SUPPLEMENTARY DATA

The structure of the 5’ end of the protein-tyrosine phosphatase PTPRJ mRNA reveals a novel mechanism for translation attenuation

Luchezar Karagyozov, Rinesh Godfrey, Sylvia-Annette Böhmer, Astrid Petermann, Sebastian Hölters, Arne Östman, Frank-D. Böhmer

Supplementary Experimental Data

A. Cloning of the extended promoter region

The sequence of the 5’ untranslated region of the PTPRJ cDNA (NM_002843) was used to BLAST search the NCBI database. A BAC genomic clone was identified (RP11-346F1) which was obtained from the BACPAC Resources Center (http://bacpac.chori.org/). The sequence of the forward cloning primer (ATATA\texttt{GCTAGCCTTGGCCTCCCGGAAGTGC}) was complementary to nucleotides 1419 - 1400 upstream of the 5’-end of the PTPRJ cDNA and the reverse cloning primer (TTAAT\texttt{AGATCTCTGGAACGTGCCCCGGAC}) was complementary to nucleotides 326 - 343 downstream of the transcription start. The PCR product was NheI and BglII digested and cloned into the firefly luciferase reporter vector pGL3-Basic (Promega, Mannheim, Germany). The clone (1762 bp, GenBank EF219146) contained sequences upstream of the transcription start and 343 nt of the 5’ leader. It ended 12 nt upstream of the ATG\textsubscript{356} coding for translation initiation (as indicated in NM_002843).

The sequence of the entire clone p1.7 (pGL3) was verified. Sequencing showed no difference between the 3’ region of this clone, the 5’ end of cDNA-HA(pcDNA3) and NM_002843. Amplification of human genomic DNA with primers flanking the transcription start site resulted in products with the same sequence as those deposited for Homo sapiens chromosome 11 genomic contig, reference assembly (NT_009237). The genomic amplification primers were: Forward - CGGGAGCGCTTCCTCTGC, Reverse – CACACGCCGGAGCTTAGC. Sequencing primer - TCCTCTGCCCGGGAG..

B. Fire-fly luciferase reporter constructs

The “Out-of- Frame” and “In-frame” constructs. Deletion constructs.

The 1762 nt PTPRJ promoter region was inserted into NheI and BglII sites of pGL3 (Promega, Mannheim, Germany). In this construct - p1.7 OutF (pGL3)- ATG\textsubscript{14} , ATG\textsubscript{191} and ATG\textsubscript{Luc} are in two different reading frames.
The p1.7(pGL3) clone was partially digested with NarI and the products were cloned into ClaI digested pBScript (Stratagene, La Jolla, CA). Two clones were obtained with the suitable orientation. One clone - pNar–Nar(pBS) - contained sequences -323 to +82 (from PTPRJ). The KpnI - Hind III fragment from this clone was inserted into the pGL3-Basic (KpnI – Hind III digested) to get the 3’deletion reporter construct pNar-Nar(pGL3). The other clone - pNar-Bgl(pBS) - contained sequences -323 to +343 from PTPRJ plus the sequences from the BglII site to the NarI site in the pGL3 vector. This clone was digested with EcoRV and BglIII and the fragment was ligated to pGL3-Basic opened with SmaI and BglIII. The clone pNar(17)_OutF(pGL3) contained 17 additional nucleotides between +343 (the BglII site) and the AUG_{Luc}.

A deletion in p1.7(pGL3) was constructed (BstAPI -NdeI region was deleted, Sebastian deletion). DNA was digested with HindIII and BglIII, blunt-ended (Klenow) and ligated to get pSebastian_InF(pGL3). The KpnI – PstI fragment from this clone was replaced either with the full-sized wild type fragment to get p1.7_InF(pGL3) or with the KpnI – PstI fragment from pNar_OutF(pGL3) to get pNar_InF(pGL3).

To prepare the 5’-end deletion clone pNde_InF(pGL3), DNA from p1.7_InF(pGL3) was digested with NheI and NdeI, blunt ended and ligated. To obtain the deletion clone pSac_InF(pGL3), DNA from p1.7_IF(pGL3), the latter was digested with SacI and ligated. These deletion clones were used to perform promoter activity analysis (see Suppl. Fig. 4).

