# Characterization of cDNA for murine tripeptidyl-peptidase II reveals alternative splicing

## **Birgitta TOMKINSON**

Department of Veterinary Medical Chemistry, Swedish University of Agricultural Sciences, Box 575, S-751 23 Uppsala, Sweden

Tripeptidyl-peptidase II (TPP II) is a cytosolic high- $M_r$  exopeptidase with an active site of the subtilisin type. This paper describes cloning of cDNA encoding murine TPP II. Four clones were isolated from a murine mastocytoma cDNA library and the 5'-end was isolated by use of 5'-RACE (rapid amplification of cDNA ends). A total of 4611 bp were isolated, including the complete coding region. The deduced amino acid sequence shows a 96% overall identity when compared with the previously cloned human TPP II. The remarkably high identity indicates that not only the catalytic domain, but almost the entire subunit,

### INTRODUCTION

Tripeptidyl-peptidase II (TPP II) (EC 3.4.14.10) is a large intracellular exopeptidase [1], with a broad tissue and species distribution [2,3]. TPP II has been purified from rat liver and human erythrocytes [2], and the cDNA encoding the human enzyme has been cloned and sequenced [4,5].

The enzyme has a rather broad substrate specificity, i.e. it removes tripeptides with little apparent similarity from a free Nterminus of larger peptides [1,2]. In addition, the peptidase appears to be structurally conserved both immunologically [3] and at the DNA level [4]. The physiological role of TPP II has yet to be established, but these facts, taken together with its broad distribution, would indicate a general role in intracellular proteolysis.

One remarkable property of the enzyme is its large size. The native enzyme occurs in a complex  $(M_r > 10^6 [1,2])$  which is built up of subunits with an  $M_r$  of 138000. The oligomeric structure has been shown to be a prerequisite for enzymic activity [6]. Human TPP II has a catalytic domain which shows 56% similarity to that of bacterial serine peptidases of the subtilisintype [4]. The function of the remainder of this large subunit is not known. Since identification of domains which are conserved between species would localize parts of the subunit which are functionally important, the cloning and sequencing of murine TPP II was undertaken.

### **MATERIALS AND METHODS**

### Screening of a cDNA library

A cDNA library containing  $4 \times 10^6$  independent clones was constructed in the  $\lambda gt$  11 vector with mRNA from murine mastocytoma [7]. A 3.9 kbp *Kpn*I fragment corresponding to the complete coding sequence of human TPP II cDNA [8] or the 1260 bp 5'-end of the fragment was labelled to high specific radioactivity with [ $\alpha$ -<sup>32</sup>P]dCTP by random priming [9] and used as a probe in the screening of the library. Hybridization was must be of functional importance. Alignment with subtilisin-like serine peptidases identified Asp<sup>44</sup>, His<sup>264</sup> and Ser<sup>449</sup> as the catalytic triad, thus defining an extra domain of  $\sim 200$  amino acids between the catalytic Asp and His in TPP II as compared with other subtilases. In addition, it was demonstrated that different polyadenylation signals can be utilized, since two different clones with untranslated 3'-ends of 155 bp and 781 bp respectively have been isolated. Finally, one of the isolated clones contains an extra 39 bp insert encoding 13 amino acids, which implies alternative splicing of the mRNA.

carried out overnight at 42 °C in the low-stringency hybridization solution used previously to detect murine genomic DNA [4]. The filters were washed twice for 5 min with  $2 \times SSC$  ( $1 \times SSC =$ 0.15 M NaCl/0.015 M sodium citrate buffer, pH 7.0)/0.2 % SDS at room temperature and twice for 25 min at 58 °C with  $0.7 \times SSC/0.5$  % SDS. Positive clones were plaque-purified and  $\lambda$ -phages were then prepared by the liquid-culture method [10]. The DNA was extracted by standard methods [10]. Fragments obtained after digestion of positive clones with *Eco*RI were subcloned into the plasmid vector pUC 19 [11].

### Nucleotide sequence determination

The nucleotide sequence was determined in alkali-denatured pUC 19 using the dideoxy-chain-termination reaction with [ $\alpha$ <sup>35</sup>S]dATP and the modified T7 DNA polymerase (Sequenase; United States Biochemical Corp., Cleveland, OH, U.S.A.) [12]. Specific or universal oligonucleotide primers (binding on either side of the pUC 19 polylinker) were used. Sequence determination was made on both strands in at least one clone and on at least one strand in all clones.

### 5'-RACE experiments

5'-RACE (rapid amplification of cDNA ends) was performed according to the manufacturer's recommendations [13], using the 5'AmpliFINDER RACE kit (Clontech, Palo Alto, CA, U.S.A.). Briefly, first-strand cDNA was synthesized using polyadenylated [poly(A)<sup>+</sup>] RNA from murine mastocytoma cells (2  $\mu$ g) as template and a gene-specific oligonucleotide primer BT24 (GCC ACA CCA AGC AGT CAT ATA CAG G). The RNA was thereafter hydrolysed and the cDNA purified. After ligation of an oligonucleotide anchor to the 3'-ends of the cDNA, an aliquot (2  $\mu$ l of a 1: 40 dilution) of the anchor-ligated cDNA was used as a template in a PCR in 10 mM Tris buffer, pH 8.3, containing 50 mM KCl, 0.01% gelatin, 2 mM MgCl<sub>2</sub>, 0.2 mM of each dNTP and 2.5 units of Taq DNA polymerase in a total volume of 50 µl. After heating for 1 min at 82 °C, the anchor primer and a nested gene-specific primer (BT 25, TTA GGA TCC TTC CGT TCT TTC TGT ATC CTT TCC) were added to a final concentration of 0.2  $\mu$ M. Amplification was started immediately and continued for 40 cycles with a temperature-cycling programme (94 °C for 45 s; 60 °C for 45 s and 72 °C for 2 min) in a Gene-ATAQ Controller (Pharmacia-LKB, Uppsala, Sweden). An aliquot (20%) of the reaction mixture was analysed by agarose gel-electrophoresis and found to contain a dominant 500 bp band (results not shown). The rest of the mixture (from two identical samples) was digested with EcoRI (site included in the anchor sequence) and BamHI (site included in BT 25). The 500 bp fragment was purified from an agarose gel by use of the Wizard PCR-preps (Promega, Madison, WI, U.S.A.) and subcloned into pUC 19. A total of nine individual clones were sequenced and shown to start at bp 1, 10, 20, 30, 43 (two clones), 47, 53 and 88 respectively.

