Research article

Genomic organization and alternative splicing of the human and mouse RPTPρ genes

Julie A Besco[1], Adrienne Frostholm[1], Magdalena C Popesco[1], Arthur HM Burghes[2] and Andrej Rotter*[1]

Address: [1]Departments of Pharmacology and and [2]Molecular and Cellular Biochemistry The Ohio State University Columbus, OH 43210, USA

E-mail: Julie A Besco - besco.1@osu.edu; Adrienne Frostholm - frostholm.1@osu.edu; Magdalena C Popesco - popesco.1@osu.edu; Arthur HM Burghes - burghes.1@osu.edu; Andrej Rotter* - rotter.1@osu.edu

*Corresponding author

Published: 8 June 2001

BMC Genomics 2001, **2**:1

Received: 29 April 2001
Accepted: 8 June 2001

This article is available from: http://www.biomedcentral.com/1471-2164/2/1

Abstract

Background: Receptor protein tyrosine phosphatase rho (RPTPρ, gene symbol PTPRT) is a member of the type IIB RPTP family. These transmembrane molecules have been linked to signal transduction, cell adhesion and neurite extension. The extracellular segment contains MAM, Ig-like and fibronectin type III domains, and the intracellular segment contains two phosphatase domains. The human RPTPρ gene is located on chromosome 20q12-13.1, and the mouse gene is located on a syntenic region of chromosome 2. RPTPρ expression is restricted to the central nervous system.

Results: The cloning of the mouse cDNA, identification of alternatively spliced exons, detection of an 8 kb 3'-UTR, and the genomic organization of human and mouse RPTPρ genes are described. The two genes are comprised of at least 33 exons. Both RPTPρ genes span over 1 Mbp and are the largest RPTP genes characterized. Exons encoding the extracellular segment through the intracellular juxtamembrane 'wedge' region are widely spaced, with introns ranging from 9.7 to 303.7 kb. In contrast, exons encoding the two phosphatase domains are more tightly clustered, with 15 exons spanning ~ 60 kb, and introns ranging in size from 0.6 kb to 13.1 kb. Phase 0 introns predominate in the intracellular, and phase 1 in the extracellular segment.

Conclusions: We report the first genomic characterization of a RPTP type IIB gene. Alternatively spliced variants may result in different RPTPρ isoforms. Our findings suggest that RPTPρ extracellular and intracellular segments originated as separate modular proteins that fused into a single transmembrane molecule during a later evolutionary period.

Background

Protein tyrosine phosphorylation regulates many important cellular functions including signal transduction, growth, differentiation, cell adhesion and axon guidance. The balance between protein tyrosine kinase and phosphatase activity is an integral part of this regulatory mechanism. A large number of protein tyrosine phosphatases have been identified, which fall into the broad categories of cytoplasmic and receptor-like molecules. All receptor-like protein tyrosine phosphatases (RPTPs) contain an extracellular region, a single transmembrane segment and at least one intracellular catalytic domain. They have been subdivided into several classes based on the structure of their extracellular segments (Figure 1). A combination of immunoglobulin-like (Ig) domains and fibronectin type III (FN-III) repeats in the ectodomain

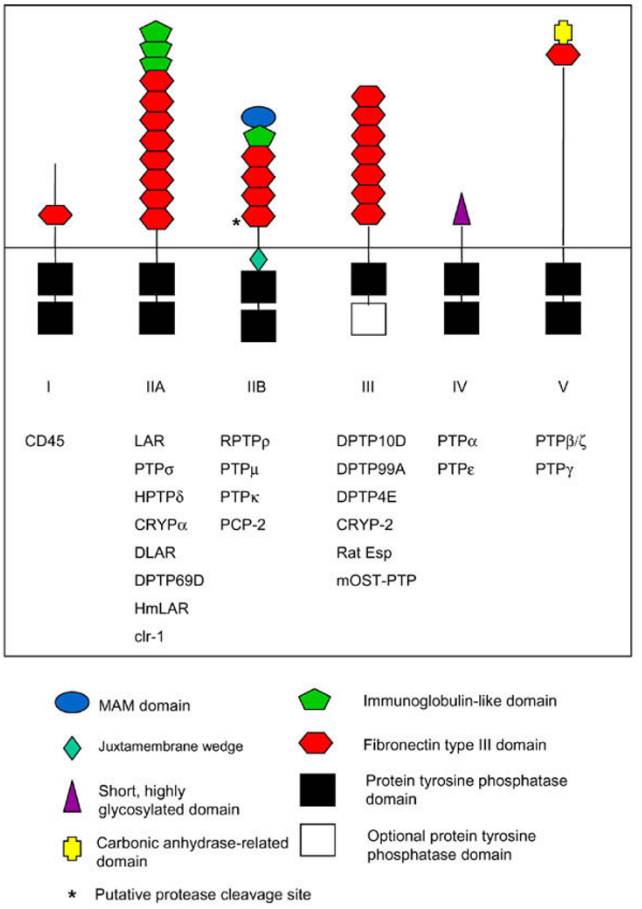


Figure 1
Domain structure of the receptor-like protein tyrosine phosphatase family. Variations in the extracellular domain structure separate the RPTP family of transmembrane proteins into five major classes (I-V). RPTPρ is a member of the type IIB subfamily of RPTPs that includes RPTPκ, μ and PCP-2.

defines the type II class of RPTPs. An additional feature of type II RPTPs is a potential proteolytic cleavage site within the membrane-proximal FN-III repeat. Upon cleavage, extracellular N-terminal and predominantly intracellular, membrane bound C-terminal segments are generated, which remain non-covalently associated [1] A subset of the type II class, identified previously as type IIB RPTPs [2], is characterized by the presence of an N-terminal MAM domain.

Currently, four type IIB phosphatases (PTPμ, PTPκ, PCP-2 and RPTPρ) have been reported. The hPCP-2 [3], hPTPκ [4], and hPTPμ [5] RPTPs are located on human chromosomes 1, 6 and 18, respectively, and hRPTPρ is located on chromosome 20 [6]. Several additional human RPTPs (PTPπ, PTPψ, hPTP-J, PTPRO) share very high sequence similarity (>98%) with PCP-2, and are likely to represent the same gene (Unigene database, [http://www.ncbi.nlm.nih.gov/unigene]). There are, in addition, several murine homologues of the four human genes: mPTPκ (Genbank #NM 008983), mPTPμ (#NM008984), mRPTPρ (#AF152556), mRPTPρ -1 and mRPTPρ -2 (# AF162856/7), mRPTPmam4 (#NM 021464), mPTPf (#D88187) and mPTPλ (#U55057). The latter two are likely to be murine homologues of hPCP-2, and mRPTPmam4 is the same gene as mRPTPρ.

RPTPρ is the most recently isolated member of the IIB family [6, 7]. Northern blot and in situ hybridization studies have shown that RPTPρ is largely restricted to the central nervous system [6]. Within the CNS, expression is developmentally regulated and, in the mouse, delineates a unique boundary region in the granule cell layer of the cerebellar cortex [7]. Motifs in the RPTPρ extracellular segment (MAM, Ig and FN-III domains) are commonly found in cell adhesion molecules. The two phosphatase domains in the intracellular segment suggest that RPTPρ, like other members of the RPTP family, is involved in signal transduction through protein tyrosine dephosphorylation.

The human RPTPρ gene has been mapped to chromosome 20q12-13.1 [6]; it is located between anchor markers D20S99 and D20S96, and is flanked by the phospholipase C gamma 1 and splicing factor SRp55-2 genes. The mouse gene maps to a syntenic region at 93 cM on mouse chromosome 2, a region closely linked to *Pltp* and flanked by the markers, D2Mit22 and D2Mit52. To date, only portions of the human RPTPκ, RPTPμ and PCP-2 genes have been sequenced, however, the region encompassing the human RPTPρ gene has been sequenced in its entirety (Chromosome 20 sequencing group, Sanger Centre), but it is not, as yet, fully assembled and annotated. The mouse chromosomal region containing the RPTPρ gene has been sequenced (Celera Discovery System), but it is also largely unassembled. In this report, we describe the cloning of the mouse cDNA, the identification of an unusually long 3' UTR, the identification of alternatively spliced exons, and the genomic organization of the human and mouse RPTPρ genes.