**Constructs with mutated ATGs**

All mutations of the ATGs were performed by PCR with thermophilic DNA polymerases Turbo-Pfu (Stratagene Europe, the Netherlands) or Phusion High Fidelity (BioCat, Heidelberg) following closely the manufacturers instructions. The PCR products were cloned and the clones were sequenced. ATG_{14} was mutated to TTG; ATG_{191} - to AGG, and ATG_{356} – to ATT.

**Constructs with fused PTPRJ and firefly luciferase**

Constructs expressing the firefly luciferase as a fusion protein with PTPRJ were prepared by PCR. The template was pNar(pGL3) and the primers (phosphorylated) were as follows:

**FU_LEFT** GCCCGCGCGCGCCCTGGGAACGTGCCC
**FU_RIGHT** ATGAAGCCGGCGGCGGAAGACGCCAAAAACA

PCR was performed with Phusion High Fidelity, the amplification product was treated with DpnI (to destroy the original template) and circularized by blunt-end ligation. The clones were screened first by restriction enzyme analysis and then by sequencing.
The resulting construct pNar_Luc_Fused(pGL3) expressed the firefly luciferase fused to the first five N-end amino acids of the signal peptide. The construct contained PTPRJ sequences from – 323 to +370, followed by the fire-fly luciferase sequences coding amino acids 2 – 550.

**Constructs with frame shift mutations**

Frame shift mutations in the region between +191 and +356, without introducing a stop codon were prepared. We inserted one A (between +200 and +201) in the 3rd codon after the AUG_{191} (plus mutant) or deleted one C (+340) in the 6th codon preceding AUG_{356} (minus mutant). The distance between the frame shifts is 139 nt, altered codons are 47. The double frame shift mutant was created by transferring a fragment BsmBI (+262) – Sall (in the pGL3 vector) from the minus to the plus mutant by standard techniques.

**Constructs containing the firefly luciferase fused to the PTPRJ 5’ leader expressed under the CMV promoter**

The entire PTPRJ-Luciferase fused region (total length 2349 nt) was cloned in pcDNA3.1(+) (Invitrogen, Kalsruhe, Germany) between HindIII and XbaI sites. From these constructs the HindIII – NotI fragment, containing the PTPRJ promoter and 170 nt of the 5’ leader was removed (DNA was restricted, ends were blunt-ended with Klenow, DNA was gel-purified and circularized by ligation). In this construct the firefly luciferase was transcribed from the strong CMV promoter and translated from the tandem codons AUG_{191} and AUG_{356}. The region from PTPRJ mRNA, which is present upstream of the luciferase codons is 201 nt long (from the NotI site to codon 5 of the signal peptide).

**D. Transfection of HEK293 cells and other cell lines with polyethylenimine (PEI)**

Cells were grown to 40 - 70 % confluence, and transfected with branched PEI (Polyethylenimine, Aldrich, Cat. No. 40872-7, MW ~ 25 kD). For transfection of cells in 35 mm dishes, 2 - 4 µg DNA was diluted in 250 µl serum-free medium at room temperature. The PEI stock (10 µg/µl in water, pH 7.2, sterile filtered) was appropriately diluted in 250 µl of serum-free medium (final ratio 2.5 µg PEI per µg DNA for HEK293 cells, 5 to 1 for other cell lines), and the DNA and PEI solutions were mixed. The mixture was incubated at room temperature for 20 – 30 min to allow formation of complexes, and then added to the culture dish containing 2 ml medium with serum under gentle agitation. Cells were incubated in a CO_{2} incubator for 24 to 48 hours prior to analysis. To prevent detachment of the HEK293 cells the dishes were coated with poly-L-lysine (Sigma, P1274). Poly-L-lysine solution (0.6 ml, 10 µg/ml in water, sterile filtered) is
added to the dish and kept for 1 hour at 37° C. Thereafter, the solution is removed and the dish dried in the microwave oven for 2 min.

For measuring density dependence of reporter expression, HT29 cells were seeded in 96-well plates (3 x 10⁴ cells per well), and transfected with reporter constructs 24 h later. Cells were harvested daily for reporter assays at day 1-5 after transfection.

