Isolation of a cDNA clone encoding a human protein-tyrosine phosphatase with homology to the cytoskeletal-associated proteins band 4.1, ezrin, and talin

(tyrosine phosphorylation/dephosphorylation/focal adhesion plaques/transformation/pp60^{src})

QING YANG AND NICHOLAS K. TONKS*

Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, NY 11724

Communicated by Edmond H. Fischer, March 28, 1991

ABSTRACT The polymerase chain reaction (PCR), from primers corresponding to conserved sequences within the catalytic domains of the protein-tyrosine phosphatases, was used to amplify protein-tyrosine phosphatase-related cDNAs from a HeLa cell library. After probing the same cDNA library with one of the PCR products, 10 positive clones were identified. The longest of these clones (3984 base pairs) contained 2739 base pairs of open reading frame and, after a stop codon, a 3' nontranslated segment of 1222 base pairs. A 4.3-kilobase transcript was detected by Northern blot analysis of HeLa cell $poly(A)^+$ RNA. The open reading frame predicts a protein of 913 amino acids (≈104 kDa), termed PTPH1. The sequence of PTPH1 can be described in terms of three segments. (i) The N-terminal segment displays homology to the domains in the cytoskeletal-associated proteins band 4.1, ezrin, and talin that direct their association with proteins at the interface between the plasma membrane and the cytoskeleton in structures such as focal adhesions. (ii) There is a central segment bearing putative phosphorylation sites for protein-serine/threonine kinases. (iii) A segment that is homologous to the members of the protein-tyrosine phosphatase family is located at the C terminus. The structure is discussed in the light of the potential role of PTPH1 in controlling cytoskeletal integrity and the possibility that overexpression of PTPH1 may reverse transformation induced by oncogenic protein-tyrosine kinases, such as the members of the src family.

Protein-tyrosine phosphorylation has been implicated in the control of many cellular processes including normal and neoplastic cell growth (1). Since the phosphorylation of tyrosyl residues *in vivo* is reversible, a complete understanding of its physiological significance must encompass the characterization of the protein-tyrosine phosphatases (PTPs; protein-tyrosine-phosphate phosphohydrolase, EC 3.1.3.48) in addition to the protein-tyrosine kinases (PTKs; ATP:protein-tyrosine *O*-phosphotransferase, EC 2.7.1.112). PTP1B, a low molecular weight PTP, was first isolated in homogeneous form from human placenta (2, 3). After determination of its amino acid sequence, homology between PTP1B and CD45 was demonstrated (4), establishing that the PTPs exist as transmembrane receptor-like proteins as well as low molecular weight cytoplasmic forms (for review, see ref. 5).

A cDNA for a second low molecular weight PTP, termed T-cell PTP (TCPTP), was isolated from a human peripheral T-cell library (6). This species has also been detected in a variety of tissues and cell lines and is not restricted to lymphocytes (unpublished observations). A number of studies involving low-stringency screening have identified further PTPs that can be classified as transmembrane or receptor-like (for review, see ref. 5). With one exception, these possess two cytoplasmic domains that are homologous to PTP1B; however, as yet, few of the proteins have been isolated or expressed and shown to be active. These receptor-like PTPs vary considerably in the structure of their extracellular segments. For leukocyte common antigen-related (LAR) molecules (7), the extracellular segment is homologous to both immunoglobulin-like and fibronectin type III-like sequences in the neural cell adhesion molecules. This suggests that, analogous to the neural cell adhesion molecules on adjacent cells may regulate adhesion phenomena by modulation of the PTP activity of their intracellular segments. At the present time, however, there are no data to suggest the identity of the ligands for the other receptor PTPs.

Recent studies involving CD45-negative T cells have established a requirement for CD45 in the stimulation of antigen-induced T-cell proliferation (8), confirming that this PTP may play a positive role in initiating signal transduction events. In addition it has been shown that the activity of $p34^{cdc2}$, a protein-serine/threonine kinase that is required for progression through both the G₁/S and G₂/M boundaries of the cell cycle, is negatively regulated by tyrosine phosphorylation (9). Dephosphorylation of a tyrosyl residue in the consensus ATP binding site (Tyr-15) activates the kinase and drives the cell into mitosis (9). These examples indicate that the PTPs should not be viewed as simply providing an "off-switch" to counter the activity of PTKs, but rather they are implicated in the positive regulation of a number of fundamental cellular processes.

An essential step to understanding fully the scope of the physiological importance of the PTPs is the identification and characterization of the members of this family. We have isolated a cDNA for a nontransmembrane protein from HeLa cells that is homologous to the PTPs. Its sequence[†] suggests that it may act at the interface between the plasma membrane and the cytoskeleton, for instance at focal adhesion plaques, and may, therefore, function in maintaining the integrity of the cytoskeleton. In the light of its structure, the potential of this enzyme, termed PTPH1, to reverse transformation by oncogenic PTKs, such as src, is also discussed.

MATERIALS AND METHODS

Polymerase Chain Reaction (PCR) Conditions. Oligonucleotide primers were synthesized based on the conserved amino acid sequences KCAQYWP [primer 1, equivalent to residues 120–126 in PTP1B (10)] and HCSAGIG [primer 2, equivalent to residues 214–220 in PTP1B (10)]. The degen-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. \$1734 solely to indicate this fact.

Abbreviations: PTP, protein-tyrosine phosphatase; PTK, protein-tyrosine kinase.

