Genomics **63,** 88–96 (2000)
doi:10.1006/geno.1999.6072, available online at http://www.idealibrary.com on IDEAL®

Genomic Organization, Mapping, Tissue Expression, and Hormonal Regulation of Trypsin-like Serine Protease (TLSP PRSS20), a New Member of the Human Kallikrein Gene Family

George M. Yousef, Andreas Scorilas, and Eleftherios P. Diamandis[1]

Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada; and Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5G 1L5, Canada

Received September 2, 1999; accepted November 16, 1999

The cDNA for the trypsin-like serine protease gene (TLSP, HGMW-approved symbol PRSS20) has been recently identified. TLSP is expressed in brain and skin tissues but little else is known about this new serine protease gene. In this paper, we describe the complete genomic organization and precise mapping of the TLSP gene. This gene spans 5.3 kb of genomic sequence on chromosome 19q13.3–q13.4. The gene consists of six exons, the first of which is untranslated. All splice junctions follow the GT/AG rule, and the intron phases are identical to those of other kallikrein-like genes, including zyme (PRSS9), NES1 (PRSSL1), and neuropsin (PRSS19). Fine-mapping of the area indicates that TLSP lies downstream from the PSA, zyme, neuropsin, and NES1 genes. Significant sequence homologies were found between TLSP and other human kallikreins. Furthermore, there is conservation of the catalytic triad (histidine, aspartic acid, serine) and of the number of coding exons (five; the same in all members of the kallikrein gene family). We thus suggest that TLSP is a new member of the human kallikrein gene family. TLSP is expressed in many tissues including cerebellum, prostate, salivary glands, stomach, lung, thymus, small intestine, spleen, liver, and uterus. TLSP expression appears to be regulated by steroid hormones in the breast carcinoma cell line BT-474.

INTRODUCTION

Serine proteases play crucial roles in many physiological processes including digestion, coagulation and fibrinolysis, apoptosis, tissue remodeling, and cell migration. A number of these proteases appear to be important for central nervous system plasticity (Yoshida *et al.*, 1990; Momota *et al.*, 1998). Kallikreins and kallikrein-like proteins are members of the serine protease enzyme family and exhibit a high degree of substrate specificity (Evans *et al.*, 1988). The biological role of these kallikreins is the selective cleavage of high-molecular-weight substrates and release of peptides with potent biological activity (Clements, 1989). Three human kallikrein genes have been described, i.e., prostate-specific antigen (PSA[2] or KLK3) (Riegman *et al.*, 1989), human glandular kallikrein 2 (KLK2) (Schedlich *et al.*, 1987), and tissue (pancreatic–renal) kallikrein (KLK1) (Angermann *et al.*, 1989). The PSA gene spans 5.8 kb of known sequence (Riegman *et al.*, 1989); the KLK2 gene is 5.2 kb, and its complete structure has also been elucidated (Schedlich *et al.*, 1987). The KLK1 gene is approximately 4.5 kb long and the exon sequences and the exon–intron junctions of this gene have been determined (Angermann *et al.*, 1989). All three human kallikrein genes map to chromosome 19q13.2–q13.4 in close proximity to one another. The distance between PSA and KLK2 is only 12 kb (Riegman *et al.*, 1992).

Some newly discovered genes also appear to be members of the kallikrein family, including the normal epithelial cell-specific 1 gene (NES1, PRSSL1) (Liu *et al.*, 1996), zyme/protease M/neurosin (PRSS9) (Little *et al.*, 1997; Anisowicz *et al.*, 1996; Yamashiro *et al.*, 1997; Yousef *et al.*, in press), prostase/KLK-L1/KLK4 (Nelson *et al.*, 1999; Yousef *et al.*, 1999b; Stephenson *et al.*, 1999), stratum corneum chymotryptic enzyme (HSCCE, PRSS6) (Lundstrom and Egelrud, 1991), KLK-L2 (Yousef and Diamandis, in press), and other kallikrein-like genes (Yousef *et al.*, 1999a).

[1] To whom correspondence should be addressed at Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. Telephone: (416) 586-8443. Fax: (416) 586-8628. E-mail: ediamandis@mtsinai.on.ca.

[2] Abbreviations used: TLSP, trypsin-like serine protease; KLK, kallikrein; KLK-L, kallikrein-like; NES1, normal epithelial cell-specific 1 gene; PCR, polymerase chain reaction; PSA, prostate-specific antigen; dNTPs, deoxynucleoside triphosphates; LLNL, Lawrence Livermore National Laboratory; RT, reverse transcription; EMSP, enamel matrix serine proteinase; BCM, Baylor College of Medicine.

0888-7543/00 $35.00

A cDNA for trypsin-like serine protease (TLSP)[3] was cloned by Yoshida *et al.* (1998) from human hippocampus cDNA using polymerase chain reaction with primers targeting conserved regions of serine protease. The putative amino acid sequence of TLSP is similar in many respects to other known serine proteases. TLSP has high sequence homology with mouse neuropsin, human kallikreins, and the mouse nerve growth factor γ subunit. Overexpressed TLSP protein in neuroblastoma cells was detected in the culture medium, suggesting that it is secreted (Yoshida *et al.*, 1998). The genomic organization and chromosomal localization of this gene were not established.

Functional and positional cloning, as well as the positional candidate gene approach, is now used widely for gene identification (Birren *et al.*, 1998). In positional cloning, the identification of the gene is made without any prior knowledge of the function of its protein product (Collins, 1992; Ruddle, 1984). Positional candidate cloning is the identification of a gene based on its map position and on the availability of candidate genes localized in the same region (Ballabio, 1993). This method is now becoming very popular because of the increasing number of cloned and mapped genes and available expressed sequence tags (ESTs).

In this study we describe the complete genomic structure of the TLSP gene and its localization in relation to four other homologous genes also mapping to the same chromosomal region. Tissue expression studies by RT-PCR indicate that this gene is widely expressed in many human tissues. We also preliminarily demonstrate that this gene is up-regulated by estrogens and glucocorticoids in the breast carcinoma cell line BT-474.

