The translation attenuating arginine-rich sequence in the extended signal peptide of the protein-tyrosine phosphatase PTPRJ/DEP1 is conserved in mammals

Lucchazar Karagyozov^{*1}, Petar Grozdanov², and Frank-D. Böhmer^{*1}

¹Institute of Molecular Cell Biology, CMB, Jena University Hospital, Jena, Germany;

²Department of Cell Biology & Biochemistry, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, USA

*) Correspondence to L.K. (lucho3ka@yahoo.com) or F-D.B. (i5frbo@uni-jena.de)

Abstract

The signal peptides, present at the N-terminus of many proteins, guide the proteins into cell membranes. In some proteins, the signal peptide contains an extended N-terminal region and a recessed hydrophobic signal sequence. Previously, it was demonstrated that the N-terminally extended signal peptide of the human *PTPRJ* contains a cluster of arginine residues, which attenuates translation. The analysis of the orthologous sequences revealed that this sequence is highly conserved among mammals. The *PTPRJ* transcripts in placentals, marsupials, and monotremes encode a stretch of 10 – 14 arginine residues, positioned 11-12 codons downstream of the initiating AUG. The remarkable conservation of the repeated arginine residues in the *PTPRJ* signal peptides points to their key role. Further, the presence of an arginine cluster in the extended signal peptides of other proteins (E3 ubiquitin-protein ligase, NOTCH3) is noted and indicates a more general importance of this cis-acting mechanism of translational suppression.

Introduction

After the start of translation, translation inhibition due to the interaction between the nascent polypeptide chain and the ribosome is reported (Ito and Chiba, 2013; Wilson et al., 2016). In eukaryotes, transient elongation arrest may be caused by consecutive prolines (Artieri and Fraser, 2014) or by an array of positively charged amino acids (Lu and Deutsch, 2008). In a limited number of proteins, the signal peptides are longer than the canonical 20 -25 residues; they consist of more than forty amino acid residues and contain an N-terminal extension and a recessed hydrophobic signal sequence (Hiss et al., 2008). This extension provides a convenient space for positioning of a translation attenuating amino acid sequence.

The human receptor-like protein tyrosine phosphatase, type J (PTPRJ, PTPReta, or DEP1) provides an example of the presence of a down-regulating sequence within the signal peptide. The human PTPRJ is a receptor-like protein tyrosine phosphatase of the R3 subtype, characterized by an extracellular region, containing several tandem fibronectin type III (FNIII) domains, a single transmembrane region, and a single cytoplasmic catalytic domain (Andersen et al., 2001). In human embryonic lung fibroblasts, the *PTPRJ* expression and activity were dramatically increased when cells approached confluence in comparison to sparse cells, suggesting a possible role in cell-density-dependent inhibition of proliferation. Thus, the name high cell density-enhanced phosphatase-1 or DEP1 was proposed (Ostman et al., 1994). *PTPRJ* is expressed in a variety of normal tissues, notably in hematopoietic cells (CD148 antigen), in epithelial tissues, including those of the digestive tract, in the vascular endothelium

(Autschbach et al., 1999). Data suggest that *PTPRJ* is a tumor suppressor in different tissues. In mice, the gene encoding PTPRJ was mapped to a colon cancer susceptibility locus (Sccl) (Ruivenkamp et al., 2002). Negative regulation of the signaling of several receptor-tyrosine kinases (RTKs), including the epidermal growth factor receptor (EGFR) (Tarcic et al., 2009), the platelet-derived growth factor receptor (Petermann et al., 2011), and Fms-like tyrosine kinase 3 (Arora et al., 2011) may be important in this context. A metabolic function of PTPRJ is indicated by the negative regulation of insulin receptor and leptin receptor signaling (Shintani et al., 2015, Krüger et al., 2015, Shintani et al., 2017). PTPRJ was also identified as an effective activator of Src-kinase in different cell types. This function of PTPRJ is important for platelet activation and thrombosis (Senis et al., 2009, Nagy et al., 2020), for efficient angiogenesis (Fournier et al., 2016), and for regulating airway hyper-responsiveness (Katsumoto et al., 2013).