^{*}To whom reprint requests should be addressed.

[†]The sequence reported in this paper has been deposited in the GenBank data base (accession no. M64572).

Proc. Natl. Acad. Sci. USA 88 (1991)

eracy of primer 1 was 64-fold and that of primer 2 was 96-fold. Phage DNA of a HeLa cell cDNA library (Stratagene; catalogue no. 936201), isolated by the plate lysate method (11), was used as template. Primers were phosphorylated with T4 polynucleotide kinase prior to the PCR and were added at a final concentration of 1 μ M to a mixture containing 10 mM Tris·HCl (pH 8.4), 50 mM KCl, 1.5 mM MgCl₂, 0.01% gelatin, all four dNTPs (each at 0.2 mM), 2.5 units of *Taq* polymerase, and 2 μ g of phage DNA. Thirty cycles of the PCR were performed; each was carried out at 94°C for 1 min, 50°C for 1 min, and 72°C for 2 min. The PCR products were analyzed on a 1% agarose gel and DNA fragments of ~0.25 kilobase (kb) were excised, eluted from the gel, subcloned into the *Sma* I site of pUC118, and sequenced.

Screening of the HeLa Cell cDNA Library. The same HeLa cDNA library was then probed with the 0.25-kb PTPH1

product of the PCR obtained above. Plaques were transferred to nitrocellulose filters (Schleicher & Schuell) and screened by hybridization at 65°C in a solution containing $2 \times SSC$ (1× SSC = 0.15 M NaCl/0.015 M trisodium citrate, pH 7.0), $5 \times$ Denhardt's solution (11), 0.1% SDS, 25 mM sodium phosphate (pH 8.0), 1% sodium pyrophosphate, 10% (wt/vol) dextran sulfate, and denatured calf thymus DNA (12.5 μ g/ ml). The filters were washed successively at the same temperature in a series of solutions containing 0.1% SDS with $4 \times$ SSC, $2 \times$ SSC, $1 \times$ SSC, and finally $0.1 \times$ SSC. Hybridizing phage were plaque-purified, DNA was prepared, and cDNA inserts were isolated and subcloned using standard techniques (11). DNA sequencing was carried out by the dideoxynucleotide chain-termination procedure (12) using either manufacturer's primers (United States Biochemical) or synthetic oligonucleotides derived from the existing sequences.