E. DIG-probes, RT-PCR primers

Preparation of DIG labeled hybridization probes

DNA from pGL3-basic and pRL-TK (Promega, Mannheim, Germany) was labelled with digoxigenin-11-dUTP, alkali-stable (Roche, Penzberg, Germany) by PCR according to the protocol of the supplier. The fire-fly luciferase specific primers (CTGCCTCATAGAACTGCCTGC and TGAGCCCATATCCTTGCTG) and the Renilla luciferase specific primers (ATTGGTATGGGCAAATCAGG and TGTTGGACGACGAACTCAC) generated products 400 bp and 414 bp long, which were used for hybridization.

Primers for cDNA amplification

The primers used for qRT-PCR for the firefly and Renilla luciferase were the same as the primers for preparation of the labeled probes. The beta- lactamase specific primers were:
Forward -       CCCAACGATCAAGCGAGTTAC
and Reverse -   CTGCAGGCACTTACTTGAC.
Supplementary Figures

Supplementary Figure 1. Genome view of the \textit{hPTPRJ} promoter region and the first exon.

Position of the cloned fragment and conservation are shown.
Supplementary Figure 2. Alignment of the nucleotide sequence of the first exon of the human and mouse *PTPRJ*.

The translation initiation codons are highlighted (green). The codons with different sense are in yellow while the synonymous codons are in grey. The amino acids encoded by the human sequence are indicated.

| Human 1 | GAC CCG AAC CGG GCC GGC AGC GGG AGC AGC --- |
| Mouse 1 | GAC CCG AAC CGG GCC GGC AGA GGG AGC AGC CCG |

Human 44

---

| Human 44 | GAC CCG AAC CGG GCC GGC AGC GGG AGC AGC |
| Mouse 44 | GAC CCG AAC CGG GCC GGC AGA GGG AGC AGC CCG |

Human 86

| Human 86 | GAC CCG AAC CGG GCC GGC AGC GGG AGC AGC --- |
| Mouse 89 | GAC CCG AAC CGG GCC GGC AGA GGG AGC AGC CCG |

Human 131

| Human 131 | GAC CCG AAC CGG GCC GGC AGC GGG AGC AGC --- |
| Mouse 134 | GAC CCG AAC CGG GCC GGC AGA GGG AGC AGC CCG |

Human 176

| Human 176 | GAC CCG AAC CGG GCC GGC AGC GGG AGC AGC --- |
| Mouse 176 | GAC CCG AAC CGG GCC GGC AGA GGG AGC AGC CCG |

Human 221

| Human 221 | GAC CCG AAC CGG GCC GGC AGC GGG AGC AGC --- |
| Mouse 221 | GAC CCG AAC CGG GCC GGC AGA GGG AGC AGC CCG |

Human 266

| Human 266 | GAC CCG AAC CGG GCC GGC AGC GGG AGC AGC --- |
| Mouse 266 | GAC CCG AAC CGG GCC GGC AGA GGG AGC AGC CCG |

Human 311

| Human 311 | GAC CCG AAC CGG GCC GGC AGC GGG AGC AGC --- |
| Mouse 311 | GAC CCG AAC CGG GCC GGC AGA GGG AGC AGC CCG |

Human 356

| Human 356 | GAC CCG AAC CGG GCC GGC AGC GGG AGC AGC --- |
| Mouse 347 | GAC CCG AAC CGG GCC GGC AGA GGG AGC AGC CCG |

Human 401

| Human 401 | GAC CCG AAC CGG GCC GGC AGC GGG AGC AGC --- |
| Mouse 392 | GAC CCG AAC CGG GCC GGC AGA GGG AGC AGC CCG |

Human 446

| Human 446 | GAC CCG AAC CGG GCC GGC AGC GGG AGC AGC --- |
| Mouse 437 | GAC CCG AAC CGG GCC GGC AGA GGG AGC AGC CCG |
Supplementary Figure 3. Deletion analysis of the *PTPRJ* promoter.