MATERIALS AND METHODS

DNA sequences on chromosome 19q13.3–q13.4. We have obtained approximately 300 kb of genomic sequences from the Web site of the Lawrence Livermore National Laboratory (LLNL) (http://www.bio.llnl.gov/genome/genome.html) encompassing a region on chromosome 19q13.3–q13.4, where the known kallikrein genes are localized. With the aid of an *Eco*RI restriction map of that area (also available from the LLNL Web site) and the computer program WebCutter (http://www.firstmarket.com/cutter/cut2.html), we performed restriction analysis of the available sequence and were able to construct an almost contiguous sequence map of the region. By using the published sequences of PSA, NES1, zyme, and neuropsin and the software "BLAST 2 sequences" (Altschul *et al.*, 1997), we were able, using alignment strategies, to identify the relative positions of these genes on the contiguous map.

Homology searching and structure analysis. Multiple alignment was performed using the Clustal X software package available at the Web site ftp://ftp.ebi.ac.uk/pub/software/dos/clustalw/clustalx/clustalx1.64b.msw.exe and the multiple alignment program available from the Baylor College of Medicine (BCM) search launcher (kiwi.imgen.bcm.tmc.edu:8808/search-launcher/launcher/html).

[3] The HGMW-approved symbol for the gene described in this paper is PRSS20.

TABLE 1

Primers Used for Reverse Transcription-Polymerase Chain Reaction Analysis

Gene	Primer name	Sequence[a]	Product size (bp)
TLSP	TLSP1	GGGAGAGACCAGGATCATCA	530
	TLSP2	TTCCTGCACGCTGGCACACA	
pS2	PS2S	GGTGATCTGCGCCCTGGTCCT	328
	PS2AS	AGGTGTCCGGTGGAGGTGGCA	
PSA	PSAS	TGCGCAAGTTCACCCTCA	754
	PSAAS	CCCTCTCCTTACTTCATCC	
Actin	ACTINS	ACAATGAGCTGCGTGTGGCT	372
	ACTINAS	TCTCCTTAATGTCACGCACGA	

[a] All nucleotide sequences are given in the 5′ → 3′ orientation.

Phylogenetic studies were performed using the Phylip software package available at http://evolution.genetics.washington.edu/phylip/getme.html. Distance matrix analysis was performed using the "Neighbor-Joining/UPGMA" program, and parsimony analysis was performed using the "Protpars" program. A hydrophobicity study was performed using the BCM search launcher program (http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html). The signal peptide was predicted using the "Signal P" www server (http://www.cbs.dtu.dk/services/signal). Protein structure analysis was performed by the "SAPS" (structural analysis of protein sequence) program (http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html).

Tissue expression of TLSP. Total RNA isolated from 26 different human tissues was purchased from Clontech (Palo Alto, CA). We prepared cDNA as described below for the tissue culture experiments and used it for PCR with the primers described in Table 1. Tissue cDNAs were amplified at various dilutions.

Breast cancer cell line and stimulation experiments. The breast cancer cell line BT-474 was purchased from the American Type Culture Collection (Rockville, MD). BT-474 cells were cultured in RPMI medium (Gibco BRL, Gaithersburg, MD) supplemented with glutamine (200 mmol/L), bovine insulin (10 mg/L), fetal bovine serum (10%), antibiotics, and antimycotics, in plastic flasks, to near confluency. The cells were then aliquoted into 24-well tissue culture plates and cultured to 50% confluency. Twenty-four hours before the experiments, the culture medium was changed into phenol red-free medium containing 10% charcoal-stripped fetal bovine serum. For stimulation experiments, various steroid hormones dissolved in absolute ethanol were added to the culture medium, at a final concentration of 10^{-8} M. Cells stimulated with ethanol were included as controls. The cells were cultured for 24 h and then harvested for mRNA extraction.

Reverse transcriptase-polymerase chain reaction. Total RNA was extracted from the breast cancer cells using Trizol reagent (Gibco BRL) following the manufacturer's instructions. RNA concentration was determined spectrophotometrically. Two micrograms of total RNA was reverse-transcribed into first-strand cDNA using the Superscript preamplification system (Gibco BRL). The final volume was 20 μl. Based on the information obtained from the predicted genomic structure of the new gene, two gene-specific primers were designed (see Table 1), and PCR was carried out in a reaction mixture containing 1 μl of cDNA, 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM dNTPs (deoxynucleoside triphosphates), 150 ng of primers, and 2.5 units of AmpliTaq Gold DNA polymerase (Roche Molecular Systems, Branchburg, NJ) on a Perkin–Elmer 9600 thermal cycler. The cycling conditions were 94°C for 9 min to activate the *Taq* Gold DNA polymerase, followed by 43 cycles of 94°C for 30 s, 69°C for 1 min, and a final extension at 69°C for 10 min. Equal

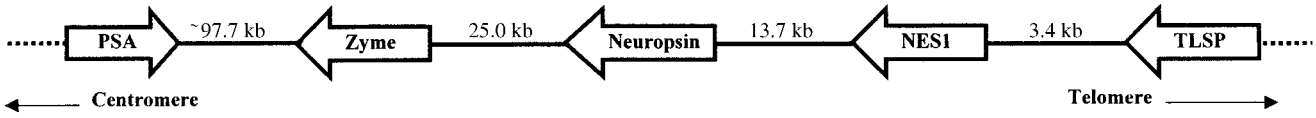

FIG. 1. Approximately 150 kb of contiguous genomic sequence around chromosome 19q13.3–q13.4. Genes are represented by arrows denoting the direction of the coding sequence. Distances between genes are shown in kilo-basepairs (kb). Figure is not drawn to scale. Additional genes between the genes shown are published elsewhere (Yousef *et al.*, 1999a).

amounts of PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. The genes PSA (up-regulated by androgens and progestins), pS2 (up-regulated by estrogen), and actin (not regulated by steroid hormones) were used as controls in the hormonal stimulation experiments. All primers for RT-PCR spanned at least two exons to avoid interference by amplification products of genomic DNA.

Cloning and sequencing of the PCR products. To verify the identity of the PCR products, they were cloned into the pCR 2.1-TOPO vector (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. The inserts were sequenced from both directions using vector-specific primers, with an automated DNA sequencer.