Previous experiments (Karagyozyov et al., 2008) showed that: (1) the human *PTPRJ* transcripts predominantly initiate translation at the first AUG in a favorable context, numbered AUG₁₉₀. This results in a PTPRJ pre-protein with an N-terminally extended signal peptide and a recessed hydrophobic signal sequence. (2) The N-terminal extension contains an unusual arginine-rich cluster (RRTGWRRRRRRR); its translation inhibits the overall PTPRJ expression.

To elucidate the importance of these findings it was of interest to examine the *PTPRJ* transcripts in mammals for sequences encoding the attenuating arginine-rich cluster. In the present paper, *PTPRJ* orthologs from placental mammals, marsupials, and monotremes were compared. The results revealed a similarity in the architecture of the *PTPRJ* transcripts. Several conserved features were noted: (1) uORFs are not present in the transcripts; (2) the *PTPRJ* transcripts encode signal peptides with N-terminal extension and a recessed hydrophobic signal sequence; (3) the extended signal peptides contain the attenuating arginine cluster. The remarkable evolutionary conservation of the attenuating sequence emphasizes the importance of suppressing the PTPRJ translation by the nascent peptide chain.

Materials and methods

Orthologous genes and data evaluation

The NCBI gene database was used to search for *PTPRJ* orthologs in mammals. The 5' end regions of the *PTPRJ* transcripts were examined for the presence of initiator AUG codons to determine the N-terminal end of the encoded proteins. The presence of a signal hydrophobic amino acid sequence and the signal-peptidase cleavage sites were determined *in silico* using the SignalP 5.0 web server at <https://services.healthtech.dtu.dk/service.php?SignalP-5.0> (Armenteros et al., 2019). In the SignalIP algorithm, the input sequence has an upper limit of 70 residues, thus, initially, the N-terminal amino acids were examined, and then - the adjoining downstream region. The distribution of the elongating and initiating ribosomes in transcripts encoded by exon 1 (450 bp) of the human *PTPRJ* on chromosome 11 was visualized by the genome browser <https://gwips.ucc.ie/> (Michel et al., 2014). BLAST, PHI-BLAST (NCBI) and Clone Manager Suite 8 (Scientific and Educational Software) were used to search the database, compare and analyze the nucleotide and amino acid sequences.

Protein-tyrosine phosphatase PTPRJ/DEP1 sequences

The analyzed 5' end regions of the *PTPRJ* transcripts were from ten species: five placental mammals, four marsupials, and one monotremes.

Placentals:

Primates - *Homo sapiens* (human), mRNA transcript variant 1: NM_002843.4, protein: NP_002834.3

Rodents - *Mus musculus* (house mouse), mRNA transcript variant 1: NM_008982.6, protein: NP_033008.4

Cetaceans - *Delphinapterus leucas* (beluga whale), mRNA: XM_030764039.1, protein: XP_030619899.1

Ruminants – *Bos taurus* (cattle), mRNA: XM_024975918.1, protein: XP_024831686.1

Carnivore - *Enhydra lutris* (sea otter), mRNA: XM_022506988.1, protein: XP_022362696.1

Marsupials:

Monodelphis domestica (gray short-tailed opossum) - mRNA: XM_016422845.1, protein: XP_016278331.1

Phascolarctos cinereus (koala) - mRNA transcript variant X1: XM_021009664.1, protein: XP_020865323.1

Sarcophilus harrisii (Tasmanian devil) - mRNA transcript variant X1: XM_031941343.1, protein: XP_031797203.1

Vombatus ursinus (common wombat) - mRNA transcript variant X1: XM_027837689.1, protein: XP_027693490.1

Monotremes:

Ornithorhynchus anatinus (platypus) – mRNA transcript variant X2: XM_029060042.1, protein: XP_028915875.1

E3 ubiquitin-protein ligase (ZNRFB3) sequences

Homo sapiens (human), mRNA transcript variant 1: NM_001206998.2, protein: NP_001193927.1