1	CTGCAGGTTATTCAGCGATAGTTATGACCTCCGGTTACGTGCGTTGCGTGGAGGAATTAATAATATATACGCACCTCGGAGTTACCCAAAGAGAAAACTCGATCAGAAGTCATTGCAGC M T S R L R A L G G R I N N I R T S E L P K E K T R S E V I C S
120	ATCCACTTITIAGATGGCGTGGTACAGACCTTTAAAGTTACTAAAGAACAAGACACTGGCCAGGTTCTTCTGGATATGGTGCACAACCACCTGGGTGTGACTGAAAAGGAATATTTGGTTA I H F L D G V V Q T F K V T K Q D T G Q V L L D M V H N H L G V T E K E Y F G L
240	CAGCATGATGACGACTCCCGTGGACTCCCTAGATGGCTGGAAGCAAGC
360	GATCCCAACACACGCCAGCAAGAACAAACAAGCCAGGCACTTGTATTTCTTACAACTGAAGATGGATATTTGCGAAGGAAG
480	GCCGTACAATCTCATTTTGGAGACTATAATTCTTCCATACATCATCATCAGGGCTATCTTTCCGGATAGTCACATTATACCCGATCAAAATGAGGACTTTTTAACAAAAGTCGAATCTCTGCAT A V Q S H F G D Y N S S I H H P G Y L S D S H F I P D Q N E D F L T K V E S L H
600	GAGCAGCACASTGGGCTAAAACAATCAGAAGCAGAATCCTGCTATATCAACATAGCGCGGGACCTCGACTTCTATGGAGTAGAACTGCACASTGGTAGGGATCTGCACAATTTAGACCTA E Q H S G L K Q S E A E S C Y I N I A R T L D F Y G V E L H S G R D L H N L D L
720	ATGATTGGAATGCGAGGTGTGCTGCTGCTGCTGCGGAAAATACATTGCACAAGTTCCTATGCATCGGGGGAACATTCTCAAAAGTTCTCTAAAAGGAAAAAGTTCCTCATACATCAG M I G I A S A G V A V Y R K Y I C T S F Y P W V N I L K I S F K R K K F F I H Q
840	CGACAGAAACAGGCTGAATCCAGGGAACATATTGTGGCCTCTAACATGCTGAATCCGATCTTGCAAAAACTTGTGGAAATCCTGTGTGAGCACCATACGTTCTTCAGGCAAAGAAG R Q K Q A E S R E H I V A F N M L N Y R S C K N L W K S C V E H H T F F Q A K K
960	CTACTACCTCAGGAAAAGAATGTTCTGTCTCAGTACTGGACTATGGGGCTCTCGGAACACCAAAAAGTCGGTAAATAACCAATATTGCAAAAAGTGATTGGCGGGATGGTGGGAACGCCA L L P Q E K N V L S Q Y W T M G S R N T K K S V N N Q Y C K K V I G G M V W N P
1080	GCCATGCGAGAGATCCTTATCAGTGGAGCACCTAGAAACCAAGAGTCTGCCTTCTCGTTCCCCTCCCATTACTCCCAACGGCGAAGTCCTCGGCTCCGGCACGAAATCCGAAAGCCACGC A M R R S L S V E H L E T K S L P S R S P P I T P N W R S P R L R H E I R K P R
1200	CACTCTTCTCCAGATAACCTTGCAAATGAACTGAACATGACATCACGGAAACGGAAGGATGTATTTTACACGTACAAGGGCTCTCTGGCCCCTCAAGACAGCAGCAGTCTGAAGTTTCTCAGAAC H S S A D N L A N E M T Y I T E T E D V F Y T Y K G S L A P Q <u>D S D S E</u> V S Q N
1320	CGAAGCCCCCACCAAGAGAGTTTATCCGAGAACAATCCGGCACAAAGCTACCTGACCCAGAAGTCATCCAGTTCTGTGTCTCCATTCTAAATGCTCCAGGCCCGGCCGCGCGCG
1440	GTIGATCAGCAGCTCTTAGATGACTICCACAGGGGGACCAAAGGGGGGCTCCACCGAGGACGACAGCAGCAGCAACTAATGATGATGATGACAGCAAGCA
1560	ATCACACCAGATGAAGATGGAAGATTTGGATTTAATCTTAAGGGAGGAGTGGATCAAAAGATGCCTCTTGTGGTATCAAGGATAAACCCAGAGTCACCTGCGGACACCTGCATTCCTAAG
1680	CTGAACGAAGGGATCAAATCGTGTTAATCAATGGCCGGGACATCTCAGAACACCACGCATGACCAAGTGGTGATGTTCATCAAAGCCAGCC
1800	GTGATCAGGAGGAGGAGGAGGAGGTGCGGCTCATTIGCTGAAGTCTGAAGGATGAACCAGGTTTGCCCGAAGCCATTITGCCCGAAGGCATTTGCGGGGGGACACTTTGGAGGGA V I R R R A V R S F A D F K <u>S E D E</u> L N Q L F P E A I F P M C P E G G D T L E G
1920	TCCATGGCACAGCTAAAGAAGGGCCTCGAAAGCGGGGCGGGGCGGATCCAGTTGGCAACTCTACAGAAAAAAGCCAGGTTTGGCCATCACGTTTGCAAAGCTGCCTCAAAGCTGCCTCAAAGCTGCCTCAAAGCTGCCACAGATTTGGAA
2040	AAAAACCGATATAAAGATGTGCTGCCTTATGACACCACCCGGGTATTATTGCAGGGAAATGAGATTATATATA
2160	AAGTACATCGCCACTCCAGGGGCCCCTGCCGCATACCTGTGCACAGTTTTGGCAGGTTGTCGCGATCAGAAGTGTCACCACATGTCACTGACGACTCTCACAGAACGAGGGCGGACC K Y I A T Q G P L P H T C A Q F W Q V V W D Q K L S L I V M L T T L T E R G R T
2280	ANATGICACCAGIACIGGCCAGAGICCCCCGAGGICAIGAACCACGGGGGCTITCACATCCAGITGAGICAGAGGACIGCACCACGCCAICGGCCAAIGCGGCCACAAAC K C H Q Y W P D P P D V M N H G G F H I Q C Q S E D C T I A Y V S R E M L V I N
2400	ACCCRACCGGGGRAGAACACACAGTGACACATCTCCAGTAGGTCGCATGGCCTGACCACGGTATACCCGATGACCTCCGGACTTTCTGAAATTGTAAGCATGTGAGGTCTCTGGAA T Q T G E E H T V T H L Q Y V A W P D H G I P D D S S D F L E F V N Y V R S L R
2520	GTGGRCAGCGAGCCTATCTCTGCAGTGCTGGAATAGGTCGAACCGGTGTGTGGTGCGAACAGCCATGTGCCTAACTGAGAGGAACCGGCCATTACCCACTGGAT V D S E P V L V H C S A G I G R T G V L V T M E T A M C L T E R N L P I Y P L D
2640	ATTOTCCGAAAAAGCGAGACCAGCGCCGCCATGATGGTGGGAGACATCAAGCCAGTACGAGGTTGGTGGGAGCGATTCTCGTGTGTATGAAGAAGAAGGTTGGGGGGCCCT I V R K M R D O R A M M V O T S S O Y K F V C E A I L R V Y E E G L V O M L D P
2760	RGTTRA GACAACTGTGAAAAAGTTCATTCCTCTTTCCCAAGGGCATCCTCCTTGAAAGAGGAGGACAGACCTCTCTGGAAGCAGCAAGAGGAACCAGTAGCTGTGGGAAAGGAATGGGCA 8 *
2880	cctctgaacccaggcactttaaacttctatagaaaagatatcgtgtacataggaactggtgtagataagcatgcat
3000	AGGATCTCAGTTTGGGGCCTGTCCTAATGCCTTCTTCCCTAACATCACCACACACCCCTGTCGGCATCCTGGAGCAATTGAGACCGGACACCCACAGAGCTGTTGTCCTCCCAGCAAC
3120	AAGATGGTGTGGTTATCTGGGTCATTGGATGTTTGTTGTTGTGTGTG
3240	GATGCTAAGGAAACCCTGACGTATGTACTAGATGGCAGGGCACTGGGGGTCAGGCTGAAGGCTGAGGCAACACCTCTCTGCCCTCCCT
3360	TCATGTTTTCTGAGAATTGTGTCCCTCTTCAGTTCCCTCTTGGTGCCTAACCTGGATTAGTAATGTGCATTCAGGTGAATTTTCAGCTGAGGCTCTGAGAACTGGTACTCTCAGTGTGTTC
3480	TGGTCATCTTGTGGGCTTAGTTGTAGAAGCAGGTGTGTCTCTTGCCTCTGCTTGCCTCCTACTGCACCTCAGCACCCAGGACTGGAATCACCGACTACTGAATCTCCTACATGTATTGCT
3600	GCTACTTCAAGCTCCTCCACTTGAAACCTTATGATTTTCCAAGGGGAGATGGGACAGTGTCATCTAAATATTCCGAATGTTTGGCCTTCTGAGAAAAGAGCTTCTAGTAATTGAACCATG
3720	GGTTTCCCAGCTTCTGGAGGGTTGGCCGTGGGCTGTGTACATGTGTGTG
3840	GGAAGAACAAATGTGTGCATTCACCCCCAGTCACAATGGTCTCCATTGCATTTCAAAGGAGAGGAGCAGCACTATCTGAATATAAACACAAATCTGATGTTAATTTATTCTAAGAACACCAT
3960	CTGTATTCATTTGATTGICCTAAA

