Isolation of a src homology 2-containing tyrosine phosphatase  

(cell signaling/protein phosphorylation/tyrosine kinases/megakaryocytes)

Jorge Flutzky*++, Benjamin G. Neel§, and Robert D. Rosenberg†$

*Division of Cardiology, Department of Medicine, The Brigham and Women's Hospital, Boston, MA 02118; †Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; and §Molecular Medicine Unit, Beth Israel Hospital, Boston, MA 02215

Communicated by Eugene Braunwald, November 4, 1991

ABSTRACT Tyrosine phosphorylation is controlled by the opposing actions of tyrosine kinases and phosphotyrosine phosphatases (PTPs). src homology 2 domains (SH2) are found in several types of signaling proteins, including some tyrosine kinases. These domains bind phosphotyrosyl proteins and thus help promote signal transduction. Using mixed oligonucleotide-directed polymerase chain reactions, two previously undescribed rat PTP cDNA fragments were generated. Through subsequent screening of rat megakaryocyte and human erythroleukemia libraries, we obtained a full-length coding sequence for one of these fragments. This CDNA, SH-PTP1, encodes a tyrosine phosphatase containing two highly conserved SH2 domains. SH-PTP1, with a 2.4-kilobase mRNA, is a predicted open reading frame of 595 amino acids, and a structure suggesting a nontransmembrane protein, is expressed primarily in hematopoietic and epithelial cells. When expressed in Escherichia coli, SH-PTP1 possesses PTP activity. The structure of SH-PTP1 establishes an additional branch of the tyrosine phosphatase family and suggests mechanisms through which tyrosine phosphatasas might participate in signal transduction pathways.

The regulation of tyrosine phosphorylation is important in both control of normal cellular processes, such as cell growth, the cell cycle, and differentiation, as well as pathological events, such as malignant transformation (1, 2). Many growth factor receptors are transmembrane tyrosine kinases, which phosphorylate themselves and other proteins in response to ligands (3). Phosphorylated growth factor receptors are linked to downstream cellular events by signaling molecules such as GTPase-activating protein (GAP; ref. 4), phospholipase C-γ (5, 6), the members of the src tyrosine kinase family (7), and the regulatory subunit of phosphatidylinositol 3-kinase (8–10). These latter proteins contain src homology 2 (SH2) domains, conserved stretches of ≈100 amino acids that appear to direct these signaling proteins to growth factor-activated phosphotyrosyl proteins (7–13). In addition, SH2 domains may regulate the activity of src-like kinases (2, 13) as well as participate in cytoskeletal interactions (14).

The level of tyrosine phosphorylation is determined by the opposing actions of tyrosine kinases and phosphotyrosine phosphatases (PTPs). PTPs are a family of conserved proteins, which have previously been grouped into two types (15). Nontransmembrane PTPs, such as PTPB (16–18) and T-cell PTP (19), have a single conserved phosphate domain. Transmembrane PTPs, such as CD45 (20) and LAR (21), usually consist of two conserved cytoplasmic phosphate domains, a transmembrane region, and a nonconserved extracellular domain (22). Analogous to receptor-linked tyrosine kinases, these extracellular domains are thought to transduce as-yet-undefined signals. The function and control of nontransmembrane PTPs remains obscure.

We were interested in isolating tyrosine phosphatases from megakaryocytes. This interest stemmed from several interesting features of these cells: (i) the property of nuclear duplication without division, implying unusual cell cycle control (23); (ii) high levels of tyrosine kinase activity (24); (iii) rapid changes in tyrosine phosphorylation seen in platelets in response to physiologic agonists such as thrombin (25, 26). Using mixed oligonucleotide primers to conserved PTP domains to direct polymerase chain reactions (PCRs), we have isolated an additional member of the nontransmembrane PTP family. This cDNA, named SH-PTP1, encodes two SH2 domains and is expressed in a tissue-restricted manner. SH-PTP1 has potentially important implications for both the regulation of nontransmembrane PTPs and mechanisms of signal transduction. 4

