[Frontiers in Bioscience 13, 754-758, January 1, 2008]

MSPL/TMPRSS13

Hiroshi Kido, Yuushi Okumura

Division of Enzyme Chemistry, Institute for Enzyme Research, The University of Tokushima, 3-18-15 Kuramoto-cho, Tokushima 770-8503, Japan

TABLE OF CONTENTS

1. ABSTRACT

Two variant cDNAs, mosaic seine protease large-form (MSPL) and transmembrane protease serine 13 (TMPRSS13), have been identified from a human lung cDNA library by polymerase chain reaction. The deduced amino acid sequences of these proteins show a type II transmembrane protein structure with a unique and long cytoplasmic tail containing tandem repeat phosphorylation motifs of protein kinases, a transmembrane domain, and a trypsin-like serine protease domain at the extracellular C-terminal side. These proteins have an identical serine protease sequence except for the C-terminal ends, and the consensus protease domain exhibits 42, 39 and 43% sequence identity with those of plasma kallikrein, hepsin and transmembrane protease serine 2 (TMPRSS2), respectively. Although both genes are widely expressed in various tissues, they are predominantly expressed in human lung, placenta, pancreas and prostate. TMPRSS13 is expressed higher than MSPL in thymus, spleen and peripheral blood lymphocytes, particularly in $CD8^+$ cells and $CD19^+$ cells. Enzymatic properties of the recombinant soluble MSPL and TMPRSS13 show that these enzymes preferentially recognize the sites consisting of paired basic amino acid residues, and are strongly inhibited by aprotinin, benzamidine and Bowman-Birk trypsin inhibitor, but poorly inhibited by α_1-antitrypsin and leupeptin. These properties raise the possibility that MSPL and TMPRSS13 play roles in the proteolytic processing of prohormones, precursors of growth factors, and also play roles in the pathogenicity of many viruses and bacteria *in vivo*.

2. INTRODUCTION

Serine proteases are the pivotal regulators in a variety of physiological and pathological conditions such as food digestion, blood coagulation, tissue remodeling, embryonic development, and tumor metastasis (1-4). Most of these enzymes involved in these events are serine proteases secreted by cells or found in biological fluids. In the past decade, due to the development of molecular technology and systematic bioinformatics, a number of cell surface anchored trypsin-like serine proteases have been identified. These proteases are classified into two groups. One is the type I or glycosylphosphatidylinositol-anchored serine proteases which possess a unique C-terminal hydrophobic extension with or without cytoplasmic domain (5, 6). The other group is the type II transmembrane serine proteases (TTSPs) including enteropeptidase (7), hepsin (8), corin (9), and matriptase (10), which have a transmembrane domain near the N-terminus and a proteolytic domain at the C-terminal side. The number of TTSP genes is increasing and over 20 family members have now been identified to date. Among them, we identified unique TTSPs with a transmembrane domain and several potential regulatory modules, named mosaic seine protease large-form (MSPL) (11) and its transcript variant TMPRSS13. We also found transcript variant of MSPL without transmembrane domain, named mosaic seine protease short-form (MSPS) (11). In this review, we summarize current knowledge of MSPL and TMPRSS13 and discuss the unique characters in structure andenzymatic properties of these variants in comparison with the other TTSPs.

3. STRUCTURE

3.1. cDNA and Gene structure

In an attempt to find a membrane-bound trypsin-type serine protease expressed in human lung, we performed a polymerase chain reaction (PCR)-based screening using a human lung cDNA library and degenerate oligonucleotides designed on the basis of the conserved catalytic motif sequences of known trypsin-type serine proteases. From the results of amplified DNA fragment sequences and the 5'- and 3'-rapid amplification of cDNA ends approaches, we have identified one novel gene MSPL with transmembrane domain and its alternatively spliced variant TMPRSS13 with an identical catalytic domain sequence. MSPL cDNA covers 2348 bp with an open reading frame (ORF) of 1743 bp and its transcript variant TMPRSS13 cDNA covers 3372 bp with ORF of 1701 bp. The human MSPL and TMPRSS13 gene variants comprise 12 exons and 13 exons, respectively and are located on chromosome 11q23.2. The murine MSPL and TMPRSS13 orthologue encompass 12 exons and 13 exons, respectively on chromosome 9A5.