FIG. 1. Nucleotide and predicted amino acid sequence of PTPH1. The open box delineates the segment of homology to the N-terminal domains of band 4.1, ezrin, and talin. The shaded box defines the segment homologous to the catalytic domains of the PTPs. In the intervening segment, seryl and threonyl residues located in sequences bearing features of sites of phosphorylation by casein kinase II are underlined. In addition seryl residues that display some of the features of potential sites of phosphorylation of $p34^{cdc2}$ are circled.

Data Base Search for Related Sequences. The GenBank data base of sequence information was searched with the FASTA program of Pearson and Lipman (13) to identify proteins with amino acid sequence similar to PTPH1. The alignment of similar sequences was optimized using the ALIGN program from the National Biomedical Research Foundation, the mutation data matrix, and a gap penalty of 10 (14). The alignment scores are expressed in units of standard deviation from the average background scores of 100 randomly generated sequences of the same composition (14).

Northern Blot Analysis. Total RNA was extracted from HeLa cells and $poly(A)^+$ RNA was selected by oligo(dT)-cellulose chromatography (11). After electrophoresis on a formaldehyde/agarose (1%) gel, the RNA was transferred to a GeneScreen*Plus* membrane and hybridized with the \approx 4-kb insert of PTPH1. The hybridization and washing conditions were as described above except that SDS was included at 1%.

RESULTS

To identify the complement of PTPs in HeLa cells, a model system for the study of the cell cycle, we utilized the PCR and primers corresponding to conserved segments within the catalytic domains to amplify PTP-related cDNAs from a HeLa cell library. The PCR products (≈ 0.25 kb) were subcloned into pUC118; of 77 independent subclones sequenced, 15 corresponded to a PTP isoform termed PTPH1. A total of 4.2×10^6 phage plaques from the same HeLa cell cDNA library were probed with the PCR product and 10 positive clones were identified. The sequence of the longest clone (3984 base pairs) is presented in Fig. 1. The open reading frame, from the first available ATG, is 2739 base pairs, which would encode a protein of 913 amino acids with a predicted molecular mass of ≈ 104 kDa. Although the sequence surrounding the putative initiator does not conform well to the Kozak consensus sequence (15), there is a purine at position -3 that is an important requirement for an initiation site. Furthermore, translation of RNA, synthesized from PTPH1 cDNA in a reticulocyte lysate, yielded a protein of ≈ 120 kDa, close to the expected size (data not shown). There is a 3' nontranslated segment of 1222 base pairs; however, the

	_	9.5	Kb	
13		7.5	Kb	
N.S.	_	4.4	Kb	
		2.4	Kb	
	_	1.4	Kb	

FIG. 2. Northern blot analysis of PTPH1 mRNA. Approximately 10 μg of poly(A)⁺ RNA from HeLa cells was subjected to electrophoresis on a formaldehyde/agarose gel and transferred to a Gene-Screen*Plus* membrane. After hybridization and washing, the membrane was exposed to film for 7 days at -70° C with intensifying screens. Numbers on the right indicate the positions of RNA markers (BRL). cDNA contains no consensus polyadenylylation signal or poly(A) tail. A transcript encoding PTPH1 was detected by Northern blot analysis of HeLa cell poly(A)⁺ mRNA, as a low-abundance message of \approx 4.3 kb (Fig. 2). The slightly larger size of the mRNA may be explained in part by the absence of a poly(A) tail in the cDNA.

A search of the GenBank data base, using the Pearson and Lipman FASTA program (13), confirmed the similarity in sequence between the C-terminal segment of PTPH1 and the catalytic domains of the members of the PTP family. This is depicted in Fig. 3A. Although the tyrosine phosphatase activity of PTPH1 remains to be demonstrated, it should be noted that in CD45 (16) and LAR (17), which share comparable levels of sequence identity to PTP1B as displayed by PTPH1, intrinsic activity has been confirmed.

A comparison of the sequence of the N-terminal 650 residues of PTPH1 with those in GenBank demonstrated similarity between residues 30 and 357 and the homologous N-terminal domains in band 4.1, ezrin, and talin that are believed to be important for localizing these proteins to junctions between the plasma membrane and the cytoskeleton (Fig. 3B).