Results and Discussion

The nucleotide sequence and domain structure of human RPTPρ

The nucleotide sequence of hRPTPρ cDNA predicts a 1463AA polypeptide containing at least eight domains. The polypeptide is comprised of extracellular and intracellular segments. The extracellular segment contains a signal peptide (AA 1-25), a MAM (meprin, A5 (neuropilin), RPTPmu) domain (AA 32-191), an Ig-like domain (AA 210-266) and four FN-III repeats (AA 286-369, 382-471, 483-576 and 593-675). A potential proteolytic cleav-

TABLE 1

exon#	domain	exon size	3' splice site		5' splice site	phase	intron size
1	25aa sig pep	272	GCC GCC CGG	ATG GCG AGC....AGC GCC CCA G M A S S A P G	gtgagtgcg	1	303715
2	MAMa	126	tgtctacag	GT GGC TGT TCC....GTG CCC ACA G G C S V P T G	gtatgtgat	1	93173
3	MAMb	272	ggtctccag	GA TCT TTC ATG.....TTC TAT CAG S F M F Y Q	gtatggcat	0	10896
4	MAMc	82	ttctatcag	GTG ATA TTT.....CAT CCA TGC A V I F H P C R	gtaagtctt	1	8666
5	Iga	116	ctctcttag	GA AAA GCA CCT.....TGG CTC CAG K A P W L Q	gtaagcatg	0	14799
6	Igb	175	gttctgcag	CAA TGG AAT.....ATC GTG AAA G Q W N I V K E	gtgttaagc	1	78304
7	FN#1	294	gtcttacag	AG CCT CCC ACG.....AAG TGT GCA G P P T K C A D	gtaagatt	1	205304
8	FN#2	297	gttttgcag	AT CCG GTA CAT.....GAG GAA GAC G P V H E E D V	gtgagcaag	1	23935
9	FN#3a	110	attctgtag	TT CCA GGA GCT.....CTC TAC GAG P G A L Y E	gtaaggagg	0	95935
10	FN#3b	202	gtcctacag	ATC AAC TAC.....AAA ATT TCA G I N Y K I S A	gtatctctg	1	1355
11	FN#4a	103	cctgtccag	CT CCA TCC ATG.....GCT CCT GTC AG P S M A P V S	gtgaggatg	2	34632
12	FN#4b prot clvg	274	cctgtccag	T GTT TAT CAG.....GCC AAT GGA V Y Q A N G	gtaagtatg	0	33199
13	FN#4c	37	tgttcacag	GAG ACC AAA.....GCT ACA AAA G E T K A T K A	gtatgttgg	1	12037
14	AS	57	tgtggtcag	CA CCA ATG GGC.....CTC ACC ACA G P M G L T T G	gtgatcacc	1	21576
15	Trans mem	136	ccttcctag	GT GCC TCC ACC.....ATC AAA AGG AG A S T I K R R	gtgagtctc	2	2429
16	AS	30	aatatatag	A AGA AAT GCT.....TCC TAT TAC TT R N A S Y Y L	gtaagtatc	2	36832
17	wedge	158	cttgcacat	G TCC CAA AGG.....AAC GGA TTC A S Q R N G F T	gtaagtaga	1	37699
18	wedge	191	ttttcgaag	CA GAT GGC AGC.....GAA TAC GAG D G S E Y E	gtaagagac	0	19350
19	D1a	88	gctttttag	GCC TTA CCA.....ATC ATA TCC T A L P I I S Y	gtaggtgcc	1	13086
20	D1b	77	tctctgcag	AC GAC CAT TCC.....TAC ATT GAC D H S Y I D	gtgagtgtc	0	8783
21	D1c	37	gtatccag	GGA TAC CAT.....GCG ACT CAA G G Y H A T Q G	gtcagtgcc	1	1435
22	D1d	98	tgttggcag	GT CCG ATG CAG.....GTG GGC AGG P M Q V G R	gtaagcctc	0	2304
22a	D1e	60	acctattag	CAC CCA TCG.....CCC GGA ATG H P S P G M	gtaatcatg	0	725
23	D1f	117	tgcccacag	GTG AAA TGT.....GTC CAG AAG V K C V Q K	gtaagcttc	0	4712
24	D1g start cat core	155	tttcctcag	AAA GGC TAC.....GTC CAC TGC AG K G Y V H C S	gtaagtcaa	2	3412
25	D1h end cat core	136	ctcttccag	T GCT GGG GCT.....CAG ACA GAG A G A Q T E	gtgagtccta	0	2068
26	D1i	150	actttgcag	GAG CAA TAT.....GAA TTT CAG E Q Y E F Q	gtatgagca	0	2272
27	D2a	174	gccccacag	ACC CTC AAC.....CTG ATG GAT T L N L M D	gtaagccga	0	3572
28	D2b	132	ctttgtcag	AGC CAC AAG.....ACT GCC CAG S H K T A Q	gtaggagga	0	12567
29	D2c	126	ttatctcag	TTC TGT ATG.....ATG GCC CGG F C M M A R	gtaagtaca	0	883
30	D2d	164	tgttctcag	CCA CAG GAT.....GTC CAC TGC CT P Q D V H C L	gtgagtact	2	2666
31	D2e	136	tgtgtttag	A AAT GGG GGA.....GAG ACC CTG N G G E T L	gtgagtatc	0	951
32	D2f	8157	gctctcag	GAA CAG TAT.....TCC TTT TAG E Q Y S F *	CTC AAT GGG		-

Table 1

Columns (left to right): Exon number, protein domain, exon size, exon/intron junctional sequences, and intron phases are shown. Amino acids (standard one-letter code) are listed below the coding nucleotides. D1 and D2 represent the first and second phosphatase domains, respectively. a - i designations indicates the individual exons within a single domain; ** intron size is not determined due to lack of contiguity of clones.

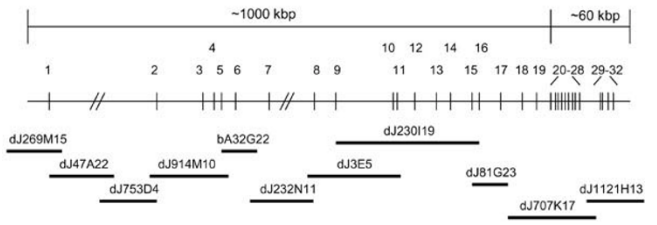


Figure 2
Genomic organization of the human RPTP ρ gene. Exons are shown as vertical bars and introns as thin horizontal lines. Thicker horizontal lines represent PAC (dJ) and BAC (b) clones (Sanger Centre, Chromosome 20 group) containing the RPTPρ gene, which extends over 1000 kbp of DNA (figure not to scale).

age site is located at AA 632-635, in the fourth fibronectin repeat. The transmembrane segment is located at AA 765-785. The intracellular region contains a juxtamembrane 'wedge' region (AA 888-920), and two highly conserved phosphatase domains (AA 1061-1162 and 1351-1456). The 11 hallmark amino acids that define the catalytic core of the first phosphatase domain are located at AA 1104-1114. The stop codon is found after residue 1463 of the amino acid sequence.