3.2. Primary amino acid sequences and domain architectures

The deduced amino acid sequences show 581 residues for MSPL and 567 residues for TMPRSS13. There are five potential *N*-glycosylation sites in both forms with the canonical Asn-X-(Ser/Thr) sequence. The serine protease domains at residues 321–581 for MSPL and 326–567 for TMPRSS13 are identical except for the C-terminal ends, *i.e.* 555–581 for MSPL and 560–567 for TMPRSS13 (Figure 1A). They have the typical catalytic triad (His, Asp and Ser) of serine proteases (12), and a characteristic disulfide bond pattern. The putative proteolytic activation site (Arg) in the Arg-Ile-Val-Gly-Gly motif is identical to those in other serine proteases. MSPL and TMPRSS13 have nine conserved cysteines that are predicted to form and stabilize the catalytic pocket. From the sequence similarity among trypsin-like serine proteases, the S1 binding pocket of MSPL is expected to be composed of Asp^{500} at its bottom, and Gly^{527} and Gly^{537} at its neck, and the corresponding amino acids in TMPRSS13 are also detected, indicating that they are clan SA/family S1-type trypsin-like serine proteases (13). The consensus protease domain of MSPL/TMPRSS13 exhibits 42, 39 and 43% identity with those of human plasma kallikrein (14), hepsin (15), and transmembrane protease serine 2 (TMPRSS2) (16), respectively. MSPL and TMPRSS13 have putative transmembrane domains (Pro^{162}–Trp^{184} and Pro^{167}–Trp^{189}, respectively) at the N-terminal sides. Cytoplasmic domain of these variants has tandem repeat motifs corresponding to the phosphorylation sites of cyclin-dependent kinase Cdc2, CaM kinase II and protein kinase C (17, 18), and is the longest among TTSPs known to date (Figure 1B). A scavenger receptor cysteine-rich (SRCR) domain, which may be involved in ligand binding (19), is found in the putative extra-cellular region adjacent to the transmembrane domain. A similar SRCR domain is also found in other TTSPs, such as enteropeptidase (7), hepsin (8), corin (9) and TMPRSS2 (16), TMPRSS3 (20), TMPRSS4 (21) and spinesin/TMPRSS5 (22). These results indicate that MSPL and TMPRSS13 are mosaic proteins with a central trypsin-type serine protease domain and various potential regulatory modules.

4. EXPRESSION PROFILES

Comparison of the mRNA expression levels between MSPL and TMPRSS13 in various human tissues is shown in Figure 2. The expressions of MSPL and TMPRSS13 were assessed by reverse transcription-polymerase chain reaction (PCR) with specific primer pairs for each C-terminal of the serine protease domains. Both genes are predominantly expressed in human lung, placenta, pancreas and prostate, but are not detectable in skeletal muscle. High expression of TMPRSS13 is observed in thymus, spleen and peripheral blood lymphocytes, especially $CD8^+$ cells and $CD19^+$ cells. In contrast to the detectable expression of TMPRSS13, MSPL expression in brain, colon and peripheral blood cells is barely detected.

For the expression of recombinant protein, the ORF of the MSPL gene was amplified by PCR, subcloned into p3xFLAG-CMV14 at the *Eco*RI and *Xba*I sites, and then transfected into HEK 293T cells (11). On Western blotting analysis, the expected protein band for MSPL is observed in the cell lysate. Although the recombinant MSPL is labeled with [^{3}H]diisopropylfluorophosphate, which reacts with the active sites of pro- and mature forms of serine proteases, the recombinant MSPL does not show any enzyme activity, suggesting that MSPL is synthesized in an inactive precursor form and required to be converted into the mature form by some protease. Then, the sequence encoding the extracellular regions of MSPL and TMPRSS13 (*i.e.* SRCR domain and serine protease domain) were amplified by PCR and subcloned into p3xFLAG-CMV9 at site immediately after the enteropeptidase recognition sequence (Asp-Asp-Asp-Asp-Lys). The resulting constructs were then transfected into HEK 293T cells and the recombinant proteins were induced in the serum free culture medium (SFCM). Under the conditions employed, purified recombinant soluble human MSPL (hMSPL) and human TMPRSS13 (hTMPRSS13) from SFCM have enzyme activities.