Firefly luciferase reporter constructs with 5’ deletions upstream of +1 were transiently transfected into HEK293 cells and activity was measured after 24 hours.
Supplementary Figure 4. Sequences of the region AUG\textsubscript{191} – AUG\textsubscript{356}, wild type and mutated. In each sequence AUG\textsubscript{191} is highlighted; codon numbers are indicated.

**Wild type.** The rarely used codons are highlighted in yellow (e.g. construct 1, Fig. 7A).

\begin{verbatim}
Wild type.
The rarely used codons are highlighted in yellow (e.g. construct 1, Fig. 7A).

1  ATG TCT CCG GGG AAG CCC GGC GGG ACG AGG CGG
   Met Ser Pro Gly Lys Pro Gly Ala Gly Ala Gly Thr Arg Arg
   Codons 1                5                  10                  15

46  ACC GCC TGG CCG AGG AGG AGG CGA AGG AGA CGG
    Thr Gly Trp Arg Arg Arg Arg Arg Arg Arg Gln Glu Ala Ala
    Codons 16              20                  25                  30

91  ACG ACG GTG CCC GGG CTC GGG CGC ACG GCG GGG CCC GAT TCG CGC
    Thr Thr Val Pro Gly Leu Gly Thr Ala Gly Pro Asp Ser Arg
    Codons 31              35                  40                  45

136  GTC CGG GGC XCG TTC CAG GGC GCC AGG GGC ATG AAG CCG GCG GCG
     Val Arg Gly Thr Phe Glu Gly Ala Arg Gly Met Lys Pro Ala Ala
     Codons 46              50                  55
     ►►► Luciferase
\end{verbatim}

**Optimized codons.** The modified codons optimized for efficient translation are in bright green (construct 4, Fig. 7A).

\begin{verbatim}
Optimized codons. The modified codons optimized for efficient translation are in bright green (construct 4, Fig. 7A).

1  ATG AGC CCC GGC AAG CCT GGC GCC GGA GGG GCC GGA ACC AGA CGG
   Met Ser Pro Gly Lys Pro Gly Ala Gly Gly Ala Gly Thr Arg Arg
   Codons 1                5                  10                  15

46  ACC GCC TGG CCG AGG AGG AGG CGA AGG AGA CGG
    Thr Gly Trp Arg Arg Arg Arg Arg Arg Arg Gln Glu Ala Ala
    Codons 16              20                  25                  30

91  ACC ACC GTG CCC GGC CTG GGC AGA ACC GCC GGA CCC GAC AGC AGA
    Thr Thr Val Pro Gly Leu Gly Arg Thr Ala Gly Pro Asp Ser Arg
    Codons 31              35                  40                  45

136  GTG CGG GGC ACC TTC CAG GGC GCC AGG GGC ATG AAG CCG GCG GCG
     Val Arg Gly Thr Phe Glu Gly Ala Arg Gly Met Lys Pro Ala Ala
     Codons 46              50                  55
     ►►► Luciferase
\end{verbatim}

**Double frame-shift mutant.** The changed amino acid residues are in red (construct 3, Fig. 7A).

\begin{verbatim}
Double frame-shift mutant. The changed amino acid residues are in red (construct 3, Fig. 7A).

1  ATG TCT CCG GAG GAA GCC CGG GGC GGG CGG AGC GGG GCC GAC GCC
   Met Ser Pro Glu Ala Arg Gly Ala Gly Ala Gly Ser Gly Asp Glu Ala
   Codons 1                5                  10                  15

46  GAC CGG CTG GCC GAG GAG GAG GCG AAG GAG ACG GCA GGA GGC GGC
    Asp Arg Leu Ala Glu Glu Glu Ala Lys Glu Thr Ala Gly Gly Gly
    Codons 16              20                  25                  30

91  GAC GAC GGT GCC CGG TGT CCG GCC CGC ACC GCC GCC CGA TGC GCC
    Asp Asp Gly Ala Arg Ala Arg Ala His Gly Gly Ala Arg Phe Ala
    Codons 31              35                  40                  45