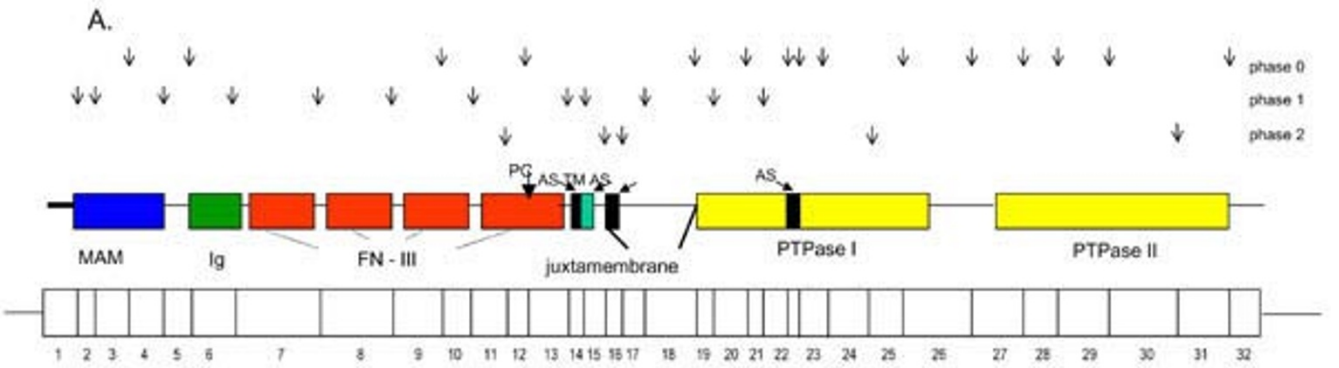
Human RPTPρ genomic organization

We have determined that the region encompassing human RPTPρ is contained within 10 contiguous PAC clones and 1 BAC clone (dJ269M15, dJ47A22, dJ753D4, dJ914M10, bA32G22, dJ232N11, dJ3E5, dJ230I19, dJ81G23, dJ707K17, and dJ1121H13; Sanger Center, chromosome 20 group) (Figure 2). We have ordered these clones by identifying RPTPρ exons within each of them. The RPTPρ gene spans a minimum of 1 Mbp, and the RPTPρ coding sequence is comprised of at least 33 exons, several of which are alternatively spliced. A prominent feature of the RPTPρ gene structure is the considerable variability of exon spacing (Figure 2). Exons 1-19 extend over the initial ~ 1000 kbp of the gene; exons 1-10 are widely separated, while exons 10-19 are more closely spaced. Of particular note are introns 1 and 7, which are ~ 300 and ~ 200 kbp long, respectively, considerably longer than the next largest intron. In contrast, exons 20-28 and 29-32 form two tight clusters, which together span approximately 60 kbp. In general, this pattern of exon organization appears to be characteristic of most RPTPs, as it is also observed in RPTPγ [8], LAR [9], CD45 [10] and RPTPα [11]. Each of these phosphatases has at least one very large intron in the 5'-region of the gene. This feature is not restricted to receptor-like phosphatases as it is also present in a number of adhesion receptor genes, including E-cadherin, N-cadherin, P-cadherin, N-CAM, deleted in colorectal cancer (DCC), axonin-1 and F11 (discussed in [12]).

The exon and intron sizes and exon/intron junctional sequences of the human RPTPρ gene are detailed in Table 6. The majority of 5' and 3' splice sites are consensus sequences. There is some variation in the length of exons, which range from 30 to 297 bp. Approximately one third of the exons are less than 100 bp, while the remaining two thirds are in the 100-300 bp range. Greater variation occurs in the size of the introns, which range from 725 to 303,715 bp. The largest number of introns (15) falls into the 10^4 to 10^5 bp bin, and somewhat fewer (12) fall into the 10^3 to 10^4 bp bin size. Only 5 introns lie outside this range: Three of these fall into the 10^2 to 10^3 bp range, and two unusually long introns in the extracellular domain are over 10^5 bp.

The RPTP extracellular segment is comprised of protein domains; the borders of these modules correspond to the boundaries of exon-clusters. There are three possible junctional phases between exons and introns: Phase 0 refers to introns with junctions between the triplet codons, whereas phase 1 and 2 introns separate within the triplet after the first and second nucleotides, respectively. Figure 3A shows the distribution of intron phases relative to the domain structure of RPTPρ. Within the RPTPρ gene, the number of phase 0 and phase 1 introns is comparable at 15 and 12, respectively. In contrast, there are only five phase 2 introns in the entire gene. A notable feature of RPTPρ gene structure is that phase 1 introns appear to be preferentially associated with the extracellular segment, where they flank each of the protein domain exon modules. The intracellular segment is almost devoid of phase 1 introns. In contrast, phase 0 introns are primarily associated with the intracellular segment, and are only infrequently represented in the extracellular region.

Recently, RPTPs have been examined in sponges [13, 14] the phylogenetically oldest extant metazoan. Although sponges are multicellular organisms, they lack the cellular cohesiveness of the higher eukaryotes. When RPTPs from yeast, sponge and human were aligned and rooted cladograms constructed, the common early ancestor of the phosphatase domains appeared to be yeast. The second phosphatase domain arose as a duplication of the first [13]. The RPTP extracellular domain was acquired during the transition from single-celled to multicellular organisms. In RPTPρ, the extracellular and intracellular exon modules are separated by phase 1 and phase 0 introns, respectively. Furthermore, intracellular introns are much smaller than those in the extracellular segment. Together, these observations suggest that the RPTPρ extracellular and intracellular segments originated



B.

Exon #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
% ID	78	90	89	93	94	88	87	85	95	92	84	91	92	100	94	97

Exon #	17	18	19	20	21	22	22a	23	24	25	26	27	28	29	30	31	32
% ID	89	91	93	90	86	93	95	94	96	92	91	89	92	93	93	93	98

Figure 3
A. Relationship between RPTPρ protein domains, corresponding exons and associated intron phases. Downward arrows indicate intron phases. Protein domains (center) show good correspondence with exon boundaries (bottom line). ***B. Percentage nucleotide identity between human and mouse exons.*** The percentage identity between human and mouse exons was calculated by dividing the number of mismatched nucleotides by the total length of the exon. AS, alternative splice site; FN-III, fibronectin type III repeats; Ig, immunoglobulin-like domain; MAM, meprin/A5/μ domain; PC, proteolytic cleavage site; PTPase I and II, protein tyrosine phosphatase domains; TM, transmembrane domain.

as separate modular proteins that evolved by exon shuffling and duplication, respectively [13, 15]. The two segments became linked to form a functional transmembrane molecule during the transition from single to multicellular organisms.

Over fifty percent of the human genome is comprised of repeat sequences [16], making it the first repeat-rich genome to be sequenced. Analysis of these numerous segments can provide important indications of the evolutionary history of a particular region, or gene. Transposon-derived elements form the largest category of repeats, and include LINEs, SINEs, LTRs and DNA elements. In the RPTPρ gene, the most common of these are: LINE1 (7.6%) and LINE2 (2.0%); the SINEs Alu (4.2%), MIR (3.6%) and THE (0.65%); LTR (0.7%); and the DNA elements MLT (2.5%), MER (2.5%), and MST (0.5%). Less common elements found in the RPTPρ gene include Tiggers in introns 2, 7 and 9 (0.5%), HAL in introns 2 and 7 (0.28%), MAD in introns 1 and 16 (0.013%), and U2 in intron 2 (0.006%). There is also a Charlie repeat in intron 7 (0.005%). In addition to the transposon-derived repeats, there is a pseudogene in intron 7, a tRNA-derived repeat in intron 30, and 133 variable length nucleotide tandem repeats (VNTRs/microsatellites) found in the gene. The G/C content of the RPTPρ gene is approximately 42%. Descriptions of the above repeat elements may be found on Repbase at [http://www.girinst.org./]

The overall percentage of the RPTPρ gene comprised of repeat sequences is lower (by 45%) than that of the entire human genome. In the human genome, LINEs comprise 21% of repetitive sequences, SINEs 13%, LTRs 8%, and DNA elements 3% [16]. In RPTPρ, LINEs comprise 9.6% of repetitive sequences, SINEs 8.4%; LTRs 0.7%; and DNA elements 6.3%. The significance of this deviation in RPTPρ from the normal range is unknown.

cDNA cloning and genomic structure of mouse RPTPρ

The mouse RPTPρ cDNA was cloned using a combination of PCR and 5'-RACE. The mouse cDNA (Genbank accession #AF152556) encodes a 1451AA polypeptide that is 96% identical to that of the human protein and