Mus musculus (house mouse), mRNA transcript variant 1: NM_001080924.2, protein: NP_001074393.1

Notch receptor 3 (NOTCH3) sequences

Homo sapiens (human), mRNA transcript: NM_000435.3, protein: NP_000426.2

Mus musculus (house mouse), mRNA: NM_008716.3, protein: NP_032742.1

Results and discussion

The *PTPRJ* orthologs in mammals

The NCBI gene database lists 123 mammalian orthologs of the human *PTPRJ*: placental mammals - 118 orthologs, marsupials - 4, monotremes - 1. We analyzed five orthologs from major groups of placentals and all orthologs from marsupials and monotremes.

In some mammalian species, different splice variants encoding different *PTPRJ* isoforms are listed. We examined all isoforms for the presence of a hydrophobic signal sequence near the N-terminus. Only isoforms with a predicted cleavable signal peptide were analyzed further.

The AUG initiator codons in mammalian *PTPRJ* transcripts

The arrangement of the AUG triplets at the 5' end of the mammalian *PTPRJ* mRNA is shown in Fig. 1. In all transcripts, the AUG codons are in-frame with the mature protein, with no intervening stop codons between them (see also Suppl. Fig. 1, 3 and 5). Thus, no uORFs exist in the mammalian *PTPRJ* transcripts.

The number of the initiating AUGs differs between mammalian subdivisions; the placentals have three AUGs, the monotremes – two, and the marsupials – only one. Remarkably, this AUG is homologous to the previously identified (Karagyozyov 2008) preferred starting site for translation in humans AUG₁₉₀ (see below).

It is firmly established that the nucleotides surrounding the AUG codons strongly affect initiation efficiency. Mutagenesis experiments with transfected COS cells (Kozak, 1986) established an optimal context sequence for initiation (RCCAAUGG, R is A or G). The nucleotides at positions -3 and +4 (the A in the AUG codon is +1) are of particular importance. The AUG context is categorized as strong (both crucial positions match the optimal sequence), favorable (only one match), or weak (no match at positions -3 and +4).

In mammalian *PTPRJ*, the nucleotide context of the AUG codons varies according to their scanning order (Table 1). In placentals, the AUG close to the 5' cap (6 – 15 nucleotides) is scanned first; it is in a weak context (CGCAAUGA). The next AUG is in a favorable context with G at -3 and U at +4 (GCCAAUGU); the context of the third AUG is also favorable, however with A at +4 (GCCAAUGA). In monotremes, the context of the AUGs is favorable (GCCAAUGU and GCCAAUGA, respectively). The marsupials possess a single AUG, which is in a favorable context (GCCAAUGU).

To summarize, in all mammals, the AUG in a favorable context, at which the scanning 40S subunits arrive first, is with U in position +4. In placentals, the transcripts contain a preceding AUG in a weak context. In monotremes and placentals, the transcripts possess a downstream AUG in a favorable context, but with A in position +4. The functional significance of these differences is unknown. One may speculate that they reflect subtle differences in expression regulation.

The efficiency of the AUG initiator codons in the human *PTPRJ* transcripts

In humans, there are three in-frame AUG codons positioned upstream of the hydrophobic signal sequence (Fig. 1). In previous experiments, the efficiency of each of the human initiator codons was tested in reporter constructs (Karagyozyov et al., 2008). Briefly, the outcomes were (1) when all three AUGs were mutated no reporter activity was detected; (2) the efficiency of the first initiator (AUG₁₃) was the weakest, and (3) the efficiency of the next two initiators (AUG₁₉₀ and AUG₃₅₅) appeared similar. Translation of the *PTPRJ* mRNA started predominantly at the first initiator in a favorable context (AUG₁₉₀).

These results are in agreement with published RiboSeq data (Fig. 2). The sequences encoded by the first exon of the human *PTPRJ* do not support appreciable non-canonical translation initiation. The first initiator (AUG₁₃), which is in a weak context, is not “tight”; it leaks scanning complexes downstream towards the second initiator. According to the profiles of the initiating ribosomes, the third initiator (AUG₃₅₅) is not active.