For the characterization of the 5'-end of the human transcript, 5'-RACE-Ready cDNA from human placenta was obtained from Clontech. Using 2  $\mu$ l of this anchor-ligated cDNA as template, PCR (30 cycles) was carried out as described above, with BT 24 and the anchor primer. A 2  $\mu$ l portion of a 1:10 dilution of the reaction mixture was used as template in a secondary PCR with BT 25 and the anchor primer. Part of the reaction mixture was analysed by agarose-gel electrophoresis. Several bands between 200 and 500 bp could be detected. The remainder of the mixture was digested with *Eco*RI and *Bam*HI, and the 500 bp fragment was purified and subcloned as described above. A total of five individual clones were sequenced and shown to start at bp 1 (two clones), 17 and 20 (two clones) respectively.

### Northern-blot analysis

Multiple tissue Northern blots with 2  $\mu$ g of poly(A)<sup>+</sup> RNA from different murine and human tissues were purchased from Clontech. A 3.8 kbp cDNA fragment, corresponding to the murine N10 clone, or a 956 bp cDNA fragment, corresponding to the 5'-end of the human A2 clone [4], were labelled by random priming [9] and used as probes. Hybridization was carried out according to the manufacturer's recommendations. The Northern blots were also hybridized with  $\beta$ -actin to ensure equal loading of RNA on the blots.

### **PCR** experiments

First-strand cDNA synthesis, using 2  $\mu$ g of poly(A)<sup>+</sup> RNA from murine mastocytoma cells and random hexamers, was performed according to the manufacturer's recommendations, using the 5'AmpliFINDER RACE kit. The murine cDNA was used without purification. 5'-RACE-Ready cDNA from human placenta (2  $\mu$ l) or 7 % of the preparation of murine cDNA was used as template in the PCR. For the detection of the N10 extra insert, oligonucleotides BT 17 (AGA TTA TAC AAT TCG ACT AC) and BT 6 (TGT CAT CCA CTG AAT TTT AA) were used. The primers were expected to give rise to two fragments of 490 and 451 bp respectively, depending on whether or not the extra insert was present. For the detection of the N17 extra insert, a degenerated antisense oligonucleotide corresponding to bp 3430-3411 [4] and BT 7 (CTA CTT AAT ACC TCC ACC AA) were used. These oligonucleotide primers were expected to give rise to products of 545 and 375 bp respectively.

For the PCR, DNA template was mixed with oligonucleotide primers (1  $\mu$ M each), dNTP (200  $\mu$ M each) and 1.3 units of *Taq* DNA polymerase in 10 mM Tris buffer, pH 9.0, containing 50 mM KCl, 0.1 % Triton X-100 and 1.5 mM MgCl<sub>2</sub>, in a total volume of 50  $\mu$ l. The reaction was run for 35 cycles, including denaturing for 45 s at 95 °C, annealing for 45 s at 45 °C and extension for 2 min at 72 °C. After PCR, the samples for detection of the N10 extra insert were precipitated with ethanol, dissolved, and digested with *PstI*.

### **RESULTS AND DISCUSSION**

# Isolation and characterization of cDNA clones encoding murine TPP II

Six independent clones were isolated when  $1 \times 10^6$  clones from a murine mastocytoma cDNA library [7] were screened with a fulllength human TPP II cDNA probe [8] as described in the Materials and methods section. An additional four clones were identified when the library ( $2 \times 10^6$  clones) was rescreened using as a probe a 1260 bp fragment corresponding to the 5'-end of the human cDNA. The relationship between the four largest clones, designated K1, K9, N10 and N17, and the previously cloned human TPP II cDNA [4,5] is outlined in Figure 1.

Since the first base-pair of the murine sequence corresponded to bp 118 in the human sequence, experiments using RACE were performed. A PCR product of about 500 bp was obtained and subcloned into pUC 19. A total of nine individual clones obtained through RACE were sequenced. Together the four clones and the 5'-RACE products encode 4611 bp, consisting of 83 bp of the 5'-untranslated region, the complete coding sequence and 781 bp of the 3'-untranslated region (Figure 2). The murine cDNA is extended 67 bp toward the 5'-end as compared with the previously isolated human cDNA [5]. Therefore 5'-RACE was used also for the isolation of the human 5'-end. The longest isolated human clone contained an untranslated 5'-region of 53 bp (Figure 2). No substantially longer PCR products could be amplified, using either murine or human cDNA. Since the leader sequences of most of the vertebrate mRNAs fall in the range of 20-100 nucleotides [15], the obtained 5'-ends could represent the true starting point for transcription. However, the variable starting points of individual clones (cf. the Materials and methods section) may indicate the presence of secondary structures in the transcripts. Therefore the final identification of the transcription starting points must await the characterization of the genes encoding TPP II.