Using the ALIGN program, the N- and C-terminal segments of PTPH1 were compared separately with each of their related sequences. The alignment scores presented in Tables

A PTPH1 PTPase 1B T-Cell PTPase Human CD45 D1 Human CD45 D2	NLDEN RYKDVLPYD TIE VLLOG N
PTPH1 PTPase 1B T-Cell PTPase Human CD45 D1 Human CD45 D2	EIPAANLVNKYIATQGPLEHITGAGEWGVVMDORLSLIVMLTHEFGRIKGHQYMP BEAQRSYILIDGGPLENITGGHEWENVMEDKSRGVMINRVNEKGSLKGAQYMPJ BEAQRSYILIDGGFLENITGGHEWENVMEDKRKAVMINRIVEKSSKGAQYMPJ FKEPRKYIAQGFRDENVDDWHWINGQKTKAVVMINRIVEKESVKGADYWPJ YWKPEVMIAAQGPLKENIGDEWGMIFQRWVVIVMLTELKHGDDETGAQYWG
PTPH1 PTPase 1B T-Cell PTPase Human CD45 D1 Human CD45 D2	-DEPDVMHGEBHIOCOSEDCIIANVORENLVYN-TOREEHITYTHOVARPDBGI Exekeifedtynkiisiadiksynymaolelen-luttografeiine hyttapedigy DECHIFKETESYKLISEDIKSYNTYHLOLEN-INGERFUTSHEHT TAPPETGY SMEEGTRAFBYVKLISEDIKSYNTYHLOLEN-INGERFUTSHEHT TAPPETGY SMEEGTRAFBYVVLIDUKKCPPUTJOKLIVJUKSKAIGREYTYYVYTTAPEDGY
PTPH1 PTPase 1B T-Cell PTPase Human CD45 D1 Human CD45 D2	DGSSDFILFUN YWRSIR VGSPVLVHCSAIGFTGVLVHRFIANCI ESPASTMULT KWRSIS PHOR VVHCSAIGFTCLAIDTLLUM ESPASTMULT KWRSGGIN PHORAVINGSAIGFTCLAIDTLLUM DGHLLUKLRRWIANFTSGTVHCSAQVERTTIGIDAMLEGA ADFKELISMIQ VWKQKLPQKNSSEGNKHHKSTRLUIIHGROGSQQTGIFCALLNLESA
PTPH1 PTPase 1B T-Cell PTPase Human CD45 D1 Human CD45 D2 ,	BRNLPTYPLDIMEKMRORAMMVOTSSOTKEVCEAILR DRRNDPSSVDIKKVLLEMMRTMMGLTOTADOLRESYLAVIE EKCODTNIKOVLLMMRKTMMGLTOTADOLRESYNAVIE BAENKVDVYGYNMLTRRORCLMVOVEAOTILIHOALVE BTEEVVDIFOVMKALMRARPCMVSTFEOYOELVDVTAS
B PTPH1 Human 4.1 Human ezrin Mouse talin	IGSIBFLDGVVOTEKVTRODIGOVLLDVVHNALGVTERERFGLOHDDSVDSBRWLBA HGKVSLLDDTVYSCVVERARGODLLRRVCERINLLEDDFGLAINONA-JEKTWLDS DVRVTTMDALEBAIOPNTGKOLFDOVKTIGLRBVWEGUHYVDNK-GFBTWLKL GITNGDESLUVRELMEERODGGGTUHKDKTLG-ROBEKVERUKCKLHTUDELNWLDH
PTPH1 Human 4.1 Human ezrin Mouse talin	SKPIRKOLKGGF-RCTLUFRVRFBIPDENTUOLEOTRH-UYFLOLKMDICEGRUTCEL Akelukgovrgv-Hwnftenakfyfpdbaculed itenation okrysadevrkenalofyrkenakfyftendaeluioditokurfelofykgeuledityeityce okrysadevrenakfyftenakfyftendaveluidditokurfelofykgeuletyce Griffiggovee-Hettulkgakfyftenoffvosrdpvolnulyfogarddilngshpysf
PTPH1 Human 4.1 Human ezrin Mouse talin	NSAYVLASYAVQSHEGDYNSSIHHEGYLSDSHEIED ATLALUGSYTIGSELGIYDEELHGVDYNSDEKLARNG ETAYLLGSXIONSFEGYNKEVHKSGELSSERIERGVNOGHKLTROWED ETAYLLGSXIVONSFEGYNKEVHKSGELDLKDELEKEYVKQKGERKIFQAHK DRACEFNGFQCQIQEGPHNEQKHKAGELDLKDELEKEYVKQKGERKIFQAHK
PTPH1 Human 4.1 Human ezrin Mouse talin	QBSGLKQSEAESCMINEARTEDFYGVELHSGRDINNLDIMIGIASAGVAVYRG Syrsmtpacadeffenterfenarkalsmygvelhkakdjegvdiilavogsgilvyndo EmrgMLNAMMerikakojolemyginypeinkskakdjegvdiilavogsgilvyndo NCGOMSEIEAKVRMVKLARSEKTYGVSFFLVKEKMKGKNKLVPRLIGIIKECVMRVDE
PTPH1 Human 4.1 Human ezrin Mouse talin	YICTSFYEWVNIL-UKISEKRKKFETHOROKOABSREHIVAENMLNYRSCRNLWRSCY Klrinrepwerv-Ukisytresfetikirgobogoyestigeklevkakkumvoyu Iterigfword-Ukisytresforterukiirokoapoyu Hyyrerikinkriloud Ktrevigewsltnikrwaasprsetlofgodyogyyvottegeqiaqliaqliatii
PTPH1 Human 4.1 Human ezrin Mouse talin	EHHTFROAKKILLOCKKNULGOYKTMESSNEEKSVNMEYCKKULGOMMAN EHHTFROAKLISTON ENHTERSOARUSTIN ENHTERSOARUSTIEVOOMKAAAREKHOMOLEROOLEROOLEROKKRETVEREKKOM KKKSKSNEGIEGEBESTEVIONKAAAREKHOMOLEROOLEROOLEROMINT