MATERIALS AND METHODS

PCRs. Degenerate mixed oligonucleotides for PCR were derived from two highly conserved regions within the PTP family. The sense primer, corresponding to the amino acid sequence KC(A/D/H)Q(F/E)YWFP, and the antisense primer, derived from the amino acid sequence VHCSAG(V/I)G, had respective degeneracies of 512 and 8192. These were synthesized on an Applied Biosystems 380B DNA synthesizer and purified by using nucelic acid purification columns (NEN). Total DNA (1 μg) from a rat megakaryocyte cDNA library (27) was placed at 100°C for 10 min and then cooled on ice. PCRs were carried out in 10 mM Tris-HCl, pH 8.3/50 mM KCl/200 μM each dNTP/10 μM each primer/2 mM MgCl2/2.5 units of Taq polymerase. DNA amplification was performed for 30 cycles of denaturation (94°C; 1 min), annealing (45°C; 2 min), and extension (72°C; 3 min), with a final extension (72°C; 10 min). DNA was purified from individual bands via low melting point agarose/glass milk (Bio 101, La Jolla, CA), subcloned into pUC19, and sequenced by the dideoxynucleotide chain-termination method (28).

Isolation of PTP Clones. Two ≈300-base-pair (bp) PCR fragments identified by sequence analysis as representing newly discovered PTPs, rat SH-PTP (rSH-PTP) and megakaryocyte clone 2 PTP (M2PTP), were labeled with [32P]dCTP (NEN) by the random-primer method (29). One million plaques each of the rat megakaryocyte cDNA library, a human lung library (Clontech), and a human erythroleukemia (HEL) library (gift of M. Poncz, University of Pennsylvania, Philadelphia), all in λgt11, were screened by standard techniques; positive clones were plaque purified and bacteriophage DNA was prepared (30). No further M2PTP clones were identified. Clones detected using the rSH-PTP PCR fragment as a probe included a partial rat cDNA clone,

Abbreviations: SH2, src homology 2; PTP, phosphotyrosine phosphatase; GAP, GTPase-activating protein; M2PTP, megakaryocyte clone 2 PTP; rSH-PTP, rat SH-PTP; hSH-PTP, human SH-PTP.

To whom reprint requests should be addressed.

The sequence reported in this paper has been deposited in the GenBank data base (accession no. M77273).
rSH-PTP1 (1747 bp), and several full-length human clones, hSH-PTP1 (2145 bp; see Results).

DNA Sequencing/ Homology Analysis. The EcoRI inserts from all bacteriophage clones identified were subcloned into pUC19 and sequenced with oligonucleotide primers. Both DNA and derived amino acid sequence data were analyzed for similarity to one another as well as to other GenBank entries (e.g., the PTP family) using Computer Genetics Group (Madison, WI) programs (FASTA and BESTFIT).

Northern/Southern Blot Analysis. Total RNA from rat tissues as well as various cell lines was prepared by the guanidinium isothiocyanate method (31). Northern blots using bacteriophage membranes (ICN) were hybridized for at least 16 hr at 65°C with labeled M2PTP and hSH-PTP1 as probes (32). The cell line Northern blot, also hybridized with hSH-PTP1, was provided by Bing Lim (Beth Israel Hospital, Boston). ATCC numbers for the cell lines used were as follows: HEL, TIB 180; KG-1, CCL 246; HL60, CCL 240; HepG2, HB 8065; Calu-1, HTB 54; SK-MEL2, HTB 68; Hs294T, HTB 140; HeLa, CCL 2. The OCR1 erythroid cell line was developed by B. Lim. The "Brown" sample was from immortalized B-lymphoid cells. Southern blot analysis was done using standard techniques (30) with radiolabeled full-length hSH-PTP1 cDNA.

Bacterial Fusion Protein Expression/PTP Assay. Full-length hSH-PTP1 coding sequence was subcloned via an EcoRI fragment into pGEX-2T (Pharmacia) to generate a fusion protein containing glutathione S-transferase plus a 10-amino acid spacer fused to the complete hSH-PTP coding region. For GEX fusion protein expression, freshly diluted overnight cultures of Escherichia coli were induced at mid-logarithmic phase with 1 mM isopropyl β-D-thiogalactopyranoside. Lysates were prepared from 1 ml of induced culture by resuspending bacterial pellets in 150 µl of phosphate-buffered saline with 1% Nonidet P-40 and sonicating (0°C) for 3 min. Cultures containing the pGEX-2T vector, expressing glutathione S-transferase alone, were treated similarly. PTP