TABLE 2

exon#	domain	exon size	3' splice site		5' splice site	phase	intron size
1	signal pep			ATG GGG AGC....AGC GCC GCA G M G S S A A G	gtgagtgcg	1	>200,000**
2	MAMa	126	tttctgtag	GT GGC TGT TCT.....GTG CCC ACA G G C S V P T G	gtatgtgat	1	85242
3	MAMb	272	ggtctccag	GG TCC TTC ATG.....TTC TAT CAG S F M F Y Q	gtatgccat	0	9739
4	MAMc	82	atgttgcag	GTG ATA TTT.....CAT CCA TGC A V I F H P C R	gtaagtcta	1	14141
5	Iga	116	gtctcttag	GA AAA GCA CCT.....TGG CTC CAG K A P W L Q	gtaagaatg	0	15049
6	Igb	175	tatatcag	CAA TGG AAT.....ATT GTG AAA G Q W N I V K E	gtgagtacc	1	91543
7	FN#1	294	ttctttcag	AG CCT CCC ACG....AAG TGT GCC G P P T K C A D	gtaaggatg	1	217065
8	FN#2	297	attttacag	AT CCC GTG CAT....GAA GAG GAT G P V H E E D V	gtgagtgag	1	>15,000**
9	FN#3a	110	ttcagacat	TT CCA GGA GCT....CTC TAT GAG P G A L Y E	gtaaggagg	0	89668
10	FN#3b	202	gtcacacag	ATT AAC TAC.....AAG ATT TCA G I N Y K I S A	gtatctctt	1	1275
11	FN#4a	103	ctttcccag	CT CCA TCA ATG.....GCC CCA GTC AG P S M A P V S	gtgaggaac	2	41274
12	FN#4b prot clvg	274	ttttttcag	T GTT TAC CAG.....GCA AAT GGA V Y Q A N G	gtaagtagg	0	31487
13	FN#4c	37	tctccctag	GAG ACA AAA....GCT ACA AAA G E T K A T K A	gtatgttga	1	14502
14	AS	57	tgtggtcag	CA CCA ATG GGC....CTC ACC ACA G P M G L T T G	gtgatcatc	1	19979
15	Trans mem	136	ccccgcaag	GT GCT TCC ACT....ATC AAA AGG AG A S T I K R R	gtgagtctc	2	10854
16	AS	30	aatctatag	A AGA AAT GCT.....TCC TAT TAC TT R N A S Y Y L	gtaagtatc	2	30765
17	wedge	158	cttgcacat	G TCC CAA AGG.....AAT GGA TTC A S Q R N G F T	gtaagtcaa	1	48771
18	wedge	191	tctttgaag	CA GAT GGC AGC....GAA TAC GAG D G S E Y E	gtaagagct	0	19697
19	D1a	88	gcattgtag	GCC TTA CCA.....ATC ATA TCT T A L P I I S Y	gtaggtttc	1	11746
20	D1b	77	tctctacag	AT GAC CAC TCT.....TAC ATT GAC D H S Y I D	gtgagtatc	0	9573
21	D1c	36	attatccag	GGG TAC CAC.....GCA ACC CAA G G Y H A T Q G	gtaagtgtc	1	1402
22	D1d	98	tgtcaccag	GT CCA ATG CAA.....GTG GGC AGG P M Q V G R	gtaagcctc	0	3433
22a	D1e	60	acctattag	CAC CCA GCG....CCC GGA ATG H P A P G M	gtaataatg	0	645
23	D1f	117	cggccacag	GTG AAG TGT.....GTC CAG AAG V K C V Q K	gtaagtttc	0	1199
24	D1g start cat core	155	tctcctcag	AAA GGC CAC.....GTC CAT TGC AG K G H V H C S	gtgagtcaa	2	3272
25	D1h end cat core	136	ctcttccag	T GCT GGA GCC....CAG ACA GAG A G A Q T E	gttagtcct	0	1536
26	D1i	150	accttgcag	GAG CAG TAC.....GAG TTT CAG E Q Y E F Q	gtatggaca	0	1811
27	D2a	174	cccccacag	ACA CTC AAC.....CTG ATG GAT T L N L M D	gtaagctga	0	4334
28	D2b	132	ttttgtcag	AGC CAC AAG.....ACT GCT CAG S H K T A Q	gtaggaggc	0	12637
29	D2c	126	ttgcggcag	CTC TGT ATG.....ATG GCT CGG L C M M A R	gtaagtaca	0	826
30	D2d	164	tgttctcag	CCA CAG GAT....GTC CAC TGC CT P Q D V H C L	gtgagtgct	2	2909
31	D2e	136	tttgtttaa	A AAT GGG GGA....GAG ACG CTG N G G E T L	gtgagcatc	0	869
32	D2f		tcctctcag	GAA CAG TAT.....TCC TTT tag E Q Y S F *			

Table 2

Columns (left to right): Exon number, protein domain, exon size, exon/intron junctional sequences, and intron phases are shown. Amino acids (standard one-letter code) are listed below the coding nucleotides. D1 and D2 represent the first and second phosphatase domains, respectively. a - i designations indicates the individual exons within a single domain; ** intron size is not determined due to lack of contiguity of clones.

predicts an analogous domain structure (Figure 3A). The Celera Discovery System mouse genomic database was used to identify clones containing RPTPρ exons. These clones were then ordered and analyzed to identify exon/intron junctions. Exon and intron sizes, exon/intron junctional sequences, and intron phases of the mouse RPTPρ gene are shown in Table 7. In general, the exon/intron splice sites in the mRPTPρ correspond to expected GT-AG intron consensus splicing sequences, and the intron phases in mouse (Table 7) are identical to those in the human gene (Table 6). Although the two species share approximately 89% nucleotide identity overall, when examined exon by exon, the degree of identity varies slightly between the extracellular and intracellular segments (Figure 3B). The overall identity of the mouse and human extracellular and intracellular segments is 89% and 92%, respectively. In general, there is slightly greater variance between the two species in the extracellular segment; for example, mouse and human exons 1 and 9 share 78% and 95% identity. Within the intracellular segment, mouse exon 21 is 86% identical to that of the human, and exon 24, which contains the first half of the catalytic core, is 96% identical. Notably, the alternatively spliced exons 14, 16 and 22a (discussed below) are 100%, 97% and 95% identical, respectively, indicating a high degree of conservation between mouse and human. In summary, the mouse and human genes are virtually identical in terms of the number and size of exons, and the exons differ only slightly with respect to the nucleotide sequence.

Exon/intron organization of the RPTPρ extracellular segment

MAM domain

The relationship between RPTPρ exon organization and protein domain boundaries is shown in Figure 3A and in Tables 6 and 7. Within the extracellular segment, exon 1 encodes the signal peptide, and exons 2, 3 and 4 encode the single N-terminal MAM domain, a distinguishing feature of all type IIB phosphatases. Although the function of the RPTPρ MAM domain is unclear, other type IIB phosphatases have shown homophilic binding properties: When heterologously expressed in non-adherent cells, both RPTPμ and RPTPκ bind homophilically to induce the formation of large, calcium-independent aggregates [17, 18]. Furthermore, when the RPTPμ MAM domain was deleted, aggregation was eliminated [19], implying that the domain had a crucial role in homophilic cellular interactions.

The three RPTPρ MAM exons differ widely in size: 126 bp (exon 2), 272 bp (exon 3) and 82 bp (exon 4). All MAM-associated introns are in phase 1, with the exception of the second internal intron, which is in phase 0. MAM domains have been identified in a variety of cell adhesion molecules. We have determined the exon structure of the MAM domain in all four human RPTP IIB genes, and in human zonadhesin and human enteropeptidase (NCBI database). The genomic organization of the MAM domain in all four IIB phosphatases is identical. In all RPTP IIB proteins (Genbank #NM 002844; NM 002845; NM 005704; NM 007050) and in human zonadhesin (Genbank #AF312032) there is a MAM domain at the N-terminus, the genomic structure of which is highly conserved. In zonadhesin, there are two additional and adjacent MAM domains. The genomic organization of the latter two domains differs from that of the first. The single MAM domain in the human enteropeptidase gene (Genbank #Y19124) is more internally located than that of RPTPρ, close to the transmembrane region. It is comprised of four exons that are 150, 135, 89 and 125 bp in length, and is unlike any of the IIB and zonadhesin MAM domains. In summary, all known MAM domains are located within the extracellular segment, but within this region, their location, exon number and exon size can vary considerably. The size and structure of exons comprising the most N-terminal MAM domain appear to be unique. Because the nucleotide sequence of the RPTPρ MAM domain predicts a protein similar to that found in the other type IIB RPTPs, it might be expected that the RPTPρ MAM domain also participates in homophilic interactions, as was shown for RPTPμ [19].

Ig domain

Adjacent to the MAM domain, the single Ig-like domain is split into two similarly sized exons (5 and 6) by one intron in phase 0 (Figure 3A). Introns flanking the Ig-like domain are in phase 1. In the majority of genes encoding Ig-like domains, only one exon encodes each domain, while in others such as N-CAM, two exons encode each domain [20]. The single Ig-like domain of the RPTPρ gene falls into the latter category, suggesting a closer relationship to N-CAM-like molecules. LAR has characteristics of both groups [9], a feature which it shares with several other genes, such as perlecan [21] and DCC [22]. Within the RPTP IIB family, the Ig-like domain appears to act in conjunction with the MAM domain to bring about homophilic cell-cell interactions [23].