FIG. 3. Alignment of amino acid residues of PTPH1 with the conserved domains in related proteins. (A) The sequence of the PTP-like domain in PTPH1 is compared with the catalytic domains of two low molecular weight PTPs, PTP1B (PTPase 1B) and TCPTP (T-cell PTPase), and a receptor-linked form (CD45). (B) The sequence of the N-terminal segment PTPH1 is compared with the homologous domains in band 4.1, ezrin, and talin. Only identities between PTPH1 and the other sequences have been highlighted and gaps (denoted by hyphens) have been inserted to optimize the alignment.

 Table 1. Evidence of homology between PTPH1 and members of the PTP family

	PTPH1	PTP1B	ТСРТР	CD45 DI	CD45 DI
PTPH1		24.8	30.2	23.4	17.6
PTP1B	39%	_	50.8	23.8	16.4
ТСРТР	42%	73%		23.9	16.4
CD45 DI	37%	35%	36%	_	20.6
CD45 DII	30%	31%	32%	35%	—

Numbers above the diagonal express the similarity between segments in terms of their alignment scores. Residues 669-900 in PTPH1 were compared pairwise with residues 40-276 in PTP1B, 42-274 in TCPTP, 491-725 in CD45 domain I (DI), and 782-1041 in CD45 domain II (DII) (10). Scores greater than 5 are indicative of homology. Numbers below the diagonal represent the percentage identity expressed as the number of identities out of possible matches between residues in the aligned sequences depicted in Fig. 3.

1 and 2 establish clearly that the sequence similarities are indicative of a homologous relationship. The lowest degree of structural similarity was observed in the comparison of the N-terminal segment of PTPH1 with the homologous structure in murine talin; in this case the species difference may contribute to reducing the alignment score. The sequence of human talin was not available for comparison.

The intervening central segment of PTPH1 (residues 358-668) did not display homology with any sequence in the data base. However, a number of putative phosphorylation sites are found in this segment. The primary structure requirements for phosphorylation by casein kinase II include the presence of acidic residues surrounding the phosphorylated serine or threonine. An aspartic or glutamic acid three residues to the C terminus of the phosphate acceptor site is particularly crucial (18). The presence of prolyl residues preceding the phosphorylation site is also common. Six potential sites are indicated in Fig. 1. In addition, servl or threonyl residues phosphorylated by p34^{cdc2} are immediately followed by a prolyl residue, frequently in a sequence, Pol-Ser/Thr-Pro-Xaa-Bas (where Pol is a polar residue and Bas is a basic residue) (19). Ser-372, -381, and -435 display some of these features. It will be of interest to ascertain whether these sites are phosphorylated and the effect of phosphorylation on activity.

DISCUSSION

The structure of PTPH1 is summarized in Fig. 4. Its sequence implies that PTPH1 may act at junctions between the membrane and the cytoskeleton, such as in focal adhesions. Focal adhesion plaques are specialized regions of the plasma membrane through which cells in culture adhere to the external substrate (22). On their internal face these structures anchor actin stress fibers, which are important in determining cell shape. Within focal adhesions, the integrins provide the transmembrane link between components of the extracellular matrix, such as fibronectin and vitronectin, and the cytoskeleton. Their intracellular segments interact indirectly with actin cables through a multiprotein complex. Talin interacts with the cytoplasmic segment of the integrin β chain and also

Table 2. Evidence of homology between PTPH1 and members of the talin family of cytoskeletal-associated proteins

	PTPH1	Band 4.1	Ezrin	Talin
PTPH1	_	33.5	17.5	9.9
Band 4.1	37%	—	18.9	7.5
Ezrin	27%	29%		9.1
Talin	20%	16%	16%	—

Residues 30–357 in PTPH1 were compared pairwise with residues 2–323 in band 4.1 (37), residues 4–339 in ezrin (39), and residues 120–455 in talin (36). See Table 1 for other details.

binds to vinculin. Vinculin has been shown to interact with α -actinin, which can bind actin directly (for review, see ref. 22). However, this picture is undoubtedly incomplete. Microtubules and intermediate filaments may also terminate at focal adhesions, but the nature of the proteins that facilitate this interaction is unknown. Furthermore, similar but less-characterized structures have been implicated in attachment between neighboring cells and adherence to the extracellular matrix *in vivo*. Thus, additional focal adhesion-associated proteins remain to be described and their functions remain to be established.