The mammalian *PTPRJ* transcripts code for signal peptides with N-terminal extension

The N-termini of the proteins targeted for secretion or membrane integration, harbor a short amino acid sequence – the signal peptide – instrumental for the translocation of the

proteins into the ER membranes. In most cases, the signal peptide is 20 – 25 amino acids long. It contains a positively charged N-terminal region (1 – 5 residues), a hydrophobic signal sequence (7 – 15 residues), and a signal-peptidase recognition site (von Heijne, 1990).

In several proteins, however, the signal peptide is N-terminally extended and the hydrophobic signal sequence is recessed (Hiss et al., 2008). Recent experiments showed that the N-terminal extension of the signal peptides is irrelevant for the signal sequence function (Meyer et al., 2018).

The signal peptides of the mammalian *PTPRJ* proteins are with an N-terminal extension (Fig. 1). In marsupials, the signal peptides are 94-103 residues long (Suppl. Fig. 4). In the platypus (monotremes), the *PTPRJ* mRNA encodes two possible signal peptides. The translation launched from the first AUG results in an N-terminally extended signal peptide of 76 amino acids (Suppl. Fig. 6). In placentals, the AUG initiating codons are three. According to the profiles of the initiating ribosomes, the third initiator (AUG₃₅₅) is not active (Fig. 2). The first and second initiators produce N-terminally extended signal peptides composed of 147 - 150 and 87 – 92 amino acid residues, respectively (Suppl. Fig. 1 and 2).

The arginine-rich sequence in the N-terminally extended signal peptides is highly conserved

The human *PTPRJ* transcripts encode an extended signal peptide, which contains repeated arginine residues (RRTGWRRRRRRR). Earlier it was demonstrated that the arginine cluster attenuates translation (Karagyozyov et al., 2008). A comparison between mammalian orthologs revealed the presence of a similar sequence in the extended signal peptides of all mammals (Fig. 3). Differences are minor. In placentals, the sequence of arginine residues is with three intervening amino acids (TGW/G). In marsupials, the string of arginine residues is interrupted by two amino acids (S/TW). In platypus, the arginine-rich cluster is 15 amino acids with one interruption (G). The arginine repeats in mammals are positioned 11-12 residues downstream of the initiating methionine.

The conservation of the composition and location of the arginine-rich cluster emphasizes the functional importance of these features. In the absence of uORFs, this seems to be a necessary mechanism to down-regulate *PTPRJ* expression.

The mechanism of translation attenuation and potential biological significance

Previous experiments showed that: (1) the translation attenuation is not due to the presence of rare codons; (2) elimination of the repeated arginine residues by frame-shift mutations (plus and minus) is sufficient to up modulate *PTPRJ* expression (Karagyozyov et al., 2008).

Most likely, the inhibition of expression is due to the positive charge of the arginine residues. Stalling of ribosomes at positively charged residues, due to electrostatic interactions with the negatively charged exit tunnel was described in model experiments (Lu and Deutsch, 2008). More recently RiboSeq data were interpreted to show that ribosomes in yeast and mammals stall at positively charged amino acids (Charneski and Hurst, 2013; Sabi and Tuller, 2015). The ribosome exit tunnel accommodates 30 – 40 amino acid residues (Ito and Chiba, 2013; Wilson et al., 2016). Therefore, it is reasonable to assume that in mammalian *PTPRJ* the translating ribosomes stall when the arginine residues of the nascent chain are in the exit tunnel. The result is translation inhibition. The high degree of conservation of this inhibitory sequence is a strong indication of its functional significance.

PTPRJ has numerous cellular substrates, such as RTKs, Src-family kinases, and others. It has been implicated in the regulation of a wide range of cellular functions. Enzymatic studies of the PTPRJ phosphatase domain revealed promiscuity with respect to substrate specificity and a very high intrinsic activity (Barr, 2009). It appears therefore plausible that high levels of PTPRJ protein may disturb cellular functions or may even be toxic. Attenuation of PTPRJ translation by virtue of the here described features of the signal peptide may serve to prevent toxic effects and to allow fine-tuning of expression at the transcriptional level.