5. ENZYMATIC CHARACTERIZATION

The substrate and inhibitor specificities of the recombinant hMSPL are shown in Tables 1 and 2. Among the substrates of trypsin-type proteases tested, MSPL hydrolyzes preferentially dibasic substrates, such as Boc-Gln-Arg-Arg-MCA, Boc-Leu-Arg-Arg-MCA, and Boc-Leu-Lys-Arg-MCA. Unlike trypsin and plasmin, however, Bz-Arg-MCA and Boc-Val-Leu-Lys-MCA are not hydrolyzed by MSPL. In addition, no significant activity is observed on the substrates for elastase and chymotrypsin-type proteases. It is noteworthy that substrates with Arg at position P1 and Arg or Lys in position P2 are effectively hydrolyzed, whereas the examined substrates with Lys in

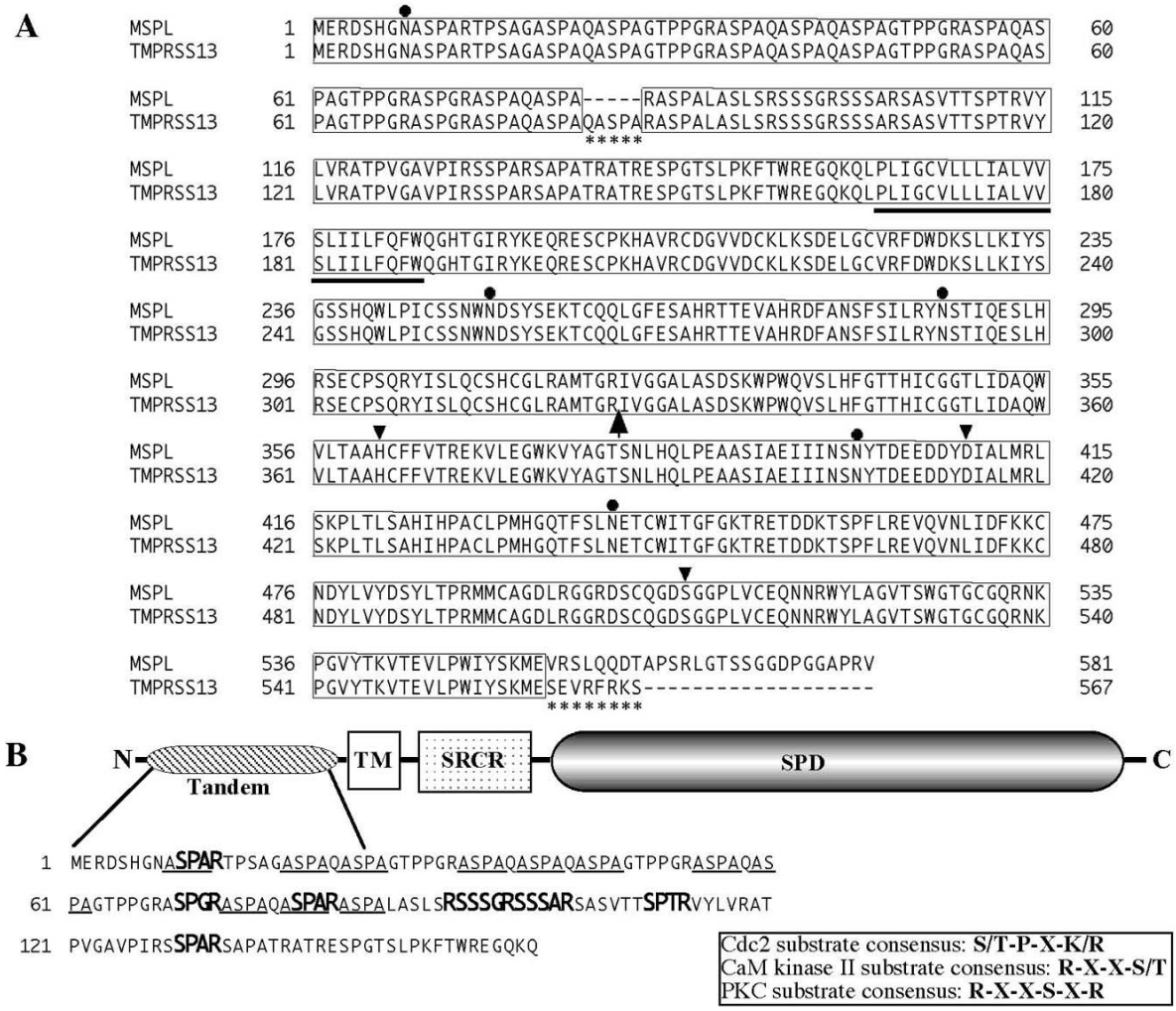

Figure 1. Alignment of the amino acid sequences of MSPL and TMPRSS13 and schematic representation of the domain structures of MSPL. (A) Alignment of the amino acid sequences of human MSPL and TMPRSS13. Numbering starts with Met^{1} of the first amino acid. The putative transmembrane domain is bold-underlined. The arrow points to the putative cleavage site for an active serine protease. The catalytic triad (His, Asp and Ser) are indicated by arrowheads. Potential *N*-glycosylation sites are indicated by closed circles. TMPRSS13 differs from MSPL by the 5 amino acids insertion in the tandem repeat sequences (from Gln^{83} to Ala^{87}) and the 8 amino acids extension at the C-terminus (from Ser^{560} to Ser^{567}) (asterisks). (B) Schematic representation of