RESULTS

Mapping and Chromosomal Localization of the TLSP Gene

Screening of 300 kb of DNA sequences around chromosome 19q13.3–q13.4 enabled us to localize known genes precisely and to determine the direction of transcription, as shown in Fig. 1. The PSA gene is the most centromeric, separated by approximately 98 kb from the zyme gene, which is transcribed in the opposite

```
gttgggagaa gcAGGAATCT GCGCTCGGGT TCCGCAGATG CAGAGGTTGA GGTGGCTGCG
GGACTGGAAG TCATCGGGCA GAGGTCTCAC AGCAGCCAgt aagtgaacag …intron-1…
gtccctcctt gttcccagAG GAACCTGGGG CCCGCTCCTC CCCCCTCCAG GCCATGAGGA
                                                        M  R  I
TTCTGCAGTT AATCCTGCTT GCTCTGGCAA CAGgtacgca ggggatgggg …intron-2…
  L  Q  L   I  L  L   A  L  A  T   G
ccccctccag GGCTTGTAGG GGGAGAGACC AGGATCATCA AGGGGTTCGA GTGCAAGCCT
           L  V  G   G  E  T   R  I  I  K   G  F  E   C  K  P
CACTCCCAGC CCTGGCAGGC AGCCCTGTTC GAGAAGACGC GGCTACTCTG TGGGGCGACG
H  S  Q  P   W  Q  A   A  L  F   E  K  T  R   L  L  C   G  A  T
CTCATCGCCC CCAGATGGCT CCTGACAGCA GCCCACTGCC TCAAGCCgtg ggtgcggggg
L  I  A  P   R  W  L   L  T  A   A  H  C  L   K  P
…intron-3… ctctcccacc tcagCCGCTA CATAGTTCAC CTGGGGCAGC ACAACCTCCA
                       R  Y   I  V  H   L  G  Q  H   N  L  Q
GAAGGAGGAG GGCTGTGAGC AGACCCGGAC AGCCACTGAG TCCTTCCCCC ACCCCGGCTT
K  E  E   G  C  E  Q   T  R  T   A  T  E   S  F  P  H   P  G  F
CAACAACAGC CTCCCCAACA AAGACCACCG CAATGACATC ATGCTGGTGA AGATGGCATC
N  N  S   L  P  N  K   D  H  R   N  D  I   M  L  V  K   M  A  S
GCCAGTCTCC ATCACCTGGG CTGTGCGACC CCTCACCCTC TCCTCACGCT GTGTCACTGC
P  V  S   I  T  W  A   V  R  P   L  T  L   S  S  R  C   V  T  A
TGGCACCAGC TGCCTCATTT CCGGCTGGGG CAGCACGTCC AGCCCCCAGT gtaggagcac
G  T  S   C  L  I  S   G  W  G   S  T  S   S  P  Q  L
…intron-4… cacgctgttt ccacagTACG CCTGCCTCAC ACCTTGCGAT GCGCCAACAT
                       R   L  P  H   T  L  R  C   A  N  I
CACCATCATT GAGCACCAGA AGTGTGAGAA CGCCTACCCC GGCAACATCA CAGACACCAT
T  I  I   E  H  Q  K   C  E  N   A  Y  P   G  N  I  T   D  T  M
GGTGTGTGCC AGCGTGCAGG AAGGGGGCAA GGACTCCTGC CAGgtcagtg tggtctccaa
V  C  A   S  V  Q  E   G  G  K   D  S  C   Q
…intron-5… tctctccccg tgcccagGGT GACTCCGGGG GCCCTCTGGT CTGTAACCAG
                       G   D  S  G  G   P  L  V   C  N  Q
TCTCTTCAAG GCATTATCTC CTGGGGCCAG GATCCGTGTG CGATCACCCG AAAGCCTGGT
S  L  Q  G   I  I  S   W  G  Q   D  P  C  A   I  T  R   K  P  G
GTCTACACGA AAGTCTGCAA ATATGTGGAC TGGATCCAGG AGACGATGAA GAACAATTAG
V  Y  T  K   V  C  K   Y  V  D   W  I  Q  E   T  M  K   N  N
ACTGGACCCA CCCACCACAG CCCATCACCC TCCATTTCCA CTTGGTGTTT GGTTCCTGTT
CACTCTGTTA ATAAGAAACC CTAAGCCAAG ACCCTCTACG AACATTCTTT GGGCCTCCTG
GACTACAGGA GATGCTGTCA CTTAATAATC AACCTGGGGT TCGAAATCAG TGAGACCTGG
ATTCAAATTC TGCCTTGAAA TATTGTGACT CTGGGAATGA CAACACCTGG TTTGTTCTCT
GTTGTATCCC CAGCCCCAAA GACAGCTCCT GGCCATATAT CAAGGTTTCA ATAAATATTT
GCTAAATGAG TG
```

FIG. 2. Genomic organization and partial sequence of the TLSP gene. Intronic sequences are not shown except for short areas around the splice junctions. Introns are shown with lowercase letters and exons with uppercase letters. The start and stop codons are encircled, and the exon–intron junctions are boxed. The translated amino acids of the coding region are shown underneath in single-letter code. The catalytic residues are inside triangles. A putative polyadenylation signal is underlined. Another potential start codon in exon 1 is shown in boldface type.

