Fig. 5) and talin (28% identical over 191 residues) (43). Ezrin was first identified as a major PTKase substrate in chicken microvilli and has been localized to microvilli in plasma membrane structures in a wide variety of cell types (32, 33). After activation of A431 epithelial carcinoma cell EGF receptors by EGF, ezrin redistributes rapidly into microvilli and membrane ruffles. The redistribution correlates with cell surface shape changes and is coincident with an increased tyrosine and serine phosphorylation (44). This suggests that ezrin is involved in cytoskeletal organization, and since ezrin is a PTKase substrate this organizational activity may be regulated by phosphorylation and dephosphorylation of tyrosine residues. Many nonreceptor PTKases are associated with membranes and some of them, including $pp60^{v\text{-}src}$, $pp120^{v\text{-}gag\text{-}abl}$, $pp90^{v\text{-}gag\text{-}yes}$, $pp80^{v\text{-}gag\text{-}yes}$, and type IV c-Abl, are bound to membrane cytoskeletons (45–47). Several cytoskeletal components have been identified as substrates for PTKases, including integrins, connexin, talin, and vinculin (48). Since PTPase MEG is related to talin and ezrin, we propose that its phosphatase activity is important in the regulation of cytoskeletal events.

We thank Theo Ross and Anne Bennett Jefferson for useful discussions and advice, Dr. Matt Thomas and Dr. J. Evan Sadler for critical review and comments, and Ann Delaney for assistance in the preparation of the manuscript. 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, A., Kapeller, R. & Soltoff, S. (1991) *Cell* **64**, 281–302.
3. Gould, K. L. & Nurse, P. (1989) *Nature (London)* **342**, 39–45.
4. Dunphy, W. G. & Newport, J. W. (1989) *Cell* **58**, 181–191.
5. Hunter, T. (1989) *Cell* **58**, 1013–1016.
6. Brugge, J., Cotton, P., Lustig, A., Yonemoto, W., Lipsich, L., Coussens, P., Barrett, J. N., Nonner, D. & Keane, R. W. (1987) *Genes Dev.* **1**, 287–296.
7. Golden, A., Nemeth, S. P. & Brugge, J. S. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 852–856.
8. Ferrell, J. E., Jr., & Martin, S. (1988) *Mol. Cell. Biol.* **8**, 3603–3610.
9. Golden, A. & Brugge, J. S. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 901–905.
10. Nakamura, S. & Yamamura, H. (1989) *J. Biol. Chem.* **263**, 7089–7091.
11. Lerea, K. M., Tonks, N. K., Krebs, E. G., Fischer, E. H. & Glomset, J. A. (1989) *Biochemistry* **28**, 9286–9292.
12. Ralph, S. J., Thomas, M. L., Morton, C. C. & Trowbridge, I. S. (1987) *EMBO J.* **6**, 1252–1257.
13. Streuli, M., Hall, L. R., Saga, Y., Schlossman, S. F. & Saito, H. (1987) *J. Exp. Med.* **166**, 1548–1566.
14. Streuli, M., Krueger, N. X., Tsai, A. Y. M. & Saito, H. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8698–8702.
15. Krueger, N. X., Streuli, M. & Saito, H. (1990) *EMBO J.* **9**, 3241–3252.
16. 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.
17. Matthews, R. J., Cahir, E. D. & Thomas, M. L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4444–4448.
18. Pingel, J. T. & Thomas, M. L. (1989) *Cell* **58**, 1055–1065.
19. Koretzky, G. A., Picus, J., Thomas, M. L. & Weiss, A. (1990) *Nature (London)* **346**, 66–68.
20. Mustelin, T., Coggeshall, K. M. & Altman, A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6302–6306.
21. Tonks, N. K., Diltz, C. D. & Fischer, E. H. (1988) *J. Biol. Chem.* **263**, 6722–6730.
22. Chernoff, J., Schievella, A. R., Jost, C. A., Erikson, R. L. & Neel, B. G. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2735–2739.
23. Guan, K., Haun, R. S., Watson, S. J., Geahlen, R. L. & Dixon, J. E. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1501–1505.
24. 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.
25. Cool, D. E., Tonks, N. K., Charbonneau, H., Fischer, E. H. & Krebs, E. G. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 7280–7284.
26. Gu, M., York, J. D., Warshawsky, I. & Majerus, P. W. (1991) *Clin. Res.* **39**, 208a.
27. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. (1985) *Science* **230**, 1350–1354.
28. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
29. Ye, R. D., Wun, T. C. & Sadler, J. E. (1987) *J. Biol. Chem.* **262**, 3718–3725.
30. York, J. D. & Majerus, P. W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9548–9552.
31. Conboy, J., Kan, Y. W., Shohet, S. B. & Mohandas, N. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9512–9516.
32. Gould, K. L., Bretscher, A., Esch, F. S. & Hunter, T. (1989) *EMBO J.* **8**, 4133–4142.
33. Gould, K. L., Cooper, J. A., Bretscher, A. & Hunter, T. (1986) *J. Cell Biol.* **102**, 660–669.
34. Devereux, J., Haeberli, P. & Smithies, D. (1984) *Nucleic Acids Res.* **12**, 387–395.
35. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruther, M. H., Neilson, T. & Turuer, D. H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9373–9377.
36. Shtivelmen, E., Lifshitz, B., Gale, R. P. & Canaai, E. (1985) *Nature (London)* **315**, 550–554.
37. Goodman, S. R., Casoria, L. A., Coleman, B. D. & Zagon, I. S. (1986) *Science* **224**, 1433–1436.
38. Cohen, C. M., Foley, S. F. & Korsgren, C. (1982) *Nature (London)* **299**, 648–650.
39. Davies, G. E. & Cohen, C. M. (1985) *Blood* **65**, 52–59.
40. Leto, T. L., Pratt, B. M. & Madri, J. A. (1986) *J. Cell. Physiol.* **127**, 423–431.
41. Correas, I., Leto, T. L., Speicher, D. W. & Marchesi, V. T. (1986) *J. Biol. Chem.* **261**, 3310–3315.
42. Correas, I., Speicher, D. W. & Marchesi, V. T. (1986) *J. Biol. Chem.* **261**, 13362–13366.
43. Rees, D. J. G., Ades, S. E., Singer, S. J. & Hynes, R. O. (1990) *Nature (London)* **347**, 685–689.
44. Bretscher, A. (1989) *J. Cell Biol.* **108**, 921–930.
45. Burridge, K., Fath, K., Kelly, T., Nuckolls, G. & Turner, C. (1988) *Annu. Rev. Cell Biol.* **4**, 487–525.
46. Drubin, D. G., Mulholland, J., Zhu, Z. & Botstein, D. (1990) *Nature (London)* **343**, 288–290.
47. Van Etten, R. A., Jackson, P. & Baltimore, D. (1989) *Cell* **58**, 669–678.
48. Bishop, J. M. (1991) *Cell* **64**, 235–248.