The 3'-untranslated region includes a possible polyadenylation signal starting at bp 4552, but no poly(A) tail. Since the polyadenylation signal is usually 10–35 nucleotides upstream from the poly(A) [16], it is possible that this signal is not utilized and that the last part of the non-coding 3'-end is missing. This is corroborated by the comparison with human TPP II cDNA, since the last base-pair in the murine cDNA corresponds to human bp 4621, which is 35 bp upstream of the poly(A) in the human cDNA (results not shown).

Two of the isolated clones, N10 and N17, contain extra inserts of 39 bp and 170 bp respectively (Figure 1), which will be discussed below.

### **Comparison of murine and human TPP II**

The nucleotide sequences for the murine and human TPP IIcDNAs [4] show 90 % identity between the coding parts (Figure 2) and 92 and 74 % identity between the untranslated 5'- and 3'ends respectively. One gap was introduced in the human untranslated 5'-end (Figure 2), and several gaps in the untranslated 3'-ends (results not shown). A number of the base substitutions within the coding region are 'silent', making the similarity between the deduced amino acid sequences even higher. On the amino acid level, an overall identity of 96 %, and a similarity of



Figure 1 Relationship between the four murine clones, the 5'-RACE product and human TPP II cDNA

Bars and lines represent coding and non-coding regions respectively. The numbers refer to start and finish of the clone compared with the assembled sequence presented in Figure 2. The N10 and N17 clones contain extra inserts of 39 bp and 170 bp respectively, which are indicated by the boxes. A indicates the poly(A) tail.

98% are obtained, since a number of the amino acid substitutions are conservative (Figure 2). This high degree of sequence similarity is remarkable for such a large protein. As a comparison it can be mentioned that murine furin (a subtilisin-like serine peptidase of 793 amino acid residues) is 94% identical with its human counterpart [17], and that murine blood-coagulation Factor VIII (2320 amino acid residues) shows 74% identity with the human enzyme, even though functionally important domains are more conserved (84–93% identity) [18].

Interestingly, there is one small part near the C-terminus, residues 1140-1183, where 13 out of 44 amino acids (i.e. 30%) are different in the two sequences. The reason for this hypervariability is not understood, but will be the subject of further investigations.

It was expected that the catalytic domain would be well conserved. However, evidently the entire subunit of the enzyme is conserved with the exception of the small domain close to the C-terminus, thus implying that practically all parts of the subunit are of importance for the function of the enzyme.

### **Comparison with other proteins**

Murine TPP II was also compared with other proteins in the Swiss database (release 27.0, containing 33 329 sequences) by use of the program BLAST [19]. Out of the 65 highest scoring proteins, 44 were serine proteinases of the subtilisin-type (subtilases). The alignment identified Asp44 as the aspartic residue of the catalytic triad (Figure 3) and not Asp<sup>211</sup>, as previously suggested [4]. Thus in TPP II there are 219 amino acids between the Asp and His of the catalytic triad, compared with 31 in subtilisin BPN' (Figure 3). This large insert makes it virtually impossible to obtain an unambiguous alignment to the corresponding area in other subtilases by use of the GAP command of the GCG program [14]. Large inserts in the catalytic domain have previously been reported for other large subtilases, e.g. the cell-wall proteinases from Lactococcus lactis have inserts of 151 amino acid residues in one of the variable regions between the His and the Ser residues of the catalytic triad [20]. However, the position of the large insert in TPP II makes it unique among the subtilases. Since TPP II is also the only hitherto-identified exopeptidase belonging to this peptidase class [20], I speculate that the extra domain may be of importance for substrate recognition, e.g. for the observed dependence of a free Nterminus [1,2].

The majority of the other similar sequences which were identified by the BLAST program recognized similarities between charged areas of TPP II (e.g. residues 999–1029, where 20 out of 30 amino acids are charged) and other proteins such as calreticulin [27] and nucleolin [28]. The function of the charged domains in the latter proteins is not known, but it has been speculated that they could be of importance for  $Ca^{2+}$  binding [27] or histone interactions [28]. However, acidic and/or basic domains can be found in a number of different proteins, and the functional significance of the observed similarity will have to be investigated experimentally.

### TPP II utilizes two different polyadenylation signals

A Northern-blot analysis shows that both murine and human TPP II are encoded by two different mRNAs, differing about 0.7 kb in size (Figure 4). When this was observed previously for the human enzyme, it was assumed to be due to the utilization of two different polyadenylation signals [4]. This has now been shown for the murine enzyme. Thus, in addition to the K9 clone, which contains a 781 bp untranslated 3'-end, one of the clones, N17, has a 155 bp untranslated 3'-end containing a short poly(A) tail (6 bp), starting at bp 3980, and a possible polyadenylation signal, ATTAAA [16], starting at bp 3955 (Figure 2). The polyadenylation signal is conserved between the human and the murine cDNA and was previously proposed to be utilized in the short human mRNA [4].

The broad tissue distribution reported previously [2] is confirmed by the Northern-blot analysis (Figure 4). Even though no reaction can be seen with pancreatic RNA in Figure 4, bands of the expected size can be detected after overexposure of the blot (results not shown). The weak reaction in pancreas could be due to the fact that a substantial fraction of the mRNA from pancreas represents organ-specific transcripts. It is also evident from Figure 4 that both mRNAs are present in about equimolar amounts in most of the tissues analysed. However, while the larger mRNA species dominates in brain, testis contains more of the short mRNA, thus implying a functional difference between the two mRNAs.