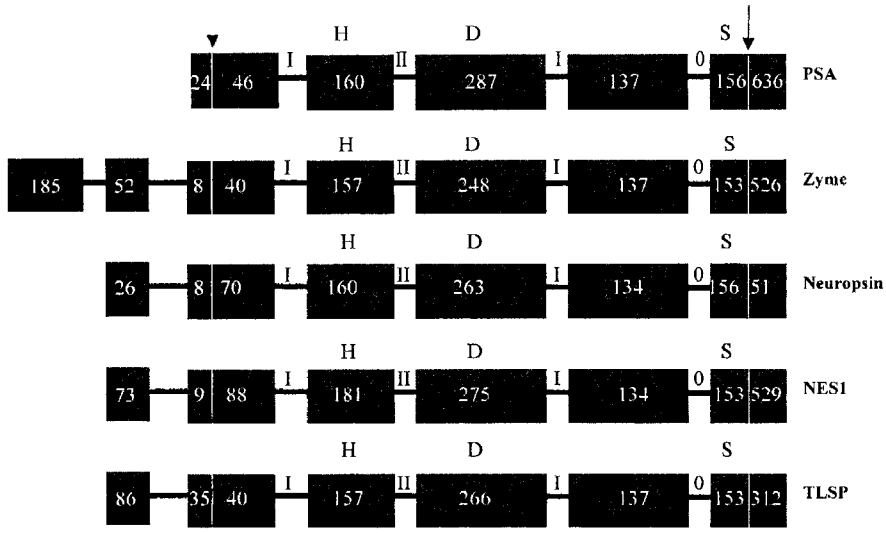

FIG. 3. Schematic diagram showing the comparison of the genomic structure of PSA, zyme, neuropsin, NES1, and TLSP genes. Exons are shown by black boxes, and introns are shown by the connecting lines. The arrowhead shows the start codon, and the arrow shows the stop codon. Letters above boxes indicate relative positions of the catalytic triad: H (histidine), D (aspartic acid), and S (serine). Roman numbers indicate intron phases. The intron phase refers to the location of the intron within the codon; I denotes that the intron occurs after the first nucleotide of the codon, II indicates that the intron occurs after the second nucleotide, and 0 indicates that the intron occurs between codons. Numbers inside boxes indicate exon lengths in basepairs. Figure is not drawn to scale.

direction. The neuropsin gene (Yoshida *et al.*, 1998) is 24,359 bp more telometric and transcribes in the same direction as zyme (telomere to centromere), followed by the NES1 gene. The distance between neuropsin and NES1 is 12,105 bp. The TLSP gene is 2488 bp more telomeric and is transcribed in the same direction. A number of other genes are dispersed within this region, as described elsewhere (Yousef *et al.*, 1999a) but their precise localization has not as yet been established.

Genomic Structure of the TLSP Gene

Aligment of the 300-kb genomic DNA sequence with the published cDNA of the TLSP gene (Yoshida *et al.*, 1998) allowed us to define the exon–intron boundaries precisely. The TLSP gene is made up of six exons and five intervening introns, spanning an area of 5308 bp of genomic sequence on chromosome 19q13.3–q13.4 (Fig. 2). The lengths of the exons are 86, 75, 157, 266, 137, and 465 bp, respectively (Fig. 3). However, the possibility of the presence of an additional 5′ untranslated exon(s) could not be excluded. The intron–exon splice sites (mGT. . ..AGm) and their flanking sequences are closely related to the consensus splicing sites (-mGTAAGT. . .CAGm-) (Iida, 1990). The protein coding region of the TLSP gene is made up of 750 bp of nucleotide sequence encoding a deduced 250-amino-acid polypeptide. There are two potential translation initiation codons (ATG) at positions 26 and 36 of the first and second exons, respectively. We suggest that the second ATG is likely the initiation codon, since (a) the flanking sequence of the second ATG matches more closely with the Kozak sequence (Kozak, 1991) (b) at this initiation codon, the putative signal sequence at the N-terminus and its length are similar to those of other trypsin-like serine proteases (see also below). A polyadenylation signal (AATAAA) is present 289 bp downstream of the termination codon and 17 bp upstream of the poly(A) tail.

Protein Product of the TLSP Gene

The TLSP protein shows a hydrophobic region in the N-terminus (Fig. 4); thus, a presumed signal peptide is present. An 18-amino-acid signal peptide is predicted with a cleavage site at the carboxyl end of Gly[18]. For better characterization of the predicted structural motifs of the TLSP protein, it was aligned with other members of the kallikrein multigene family. Sequence alignment supports, by analogy, the presence of an additional cleavage site at the carboxyl end of Arg[20], which is the exact site predicted for cleavage of the activation peptide of all the other kallikreins aligned in Fig. 5. Interestingly, the starting amino acid sequence of the mature protein [I, I, (K), G] is conserved in the prostase and enamel matrix serine proteinase 1 (EMSP) (Nelson *et al.*, 1999) and KLK-L2 (Yousef and Diamandis, in press) genes. Thus, like other kallikreins, the TLSP protein appears to be synthesized as a pre-proenzyme that contains an N-terminal signal peptide (prezymogen) followed by an activation peptide and the enzymatic domain at the C-terminal end. The phylogenetic tree of the kallikreins grouped TLSP with KLK-L2, EMSP, and prostase/KLK-L1 (Fig. 4).

The presence of aspartate (D) at position 197 suggests that TLSP has trypsin-like enzymatic activity, like most of the other kallikreins (e.g., KLK1, KLK2, neuropsin, zyme, prostase, and EMSP) but different