predicted domain structure of human MSPL. TM, transmembrane domain; SRCR, scavenger receptor cysteine-rich domain; SPD, serine protease domain. The putative kinase phosphorylation sites found in the N-terminal side tandem repeat sequence. The sequence motifs matching the consensus sequences, such as S/T-P-X-K/R of Cdc2, R-X-X-S/T of CaM kinase II and R-X-X-S-X-R of protein kinase C, are in boldface. Homologous sequences, ASPA, are underlined.

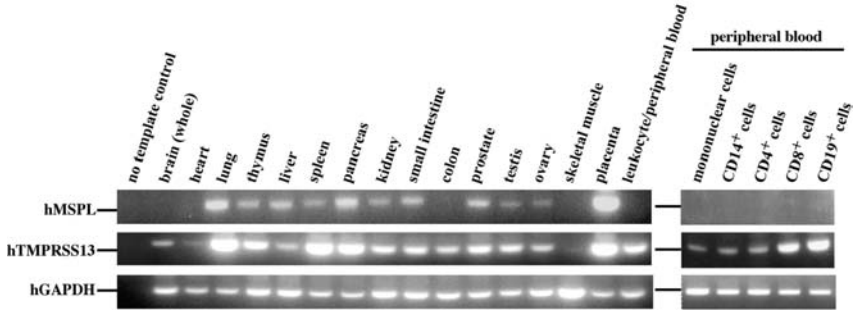

Figure 2. Gene expression profiles of MSPL and TMPRSS13 in various human tissues and cells. RT-PCR analysis was carried out by using MSPL and TMPRSS13 specific primers in various tissues and cells as indicated. As a template, Human Multiple Tissue cDNA Panels were used. The expression level of GAPDH mRNA in each sample was used as an internal control.

Table 1. Substrate specificities of recombinant soluble hMSPL

Substrate (100 μM)	Relative activity (%) hMSPL
Boc-Gln-Ala-Arg-MCA	100.0 ± 1.8
Boc-Gln-Gly-Arg-MCA	20.0 ± 0.2
Boc-Leu-Arg-Arg-MCA	350.0 ± 2.3
Boc-Gln-Arg-Arg-MCA	707.7 ± 5.9
Boc-Glu-Lys-Lys-MCA	57.8 ± 0.8
Boc-Leu-Lys-Arg-MCA	192.3 ± 1.9
Boc-Val-Leu-Lys-MCA	0.0 ± 0.1
Boc-Phe-Ser-Arg-MCA	33.3 ± 0.2
Boc-Leu-Thr-Arg-MCA	18.9 ± 0.1
Boc-Leu-Ser-Thr-Arg-MCA	46.7 ± 0.3
Bz-Arg-MCA	0.0 ± 0.1
Boc-Glu(OBzl)-Ala-Arg-MCA	66.7 ± 1.1
Boc-Glu(OBzl)-Gly-Arg-MCA	11.1 ± 0.1
Pro-Phe-Arg-MCA	12.2 ± 0.1
Suc-Ala-Pro-Ala-MCA	0.0 ± 0.1
Suc-Ala-Ala-Pro-Phe-MCA	0.0 ± 0.1
Suc-Leu-Leu-Val-Tyr-MCA	15.4 ± 0.1