FN-III domains

Following the Ig domain are four FN-III repeats (Figure 3A), each of which begins with a highly conserved proline residue. FN-III domains are found in a wide range of proteins, and recently, have been shown to be involved in retinal axon target selection [24]. As a general rule, FN-III domains are encoded either by 1 or 2 exons [25]. Within genes that encode multiple FN-III domains, exon organization may be of one type, or a combination of the two types. For example, N-CAM has 2 exons for each FN-III domain [26], whereas tenascin [27] and LAR [9] have

a mixture of both types. In the RPTPρ gene, there is a good correlation between exon structure and FN-III boundaries (Figure 3A), although there is some variation in the number of exons per domain: Each of the first two FN-III repeats is encoded by a single exon (exons 7 and 8, respectively). In contrast, the third FN-III repeat is encoded by two exons (9 and 10). Somewhat atypically, the fourth FN III repeat is encoded by three exons (11, 12 and 13). This domain contains a putative proteolytic cleavage site. RPTPρ FN-III repeats share high sequence similarity with those of N-CAM, but only the third FN-III domain in RPTPρ is encoded by two exons. In contrast to the type IIA phosphatase LAR, the RPTPρ gene does not contain exons encoding more than one fibronectin domain; however, like LAR, it has a FN-III domain encoded by three exons.

In the majority of known cases, the exon/intron junctions corresponding to the FN-III domain boundaries are in phase 1. When two exons encode a FN-III domain, an intron interrupts the coding region in a central, relatively non-conserved, part of the domain, and the exon/intron junction may be in any phase. In the RPTPρ gene, introns separating the individual FN-III repeats are in phase 1; the intron internal to the third repeat is in phase 0, and introns internal to the fourth FN-III repeat are in phase 2 and 0, respectively.

Exon/intron organization of the RPTPρ intracellular segment

Juxtamembrane region

Following the transmembrane segment (exon 15), exons 16-18 encode the juxtamembrane region (Figure 3A, Tables 6 and 7). This segment of the RPTPρ protein is similar to the membrane proximal region in the type IV phosphatase, murine RPTPα, for which the crystal structure has been determined [28]. RPTPα exists as a dimer in which the catalytic site of one molecule is blocked by contact with a 'wedge' from the other. Specifically, the 'turn' part of the helix-turn-helix motif is inserted into the active site, which maintains the WpD loop in the open state [28]. In other phosphatases [29], the WpD loop undergoes a conformational shift upon substrate binding, which appears to be crucial for catalysis. Thus, it is very likely that the dimeric form of RPTPα is unable to bind tyrosine-phosphorylated substrates, rendering it catalytically inactive. The negative charge of two adjacent residues within a highly conserved sequence in the juxtamembrane region appears to be crucial for inhibition [28, 30]. In RPTPα, these two residues are negatively charged aspartates. In type IIB RPTPs, the first residue is changed to an alanine in PCP-2 and RPTPμ, and to a serine in RPTPκ. The second residue is retained as either a glutamate in PCP-2 and RPTPμ, or an aspartate in RPTPκ. These single amino acid changes may indicate a somewhat weaker level of inhibition. This is supported by the examination of the crystal structure of RPTPμ, which shows that although a wedge is formed, catalytic activity is not inhibited by its insertion into the active site on the adjacent monomer [31]. However, in the case of RPTPρ, the first residue is a glycine, and the second is the large basic residue, glutamine. Thus, the RPTPρ juxtamembrane catalytic region is likely to have a different conformation to that of the other phosphatases and a net positive charge, making the regulation of phosphatase activity by dimerization-induced wedge inhibition unlikely.

Phosphatase domains

Although the extracellular regions of receptor-like phosphatases are highly variable, the intracellular tandem phosphatase domains appear quite closely related. The structure of the CD45 gene indicates that both protein tyrosine phosphatase (PTPase) domains have a very similar exon/intron organization, which probably arose by duplication [10]. In RPTPρ, the first and second phosphatase domains are encoded by exons 19-26 and 27-32, respectively (Figure 3A). The exon structure of the RPTPρ phosphatase domains, and that of homologous domains in PCP-2 (NM_005704), RPTPκ (NM 002844), RPTPμ (NM 002845), LAR [9], CD45 [10] RPTPα [11], RPTPγ [8] and rat Esp/mOST-PTP [32, 33], are compared in Figure 4. We have deduced the genomic structure of RPTPκ, RPTPμ and PCP-2 by comparing known cDNA sequences with human genomic clones (NCBI). The positions of the exon boundaries in the phosphatase domains of RPTPρ, RPTPκ, RPTPμ and PCP-2 coincide exactly, and correspond well with the five other phosphatases. LAR is somewhat anomalous in that, although the exon/intron structure of the second phosphatase domain is generally similar to that of the other RPTPs, exons in the first phosphatase domain are fewer in number, but greater in size. The final exon in all nine genes encodes the end of the second phosphatase domain, the short C-terminus and the entire 3'-untranslated region.

A striking similarity among the RPTP genes is the conservation of exon/intron junction 24/25 in the first phosphatase domain. In LAR, CD45 and RPTPα, this junction interrupts the highly conserved sequence VHCSAGV, part of the catalytic core of the phosphatase [34, 35]. Although this exon/intron junction in the IIB phosphatases corresponds exactly, there is a change in the last amino acid from a valine to an alanine. Interestingly, an exon/intron junction is not observed at this position in the cytoplasmic PTPase PTP1B [36], an observation that may indicate an early evolutionary divergence of the cytoplasmic and transmembrane PTPases [37].

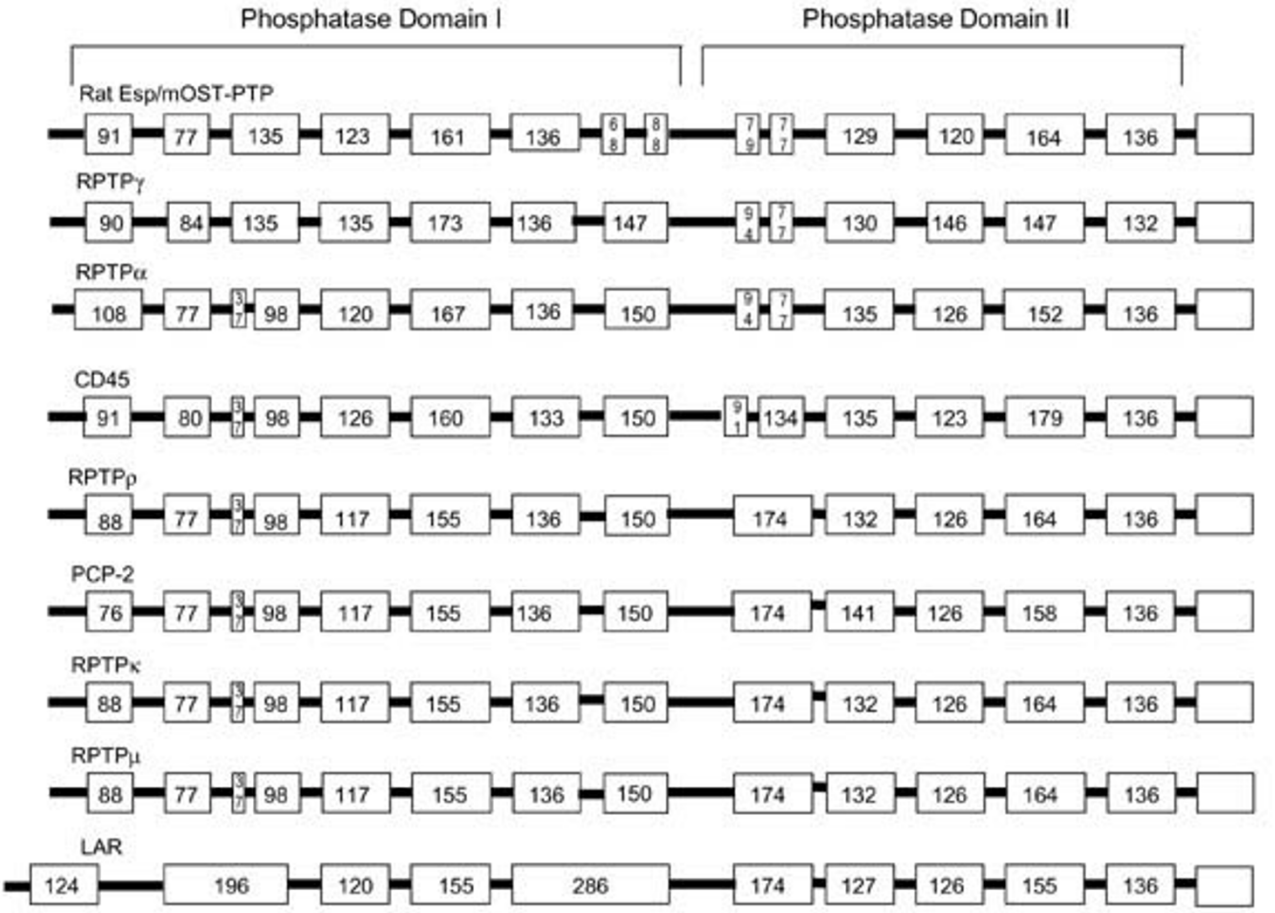


Figure 4

Genomic organization of the two phosphatase domains in nine RPTPs. Boxed numbers indicate the number of nucleotides in each exon; interconnecting horizontal lines represent introns (neither are to scale). Note that exon 22a is not shown in order to preserve alignment among the type IIB RPTPs.