Oncogenic transformation is frequently associated with a less-adherent rounded cellular morphology that results from a disruption of cytoskeletal integrity and reduction in the number of focal adhesions (23). Furthermore, residual adhesion plaques are invariably associated with the transforming PTK, for example, src, yes, or abl (23). In Rous sarcoma virus-transformed cells, for instance, it has been postulated that a contributing factor to the generation of the transformed phenotype is the aberrant phosphorylation by pp60^{v-src} of tyrosyl residues in key focal adhesion proteins (24-28). The phosphorylation of tyrosyl residues has been demonstrated in the β subunit of integrin (27), vinculin (29), talin (30), ezrin (31), and paxillin (32). Although a correlation has yet to be established between such phosphorylation and the appearance of the morphology of a transformed cell, evidence is accumulating to suggest that phosphorylation of the integrin β subunit is important in this regard (33). Staining with antibodies to phosphotyrosine has indicated that tyrosine phosphorylation of focal adhesion (34) and apical junction (35) proteins also occurs in nontransformed cells, suggesting that PTKs act at these sites during normal cell function. However, the kinases involved have yet to be identified. In view of the potentially disastrous effects of aberrant tyrosine phosphorylation, it is expected that the activity of these PTKs is tightly controlled. The localization of a PTP to these structures is one means by which such regulation could be achieved. We propose that PTPH1 may perform such a function.

Protein and cDNA sequence analyses have defined a family of proteins, including talin (36), band 4.1 (37), and ezrin (38, 39), that participate in the interaction between the membrane and cytoskeleton. They possess a homologous N-terminal domain that appears to associate with protein components in the plasma membrane. In band 4.1, which

	TM 37% 30%	
	42%	
	39%	
	PTPH1	
37%	Human band 4.1	
27%	Human ezrin	
20%		Mouse talin

FIG. 4. Schematic diagram illustrating the structure of PTPH1 and its relationship to other proteins. PTPH1 comprises three segments: an N-terminal segment (shaded box) with homology to the membrane localization domains in band 4.1, ezrin, and talin, a putative regulatory segment in the middle of the protein (open box), and a C-terminal PTP-related segment (solid box). TM denotes the transmembrane domain in CD45. The length of each protein is shown in proportion to the number of amino acid residues. For PTP1B, the full-length 435-residue protein (20, 21) is depicted. The molecule originally isolated from human placenta (2, 3) was a C-terminally truncated form. The homology among domains is indicated by percentage identity. The segments are defined in the legend to Table 1. promotes association of actin and spectrin in erythrocytes, this domain interacts with the transmembrane protein glycophorin (40). A similar model has been proposed for the interaction of talin with integrin (36). For ezrin, which displays a submembranous localization in brush border cells, the details of its interaction with other proteins remain to be established. The presence of a homologous domain in PTPH1 suggests that this PTP will also be located in submembranous structures, such as focal adhesions, where it may modulate the level of phosphotyrosine in proteins.

Considering the apparent correlation between Rous sarcoma virus-induced transformation and the cytoskeletal association of $pp60^{src}$ (24–28), we propose that PTPH1 is also an excellent candidate PTP with which, through overexpression, to achieve a reversion of src-induced transformation and, furthermore, to delineate the precise role of tyrosine phosphorylation in the morphological changes induced by src. In addition this PTP clearly has potential to function as a growth suppressor. It has been noted that treatment of NRK cells with vanadate, the actions of which include the inhibition of PTPs, enhanced the level of cellular phosphotyrosine and lead to production of a transformed phenotype (41). Thus, the inactivation or deletion of PTPH1 could conceivably in itself be sufficient to generate a transformed cell.

The PTPs form a rapidly expanding family (5) and it is anticipated that the various isoforms will have specific functions in vivo. The structure of PTPH1 may illustrate a general theme among the PTPs; within the protein, distinct structural motifs may, at least in part, control specificity by restricting intracellular localization. We propose that PTPH1 acts at the junction between the cytoskeleton and the plasma membrane and plays a role in controlling cytoskeletal integrity. The occurrence of a signal peptide sequence and a transmembrane domain in CD45 directs it to span the membrane (4). The C-terminal noncatalytic segment of the low molecular weight cytoplasmic PTPs also appears to direct association with the particulate fraction of cell extracts (42). In addition such structural motifs may determine how the activity of the catalytic domain is controlled. Thus, for the receptor-like forms the binding of ligands to the extracellular segments may modulate activity. The C-terminal segment of the low molecular weight PTP appears to repress the activity of the catalytic domain (42). If PTPH1 is localized to focal adhesions, it should also be regulated to permit normal tyrosine phosphorylation at such sites in nontransformed cells. Phosphorylation of serine/threonine residues in the central segment of the protein (Fig. 1) may directly modulate activity. In addition, by analogy with band 4.1 whose affinity for glycophorin is regulated by phosphatidylinositol 4,5-bisphosphate (43) it is possible that the localization of PTPH1 may be altered with phosphatidylinositol turnover.

We are very grateful to Bruce Stillman for advice and critically reading the manuscript. Technical assistance was provided by Peter Guida and Jin Yang. We also thank Harry Charbonneau, Yue Xiong, and Catriona Simpson for helpful discussions. This work was supported in part by funds from the Long Island Biological Association.