The amidolytic activities of the purified hMSPL were analyzed toward various synthetic peptides in 0.1 M Tris-HCl, pH 8.0 at 37°C. The amount of 7-amino-4-methylcoumarin liberated from the substrate was determined fluorimetrically with excitation and emission wavelengths of 370 and 460 nm, respectively (23). One unit of enzyme activity was defined as the amount that degraded 1 μmol of substrate per min. Data show mean values ± S.D. of three separate experiments and are expressed as a percentage of activity toward Boc-Gln-Ala-Arg-MCA.

Table 2. Inhibitor specificities of recombinant soluble hMSPL

Inhibitor (final concentration, in mM)	Residual activity (%) hMSPL
None	100.0 ± 1.2
PMSF (5.0)	77.9 ± 0.8
DFP (1.0)	104.2 ± 1.2
Aprotinin (0.01)	0.0 ± 0.1
Leupeptin (0.01)	91.4 ± 1.2
Benzamidine (1.0)	9.5 ± 0.1
Soybean trypsin inhibitor (0.01)	13.7 ± 0.1
Bowman-Birk trypsin inhibitor (0.01)	2.7 ± 0.1
Elastatinal (0.01)	92.1 ± 0.3
Chymostatin (0.01)	103.9 ± 1.2
E-64c (0.01)	96.1 ± 0.4
Pepstatin A (0.01)	100.0 ± 0.8
Phosphoramidon (0.01)	86.8 ± 0.6
EDTA (1.0)	92.9 ± 0.8
UTI (0.01)	32.2 ± 0.3
α_1-Antitrypsin (0.01)	109.0 ± 1.1
SLPI (0.01)	27.5 ± 0.2

Purified hMSPL was preincubated for 5 min with various inhibitors indicated in 0.1 M Tris-HCl, pH 8.0 at 37°C, and the residual enzyme activities were measured toward Boc-Gln-Arg-Arg-MCA, one of the efficient substrates tested. Data show mean values ± S.D. of three separate experiments and are expressed as a percentage of activity toward Boc-Gln-Arg -Arg-MCA in the absence of inhibitor.

the P1 position are less hydrolyzed. Recombinant TMPRSS13 shows similar substrate specificity to that of MSPL. The active site mutants, MSPLS506A and TMPRSS13S511A, do not hydrolyze any substrates at all.

The effects of various protease inhibitors on the activity of MSPL are shown in Table 2. Aprotinin, benzamidine and Bowman-Birk trypsin inhibitor significantly inhibit the enzymatic activity. However, unlike trypsin, α_1-antitrypsin does not inhibit the enzyme activity. No effect of any other types of protease inhibitors, such as cysteine-, aspartic- and metalloproteinase inhibitors, is observed. Recombinant TMPRSS13 shows similar inhibitor specificity to that of MSPL.

6. SUMMARY AND PERSPECTIVE

We report a cDNA isolated from a human lung cDNA library that encodes a novel type II transmembrane mosaic serine protease, named MSPL, and its transcript variant that encodes a shorter version TMPRSS13. The primary amino acid sequences of MSPL and TMPRSS13 encode a typical TTSP structure with a transmembrane domain near the N-terminus and a trypsin-like serine protease in the extracellular domain at the C-terminal side. Compared with other TTSPs for protein and enzymatic characteristics, MSPL and TMPRSS13 have following unique characters. One is the longest cytoplasmic domain of MSPL/TMPRSS13 with tandem repeat phosphorylation motifs of various protein kinases. Studies on the roles of each phosphorylation motif in signaling events are now under investigation. The other is the substrate specificity of MSPL/TMPRSS13. Active MSPL/TMPRSS13 preferentially recognizes and hydrolyzes at the carboxy-terminal arginine (Arg) residue in the paired basic residues of the substrates, like furin and the proprotein convertases (24-26). No other TTSPs having similar substrate specificities have been reported. These results suggest that MSPL and TMPRSS13 on the membrane play roles in the proteolytic activation of prohormones and precursors of growth factors, and also play roles in the pathogenicity of many viruses and bacteria.