Other transcripts, encoding N-terminally extended signal peptides, may use a similar cis-acting mechanism to attenuate expression. An analysis of the extended signal peptides in the precursors of the human E3 ubiquitin-protein ligase ZNRF3 and NOTCH3 revealed the presence of arginine clusters and stretches of proline residues. Both structures may cause ribosome pausing as the nascent chain is synthesized (Fig. 4). In these cases, the efficiency of the nascent chains to throttle translation remains to be elucidated.

Acknowledgments

One of the authors (LK) highly appreciates the help of I. Stancheva for critically reading the manuscript and helpful suggestions.

References

- Andersen, J.N., Del Vecchio, R.L., Kannan, N., Gergel, J., Neuwald, A.F. and Tonks, N.K., 2005. Computational analysis of protein tyrosine phosphatases: practical guide to bioinformatics and data resources. *Methods*, 35(1), pp.90-114.
- Armenteros, J.J.A., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G. and Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnol*, 37(4), pp.420-423.
- Arora, D., Stopp, S., Böhmer, S.A., Schons, J., Godfrey, R., Masson, K., Razumovskaya, E., Rönnstrand, L., Tänzer, S., Bauer, R., Böhmer, F.D., and Müller, J.P., 2011. Protein-tyrosine phosphatase DEP-1 controls receptor tyrosine kinase FLT3 signaling. *J Biol Chem*, 286(13), pp.10918-10929
- Artieri, C.G. and Fraser, H.B., 2014. Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. *Genome Res*, 24(12), pp.2011-2021.
- Autschbach, F., Palou, E., Mechttersheimer, G., Rohr, C., Piroto, F., Gassler, N., Otto, H.F., Schraven, B. and Gaya, A., 1999. Expression of the membrane protein tyrosine phosphatase CD148 in human tissues. *Tissue Antigens*, 54(5), pp.485-498.
- Barr, A.J., Ugochukwu, E., Lee, W.H., King, O.N., Filippakopoulos, P., Alfano, I., Savitsky, P., Burgess-Brown, N.A., Müller, S. and Knapp, S., 2009. Large-scale structural analysis of the classical human protein tyrosine phosphatome. *Cell*, 136(2), pp.352-363.

Charneski, C.A. and Hurst, L.D., 2013. Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biology*, 11(3).

Fournier, P., Dussault, S., Fusco, A., Rivard, A. and Royal, I., 2016. Tyrosine phosphatase PTPRJ/DEP-1 is an essential promoter of vascular permeability, angiogenesis, and tumor progression. *Cancer Res*, 76(17), pp.5080-5091.

von Heijne, G., 1990. The signal peptide. *J Membrane Biol*, 115(3), pp.195-201.

Hiss, J.A., Resch, E., Schreiner, A., Meissner, M., Starzinski-Powitz, A. and Schneider, G., 2008. Domain organization of long signal peptides of single-pass integral membrane proteins reveals multiple functional capacity. *Plos One*, 3(7), p.e2767.

Ito, K. and Chiba, S., 2013. Arrest peptides: cis-acting modulators of translation. *Ann Rev Biochem*, 82, pp.171-202.

Karagyozov, L., Godfrey, R., Böhmer, S.A., Petermann, A., Hölters, S., Östman, A. and Böhmer, F.D., 2008. The structure of the 5'-end of the protein-tyrosine phosphatase PTPRJ mRNA reveals a novel mechanism for translation attenuation. *Nucleic Acids Res*, 36(13), pp.4443-4453

Katsumoto, T.R., Kudo, M., Chen, C., Sundaram, A., Callahan, E.C., Zhu, J.W., Lin, J., Rosen, C.E., Manz, B.N., Lee, J.W. and Matthay, M.A., 2013. The phosphatase CD148 promotes airway hyperresponsiveness through SRC family kinases. *J Clin Invest*, 123(5), pp.2037-2048.

Kozak, M., 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, 44(2), pp.283-292.