### Investigation of extra inserts and evidence for alternative splicing

The N17 clone contains an extra insert of 170 bp between bp 3294 and 3295 (results not shown). This insert is tentatively considered as a cloning artefact, since (i) it does not contain an open reading frame, (ii) it does not have the characteristics of an intron sequence [29] and (iii) the expected 545 bp product cannot be demonstrated by PCR experiments in either murine or human

	10	30	50	70	90	110	00
1 1 1	gggtagccgggtgtcctcgcgccg	a a ctcgtccgcgcgctgcctggc	g cogtttgcctcttcctcgtc	- t c tgtcctccagcttgcgtccat M	t t tggccaccgccgcgaccgagg A T A A T E E	agccctttcctttcc PFPFH	90 120 13
91 121 14	130 cgg acggtcttctaccaaagaaggaga GLPKKET	150 a g c ccgggggcctcctccttcctgt G A S S F L C A	170 tgccgctacccggagtatga C R Y P E Y D	190 t g cgggcgcgggggtgctcatcgc G R G V L I A	210 a g ccgtcctggacacaggggtcg V L D T G V D A	230 cgt atcccggggccccgg PGAPG	210 240 53
211 241 54	250 gcatgcaggttacaactgatggaa M Q V T T D G K	270 g aaccaaaaaatcattgatatca PKIIDII V	290 httgatacaacaggaagtgg [ D T T G S G	310 c g tgatgtaaatactgctacaga D V N T A T E	330 g g aagtagaaccaaaagatggtg V E P K D G E	350 g g c laaattattggtcttt I G L S V	330 360 93
331 361 94	370 a ctggaagagtgcttaagattcctg G R V L K I P A	390 g cc ccaaactggacaaatcctttag N W T N P L C S S	410 c ggaaaatatcatattggcat 5 K Y H I G I	430 a c taaaaatggttatgacttcta K N G Y D F Y	450 t a ptccaaaggctctcaaggaaa PKALKER	470 Iggatacagaaagaac IQKER	450 480 133
451 481 134	490 c tg ggaaggaaaaaatctgggatccaa K E K I W D P I V V	510 g c ttcacagagttgcacttgca H R V A L A I	530 c g gaagcttgtagaaaacaaga E A C R K Q E	550 g agaatttgatattgccaacaa E F D I A N N V	570 c t a atggctcttcccaagccaata G S S Q A N K	590 g aaactaatcaaggaag C L I K E E	570 600 173
571 601 174	610 act g agttgcaaagtcaagtggaattac LQSQVELL	630 g a ttaattcttttgagaaaaag NSFEKK	650 c c t tatagtgatccaggccctgt Y S D P G P V	670 a atatgactgcttggtgtggcd Y D C L V W H	690 c agt atgatggtgagacctggagag D G E T W R A V	710 ca t jcctgtgttgattcaa C V D S N I	690 720 213
691 721 214	730 ag atgagaatggggacttgagtaaat ENGDLSKC D S S	750 c a g gtgccgtattgagaaactaca A V L R N Y I T	770 a atg aaagaggcccaagagtacag K E A Q E Y S G	790 c lttcttttggcacagctgaga S F G T A E M	810 t ctt tgctgaattactctgtgaaca LNYSVNI	830 a c t a Itttatgacgatggga C Y D D G N	810 840 253
811 841 254	850 acctgctctccattgtgaccagcg L L S I V T S G	870 g a gaggagctcatggaacccat G A H G T H V A	890 t t t gtagcaagtatagccgcagg V A S I A A G	910 a c a Igcattttccagaagagcctg H F P E E P E	930 g a aacggaatggagttgctcctg R N G V A P C	950 t ggtgctcaaattctat 5 A Q I L S	930 960 293
931 961 294	970 c a a ccattaagattggtgatacacggc I K I G D T R L	990 a taagcactatggaaacaggc S T M E T G	1010 a acaggcctcatcagagctat T G L I R A M	1030 gatagaagttataaatcata I E V I N H K	1050 agtgtgatcttgtcaactaca C D L V N Y S	1070 c ngttatggagaagcaa 5 Y G E A T	1050 1080 333