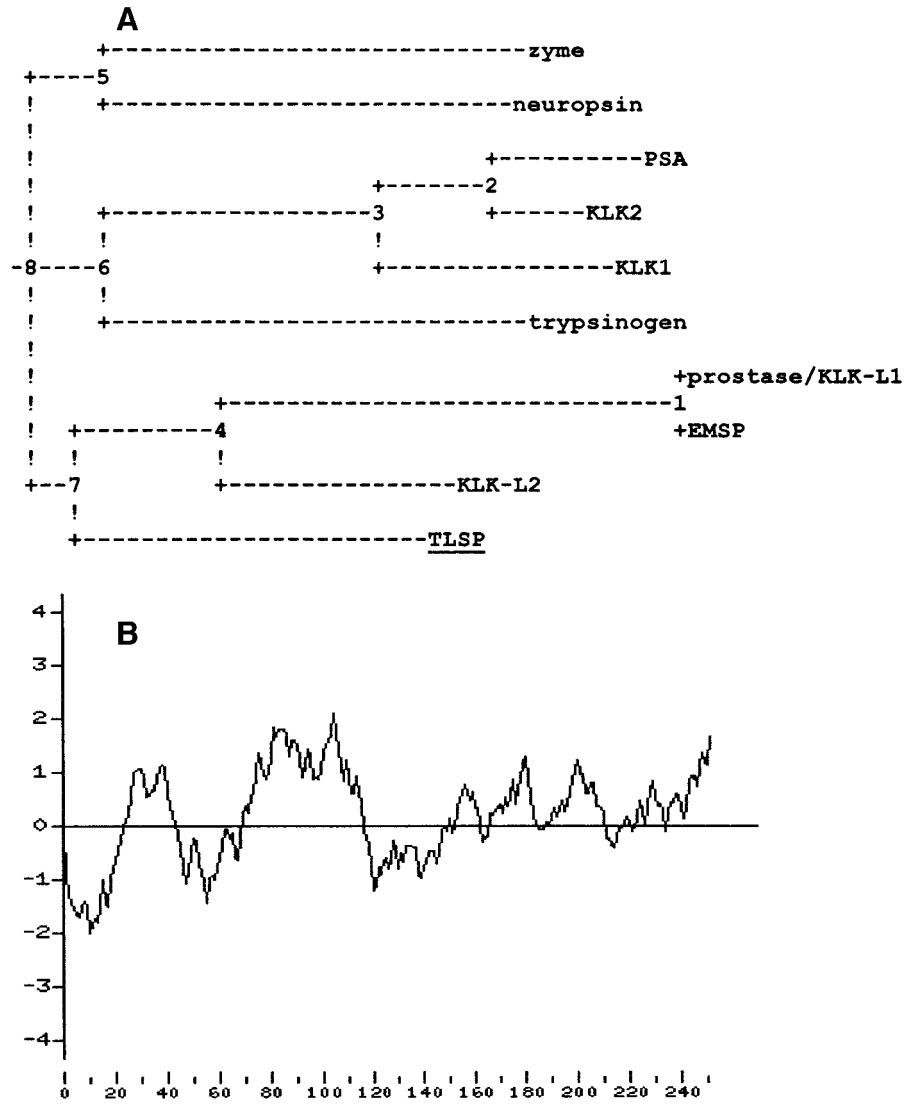

FIG. 4. **(A)** Dendrogram of the predicted phylogenetic tree for some kallikrein genes. Neighbor-joining/UPGMA method was used to align TLSP with other members of the kallikrein gene family. The tree grouped the classical kallikreins (KLK1, KLK2, and PSA) together and aligned the TLSP gene in one group with EMSP, prostase, and KLK-L2. **(B)** Plot of hydrophobicity and hydrophilicity of TLSP protein.

from PSA, which has a serine (S) residue in the corresponding position and has chymotrypsin-like activity (Fig. 5). There is a potential N-glycosylation site (Asn-Asn-Ser) at position 99 similar to several other serine proteases (Fig. 5, dotted line). The boldface italic region just upstream from the catalytic triad amino acid D^{110} (Fig. 5) indicates a 10-amino-acid loop characteristic of the classical kallikreins (PSA, KLK1, and KLK2), which is not found in TLSP or other members of the kallikrein-like gene family.

Comparison of TLSP with the Kallikreins

Amino acid identity between TLSP and other kallikreins was found to be as follows: human neuropsin, 51%; KLK-L2, 50%; hK2, trypsinogen, zyme, 44%; human pancreatic/renal kallikrein, prostase/KLK-L1, EMSP, 42%; and PSA, 41%. A multiple alignment study shows that the typical catalytic triad of serine proteases is conserved in the TLSP gene (His^{62}, Asp^{110}, and Ser^{203}). As is the case with all other kallikreins, a well-conserved peptide motif is found around the amino acid residues of the catalytic triad (i.e., histidine WLLTAA***H***C, aspartate ***D***IMLV, and serine GD***S***GGPL].

Twelve cysteine residues are present in the putative mature TLSP protein; 10 of them are conserved in all the serine proteases that are aligned in Fig. 5 and would be expected to form disulfide bridges. The other two (C^{135} and C^{237}) are not found in PSA, KLK1, KLK2, and trypsinogen; however, they are found in similar positions in prostase, EMSP1, zyme, neuropsin, and KLK-L2 genes and are expected to form an additional disulfide bond. Twenty-nine "invariant" amino acids surrounding the active site of serine proteases have been described (Dayhoff and Orcutt, 1979). Of these, 25 are conserved in TLSP. One of the unconserved amino acids (Leu^{57} instead of Val) is also found in the KLK-L2