Although the exon/intron structure of the two phosphatase domains was remarkably similar in each of the nine RPTPs examined, there were variations in exon size and number, primarily in those close to the transmembrane domain. For example, the third exon (135 nt) in the first phosphatase domain of rat Esp/mOST-PTP and RPTPγ is replaced by two smaller exons (37 and 98 nt) in RPTPα, CD45, RPTPρ, PCP-2, RPTPκ, and RPTPμ . Two smaller exons replace a single exon at the C-terminal end of the first phosphatase domain of rat Esp/mOST-PTP. Similarly, at the start of the second phosphatase domain, the first exon (174nt) in RPTPρ, PCP-2, RPTPκ, RPTPμ and LAR is replaced by two smaller exons in rat Esp/mOST-PTP, RPTPα, RPTPγ and CD45. In each case, the total number of nucleotides in the two smaller exons is virtually identical to that of the single larger exon at the same position. It is unclear whether these changes in exon number resulted from intron gain or exon fusion.

RPTPρ 3' untranslated region

Following the second phosphatase domain, there is a long (8.0 kb) 3' untranslated sequence. BLAST comparisons identified a region on the KIAA0283 gene (Genbank accession #AB006621) that showed 99% identity to nucleotides 3181 to 4437 of the hRPTPρ sequence. Thus, the 3'-UTR of hRPTPρ, which is contained in exon 32, was identified as KIAA0283. Polyadenylation signals were found at 12425 nt and 12663 nt (NM_007050).

Alternative splicing of mouse and human RPTPρ genes

Comparison of the four RPTP type IIB (RPTPμ, RPTPκ, RPTPρ, PCP-2) nucleotide sequences predicted that, at least, two exons (14 and 16) are likely to be alternatively

spliced. In addition, the presence of a segment (AA 826-850) in xenopus RPTPρ that is absent in the majority of other type IIB RPTPs, raised the possibility of an alternatively spliced exon between exons 17 and 18. Human fetal brain, mouse neonatal brain, and several regions (cortex, forebrain, brainstem, and cerebellum) of adult C57BL/6 mouse brain were examined for the presence of alternatively spliced regions. PCR primers were designed to amplify the regions encapsulating exons 14 and 16, and the region between exons 17 and 18. An additional region between exons 22 and 23 was also examined. The identity of all PCR products was verified by sequencing.

The RPTPρ exon 14 primers yielded two products of 257 and 200 bp (Figure 5A and 5B), indicating a 57 nt alternatively spliced region at 2177 to 2233 nt. This 19AA segment is encoded by exon 14. Both splice forms were observed in human fetal, and in neonatal and adult mouse brain mRNA. We have obtained similar results for RPTPμ (data not shown), in which exon 14 was reported to be absent (NM_002845). The RPTPρ exon 16 primers yielded two bands of 356 and 326 bp (Figure 5C and 5D). This indicates an additional 10 AA alternatively spliced region, located between the transmembrane and the first phosphatase domain (2370-2399 nt). Both transcripts were present in mouse and human brain, and were observed in all brain regions analyzed. PCR of the same region in RPTPμ yielded only one product that did not contain the exon 16 sequence (data not shown). A third alternatively spliced exon (22a) was identified in the first phosphatase domain between exons 22 and 23. Exon 22a was inserted after nucleotide 3172 in mouse, and after nucleotide 3232 in human RPTPρ, predicting an additional alternatively spliced region 20 AA in length. In each case, primers yielded two bands of 93 and 152 bp (Figure 5E and 5F) in all brain regions examined. It remains to be determined if other members of the type IIB subfamily also contain this exon, or whether the region is unique to RPTPρ.

Comparison of xenopus, mouse and human type IIB RPTP nucleotide sequences indicated the possibility of a fourth alternatively spliced region located 3' to exon 17, within the wedge domain. This 75 nt segment is present in the reported sequence of human RPTPμ (2445-2520 nt) and in xenopus RPTPρ (2448-2523 nt). It is absent in the reported sequences of human and mouse RPTPκ, RPTPρ and PCP-2. The exon 17/18 primers were designed to amplify two potential products of 209 and 134 nt. However, only a single product of 134 nt was observed in human and mouse brain regions (data not shown). This sequence appears to be unique to human RPTPμ and xenopus RPTPρ and is unlikely to represent an alternatively spliced exon in any of the RPTP IIB genes.

Both splice variants of exons 14, 16 and 22a were present in human and mouse brain, at all ages and in all brain regions examined. Although the RPTPρ protein products encoded by the alternatively spliced exons do not appear to encode any known motifs, different isoforms of the phosphatase, with as yet unknown functions, are likely to be present. Alternatively spliced isoforms of the related RPTPs, LAR [38] and RPTPβ/ζ [39], are spatially and temporally distinct in the central nervous system, and there is evidence that alternatively spliced exons can influence ligand binding, as is the case with LAR [9].

Conclusions

We describe the cloning of the mouse RPTPρ cDNA, the genomic structure and alternative splicing of the mouse and human genes, and the presence of an 8 kb 3'-UTR in human RPTPρ. RPTPρ is the largest RPTP gene characterized to date, extending over more than 1 megabase pairs of genomic DNA. Its considerable length is due, primarily, to expanded introns in the extracellular region. The protein domains of the extracellular segment are encoded by 1 to 3 exons, which form modules that are flanked by phase 1 introns. The majority of introns in the intracellular segment are in phase 0, and are relatively small. These data suggest that the ectodomain and the phosphatase domain arose separately by exon shuffling and duplication and fused at a later evolutionary period. The MAM domain, the region characterizing type IIB phosphatases, possesses a unique genomic structure common to all such domains when located at the N-terminus. The fourth fibronectin repeat in RPTPρ is encoded by three exons, an additional feature found only in type II phosphatases. At least two alternatively spliced exons flank the transmembrane domain, the region showing the greatest variability between the four IIB phosphatases. An additional alternatively spliced exon precedes the catalytic core of the first phosphatase domain. Comparison of the genomic structure of representative members of the RPTP family (types I-V) indicates that the intron/exon organization of both phosphatase domains is highly conserved. There is considerable variation in the length of the 3' UTR in the RPTPs; at 8 kb, the RPTPρ 3' UTR is the longest characterized to date. Our results provide the first characterization of the genomic structure of an RPTP type IIB gene. This information will facilitate future studies of promoter and other regulatory elements responsible for the tissue specificity of gene expression.

Materials and Methods

Cloning of mouse RPTPρ cDNA

The mouse RPTPρ cDNA was obtained using a combination of 5'-RACE and PCR by methods described in [40]. Total RNA was isolated (RNAzol, Tel-Test, Friendswood, TX) from C57BL/6 mouse brain and used to synthesize