1051 1081 334	1090 c ctcattggccaaattctgggaga H W P N S G R	1110 atttgtgaagtaattaatga I C E V I N E	1130 g agcagtatggaaacataata A V W K H N T I	1150 t caatttatgtttcaagtgct I Y V S S A	1170 ggaaataatggtccatgcctt G N N G P C L	1190 g t ttctacagtgggttgtc S T V G C P	1170 1200 373
1171 1201 374	1210 t a caggaggaactacatccagtgtg G G T T S S V	1230 t ataggtgttggagcttatgt I G V G A Y V	1250 t ttcccctgatatgatggttg S P D M M V A	1270 t gcagagtattcactgagagag E Y S L R E	1290 t a aaactgcctgcaaatcaatat K L P A N Q Y	1310 t cacatggtcttctagag T W S S R G	1290 1320 413
1291 1321 414	1330 a t c g t gcccaagtgctgatggagccctc P S A D G A L	1350 : t g ggtgtgagcatcagtgcacc G V S I S A P	1370 a c aggaggtgctattgcttctg G G A I A S V	1390 t c a jtgcctaactggacattgagg / P N W T L R	1410 g g c gggactcagctaatgaatggg G T Q L M N G	1430 t gacatcaatgtcttccc T S M S S P A	1410 1440 453
1411 1441 454	1450 a a ccaatgcctgtggtggcattgcc N A C G G I A	1470 ac t cctggtactttcagggctgaa LVLSGLK I	1490 t ca c agcaaataatgttgactata A N N V D Y T I	1510 a t t actgtacactcagtcagaaga V H S V R R	1530 c g g gctctagaaaatactgcaata A L E N T A I V	1550 3 g aaaagctgacaatatag K A D N I E	1530 1560 493
1531 1561 494	1570 t t aagtatttgcccaaggacatgga V F A Q G H G	1590 tattattcaggttgacaaagc I I Q V D K A	1610 c g g ttatgactacctcattcaaa Y D Y L I Q M V	1630 t a patacatcatttgctaacaga N T S F Å N R K	1650 t ttaggttttacagttactgtt L G F T V T V	1670 c tggaaataaccgtggta G N N R G I	1650 1680 533
1651 1681 534	1690 t tctacctccgagatcctgtccag Y L R D P V Q	1710 a gtggctgctccttcagatca V A A P S D H	1730 c a tggtgttggcattgagcctg G V G I E P V	1750 g gtatttccagagaacacagaa / F P E N T E	1770 aactctgaaaaaatatcctti N S E K I S F L	1790 c tcagcttcatttagctt Q L H L A L	1770 1800 573
1771 1801 574	1810 g taacttcaaattcatcttgggtt T S N S S W V	1830 cagtgtcccagccatttgga Q C P S H L E	1850 a a a actcatgaatcagtgtcggc L M N Q C R H	1870 tacataaacatacgtgtggac i I N I R V D	1890 cccaggggcttaagagaaggg P R G L R E G	1910 9 gttacattatacagagg L H Y T E V	1890 1920 613
1891 1921 614	1930 tatgtggctatgatatagcatco C G Y D I A S	1950 t c gc c cccaatgcaggtcctttatt P N A G P L F	1970 g cagagttccaatcactgcag R V P I T A \	1990 gttatagcagcaaaagtaaat / I A A K V N	2010 a gagtcatcacattatgatcta E S S H Y D L	2030 agcctttacagatgtac A F T D V H	2010 2040 653
2011 2041 654	2050 c a attttaaacctggtcagattcga F K P G Q I R	2070 g a nagacattttgttgaggttcc R H F V E V P I	2090 g t a tgaagggggcaacctgggctg E G A T W A E	2110 g a g gaagttaccgtgtgttcatgt E V T V C S C	2130 g a g tcttctgaggtatcggcaaa S S E V S A K	2150 g a atttgttcttcatgcag F V L H A V	2130 2160 693
2131 2161 694	2170 a tacagcttgtgaagcagagagca Q L V K Q R A	2190 c c atatcgaagtcatgaattta Y R S H E F Y	2210 t taagttttgttcccttccaa K F C S L P i	2230 g g c gaaaaaggaacacttattgaa E K G T L I E T	2250 cc a t gcttttcctgttttgggcgg A F P V L G G	2270 aaaagcaattgaatttt K A I E F C	2250 2280 733