7. ACKNOWLEDGEMENTS

We thank Etsuhisa Takahashi for the excellent technical support. This work was supported by a Grant-in-Aid from the Ministry of Education, Science and Culture of Japan (No. 17790213) and the Program for Fundamental Studies in Health Sciences of NIBIO.

8. REFERENCES

1. Dahlbäck, B: Blood coagulation. *Lancet* 355, 1627-1632 (2000)
2. Lijnen, H. R: Plasmin and matrix metalloproteinases in vascular remodeling. *Thromb Haemost* 86, 324-333 (2001)
3. LeMosy, E. K., Y-Q. Tan & C. Hashimoto: Activation of a protease cascade involved in patterning the Drosophila embryo. *Proc Natl Acad Sci USA* 98, 5055-5060 (2001)
4. Mignatti, P. & D. B. Rifkin: Biology and biochemistry of proteinases in tumor invasion. *Physiol Rev* 74, 161–195 (1993)
5. Wong, G. W., Y. Tang, E. Feyfant, A. Sali, L. Li, Y. Li, C. Huang, D. S. Friend, S. A. Krilis & R. L. Stevens: Identification of a new member of the tryptase family of mouse and human mast cell proteases which possesses a

novel COOH-terminal hydrophobic extension. *J Biol Chem 274*, 30784-30793 (1999)
6. Chen, L. M., M. L. Skinner, S. W. Kauffman, J. Chao, L. Chao, C. D. Thaler & K. X. Chai: Prostasin is a glycosylphosphatidylinositol-anchored active serine protease. *J Biol Chem* 276, 21434-21442 (2001)
7. Kitamoto, Y., R. A. Veile, H. Donis-Keller & J. E. Sadler: cDNA sequence and chromosomal localization of human enterokinase, the proteolytic activator of trypsinogen. *Biochemistry* 34, 4562-4568 (1995)
8. Leytus, S. P., K. R. Loeb, F. S. Hagen, K. Kurachi & E. W. Davie: A novel trypsin-like serine protease (Hepsin) with a putative transmembrane domain expressed by human liver and hepatoma cells. *Biochemistry* 27, 1067-1074 (1988)
9. Yan, W., N. Sheng, M. Seto, J. Morser & Q. Wu: Corin, a mosaic transmembrane serine protease encoded by a novel cDNA from human heart. *J Biol Chem* 274, 14926-14935 (1999)
10. Lin, C. Y., J. Anders, M. Johbson, Q. A. Sang & R. B. Dickson: Molecular cloning of cDNA for matriptase, a matrix-degrading serine protease with trypsin-like activity. *J Biol Chem* 274, 18231-18236 (1999)
11. Kim, D. R., S. Sharmin, M. Inoue & H. Kido: Cloning and expression of novel mosaic serine proteases with and without a transmembrane domain from human lung. *Biochem Biophys Acta* 1518, 204-209 (2001)
12. Hartley, B. S., J. R. Brown, D. L. Kauffman & L. B. Smillie: Evolutionary similarities between pancreatic proteolytic enzymes. *Nature* 207, 1157–1159 (1965)
13. Halfon, S. & C. S. Craik: Family S1 of trypsin (clan SA). In: Handbook of Proteolytic Enzymes. Eds: Barrett, A. J., N. D. Rawlings, & J. F. Woessner, Academic Press, San Diego, CA, 12–21 (1998)
14. Beaubien, G., I. Rosinki-Chupin, M. G. Mattei, M. Mbikay, M. Chrestien & N. G. Seidah: Gene structure and chromosomal localization of plasma kallikrein. *Biochemistry* 30, 1628–1635 (1991)
15. Tsuji, A., A. Torres-Rosado, T. Arai, M. M. LeBeau, R. S. Lemons, S. H. Chou & K. Kurachi: Hepsin, a cell membrane-associated protease. Characterization, tissue distribution, and gene localization. *J Biol Chem* 266, 16948–16953 (1991)
16. Paoloni-Giacobino, A., H. Chen, M. C. Peitsch, C. Rossier & S. E. Antonarakis: Cloning of the TMPRSS2 gene, which encodes a novel serine protease with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3. *Genomics* 44, 309-320 (1997)