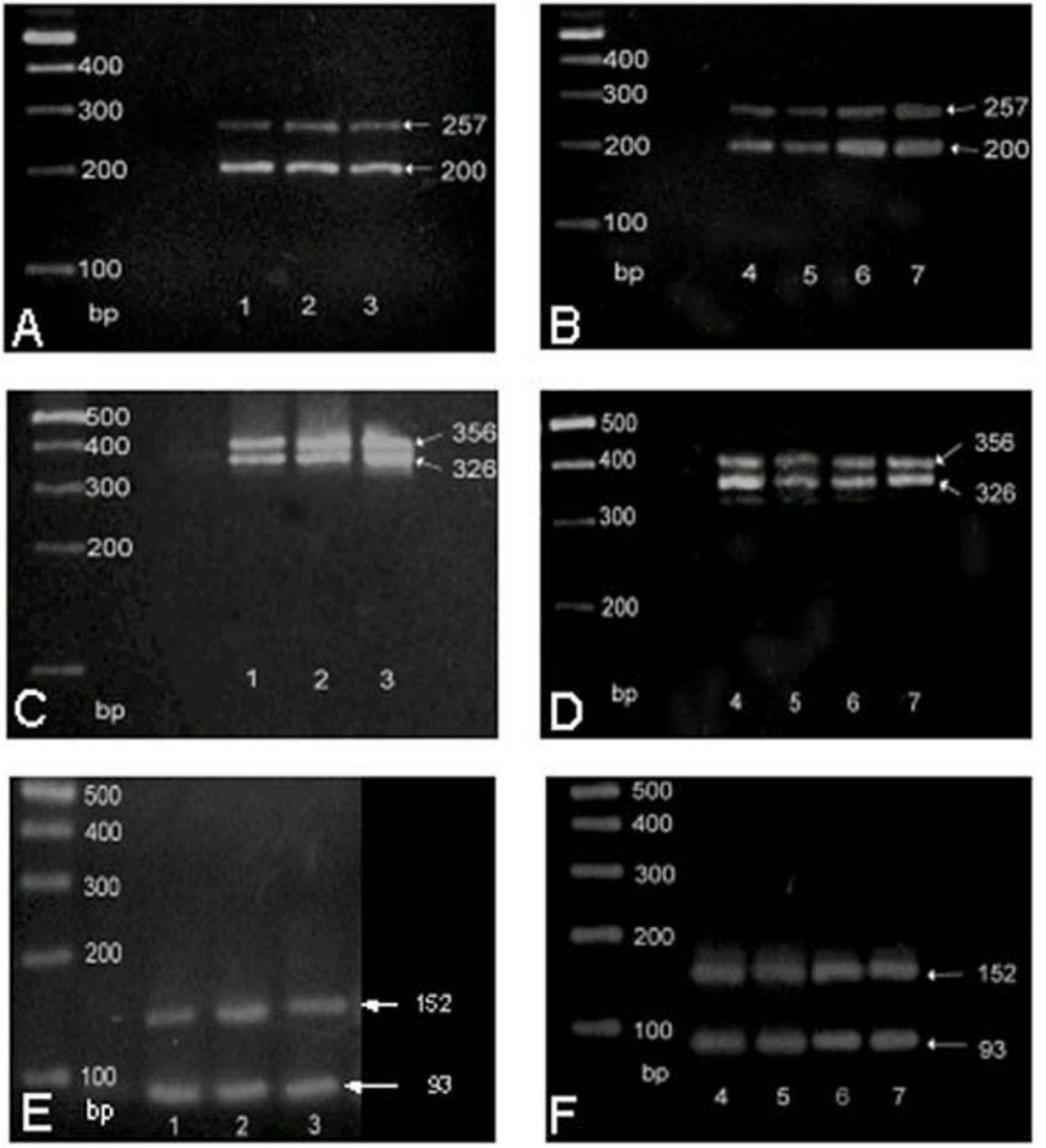


Figure 5
Alternative splicing of exons 14 and 16. RT-PCR products were amplified using primers flanking exon 14 (panels A and B), exon 16 (panels C and D) and exon 22a (panels E and F). Left panels: bands in lanes 1, 2, and 3 are from human fetal brain, mouse P1 brain, and mouse P60 brain total RNA, respectively. Right panels: bands in lanes 4, 5, 6 and 7 contain total RNA from cerebellum, brain stem, basal forebrain and cortex (P23), respectively. Transcripts containing both splice forms of exons 14, 16 and 22a were found in all lanes.

first strand cDNA (AMV-RT, Roche Molecular Biochemicals, Indianapolis), which was then amplified by PCR using degenerate primers based on the human RPTPρ sequence. PCR products were analyzed on 1% agarose gels and subcloned into the TOPO2.1 vector (Invitrogen, Carlsbad, CA). Each strand was sequenced at least twice. Sequence analysis and assembly were performed using Vector NTI Suite (Informax, Bethesda, MD). Murine RPTPρ sequences were identified by BLAST [41] using blastn, on the nr database, with all parameters set to default values. An initial 923 nt fragment was obtained, which spanned the region from the 4th FN-III repeat through the first phosphatase domain. Additional PCR was performed using new gene specific primers based on the newly isolated murine RPTPρ sequence (Genbank #AF152556), and degenerate primers based on the hRPTPρ sequence (Genbank #NM 007050).

Alternative splicing

First strand cDNA was made from total RNA from human fetal brain (16-24 weeks; Clontech, Palo Alto, CA) and from neonatal (P1) and adult (P60) mouse whole brain using Superscript II Reverse Transcriptase (Gibco BRL, Rockville, MD). In addition, cDNA was made from cerebellum, brainstem, forebrain and cortex (P23). The reverse primer, 5' CACGCACACAGTTGAAGATGTCC, which is complementary to a region near the end of the first phosphatase domain (3580 to 3602 nt; NM_007050), was used in all cDNA synthesis. PCR was performed (Expand Long Template PCR system, Roche Molecular Biochemicals, Indianapolis) as recommended by the manufacturer. Primers were as follows: *Exon 14*: forward primer, 5' CAACTGTGTTCGTCTGGCTAC (AS1); reverse primer, 5' GCCCAGGAGAATGATGATGAAC (Ex15rv2). *Exon 16*: forward primer, 5'GAGAAGCAGGTGG ACAACACCGTG (AS2fw); reverse primer, 5' GCTCATCTCCACAGGGTCAC (Exrv). *Exon 17/18*: forward primer, 5' CGCA ATGATGAAGGCTTCTC (Exfw); reverse primer, 5' GCTCATCTCCACAGGGTCAC (Exrv).). *Exon 22a*: forward primer, 5' CTCTGCCAGCATCGTCATGGTCAC (Ex22fw); reverse primer, 5' GTCTCCATAGACCTCTGTGTCATCTGGCC (Ex23rv). All primers were used at a final concentration of 250 nM. An Eppendorf Mastercycler Gradient was used with the following cycling parameters: 2 minutes at 94°C, 35 cycles of 15 seconds at 94°C, 30 seconds at 56°C (exon 14), 55°C (exon 16), 57°C (exon 17/18) or 60°C (exon22/23), 45 seconds at 72°C, and a final extension step (2 minutes at 72°C). The PCR products were run on 4% NuSieve GTG agarose (FMC, Rockville) gels and confirmed by sequencing.

Human and mouse nucleotide sequence analysis

The human RPTPρ cDNA sequence was used to search the Sanger Center's chromosome 20 database for genomic clones encoding RPTPρ exons. The chromosomal region containing the human RPTPρ gene was represented within PAC and BAC clones of chromosome 20, contig 125. The mouse cDNA sequence (Genbank accession #AF152556) was used to search the Celera Discovery System mouse genomic database for clones containing RPTPρ exons.

Abbreviations

AS, alternative splice site; BAC, bacterial artificial chromosome; Ig, immunoglobulin-like domain; FN-III, fibronectin type III repeats; MAM, meprin/A5/μ domain; nt, nucleotide; PAC, P1 artificial chromosome; PCR, polymerase chain reaction; PTPase, protein tyrosine phosphatase; PC, proteolytic cleavage site; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcriptase polymerase chain reaction; RPTP, receptor-like protein tyrosine phosphatase; TM, transmembrane domain; UTR, untranslated region.

Acknowledgements

This paper includes sequence data that were produced by the Chromosome 20 Sequencing Group at the Sanger Center, and can be obtained from ftp://ftp.sanger.ac.uk/pub/human/chr. Mouse genomic sequence data were obtained through use of the Celera Discovery System and Celera's associated databases. JB and MP are members of The Ohio State University Biochemistry Program. The work was supported by NIH grant MH57415 to AR.