Fig. 1. Sequence of a PTP containing two SH2 domains. (A) Amino acid sequence comparison between SH-PTP1 (rat and human) and M2PTP phosphatase domains and representative domains of the PTP family. Bold lines indicate identical residues; conservative substitutions, marked by double dots, were generated by the program BESTFIT. In proteins with two PTP domains, the N-terminal domain was used. Invariable PTP residues are indicated by boldface type, and highly conserved residues are indicated by asterisks. Alignment is as described (22). (B) Nucleotide and predicted protein sequence of SH-PTP1. The DNA sequence is numbered on the left and the corresponding predicted amino acid sequence is numbered on the right. The two SH2 domains are underlined. Positions of the oligonucleotides used to prime the rSH-PTP1 PCR fragment are in boldface type. A single potential phosphorylation site for cdc2 kinase is indicated by double underlining (amino acids 160–163). Two potential tyrosine phosphorylation sites are indicated by jagged underlining (amino acids 533–536 and 561–565) (35–37). Asterisks denote the stop codon. The polyadenylation consensus signal is marked by arrowheads. (C) Schematic diagram of predicted protein structure of SH-PTP1. Relative positions of the two SH2 and the PTP domain are indicated. Numbers refer to positions in the predicted amino acid sequence.
assays were performed by using as model substrates either Raytide (Oncogene Science, Uniondale, NY) or myelin basic protein (Sigma). These substrates were phosphorylated on their tyrosine residues using recombinant p42 MAP kinase (Oncogene Science) and [γ-32P]ATP (200 μCi; 1 Ci = 37 GBq). Trichloroacetic acid precipitated and washed (33). PTP assays [in 25 mM imidazole, pH 7.2/bovine serum albumin (1 mg/ml)/10 mM dithiothreitol/100 mM phosphorylated substrate/bacterial protein as indicated] were incubated at 30°C for various times (see Fig. 4). Dephosphorylation was quantitated by a charcoal binding assay described by Streuli et al. (33).

RESULTS

Cloning of SH-PTP1. DNA from a rat megakaryocyte cDNA library (27) was used as a template in PCRs primed with degenerate oligonucleotides derived from two regions highly conserved among known PTPs. These reactions resulted in multiple bands, which were individually isolated, cloned, and sequenced. Products with PTP similarity were derived exclusively from the 3′ end of PCR band, corresponding to the expected intramembrane distance in known PTPs (data not shown). These products included sequences corresponding to known PTPs such as CD45, LAR, LRP (22), and two previously undescribed fragments, termed rSH-PTP and M2PTP, with similarity to one another as well as to the PTP family (Fig. 1A). These fragments served as probes to screen the rat megakaryocyte cDNA library. Although no larger M2PTP clones were obtained, a partial rSH-PTP cDNA clone (rSH-PTP1; 1747 bp) was identified. Based on tissue distribution studies (see below), we used the rat clone to screen human lung and human erythroleukemia cDNA libraries. After plaque purification and subcloning, full-length sequence information for this gene, now termed hSH-PTP1, was obtained via overlapping clones from each library.

The sequence of hSH-PTP1 is shown in Fig. 1B. There is a strong consensus site (34) for translation initiation (AG-GATGG), with upstream termination codons in all frames. This is followed by a single open reading frame predicting a protein of 595 amino acids. A consensus signal for polyadenylation (AUA AAA) is present at nucleotide 2126.

The predicted protein sequence has several notable features. There is a highly conserved PTP domain, retaining all 26 invariant and many of the conserved amino acids of the PTP family (Fig. 1B). The rat and human PTP amino acid domains are >95% identical, indicating strong evolutionary conservation (Fig. 1A). hSH-PTP1 lacks a signal sequence or potential transmembrane region, indicating that it is a nontransmembrane PTP. However, its PTP domain has greatest amino acid sequence similarity to CD45 (54%), a transmembrane PTP, rather than the nontransmembrane PTP1B (23%). This suggests that hSH-PTP1 represents a separate branch of the PTP family. More importantly, the 5′ coding sequence contains two regions with strong similarity to SH2 domains (Fig. 1B and C). The homology between these two regions of hSH-PTP1 and