17. Kennelly, P. J. & E. G. Krebs: Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. *J Biol Chem* 266, 15555–15558 (1991)
18. Wang, Y. & P. J. Roach: Purification and assay of mammalian protein (serine/threonine) kinases. In: Protein Phosphorylation: A Practical Approach. Eds: Hardie, D. G., Oxford University Press, Oxford, NY, 121–144 (1993)
19. Freeman, M., J. Ashkenas, D. J. Rees, D. M. Kingsley, N. G. Copeland, N. A. Jenkins & M. Krieger: An ancient, highly conserved family of cysteine-rich protein domains revealed by cloning type I and type II murine macrophage scavenger receptors. *Proc Natl Acad Sci USA* 87, 8810-8814 (1990)
20. Scott, H. S., J. Kudoh, M. Wattenhofer, K. Shibuya, A. Berry, R. Chrast, M. Guipponi, J. Wang, K. Kawasaki, S. Asakawa, S. Minoshima, F. Younus, S. Q. Mehdi, U. Radhakrishna, M. P. Papasavvas, C. Gehrig, C. Rossier, M. Korostishevsky, A. Gal, N. Shimizu, B. Bonne-Tamir & S. E. Antonarakis: Insertion of b-satellite repeats identifies a transmembrane protease causing both congenital and childhood onset autosomal recessive deafness. *Nat Genet* 27, 59-63 (2001).
21. Wallrapp, C., S. Hähnel, F. Müller-Pillasch, B. Burghardt, T. Iwamura, M. Ruthenbürger, M. M. Lerch, G. Adler & T. M. Gress: A novel transmembrane serine protease (TMPRSS3) overexpressed in pancreatic cancer. *Cancer Res* 60, 2602-2606 (2000)
22. Yamaguchgi, N., A. Okui, T. Yamada, H. Nakazato & S. Mitsui: Spinesin/TMPRSS5, a novel transmembrane serine protease, cloned from human spinal cord. *J Biol Chem* 277, 6806-6812 (2002)
23. Okumura, Y., M. Hayama, E. Takahashi, M. Fujiuchi, A. Shimabukuro, M, Yano & H. Kido. Serase-1B, a new splice variant of polyserase-1/TMPRSS9, activates urokinase-type plasminogen actiovator and the proteolytic activation is negatively regulated by glycosaminoglycans. *Biochem J* 400, 551-561 (2006)
24. Parish, D. C., R. Tuteja, M. Altstein, H. Gainer & Y. P. Loh: Purification and characterization of a paired basic residue-specific prohormone-converting enzyme from bovine pituitary neural lobe secretory vesicles. *J Biol Chem* 261, 14392–14397 (1986)
25. Takahashi, K., Y. Tamanoue, M. Yanagida, Y. Sakurai, T. Takahashi & K. Sutoh: Specific cleavages of arginyl peptide bonds at basic amino acid pairs by a serine proteinase from the microsomal membranes of rat liver. *Biochem Biophys Res Commun* 175, 1152-1158 (1991)
26. Thomas, G: Furin at the cutting edge: from protein traffic to embryogenesis and disease. *Nature review Mol Cell Biol* 3, 753-766 (2002)

Abbreviations: Boc, t-butoxycarbonyl; B_z, benzoyl; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; HEK, human embryonic kidney; MCA, 4-methylcoumaryl-7-amide; OB_zl, benzyloxy; SLPI, secretory leukoprotease inhibitor; UTI, urinary trypsin inhibitor

Key Words: type II transmembrane serine protease, TTSP, Mosaic Protein, Tandem Repeat Motifs, Paired Basic Residue Cleaving Enzyme, Review

Send correspondence to: Dr Hiroshi Kido, Division of Enzyme Chemistry, Institute for Enzyme Research, The University of Tokushima, Kuramoto-cho 3-18-15, Tokushima 770-8503, Japan, Tel: 81-88-633-7423, Fax: 81-88-633-7425, E-mail: kido@ier.tokushima-u.ac.jp

http://www.bioscience.org/current/vol13.htm