2254	2290	2310	2330	2350	2370	2390	
2251 2281 734	c gtattgctcgttggtgggcaagtc I A R W W A S L	c c ttagtgatgtcaatattgatt SDVNID	t t tataccatatcattccatgg / T I S F H G	t t gatagtgtgtactgcaccac IVCTAPQ	g agttaaacattcatgcatctg L N I H A S I	c c gaaggaatcaatcgtt E G I N R F	2370 2400 773
2271	2410	2430	2450	2470	2490	2510	
2371 2401 774	cga ttgatgttcagtcctctttaaagt DVQSSLKY	c c atgaagatctggctccttgcc EDLAPC]	a ataactttgaagagctgggt [ T L K S W V N	c a g g gcaaacgctacgcccagtaa Q T L R P V N S	g aa atgcaaaaaccagacctttag AKTRPL( K	ggatcaagagatgttt G S R D V L	2490 2520 813
2491	2530	2550	2570	2590	2610	2630	7610
2521 814	tgccaaataatcgccagctttatg PNNRQLYE	agatggtccttacatacagc M V L T Y S I N	tttcatcagcccaagagcgg F H Q P K S G	agaagtaacacctagttgtc E V T P S C P	cactcctttgtgaattgtta L L C E L L Y	tatgagtcagaatttg Y E S E F D	2640 853
2611	2650	2670	2690 +	2710	2730	2750	7730
2641 854	acagtcagttgtggattattttcg S Q L W I I F D	accagaacaaaagacagatgg ) Q N K R Q M (	<b>ggctcaggcga</b> tgcctatcc G S G D A Y P	acatcagtattctctgaaat H Q Y S L K L	tggagaaaggagattatacaa E K G D Y T I	attcgattacagattc I R L Q I R	2760 893
2731	2770 c	2790 act c a	2810 t a	2830	2850 ti	2870 t g	2850
2761 894	gtcatgagcaaatcagtgatttgg H E Q I S D L D E	atcgtctcaaagatcttccg RLKDLP	ttattgtttcacataggtt F I V S H R L	gtctaataccttgagcttag S N T L S L D	atattcatgaaaatcatagc IHENHS 	cttgcacttctaggāa L A L L G K F	2880 933
2851	2890 at a	2910 t c	2930	2950	2970	2990	2070
2881 934	agaagaaatcaagcagcttaacat K K S S S L T L N	taccacccaaatacaatcag PPKYNQI	ccattctttgttacttcctt P F F V T S L	acctgatgataaaataccta PDDKIPK	aggggggcaggacccggatgc G A G P G C Y	taccttgcaggctcct Y L A G S L	3000 973
	3010	3030	3050	3070	3090	3110	
2971 3001 974	a a a tgacattgtcaaagactgagcttg T L S K T E L G	↓ a Igaaagaaagctgatgtgatca i K K A D V I I	ct a cctgttcattactatctcat P V H Y Y L I	c acctccaccaacaagatta PPPTKIK T	g agaatggcagcaaagataaa NGSKDKI	a gaaaaggattcagaaa E K D S E K	3090 3120 1013
	3130	3150	3170	3190	3210	3230	
3091 3121 1014	t a aagagaaagacttgaaagaagagt EKDLKEEF	a ttactgaagcattacgcgata TEALRDI	t ctcaaaattcagtggatgac L K I Q W M T	a g g caagcttgattctactgaca K L D S T D I S	t c tttacaatgaattgaaagaa YNELKE	aa t acatatcctgcttacc TYPAYL N	3210 3240 1053
	3250	3270	3290	3310	3330	3350	
3241 1054	ttcctttgtatgttgcacgtcttc PLYVARLH	t atcaactggatgctgaaaag VQLDAEKI	gaacgaatgaaaagacttaa E R M K R L N	g a tgaaattgttgatgctgcca E I V D A A N	t a atgctgttatttctcacatc A V I S H I I	a c a gatcaaaccgctcttg D Q T A L A	3330 3360 1093
3331	3370 t	3390 t	3410	3430	3450	3470 +	3450
3361 1094	cagtttacattgcaatgaagactg V Y I A M K T D	accccaggcctgatgcagcta ) P R P D A A	actataaaaaatgatatgga T I K N D M D	caagcagaaatccacctga KQKSTLI V	tagatgccctctgcaggaaa D A L C R K	ggatgtgccctggcag G C A L A D	3480 1133
3451	3490 ct cc q	3510 a a cattt cact	3530 g g g	3550 a ctt	3570 t aga at	3590 ta	3570
3481 1134	atcaccttcttcatacacagcctc H L L H T Q P H A Q	acgacgggggcagcagctgga   D G A A A G I   I S T	gatgctgaagcaaaggaaga D A E A K E E G	agaaggagaaagtaccatgg E G E S T M E P L D	aatctctatcggaaacctat S L S E T Y A F	tgggaaactacaaagt W E T T K W	3600 1173
3571	3610	3630	3650	3670	3690	3710	2600
3601 1174	ggacagatctttttgacactaagg T D L F D T K V N	jttttgatatttgcatacaag / L I F A Y K I T .	catgcattagtaaataagat H A L V N K M	gtatgggagagggccttaagt Y G R G L K F	ttgcaaccaaactcgtagaa A T K L V E	gaaaaaccaacaaaag E K P T K E	3720 1213
3691	3730	3750	3770	3790	3810	3830	3803
3721 1214	aaaactggaaaaattgtattcaac N W K N C I Q L	tčatgaaattacttggatgg . M K L L G W	acccactgtgcatcttttac T H C Å S F T	tgaaaactggctccccatca ENWLPIN	tgtatcctcctgattattgt Y P P D Y C	gtattctaaaataaga V F *	3840 1249
3841	3850 accaaaactttacatttcaaaaca	3870 Igaaaattttatagtgaatgg	3890 atatatgaagaattgtgtgg	3910 catttttagtctgactaatg	3930 catgttttcatctaccattc	3950 aatactgatc <u>attaaa</u>	3960
3961	3970 actatatgtatttatcagagaact	3990 cagtggcgtgtggcttaata	4010 catgtaatacagacctctga	4030 catcatgctgctttcctact	4050 gcctccgacactcggcaagt	4070 agagagcactccctgc	4080
4081	4090 ctgccagccccacctgagttggat	4110 gcagctgcacactgcatgag	4130 aggttttcatgttttaattt	4150 tagctgtaaaatgtcatcct	4170 gttttttatttataaaatca	4190 caaaggttaaaacatg	4200
4201	4210 ctaaatttttccaattatatagag	4230 Jgccttaaaaatgctacatta	4250 ggtgtagctaaattatttat	4270 ttgactaaaaaatatgggaa	4290 cataatttccagacttttt	4310 tctaaaataaattca	4320
4321	4330 attagtggggatggtggggacat	4350 taaacaaatggcaactggaaa	4370 aagaaacatgttaaggaaat	4390 ctccataaggaatcataaca	4410 aaaacaaggtatttttaaaa	4430 caccctcaattacata	4440
	4450	4470	4490	4510	4530	4550	
4441	gcatcagaatgagttaaaatttct	ttgtgttggtgacagacagct	gaatctatctggttttgtaa 4610	agaaatacacagtatgttta	taacattgaaatcatgtaaa	atacat <u>gaataaa</u> cgt	4560
4561	gcaaaaccacaggcacagtacaca	catatgcactctgatacctta	tttttt 4611				