References

1. Streuli M, Krueger NX, Ariniello PD, Tang M, Munroe JM, Blattler WA, Adler DA, Disteche CM, Saito H: **Expression of the receptor-linked protein tyrosine phosphatase LAR: proteolytic cleavage and shedding of the CAM-like extracellular region.** *EMBO J* 1990, **9**:2399-2407
2. Brady-Kalnay SM, Tonks NK: **Protein tyrosine phosphatases as adhesion receptors.** *Curr Opin Cell Biol* 1995, **7**:650-657
3. Wang HY, Lian ZR, Lerch MM, Chen ZJ, Xie WF, Ullrich A: **Characterization of PCP-2, a novel receptor protein tyrosine phosphatase of the MAM domain family.** *Oncogene* 1996, **12**:2555-2562
4. Yang Y, Gil MC, Choi EY, Park SH, Pyun KH, Ha H: **Molecular cloning and chromosomal localization of a human gene homologous to the murine R-PTP-kappa, a receptor-type protein tyrosine phosphatase.** *Gene* 1997, **186**:77-82
5. Gebbink MFBG, Vanetten I, Hateboer G, Suijkerbuijk R, Beijersbergen RL, Vankessel AG, Moolenaar WH: **Cloning, expression, and chromosomal localization of a new putative receptor-like protein tyrosine phosphatase.** *FEBS Lett* 1991, **290**:123-130
6. McAndrew PE, Frostholm A, White R, Rotter A, Burghes AHM: **Identification and characterization of RPTPρ, a novel μ/κ -like receptor protein tyrosine phosphatase whose expression is restricted to the central nervous system.** *Mol Brain Res* 1998, **56**:9-21
7. McAndrew PE, Frostholm A, Evans JE, Zdilar D, Goldowitz D, Chiu I-M, Burghes AHM, Rotter A: **Novel receptor protein tyrosine phosphatase (RPTPρ) and acidic fibroblast growth factor (FGF-1) transcripts delineate a rostrocaudal boundary in the granule cell of the murine cerebellar cortex.** *J Comp Neurol* 1998, **391**:444-455
8. Kastury K, Ohta M, Lasota J, Moir D, Dorman T, LaForgia S, Druck T, Huebner K: **Structure of the human receptor tyrosine phosphatase gamma gene (PTPRG) and relation to the familial RCCt (3;8) chromosome translocation.** *Genomics* 1996, **32**:225-235
9. O'Grady P, Krueger NX, Streuli M, Saito H: **Genomic organization of the human LAR protein tyrosine phosphatase gene and al-**

ternative splicing in the extracellular fibronectin type-III domains. *J Biol Chem* 1994, **269**:25193-25199
10. Hall LR, Streuli M, Schlossman SF, Saito HL: **Complete exon-intron organization of the human leukocyte common antigen (CD45) gene.** *J Immunol* 1988, **141**:2781-2787
11. Wong ECC, Mullersman JE, Thomas ML: **Leukocyte common antigen-related phosphatase (LRP) gene structure: conservation of the genomic organization of transmembrane protein tyrosine phosphatases.** *Genomics* 1993, **17**:33-38
12. Plagge A, Brummendorf T: **The gene of the neural cell recognition molecule F11: conserved exon-intron arrangement in genes of neural members of the immunoglobulin superfamily.** *Gene* 1997, **192**:215-225
13. Müller CI, Blumbach B, Krasko A, Schröder HC: **Receptor protein-tyrosine phosphatases: origin of domains (catalytic domain, Ig-related domain, fibronectin type III module) based on the sequence of the sponge Geodia cydomium.** *Gene* 2001, **262**:221-230
14. Ono K, Suga H, Iwabe N, Kuma K-I, Miyata T: **Multiple protein tyrosine phosphatases in sponges and explosive gene duplication in the early evolution of animals before the parazoan-eumetazoan split.** *J Mol Evol* 1999, **48**:654-662
15. Patthy L: **Protein evolution by exon-shuffling.** *Heidelberg, Springer-Verlag,* 1995
16. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, *et al*: **The sequence of the human genome.** *Science* 2001, **291**:1304-1351
17. Brady-Kalnay SM, Flint AJ, Tonks NK: **Homophilic binding of PTPμ, a receptor-type protein tyrosine phosphatase, can mediate cell-cell aggregation.** *J Cell Biol* 1993, **122**:961-972
18. Sap J, Jiang YP, Freidlander D, Grumet M, Schlessinger J: **Receptor tyrosine phosphatase R-PTP-κ mediated homophilic binding.** *Mol Cell Biol* 1994, **14**:1-9
19. Zondag GCM, Koningstein GM, Jiang YP, Sap J, Moolenaar WH, Gebbink MBFG: **Homophilic interactions mediated by receptor tyrosine phosphatases μ and κ : a critical role for the novel extracellular MAM domain.** *J Biol Chem* 1995, **270**:14247-14250
20. Williams AF, Barclay AN: **The immunoglobulin superfamily: Domains for cell surface recognition.** *Ann Rev Immunol* 1988, **6**:381-405
21. Cohen IR, Grassel S, Murdoch AD, Iozzo RV: **Structural characterization of the complete human perlecan gene and its promoter.** *Proc Natl Acad Sci, USA* 1993, **90**:10404-10408
22. Cho KR, Oliner JD, Simons JW, Hedrick L, Fearon ER, Preisinger AC, Hedge P, Silverman GA, Vogelstein B: **The DCC gene - Structural analysis and mutations in colorectal carcinomas.** *Genomics* 1994, **19**:525-531
23. Brady-Kalnay SM, Tonks NK: **Identification of the homophilic binding site of the receptor protein tyrosine phosphatase PTPμ .** *J Biol Chem* 1994, **269**:28472-28477
24. Garrity PA, Lee C-H, Salecker I, Robertson HC, Desai CJ, Zinn K, Zipursky SL: **Retinal axon target selection in Drosophila is regulated by a receptor protein tyrosine phosphatase.** *Neuron* 1999, **22**:707-717
25. Dufour S, Duband J-L, Kornblihtt AR, Thierry JP: **The role of fibronectins in embryonic cell migration.** *Trends Genet* 1988, **4**:198-203
26. Owens GC, Edelman GM, Cunningham BA: **Organization of the neural cell-adhesion molecule (N-CAM) gene - alternative exon usage as the basis for different membrane-associated domains.** *Proc Natl Acad Sci, USA* 1987, **84**:294-298
27. Gulcher JR, Nies DE, Alexakos MJ, Ravikant NA, Sturgill ME, Marton LS, Stefansson K: **Structure of the human hexabrachion (tenascin) gene.** *Proc Natl Acad Sci, USA* 1991, **88**:9438-9442
28. Bilwes AM, den Hertog J, Hunter T, Noel JP: **Structural basis for inhibition of receptor protein-tyrosine phosphatase-alpha by dimerization.** *Nature* 1996, **382**:555-559
29. Barford D, Keller JC, Flint AJ, Tonks NK: **Purification and crystallization of the catalytic domain of the human protein tyrosine phosphatase 1B expressed in Escherichia coli.** *J Mol Biol* 1994, **239**:726-730
30. Majeti R, Bilwes AM, Noel JP, Hunter T, Weiss A: **Dimerization-induced inhibition of receptor protein tyrosine phosphatase function through an inhibitory wedge.** *Science* 1998, **279**:88-91
31. Hoffmann KMV, Tonks NK, Barford D: **The crystal structure of domain 1 of receptor protein tyrosine phosphatase μ .** *J Biol Chem* 1997, **272**:27505-27508
32. Lathrop W, Jordan J, Eustice D, Chen D: **Rat osteotesticular phosphatase gene (Esp): genomic structure and chromosome location.** *Mamm Genome* 1999, **10**:366-370
33. Morrison DF, Mauro LJ: **Structural characterization and chromosomal localization of the mouse cDNA and gene encoding the bone tyrosine phosphatase mOST-PTP.** *Gene* 2000, **257**:195-208
34. Cho H, Ramer SE, Itoh M, Kitas E, Bannwarth W, Burn P, Saito H, Walsh CT: **Catalytic domains of the LAR and CD45 protein tyrosine phosphatases from Escherichia coli expression systems: purification and characterization for specificity and mechanism.** *Biochemistry* 1992, **31**:133-138
35. Pot DA, Dixon JE: **Active site labeling of a receptor-like protein tyrosine phosphatase.** *J Biol Chem* 1992, **267**:140-143
36. Brown-Shimer S, Johnson KA, Lawrence JB, Johnson C, Bruskin A, Green NR, Hill DE: **Molecular cloning and chromosome mapping of the gene encoding protein phosphotyrosyl phosphatase 1B.** *Proc Natl Acad Sci, USA* 1990, **87**:5148-5152
37. Krueger NX, Streuli M, Saito H: **Structural diversity and evolution of human receptor-like protein tyrosine phosphatases.** *EMBO J* 1990, **9**:3241-3252
38. Honkaniemi J, Zhang JS, Yang T, Zhang C, Tisi MA, Longo FM: **LAR tyrosine phosphatase receptor: proximal membrane alternative splicing is coordinated with regional expression and intraneuronal localization.** *Mol Brain Res* 1998, **60**:1-12
39. Nishiwaki T, Maeda N, Noda M: **Characterization and developmental regulation of proteoglycan-type protein tyrosine phosphatase zeta/RPTP beta isoforms.** *J Biochem* 1998, **123**:458-467
40. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K: **Current Protocols in Molecular Biology.** *New York, Greene Pub. Associates and Wiley-Interscience,* 1998
41. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nuc Acids Res* 1997, **25**:3389-3402