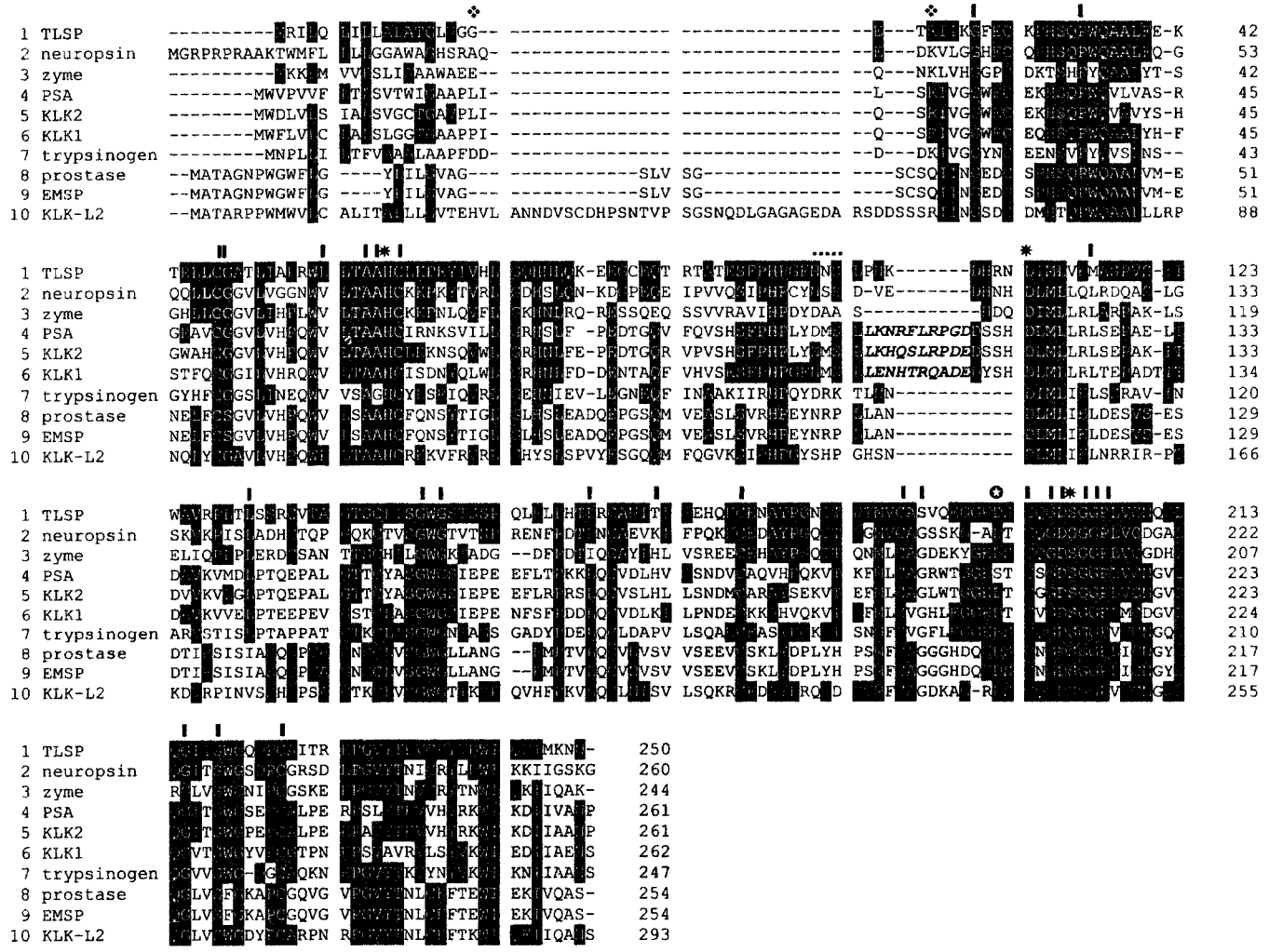

FIG. 5. Alignment of the deduced amino acid sequence of TLSP with members of the kallikrein multigene family. The residues of the catalytic triad are indicated by ∗ and the 29 invariant serine protease residues by | or ∗. The predicted cleavage sites are indicated by ❖, and further explanations are in the text. The dotted line represents the putative glycosylation site. The 10-amino-acid kallikrein loop lies between the glycosylation signal and the catalytic D amino acid and is boldface italics; it is present only in KLK1, KLK2, and PSA genes. Black background highlights amino acids common in TLSP and other kallikrein-like genes. The trypsin-like cleavage pattern is indicated by ✪.

protein. The other three unconserved amino acids are Met^{116}, Thr^{167}, and S^{190}. According to protein evolution studies, each of these changed amino acids represents a conserved evolutionary change to a protein of the same group (Miyata *et al.*, 1979; Simmer *et al.*, 1998).

Tissue Expression of the TLSP Gene

Tissue expression of the TLSP gene was assessed by RT-PCR. Actin was used as a positive control, and PSA was used as another positive control that is expressed in only a few tissues. The PSA gene was found to be highly expressed in the prostate, as expected, and to a lesser extent in mammary, thyroid gland, and salivary gland, as also expected from recent literature reports (Diamandis *et al.*, 1994; Ishikawa *et al.*, 1998). The tissue expression of TLSP is summarized in Table 2 and Fig. 6. This protease is primarily expressed in the prostate, salivary gland, stomach, lung, spleen, uterus, thymus, liver, small intestine, and cerebellum while lower levels of expression are found in many other tissues. To verify the RT-PCR specificity, the PCR products were cloned and sequenced.

Hormonal Regulation of the TLSP Gene

We have used the steroid hormone receptor-positive breast carcinoma cell line BT-474 as a model system to verify whether the TLSP gene is under steroid hormonal regulation. PSA was used as a control known to be up-regulated by androgens and pS2 as an estrogen up-regulated control. As shown in Fig. 7, the TLSP gene appears to be up-regulated by estrogens and glucocorticoids.

DISCUSSION

In this study we found that the TLSP gene is made up of six exons and five intervening introns. The first

TABLE 2

Tissue Expression Levels of TLSP by RT-PCR Analysis

High-level expression	Medium-level expression	Low-level expression	No expression
Salivary gland	Heart	Fetal brain	Trachea
Stomach	Fetal liver	Brain	Colon
Uterus	Mammary gland	Bone marrow	Testis
Lung	Thyroid		Adrenal
Thymus	Skeletal muscle		Placenta
Prostate	Spinal cord		Pancreas
Spleen			Kidney
Cerebellum			
Liver			
Small intestine			

exon is noncoding, similar to NES1, neuropsin, and zyme (Fig. 3). The gene spans 5308 bp on chromosome 19q13.3–q13.4. All splice junctions agree with the consensus sequence GT-AG. The possibility of an additional 5′ untranslated exon(s) could not be excluded. No TATA box or CAAT sequences were found in the 5′ flanking region of the TLSP gene. The same observation was made for some other kallikrein-like genes including neuropsin (Yoshida and Shiosaka, 1999) and prothrombin (Irwin *et al.*, 1988). Significant similarities were found between TLSP and other serine proteases of the human kallikrein family, suggesting that this gene belongs to the same family. The complete genomic structure of TLSP can be found in GenBank (Accession No. AF164623). When the protein coding sequence of TLSP was compared with that of other kallikrein-like proteins (Fig. 5), it was found that neuropsin has 51% exact matches and 69% matches with conserved changes. KLK-L2 has 50% exact homology and 68% matches with conservative changes. The presence of aspartate at position 197 indicates that the protein is a trypsin-like enzyme. TLSP contains 12 cysteine residues. Ten of these are conserved in many other kallikrein-like proteins (Fig. 5) and trypsin. Twenty-five of 29 invariant amino acids surrounding the active site of serine proteases are conserved in TLSP, and one of the nonconserved amino acids (Leu^{57} instead of Val) is also found in KLK-L2.