136  CGT CCG GCC ACC GTT CAG GCC CGG CGG GCC ATG AAG CCG GCC CC
     Arg Pro Gly His Val Glu Gly Ala Arg Gly Met Lys Pro Ala Ala
     Codons 46              50                  55
     ►►► Luciferase
\end{verbatim}
**Deletion mutant.** The amino acid residues between AUG\textsubscript{191} and AUG\textsubscript{356} are shown (construct 5, Fig. 7A).

\begin{verbatim}
1  ATG TCT CCG GAG GGC GCG CGG ATG AAG CCG GCG GCG Met Ser Pro Glu Gly Ala Arg Gly Met Lys Pro Ala Ala Codons 1               5                   10

►►► Luciferase
\end{verbatim}

**Supplementary Figure 5.** Sequences of the region AUG\textsubscript{191} – CGG\textsubscript{266} (Arg\textsubscript{266}) - wild type and mutated. In each sequence AUG\textsubscript{191} is highlighted; codon numbers are indicated.

**Wild type codons.** The triplets coding Arg are in grey (construct 2, Fig.7B).

\begin{verbatim}
1  ATG TCT CCG GGG AAG CCC GGG GCG GGC GGA GCG GGG ACG AGG CGG Met Ser Pro Gly Lys Pro Gly Ala Gly Gly Ala Gly Thr Arg Arg Codons 1                5                  10                  15

46  ACC GGC TGG CGG AGG AGG AGG AGG CGA AGG AGA CGG CAT AAG CCG Thr Gly Trp Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg His Lys Pro Codons 16              20                  25
\end{verbatim}

**Frame-shifted codons at the N-end of the protein sequence.** The triplets coding Arg are highlighted in grey. The changed amino acid residues are in red (construct 3, Fig. 7B).

\begin{verbatim}
1  ATG TCT CCG GAG GAA GCC CGG GGC GGG CGG AGC GGG GCG AGG CGG Met Ser Pro Glu Glu Ala Arg Gly Gly Arg Ser Gly Ala Arg Arg Codons 1                5                  10                  15

46  ACC GGC TGG CGG AGG AGG AGG AGG CGA AGG AGA CGG CAT AAG CCG Thr Gly Trp Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg His Lys Pro Codons 16              20                  25
\end{verbatim}

**Extended frame-shifted region.** The changed amino acid residues are in red (construct 4, Fig. 7B).

\begin{verbatim}
1  ATG TCT CCG GAG GAA GCC CGG GGC GGG CGG AGC GGG GCG AGG CGG Met Ser Pro Glu Glu Ala Arg Gly Gly Arg Ser Gly Ala Asp Glu Ala Codons 1                5                  10                  15

46  GAC CGG CTG GCG GAG GAG GAG GCG AAG GAG ACG GCT AAG CCG Asp Arg Leu Ala Glu Glu Ala Lys Glu Thr Ala Lys Pro Codons 16              20                  25
\end{verbatim}
Supplementary Figure 6. Folding of 5’ end of PTPRJ mRNA (nt 1 - 358).

Program RNA mfold, version 3.2. (Zuker, 2003). Arrangement of nucleotides is clockwise, the AUGs are highlighted. Free energy dG = -186.70
Supplementary Figure 7. *PTPRJ* reporter expression at different cell densities.

HT29 cells were transfected with the indicated firefly luciferase reporters (vector pGL3, core *PTPRJ* promoter, wildtype construct as in Fig.5; double frame-shift analogous to Fig.7), or with a pGL2 control vector (luciferase expression driven by the SV40 promoter), and cotransfected with pRL-TK. Cells were cultivated for different length of time, and the relative firefly luciferase activity (normalized to *Renilla* luciferase) was measured. Note that activity of the *PTPRJ* promoter is increasing with increasing cell density, while the SV40 promoter activity decreases under these conditions. Elevated activity of the construct with altered amino acid sequence between AUG$_{191}$ and AUG$_{356}$ (caused by double frame-shift) is maintained at all time points.
### Supplementary Table 1. Promoter and leader sequences of the human receptor like PTPs.