Krüger, J., Brachs, S., Trappiel, M., Kintscher, U., Meyborg, H., Wellnhofer, E., Thöne-Reineke, C., Stawowy, P., Östman, A., Birkenfeld, A.L. and Böhmer, F.D., 2015. Enhanced insulin signaling in density-enhanced phosphatase-1 (DEP-1) knockout mice. *Mol Metabolism*, 4(4), pp.325-336.

Lu, J. and Deutsch, C., 2008. Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J Mol Biol*, 384(1), pp.73-86.

Meyer, C., Barniol, L., Hiss, J.A. and Przyborski, J.M., 2018. The N-terminal extension of the *P. falciparum* GBP130 signal peptide is irrelevant for signal sequence function. *Int J Med Microbiol*, 308(1), pp.3-12.

Michel, A.M., Fox, G., M. Kiran, A., De Bo, C., O'Connor, P.B., Heaphy, S.M., Mullan, J.P., Donohue, C.A., Higgins, D.G. and Baranov, P.V., 2014. GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res*, 42(D1), pp.D859-D864

Nagy, Z., Mori, J., Ivanova, V.S., Mazharian, A. and Senis, Y.A., 2020. Interplay between the tyrosine kinases Chk and Csk and phosphatase PTPRJ is critical for regulating platelets in mice. *Blood*, 135(18), pp.1574-1587.

Östman, A., Yang, Q. and Tonks, N.K., 1994. Expression of DEP-1, a receptor-like protein-tyrosine-phosphatase, is enhanced with increasing cell density. *Proc Nat Acad Sci USA*, 91(21), pp.9680-9684.

Petermann, A., Haase, D., Wetzel, A., Balavenkatraman, K.K., Tenev, T., Gührs, K.H., Friedrich, S., Nakamura, M., Mawrin, C. and Böhmer, F.D., 2011. Loss of the protein-tyrosine phosphatase DEP-1/PTPRJ drives meningioma cell motility. *Brain Pathol*, 21(4), pp.405-418.

Ruivenkamp, C.A., van Wezel, T., Zanon, C., Stassen, A.P., Vlcek, C., Csikós, T., Klous, A.M., Tripodis, N., Perrakis, A., Boerrigter, L. and Groot, P.C., 2002. Ptpnj is a candidate for the mouse colon-cancer susceptibility locus Sccl and is frequently deleted in human cancers. *Nature Genetics*, 31(3), pp.295-300.

Sabi, R. and Tuller, T., 2015. A comparative genomics study on the effect of individual amino acids on ribosome stalling. *BMC Genomics*, 16(S10), p.S5.

Senis, Y.A., Tomlinson, M.G., Ellison, S., Mazharian, A., Lim, J., Zhao, Y., Kornerup, K.N., Auger, J.M., Thomas, S.G., Dhanjal, T. and Kalia, N., 2009. The tyrosine phosphatase CD148 is an essential positive regulator of platelet activation and thrombosis. *Blood*, 113(20), pp.4942-4954.

Shintani, T., Higashi, S., Suzuki, R., Takeuchi, Y., Ikaga, R., Yamazaki, T., Kobayashi, K. and Noda, M., 2017. PTPRJ inhibits leptin signaling, and induction of PTPRJ in the hypothalamus is a cause of the development of leptin resistance. *Scientific Rep*, 7(1), pp.1-14.

Shintani, T., Higashi, S., Takeuchi, Y., Gaudio, E., Trapasso, F., Fusco, A. and Noda, M., 2015. The R3 receptor-like protein tyrosine phosphatase subfamily inhibits insulin signalling by dephosphorylating the insulin receptor at specific sites. *J Biochem*, 158(3), pp.235-243.

Tarcic, G., Boguslavsky, S.K., Wakim, J., Kiuchi, T., Liu, A., Reinitz, F., Nathanson, D., Takahashi, T., Mischel, P.S., Ng, T. and Yarden, Y., 2009. An unbiased screen identifies DEP-1 tumor suppressor as a phosphatase controlling EGFR endocytosis. *Curr Biology*, 19(21), pp.1788-1798.