Serine protease genes can be classified into five different groups according to intron position (Irwin *et al.*, 1988; Luo *et al.*, 1998). The kallikreins, trypsinogen, and chymotrypsinogen belong to a group that has (a) an intron just downstream from the codon for the active site histidine residue, (b) a second intron downstream from the exon containing the codon for the active site aspartic acid residue, and (c) a third intron just upstream from the exon containing the codon for the active site serine residue. As seen in Fig. 3, the genomic organization of TLSP is very similar to that of this group of genes, and intron phases are completely conserved.

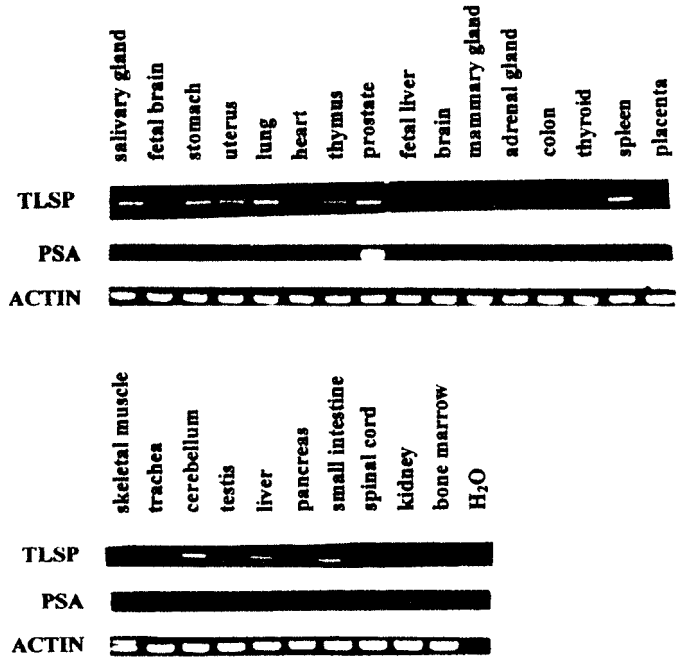

FIG. 6. Tissue expression of the TLSP gene as determined by RT-PCR. Actin and PSA are control genes. Qualitative interpretations are presented in Table 2.

Kallikreins are a subgroup of serine proteases and are known to play important roles in diverse physiological processes (Clements, 1997). Recently, many kallikrein-like genes have been implicated in the development and/or progression of different malignancies. Human kallikrein 3 (hK3/PSA) is the best diagnostic and prognostic marker for prostate cancer to date (Diamandis, 1998). Recombinant hK2 protein has been shown to activate PSA *in vitro* (Takayama *et al.*, 1997), and the combination of hK2 and free PSA has been recently found to increase the discrimination between prostate cancer and benign prostatic hyperplasia in patients with moderately elevated total PSA levels (Stenman, 1999). NES1 was found to be a tumor suppressor gene (Goyal *et al.*, 1998), and zyme/protease M/neurosin is suggested to be important in establishing breast and ovarian tumors and may function later in progression as a potential metastatic inhibitor

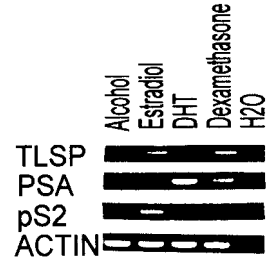

FIG. 7. Hormonal regulation of the TLSP gene in the BT-474 breast carcinoma cell line. Steroids were at 10^{-8} M final concentrations. Actin (not regulated by steroid hormones), PSA (up-regulated by androgens), and pS2 (up-regulated by estrogens) are control genes. TLSP is up-regulated by estrogens and gluccorticoids. For more details see text.

(Anisowicz *et al.*, 1996). More recent data suggest that PSA may be a potent antiangiogenic molecule (Fortier *et al.*, 1999) and an inducer of apoptosis and reduced cell proliferation (Balbay *et al.*, 1999; Lai *et al.*, 1996). Also, an alternatively spliced form of neuropsin and of the HSCCE is found to be elevated in ovarian carcinomas (Lowell *et al.*, 1999). All these data allow us to speculate that at least some members of the human kallikrein gene family may be involved in diverse malignancies and may serve as markers of the disease. Little is known about the role of TLSP in human malignancies. More research will be necessary to establish further the role of this new gene in human pathobiology.

REFERENCES

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, Z., Miller, W., *et al.* (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.

Angermann, A., Bergmann, C., and Appelhans, H. (1989). Cloning and expression of human salivary-gland kallikrein in *Escherichia coli*. *Biochem J.* **262:** 787–793.

Anisowicz, A., Sotiropoulou, G., Stenman, G., Mok, S. C., and Sager, R. (1996). A novel protease homolog differentially expressed in breast and ovarian cancer. *Mol. Med.* **2:** 624–636.

Balbay, M. D., Juang, P., Llansa, N., *et al.* (1999). Stable transfection of human prostate cancer cell line PC-3 with prostate specific antigen induces apoptosis both in-vivo and in-vitro. *Proc. Am. Assoc. Cancer Res.* **40:** 225–226 [Abstract].

Ballabio, A. (1993). The rise and fall of positional cloning? *Nat. Genet.* **7:** 513–520.

Birren, B., Green, E. D., Klaphoz, S., *et al.* (1998). "Genome Analysis: A Laboratory Manual," Cold Spring Harbor Laboratory Press, Plainview, NY.

Clements, J. (1997). The molecular biology of the kallikreins and their roles in inflamation. *In* "The Kinin System" (S. Farmer, Ed.), pp. 71–97, Academic Press, New York.

Clements, J. A. (1989). The glandular kallikrein family of enzymes: Tissue-specific and hormonal regulation. *Endocr. Rev.* **10:** 393–419.

Collins, F. S. (1992). Positional cloning: Let us not call it reverse anymore. *Nat. Genet.* **1:** 3–6.

Dayhoff, M. O., and Orcutt, B. C. (1979). Methods for identifying proteins by using partial sequences. *Proc. Natl. Acad. Sci. USA* **76:** 2170–2174.

Diamandis, E. P. (1998). Prostate specific antigen: Its usefulness in clinical medicine. *Trends Endocrinol. Metab.* **9:** 310–316.