| Name   | Gene symbol | cDNA Accession | CpG island promoter | 5’UTR, length (nt) | uAUGs, position from 5’end (nt) | uORF, length (codons) |
|--------|-------------|----------------|---------------------|-------------------|---------------------------------|------------------------|
| 1.     | hCD45       | PTPRC          | No                  | 181               | 55                              | 10                     |
|        |             |                |                     |                   |                                 |                        |
|        |             |                |                     |                   | 140                             | 12                     |
|        |             |                |                     |                   | 175                             | No stop, in-frame§      |
| **Subtype R1** |       |                |                     |                   |                                 |                        |
| 2.     | hPTPalpha   | PTPRA          | Yes                 | 676               | 36, 147, 168, 393                | All four uORF in-frame; one stop |
|        |             |                |                     |                   | 241, 286                         | Both uORF in frame; one stop |
|        |             |                |                     |                   | 340, 361                         | Both uORF in frame; one stop |
|        |             |                |                     |                   | 604, 607                         | Both uORF in frame; one stop |
|        |             |                |                     |                   | 463                             | 45                     |
|        |             |                |                     |                   | 579                             | 4                      |
|        |             |                |                     |                   | 600                             | No stop; not-in-frame§§ |
|        |             |                |                     |                   | 632                             | 6                      |
| **Subtype R4** |       |                |                     |                   |                                 |                        |
| 3.     | hPTPepsilon | PTPRE          | Yes                 | 280               | 197                             | 19                     |
|        |             |                |                     |                   |                                 |                        |
| Subtype R2B | 4. hPTPmu | PTPRM | NM_002845 | Yes | No data | No data | No Data |
|---|---|---|---|---|---|---|---|
| 5. hPTPkappa | PTPK | NM_002844 | Yes | 221 | 6 | 19 |
| 6. hPTPrho | PTPRT | NM_007050 | Yes | 185 | 106 | 18 |
| 7. hPTPlambda | PTPRU | NM_005704 | Yes | 111 | None |

| Subtype R2A | 8. hLAR | PTPRF | NM_002840 | Yes | 341 | 188 | 27 |
|---|---|---|---|---|---|---|---|
| 9. hPTPsigma | PTPRS | NM_002850 | Yes | 235 | 79 | 25 |
| 10. hPTPdelta | PTPRD | NM_002839 | Yes | 712 | 307 | 16 |
| | | | | | | 328 | 9 |
| | | | | | | 370 | 32 |
| | | | | | | 476 | 6 |
| | | | | | | 479 | 5 |

| Subtype R5 | 11. hPTPgamma | PTPRG | NM_002841 | Yes | 718 | 1 | 153 |
|---|---|---|---|---|---|---|---|
| | | | | | | 15 | 37 |
| | | | | | | 23 | 25 |
| | | | | | | 225 | 14 |
| | | | | | | 702 | No stop, not-in-frame§§ |
| 12. hPTPzeta | PTPRZ1 | NM_002851 | Yes | 369 | 111 | 4 |

| Subtype R3 | 13. hSAP1 | PTPRH | NM_002842 | No | 42 | None | None |
|---|---|---|---|---|---|---|---|
| 14. hPTP beta | PTPRB | NM_002837 | Yes | 31 | None | None |
| 15. hDEP1 | PTPRJ | NM_002843 | Yes | 356 | 14, 191 | No stop, both in-frame§§ |
| 16. hGLEPP1 | PTPRO | NM_030667 | Yes | 175 | 50 | 12 |
### Subtype R7

|   | Protein | Accession | Start | Stop | In-frame? | uAUG 5′UTR |
|---|---------|-----------|-------|------|-----------|-----------|
| 17. | hPCPTP1 | PTPRR     | NM_002849 | No | 417 | 69 | 56 |
|    |         |           |       |      |           | No stop; not-in-frame |

### Subtype R8

|   | Protein | Accession | Start | Stop | In-frame? | uAUG 5′UTR |
|---|---------|-----------|-------|------|-----------|-----------|
| 18. | hIA2   | PTPRN     | NM_002846 | Yes | 90 | None | None |
| 19. | hIA2beta | PTPRN2   | NM_002847 | Yes | 58 | None | None |

Classification of the receptor type PTPs is according to Andersen JN, RL Del Vecchio, N Kannan, J Gergel, AF. Neuwald and NK. Tonks (2005) Computational analysis of protein tyrosine phosphatases: practical guide to bioinformatics and data resources, Methods 35: 90-114.