## Figure 2 Nucleotide sequence and predicted amino acid sequence of murine TPP II and comparison with human TPP II

The nucleotide sequence and the deduced amino acid sequence of murine and human TPP II [4,5] were compared by use of the command GAP of the GCG program [14]. Nucleotide differences in the human untranslated 5'-region and coding sequence are indicated above the murine nucleotide sequence. A gap introduced in the human nucleotide sequence is represented by '-'. Amino acid differences in the human sequence are indicated below the murine amino acid sequence. Asp, His and Ser of the catalytic triad are marked by  $\Delta$ . Two possible polyadenylation signals are underlined. The asterisk indicates termination codon. The arrow indicates the extra insert in clone N10 (cf. Figure 5).

TPP II	32	EYDGRGVLIA VLDTGVDPGA PGMOVTTDGK PKIID 66
Bacs9	133	TSG <u>GGGINIA VLDTGV</u> NTNH <u>P</u> DLRNNVEQC KDFTV 167
Isp1	38	GVKGKNIKVA VLDTGCDTSH PDLKNQIIGG KNFSD 72
LLP	205	KYKGEGTVVS VIDSGIDPTH KOMRLSDDKD VKLTK 239
BPN'	127	GYTGSNYKVA VIDSGIDSSH POLKVAGGAS MVPSE 161
ISP	37	OTRGRGYKVA VLDTGCDADH POLKARIIGG RNFTD 71
MPC3	143	GLTGRGVVIS ILDDGIEKDH PDLWANYDPL ASYDF 177
		•
TPP II	250	DDGNLLSIVT SGGAHGTHVA SIAAGHFP EEPER-NGVA PGAOILSIKI 296
Bacs9	168	GTTYTNNSCT DROGHGTHVA GSALADGG TGNGV-YGVA PDADLWAYKV 214
Isp1	73	DDGGKEDAIS DYNGHGTHVA GTIAAND SNGGI-AGVA PEASLLIVKV 118
LLP	267	DNNDTITDDT VDEOHGMHVA GIIGANGTGD DPAKSVVGVA PEAOLLAMKV 316
BPN'	157	MVPSETNPFO DNNSHGTHVA GTVAALN NSIGV-LGVA PSASLYAVKV 202
ISP	72	DDEGDPEIFK DYNGHGTHVA GTIAATE NENGV-VGVA PEADLLIIKV 117
MPC3	182	PDPOPRYTPN DENRHGTRCA GEVSATAN NGECG-AGVA ENARIGGVRM 228
		•
TPP II	422	<u>VSISAPGGAI ASVPNWTLRG TQLMNGTSMS SPNACGGIAL VLS</u> 464
Bacs9	335	VEISAPGAAI YSTWFDGG YATISGTSMA SPHAAGLAAK IWA 375
Isp1	221	IDLVAPGENI LSTLP NKK YGKLTGTSMA APHVSGALAL IKS 261
LLP	595	PDITAPGGNI WSTONNNG YTNMSGTSMA SPFIAGSOAL LKO 635
BPN '	303	LOVMAPGVSI OSTLPGNK YGAYNGTSMA SPHVAGAAAL ILS 343
ISP	219	VDLVAPGEDI LSTVPGGK YATFSGTSMA TPHVAGALAL IKO 259
MPC3	344	SSGVVTDPOI VTTDLH-HOC TDKHTGTSAS APLAAGMIAL ALE 385

#### Figure 3 Sequence similarities at the active site between murine TPP II and other subtilases

Sequences were aligned by use of the BLAST program [19]. Gaps (indicated by '-') are introduced in variable regions as described by Siezen et al. [20]. Amino acids identical with those in the TPP II sequence are underlined. Abbreviations: Bacs9, subtilisin from *Bacillus TA39* [21]; Isp1, major intracellular serine proteinase [22]; LLP, cell-wall-associated proteinase from *Lactococcus lactis* [23]; BPN', subtilisin BPN' [24]; ISP, intracellular serine proteinase from *B. polymyxa* [25]; MPC3, murine prohormone convertase 3 [26]. Asp, His and Ser of the catalytic triad are marked with asterisks.

cDNA (Figure 5b, lanes 1 and 4), even after hybridization of the PCR products with murine cDNA (results not shown).

The N10 clone contains an extra insert of 39 bp at bp 3036 (Figure 5a), compared with the other clones (Figure 1). Using PCR, it could be demonstrated that this extra exon is present in mRNA from both murine mastocytoma and human placenta, since both a 451 bp and a 490 bp fragment could be amplified (Figure 5b). The true identities of the PCR products were



# Figure 4 Northern-blot analysis of mRNA from different murine (a) or human (b) tissues

Northern blots were hybridized with a 3.8 kbp murine cDNA fragment (a) or a 956 bp human cDNA fragment (b), as described in the Materials and methods section. The sizes of the RNA ladder markers are shown in kb.



<u>aa</u> K

Pst									ggaaaatttaaaaag:gatgtaatc							A2	Human		
gaaa	agct	:gg	gca	gtc	tgc	ago	aaa	acg	aca	agg	aaa	att	taa	laaa	g:ga	tat	gatc	N10	Murine
K	A	G	0	S	A	A	K	R	0	G	K	F	K	K	D	V	I		



### Figure 5 Investigation of extra inserts in clones N10 and N17

(a) Nucleotide sequence and deduced amino acid sequence of the extra insert in clone N10. The underlined nucleotides are also present in the other clones. The comparison with the human A2 clone [4] is shown above the nucleotide sequence. The *Pst* site used for the identification of PCR fragments containing the extra insert is indicated above the sequence. (b) PCR experiments for the investigation of the extra inserts. Murine (lanes 1–3) or human (lanes 4–6) cDNA was used as template in PCR experiments with primers binding on either side of the N17 extra insert (lanes 1 and 4) or the N10 extra insert (lanes 2, 3, 5 and 6). The experiment was performed as described in the Materials and methods section. The PCR products formed were precipitated with ethanol, dissolved and digested with *Pst* (lanes 3 and 6).