![Fig. 3. Tyrosine phosphatase activity of hSH-PTP1 expressed in bacteria: Time course of dephosphorylation of Raytide. Equal amounts of total protein (500 ng for GEX and GEX–SH-PTP; 50 ng for GEX–SH-PTP, 1:10) from E. coli expressing glutathione S-transferase (GEX) or the glutathione S-transferase–SH-PTP1 fusion (GEX–SH-PTP) were incubated with 32P Raytide. At the indicated times, triplicate points were assayed for 32P release by a charcoal binding assay (33). The mean ± SEM is indicated for each time point.](image)

known SH2 proteins spans the entire SH2 domain (Fig. 2), clustering within the five conserved subdomains described by Pawson and coworkers (13). As with other SH2-containing proteins, the conserved subdomains of hSH-PTP1 are interrupted by glycine/proline-rich variable regions. Three invariant residues with conserved locations have been described in all SH2 proteins; all three are present at their appropriate positions in hSH-PTP1 (Fig. 2). Moreover, two of the three basic amino acid residues proposed as possible phosphotyrosine recognition sites are also present (Fig. 2). Overall, the two SH2 domains in hSH-PTP1 are most similar to the corresponding SH2 domains of GAP (45% identity). There is no similarity to SH3 domains, regions found in some other SH2-containing proteins (13). Finally, hSH-PTP1 contains several potential phosphorylation sites (Fig. 1B). There is a single site, located in the second SH2 domain, that conforms to the consensus target sequence for the cell cycle-regulated kinase cdc2 (35, 36). Numerous possible phosphorylation sites for other serine/threonine kinases exist throughout the molecule. There are also two potential tyrosine phosphorylation sites near the C terminus (37).

Phosphotyrosine Phosphatase Activity. To confirm that hSH-PTP1 encodes an active PTP, the hSH-PTP1 coding sequence was introduced into the bacterial expression vector pGEX-2T (Fig. 3). Recombinant clones expressing the predicted glutathione S-transferase–SH-PTP1 fusion protein (GEX–SH-PTP) were obtained. GEX–SH-PTP lysates displayed substantially increased phosphatase activity against two model substrates, the synthetic peptide Raytide (Fig. 3) and myelin basic protein (data not shown), as compared to controls expressing glutathione S-transferase alone (Fig. 3). PTP activity was linear over time and proportional to protein
Similar results were obtained with fusion protein partially purified on glutathione agarose beads (data not shown). As predicted for a tyrosine phosphatase, GEX-SH-PTP activity was strongly inhibited by 200 μM sodium orthovanadate, but not 50 mM sodium fluoride; there was no significant activity against the serine-phosphorylated substrate Kemptide (Sigma; data not shown).

**SH-PTP1 Gene Structure and Expression.** Southern blots of human DNA hybridized with hSH-PTP1 cDNA probe revealed a simple pattern consistent with a relatively small, single copy gene (Fig. 4A). Northern blot analysis revealed a 2400-bp mRNA. hSH-PTP1 is expressed at high levels in all hematopoietic lineages as evidenced by strong signals in erythroid (OCR1), erythromegakaryoblastic (HEL), lymphoid (Brown), and myeloid (KG-1, HL60) cell lines (Fig. 4B). Lower but significant signals are seen in several epithelial cell lines (HepG2, HeLa, Calu-1; Fig. 4B). Analysis of specific tissues revealed strong signals in megakaryocytes/platelets and, surprisingly, in lung; expression in other tissues was substantially lower (Fig. 4C). hSH-PTP1 is also expressed in primary tracheobronchial epithelial cells (data not shown). M2PTP message is shown in the tissue distribution blots (Fig. 4C); high level expression of this PTP appears more restricted to megakaryocytes.

**DISCUSSION**

Our aim was to isolate PTPs from megakaryocytes. By using probes generated by degenerate oligomer-based PCR for traditional library screening, we identified two previously undescribed PTPs. M2PTP, a PCR fragment, corresponds to a message that appears restricted to the megakaryocytic lineage; a full-length cDNA was found for hSH-PTP1, which is expressed primarily in hematopoietic cells. The hSH-PTP1 gene encodes two SH2 domains. The existence of SH2 domains in a tyrosine phosphatase has implications for understanding the PTP and SH2 gene families, as well as the roles SH-PTP1 might play in signal transduction.

Sequence features of hSH-PTP1 suggest that the simple division of the PTP family into transmembrane and nontransmembrane forms should be revised. Although hSH-PTP1 is a member of the nontransmembrane class, its greater degree of sequence similarity to transmembrane PTPs, as well as the presence of SH2 domains, establishes a discernible new branch of the tyrosine phosphatase family. This is supported by our recent isolation of another distinct SH2-encoding PTP (R. M. Freeman, J.P., and B.G.N., unpublished data). Future studies using mixed oligomers to SH2 and PTP domains should address the possibility of yet other SH2-containing PTPs. In addition, hSH-PTP1 further strengthens the remarkable structural analogies between the tyrosine phosphatase and tyrosine kinase families (transmembrane and nontransmembrane proteins; SH2- and non-SH2-containing proteins; see refs. 1 and 2).