1) The search for CpG island promoters was performed in the UCSC Human Genome Browser at [http://genome.ucsc.edu/](http://genome.ucsc.edu/)

2) The search for uORF longer than 4 codons was performed on Clone Manager Professional Suite, version 8.

§) In-frame – uAUGs without stop codon in the 5′UTR. The uAUG and the AUG of the main protein are in the same reading frame.

§§) Not-in-frame – uAUG without stop codon in the 5′UTR. The uAUG and the AUG of the main protein are in different reading frames.
### Supplementary Table 2. Promoter and leader sequences of the human non-transmembrane PTPs

| Name | Gene symbol | cDNA Accession | CpG island promoter\(^1\) | 5’UTR, length (nt) | uAUGs, position from 5’end (nt) | uORF, length \(^2\) (codons) |
|------|-------------|----------------|-----------------------------|-------------------|-------------------------------|-----------------------------|
| 1. hPTP1B | PTPN1 | NM_002827 | yes | 175 | 4 | 9 |
| 2 | hTCPTP | PTPN2 | NM_002828 | yes | 195 | 101 | 14 |
| **Subtype NT2** | | | | | | |
| 3 | hSHP1 | PTPN6 | NM_002831 | No | 234 | None | None |
| 4 | hSHP2 | PTPN11 | NM_002834 | Yes | 381 | 29 | 25 |
| **Subtype NT3** | | | | | | |
| 5 | hMEG2 | PTPN9 | NM_002833 | Yes | 509 | 394 | 8 |
| **Subtype NT4** | | | | | | |
| 6 | hBDP1 | PTPN18 | NM_014369 | Yes | 63 | None | None |
| 7 | hLyPTP | PTPN22 | NM_015967 | No | 90 | 27 | 4 |
| | | | | | | | 43 | No stop; not-in-frame\(^3\) |
| 8 | hPEST | PTPN12 | NM_002835 | Yes | 30 | None | None |
| **Subtype NT5** | | | | | | |
| 9 | hMEG1 | PTPN4 | NM_002830 | Yes | 772 | 145 | 8 |
| | | | | | | | 158 | 31 |
| | | | | | | | 583 | 22 |
| 10 | hPTPH1 | PTPN3 | NM_002829 | No | 24 | None | None |
| Subtype NT6 |  |
|-------------|--|
| 11. hPTPD1  | PTPN21  | NM_007039 | Yes | 332 | 106 | 25 |
|  |  |  |  |  |  | |
|  |  |  |  |  |  | |
|  |  |  |  |  |  | |
| 12. hPTPD2  | PTPN14  | NM_005401 | Yes | 272 | 144 | 22 |
| Subtype NT7 |  |
| 13. hPTPBAS | PTPN13  | NM_006264 | Yes | 64 | None | None |
| Subtype NT8 |  |
| 14. hHDPTP  | PTPN23  | NM_015466 | Yes | 97 | None | None |
| Subtype NT9 |  |
| 15. hPTPTyp | PTPN20* | NM_015605 | Yes | 188 | 177 | No stop, not-in-frame
§§ |
| Subtype NT unclassified |  |
| 16. hHePTP  | PTPN7   | NM_002832 | No | 132 | None | None |
| 17. hSTEP   | PTPN5   | NM_032781 | Yes | 387 | 355 | 11 |

Classification of the receptor type PTPs is according to Andersen JN, RL Del Vecchio, N Kannan, J Gergel, AF. Neuwald and NK. Tonks (2005) Computational analysis of protein tyrosine phosphatases: practical guide to bioinformatics and data resources, Methods 35: 90-114.

1) The search for CpG island promoters was performed in the UCSC Human Genome Browser at [http://genome.ucsc.edu/](http://genome.ucsc.edu/)

2) The search for uORF longer than 4 codons was performed on Clone Manager Professional Suite, version 8.

§§) Not-in-frame – uAUG without stop codon in the 5’UTR The uAUG and the AUG of the main protein are in different reading frames