confirmed, not only by their size but also by hybridization to murine cDNA (results not shown). In addition, the larger product could also be identified by the presence of a PstI site (Figure 5b, lanes 3 and 6), dividing the 490 bp product into two fragments of 156 bp and 334 bp. It is noteworthy that when the human cDNA was cloned, one of the isolated clones (A2) contained a 'divergent' 5'-end [4], but it could not at that time be determined whether this was due to alternative splicing or an intron sequence left by incorrect splicing [4]. It is now evident that the last 15 bp of the extra insert in the N10 clone is identical with that divergent end (Figure 5a). Taken together, these results demonstrate that the extra insert is due to alternative splicing. The extra exon represents an open reading frame encoding 13 amino acids. The physiological importance of this extra domain is not known and will be the subject of future investigations.

The technical assistance of Ms. Parivash Joukar, Ms. Ritva Forssell and Ms. Maria Nyström is gratefully acknowledged. The cDNA library and murine mastocytoma mRNA were generously provided by Ms. Inger Eriksson and Dr. Lena Kjellén at this Department. This work was supported by the Swedish Medical Research Council (Project 03X-09914), Stiftelsen Lars Hiertas Minne and Magn. Bergvalls Stiftelse.

### REFERENCES

- 1 Bålöw, R.-M., Ragnarsson, U. and Zetterqvist, Ö. (1983) J. Biol. Chem. 258, 11622–11628
- 2 Bålöw, R.-M., Tomkinson, B., Ragnarsson, U. and Zetterqvist, Ö. (1986) J. Biol. Chem. 261, 2409–2417
- 3 Bålöw, R.-M. and Eriksson, I. (1987) Biochem. J. 241, 75-80
- 4 Tomkinson, B. and Jonsson, A.-K. (1991) Biochemistry 30, 168–174
- 5 Tomkinson, B. (1991) Biomed. Biochim. Acta 50, 727-729

- 6 Macpherson, E., Tomkinson, B., Bålöw, R.-M., Höglund, S. and Zetterqvist, Ö. (1987) Biochem. J. 248, 259–263
- 7 Eriksson, I., Sandbäck, D., Ek, B., Lindahl, U. and Kjellén, L. (1994) J. Biol. Chem. 269, 10438–10443
- 8 Martinsson, T., Vujic, M. and Tomkinson, B. (1993) Genomics 17, 493-495
- 9 Feinberg, A. P. and Vogelstein, B. (1984) Anal. Biochem. 137, 266-267
- 10 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- 11 Vieira, J. and Messing, J. (1982) Gene 19, 259-268
- 12 Tabor, S. and Richardson, C. C. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 4767-4771
- 13 Apte, A. N. and Siebert, P. D. (1993) BioTechiques 15, 890-893
- 14 Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387-395
- 15 Kozak, M. (1987) Nucleic Acids Res. 20, 8125-8132
- 16 Wahle, E. and Keller, W. (1992) Annu. Rev. Biochem. 61, 419-440
- 17 Hatsuzawa, K., Hosaka, M., Nakagawa, T., Nagase, M., Shoda, A., Murakami, K. and Nakayama, K. (1990) J. Biol. Chem. **265**, 22075–22078
- 18 Elder, B., Lakich, D. and Gitschier, J. (1993) Genomics 16, 374-379
- 19 Altshul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) J. Mol. Biol. 215, 403–410

Received 23 May 1994/7 July 1994; accepted 7 July 1994

- 20 Siezen, R. J., de Vos, W. M., Leunissen, J. A. M. and Dijkstra, B. W. (1991) Protein Eng. 4, 719–737
- 21 Narinx, E., Davail. S., Feller, G. and Gerday, C. (1992) Biochim. Biophys. Acta 1131, 111–113
- 22 Koide, Y., Nakamura, A., Uozumi, T. and Beppu, T. (1986) J. Bacteriol. 167, 110–116
- 23 Kok, J., Leenhouts, K. J., Haandrikman, A. J., Ledeboer, A. M. and Venema, G. (1988) Appl. Environ. Microbiol. 54, 231–238
- 24 Wells, J. A., Ferrari, E., Henner, D. J., Estell, D. A. and Chen, E. Y. (1983) Nucleic Acids Res. 11, 7911–7925
- 25 Takekawa, S., Uozumi, N., Tsukagoshi, N. and Udaka, S. (1991) J. Bacteriol. 173, 6820–6825
- 26 Nakayama, K., Kim, W. S., Torii, S., Hosaka, M., Nakagawa, T., Ikemizu, J., Baba, T. and Murakami, K. (1992) J. Biol. Chem. 267, 5897–5900
- 27 McCauliffe, D. P., Lux, F. A., Lieu, T.-S., Sanz, I., Hanke, J., Newkirk, M. M., Bachinski, L. L., Itoh, Y., Siciliano, M. J., Reichlin, M., Sontheimer, R. D. and Capra, J. D. (1990) J. Clin. Invest. 85, 1379–1391
- 28 Srivastava, M., McBride, O. W., Fleming, P. J., Pollard, H. B. and Burns, A. L. (1990) J. Biol. Chem. 265, 14922–14931
- 29 Mount, S. M. (1982) Nucleic Acids Res. 10, 459-472