The role of SH2 domains in other signaling proteins suggests several possibilities for their function in hSH-PTP1. SH2 domains appear to direct two types of protein–protein interactions. First, they target signaling proteins such as GAP and phosphatidylinositol 3-kinase to autophosphorylation sites of activated growth factor receptors (2, 13). The resultant assembled complex may activate signaling molecules or make them accessible to specific substrates. Second, the SH2 domains of src-like kinases are thought to interact with their own tyrosine phosphorylation sites to negatively regulate enzyme activity (2, 13, 38–42).
The SH2 domains of PTPs may participate in similar regulatory mechanisms. It seems unlikely that these domains simply localize hSH-PTP1 to its phosphotyrosine substrates since all previously identified, catalytically active PTPs lack such a region. Instead, this domain could target PTPs to signal transduction complexes for signal termination or amplification. Termination could occur by reversing the tyrosine phosphorylation of receptor-activated signaling proteins. Amplification might take place by releasing already bound signaling molecules, freeing them to find other substrates within the cell. Alternatively, SH2 domains could allow PTPs to interact with src-like kinases to dephosphorylate their negative regulatory tyrosine residues. Given the potential tyrosine phosphorylation sites in the C terminus of hSH-PTP1, this possibility is of particular interest since specific coupling between hSH-PTP1 and an SH2-containing tyrosine kinase could thus occur.

The biologic role of hSH-PTP1 exerted via these or other mechanisms is unknown. The potential cdc2 phosphorylation site in hSH-PTP1 suggests its activity may be cell cycle regulated. PTPs have already been implicated in cell cycle control; transition of cells from G2 to M phase appears to require dephosphorylation of a tyrosine in p34<sup>cdk2</sup> (43). Overexpression of a truncated form of T-cell PTP in baby hamster kidney cells apparently affects cytokinesis, resulting in multinucleated cells (44). The ability to induce multinucleation would be of particular interest in megakaryocytes, given their higher ploidy states. Alternatively, the restriction of hSH-PTP1 expression to tissues that are actively replenished from stem cells (hematopoietic and epithelial cells) suggests a possible role in cell growth or differentiation. The potential binding of SH-PTP1 to activated growth factor receptors through its SH2 domain could have such an implication. The ability to independently manipulate each SH2 domain and the PTP domain of hSH-PTP1 suggests straightforward approaches to resolving these issues.

Note. While this paper was in review, a human cDNA clone (PTP1C) similar to hSH-PTP1 was reported by Shen et al. (45). Three differences are noteworthy. First, PTP1C was isolated from a human breast carcinoma cDNA library (ZR-75-1). Although no tissue distribution data were provided by those authors, our studies indicate its predominant expression in hematopoietic cells. Second, the reported PTP1C sequence differs from SH-PTP1 in its absence of a guanine at base pair 1912. The predicted protein would be 13 amino acids longer than SH-PTP1, differing in the last 18 amino acids. Our human and rat SH-PTP1 cDNA sequences are highly conserved and both contain a stretch of similarity at base 2062; this conservation suggests that the 101 sequence reported here is correct. Third, comparison of PTP1C to SH-PTP1 from the 5′ untranslated sequence through to the first SH2 domain (base pair 154) reveals a completely different sequence (except for the start site). Absolute identity begins with the start of the SH2 domain (amino acid 5 of SH-PTP1). These sequence differences suggest that the PTP1C cDNA may represent a sequence variant, an extremely homologous but distinct gene, or an altered version of SH-PTP1 in the breast carcinoma line ZR-75-1. The latter possibility is of interest given its potential pathogenic significance.

We thank Pamela Beahm and David Beeler for excellent technical assistance and Katya Ravid for helpful suggestions. Gifts from Dr. M. Auger of the Hsu library and Dr. Bing Lim of the Northern blots were appreciated. This work was supported by National Institutes of Health Grant RO1-CA 49152-02 to B.G.N. and National Institutes of Health Grants HL 33014 and R37 HL 39753 to R.D.R.