- 1. Hunter, T. (1989) Curr. Opin. Cell Biol. 1, 1168-1181.
- Tonks, N. K., Diltz, C. D. & Fischer, E. H. (1988) J. Biol. Chem. 263, 6722-6730.
- Tonks, N. K., Diltz, C. D. & Fischer, E. H. (1988) J. Biol. Chem. 263, 6731-6737.
- Charbonneau, H., Tonks, N. K., Walsh, K. A. & Fischer, E. H. (1988) Proc. Natl. Acad. Sci. USA 85, 7182–7186.
- 5. Tonks, N. K. (1990) Curr. Opin. Cell Biol. 2, 1114-1124.
- 6. 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.

- Streuli, M., Krueger, N. X., Hall, L. R., Schlossman, S. F. & Saito, H. (1988) J. Exp. Med. 168, 1523-1530.
- Koretzky, G. A., Picus, J., Thomas, M. L. & Weiss, A. (1990) Nature (London) 346, 66–68.
- Gould, K. L., Moreno, S., Tonks, N. K. & Nurse, P. (1990) Science 250, 1573–1576.
- Charbonneau, H., Tonks, N. K., Kumar, S., Diltz, C. D., Harrylock, M., Cool, D. E., Krebs, E. G., Fischer, E. H. & Walsh, K. A. (1989) Proc. Natl. Acad. Sci. USA 86, 5252-5256.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
- 12. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- Pearson, W. R. & Lipman, D. J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444–2448.
- 14. Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) Methods Enzymol. 91, 524-545.
- 15. Kozak, M. (1989) J. Cell Biol. 108, 229-241.
- 16. Tonks, N. K., Charbonneau, H., Diltz, C. D., Fischer, E. H. & Walsh, K. A. (1988) *Biochemistry* 27, 8695-8701.
- Streuli, M., Kreuger, N. X., Thai, T., Tang, M. & Saito, H. (1990) EMBO J. 9, 2399–2407.
- Kuenzel, E. A., Mulligan, J. A., Sommercorn, J. & Krebs, E. G. (1987) J. Biol. Chem. 262, 9136–9140.
- 19. Moreno, S. & Nurse, P. (1990) Cell 61, 549-551.
- Chernoff, J., Schievella, A. R., Jost, C. A., Erikson, R. L. & Neel, B. G. (1990) Proc. Natl. Acad. Sci. USA 87, 2735–2739.
- Brown-Shimer, S., Johnson, K. A., Lawrence, J. B., Johnson, C., Bruskin, A., Green, N. R. & Hill, D. E. (1990) Proc. Natl. Acad. Sci. USA 87, 5148-5152.
- 22. Burridge, K., Fath, K., Kelly, C., Nuckolls, G. & Turner, C. (1988) Annu. Rev. Cell Biol. 4, 487-525.
- 23. Kellie, S. (1988) BioEssays 8, 25-30.
- 24. Burr, J. G., Dreyfuss, G., Penman, S. & Buchanan, J. M. (1980) Proc. Natl. Acad. Sci. USA 77, 3484–3488.
- 25. Shriver, K. & Rohrschneider, L. R. (1981) Cold Spring Harbor Conf. Cell. Proliferation 8, 1247-1262.
- 26. Hamaguchi, M. & Hanafusa, H. (1987) Proc. Natl. Acad. Sci. USA 84, 2312-2316.
- Tapley, P., Horwitz, A. S., Buck, C., Duggan, K. & Rohrschneider, L. (1989) Oncogene 4, 325-333.
- Glenney, J. R., Jr., & Zokas, L. (1989) J. Cell Biol. 108, 2401-2408.
- Sefton, B. M., Hunter, T., Ball, E. H. & Singer, S. J. (1981) Cell 24, 165–174.
- Pasquale, E. B., Maher, P. A. & Singer, S. J. (1986) Proc. Natl. Acad. Sci. USA 83, 5507-5511.
- 31. Kamps, M. P., Buss, J. E. & Sefton, B. M. (1986) Cell 45, 105-112.
- Turner, C. E., Glenney, J. R., Jr., & Burridge, K. (1990) J. Cell Biol. 111, 1059–1068.
- 33. Horvath, A. V., Elmore, M. A. & Kellie, S. (1990) Oncogene 5, 1349–1357.
- Maher, P. A., Pasquale, E. B., Wang, J. Y. J. & Singer, S. J. (1985) Proc. Natl. Acad. Sci. USA 82, 6576–6580.
- 35. Takata, K. & Singer, S. J. (1988) J. Cell Biol. 106, 1757-1764.
- Rees, D. J. G., Ades, S. E., Singer, S. J. & Hynes, R. O. (1990) Nature (London) 347, 685-689.
- Conboy, J., Kan, Y. W., Shohet, S. B. & Mohandas, N. (1986) Proc. Natl. Acad. Sci. USA 83, 9512–9516.
- Turunen, O., Winqvist, R., Pakkanen, R., Grzeschik, K. H., Wahlstrom, T. & Vaheri, A. (1989) J. Biol. Chem. 264, 16727– 16732.
- Gould, K. L., Bretscher, A., Esch, F. S. & Hunter, T. (1989) EMBO J. 8, 4133-4142.
- 40. Bennet, V. (1989) Biochim. Biophys. Acta 988, 107-121.
- 41. Klarlund, J. K. (1985) Cell 41, 707-717.
- Cool, D. E., Tonks, N. K., Charbonneau, H., Fischer, E. H. & Krebs, E. G. (1990) Proc. Natl. Acad. Sci. USA 87, 7280– 7284.
- 43. Anderson, R. A. & Marchesi, V. T. (1985) Nature (London) 318, 295-298.