Wilson, D.N., Arenz, S. and Beckmann, R., 2016. Translation regulation via nascent polypeptide-mediated ribosome stalling. *Curr Opinion Structural Biol*, 37, pp.123-133.

Figure 1.

Sequence pattern of the 5'-end of the mammalian *PTPRJ* transcripts.

The AUG codons are indicated. The position of the transcripts is adjusted to align the first initiating AUGs .

Arg-cluster – repeated Arg residues; Sig Seq – region of nonpolar amino acids;

PTPRJ – the mature protein

Figure 2.

Ribosome profiling of *hPTPRJ*, exon 1.

Top – the AUGs in reading frame 2 are shown (green). Middle - RiboSeq data, initiating ribosomes. The positions of the Arg-cluster (yellow) and the signal sequence (grey) are indicated. Bottom – the RefSeq gene, exon 1.

Placentals	1	MSPGKPGAGGAGT	RR	TGW	RRRRRRRR	LEAATRAPGLGRTA
Marsupials	1	MSPGKPGAGGA	RRRR	SW	RRRRRRRR	PRPRPPAPA-----
Monotremes	1	MSPGKPGAGETPP	RRRRRR	G	RRRRRRRRRR	PQPGPATTKAAA

Figure 3.

Comparison of the arginine cluster in the N-terminally extended signal peptides in mammals (placentals and marsupials - consensus sequences, monotremes – platypus, see Suppl. Fig. 2, 4 and 6).

The initiating Met residues (green), the conserved Arg-residues (yellow) and the intervening amino acids (grey) are marked.

E3 ubiquitin-protein ligase ZNRF3 isoform 1 precursor

Human	M	R	P	R	S	G	G	R	P	G	A	T	G	R	R	R	R	L	R	R	R	P	R	G	L	R	C	S	R	L	P	P	P	P	L	P	L	L	L	G	L	L	L	A	A	G	P	G	A	A	R	A	▼	K	E
Mouse	M	R	P	R	S	G	G	R	P	G	A	P	G	R	R	R	R	L	R	R	G	P	R	G	-	-	-	R	R	L	P	P	P	P	L	P	L	L	L	G	L	L	L	A	A	G	P	G	A	A	R	A	▼	K	E

Notch receptor 3 (NOTCH3)

Human	M	G	P	G	A	R	G	R	R	R	R	R	P	M	S	P	P	P	P	P	P	P	V	R	A	L	P	L	L	L	L	L	L	A	G	P	G	A	▼	A	P
Mouse	M	G	L	G	A	R	G	R	R	R	R	R	L	M	A	L	P	P	P	P	P	P	M	R	A	L	P	L	L	L	L	L	L	A	G	L	G	A	▼	A	P

Figure 4.

The N-terminally extended signal peptide of the E3 ubiquitin-protein ligase ZNRF3 and NOTCH3 in humans and mouse.

The initiating Met residues (green), the Arg-cluster (yellow), the consecutive Pro residues (underlined) and the nonpolar signal sequences (grey) are shown. The signal peptidase cleavage site (▼) was predicted *in silico* by the SignalP 5.0 Server.

Table 1.

Context of the AUG codons in the 5' end of the *PTPRJ* transcripts.

(The AUG codons are arranged in 5' to 3' direction according to the movement of the scanning complexes).

Placentals.

Organism	1 st AUG	2 nd AUG	3 rd AUG
Consensus	CGC <u>AUG</u> A	GCC <u>AUG</u> U	GGC <u>AUG</u> A

Monotremes.

Organism	1 st AUG	2 nd AUG
Platypus	GCC <u>AUG</u> U	GGC <u>AUG</u> A

Marsupials.

Organism	1 st AUG
Consensus	GCC <u>AUG</u> U

The optimal initiator sequence is RCCAUGG, R is A or G (Kozak, 1986). Critical positions are R at -3 and G at +4 (the A in the AUG codon is +1). The nucleotides surrounding the AUGs are from Suppl. Fig. 1, 3, and 5.