Diamandis, E. P., Yu, H., and Sutherland, D. J. (1994). Detection of prostate-specific antigen immunoreactivity in breast tumours. *Breast Cancer Res. Treat.* **32:** 301–310.

Evans, B. A., Yun, Z. X., Close, J. A., Tregear, G. W., Kitamura, N., Nakanishi, S., *et al.* (1988). Structure and chromosomal localization of the human renal kallikrein gene. *Biochemistry* **27:** 3124–3129.

Fortier, A. H., Nelson, B. J., Grella, D. K., and Holaday, J. W. (1999). Antiangiogenic activity of prostate-specific antigen. *J. Natl. Cancer Inst.* **91:** 1635–1640.

Goyal, J., Smith, K. M., Cowan, J. M., *et al.* (1998). The role of NES1 serine protease as a novel tumor suppressor. *Cancer Res.* **58:** 4782–4786.

Iida, Y. (1990). Quantification analysis of 5′-splice signal sequences in mRNA precursors. Mutations in 5′-splice signal sequence of human beta-globin gene and beta-thalassemia. *J. Theor. Biol.* **145:** 523–533.

Irwin, D. M., Robertson, K. A., and MacGillivray, R. T. (1988). Structure and evolution of the bovine prothrombin gene. *J. Mol. Biol.* **200:** 31–45.

Ishikawa, T., Kashiwagi, H., Iwakami, Y., Hirai, M., Kawamura, T., Aiyoshi, Y., *et al.* (1998). Expression of alpha-fetoprotein and prostate specific antigen genes in several tissues and detection of mRNAs in normal circulating blood by reverse transcriptase-polymerase chain reaction. *Jpn. J. Clin. Oncol.* **28:** 723–728.

Kozak, M. (1991). An analysis of vertebrate mRNA sequences: Intimations of translational control. *J. Cell Biol.* **115:** 887–892.

Lai, L. C., Erbas, H., Lennard, T. W. J., and Peaston, R. T. (1996). Prostate-specific antigen in breast cyst fluid: Possible role of prostate-specific antigen in hormone-dependent breast cancer. *Int. J. Cancer* **66:** 743–746.

Little, S. P., Dixon, E. P., Norris, F., Buckley, W., Becker, G. W., Johnson, M., *et al.* (1997). Zyme, a novel and potentially amyloidogenic enzyme cDNA isolated from Alzheimer's disease brain. *J. Biol. Chem.* **272:** 25135–25142.

Liu, X. L., Wazer, D. E., Watanabe, K., and Band, V. (1996). Identification of a novel serine protease-like gene, the expression of which is down-regulated during breast cancer progression. *Cancer Res.* **56:** 3371–3379.

Lowell, J., *et al.* (1999). Cloning of tumor-associated differentially expressed gene-14, a novel serine protease overexpressed by ovarian carcinoma. *Cancer Res.* **59:** 4435–4439.

Lundstrom, A., and Egelrud, T. (1991). Stratum corneum chymotryptic enzyme: A proteinase which may be generally present in the stratum corneum and with a possible involvement in desquamation. *Acta Derm. Venereol.* **71:** 471–474.

Luo, L., Herbrick, J. A., Scherer, S. W., Beatty, B., Squire, J., and Diamandis, E. P. (1998). Structural characterization and mapping of the normal epithelial cell-specific 1 gene. *Biochem. Biophys. Res. Commun.* **247:** 580–586.

Miyata, T., Miyazawa, S., and Yasunaga, T. (1979). Two types of amino acid substitutions in protein evolution. *J. Mol. Evol.* **12:** 219–236.

Momota, Y., Yoshida, S., Ito, J., Shibata, M., Kato, K., Sakurai, K., *et al.* (1998). Blockade of neuropsin, a serine protease, ameliorates kindling epilepsy. *Eur. J. Neurosci.* **10:** 760–764.

Nelson, P. S., *et al.* (1999). Molecular cloning and characterization of prostase, an androgen-regulated serine protease with prostate-restricted expression. *Proc. Natl. Acad. Sci. USA* **96:** 3114–3119.

Riegman, P. H., Vlietstra, R. J., Suurmeijer, L., Cleutjens, C. B. J. M., and Trapman, J. (1992). Characterization of the human kallikrein locus. *Genomics* **14:** 6–11.

Riegman, P. H. J., Vlietstra, R. J., van der Korput, J. A. G. M., Romijn, J. C., and Trapman, J. (1989). Characterization of the prostate-specific antigen gene: A novel human kallikrein-like gene. *Biochem. Biophys. Res. Commun.* **159:** 95–102.

Ruddle, F. H. (1984). Reverse genetics and beyond. *Am. J. Hum. Genet.* **6:** 944–953.

Schedlich, L. J., Bennetts, B. H., and Morris, B. J. (1987). Primary structure of a human glandular kallikrein gene. *DNA* **6:** 429–437.

Simmer, J. P., Fukae, M., Tanabe, T., Yamakoshi, Y., Uchida, T., Xue, J., *et al.* (1998). Purification, characterization, and cloning of enamel matrix serine proteinase 1. *J. Dent. Res.* **77:** 377–386.

Stenman, U. (1999). New ultrasensitive assays facilitate studies on the role of human glandular kallikrein (hk2) as a marker for prostatic disease. *Clin. Chem.* **45:** 753–754.

Stephenson, S. A., Verity, K., Ashworth, L. K., and Clements, J. A. (1999). Localization of a new prostate-specific antigen-related serine protease gene, KLK4, is evidence for an expended human kallikrein gene family cluster on chromosome 19q13.3–13.4. *J. Biol. Chem.* **274:** 23210–23214.

Takayama, T. K, Fujikawa, K., and Davie, E. W. (1997). Characterization of the precursor of prostate-specific antigen. Activation by

trypsin and by human glandular kallikrein. *J. Biol. Chem.* **272:** 21582–21588.

Yamashiro, K., Tsuruoka, N., Kodama, S., Tsujimoto, M., Yamamura, Y., Tanaka, T., *et al.* (1997). Molecular cloning of a novel trypsin-like serine protease (neurosin) preferentially expressed in brain. *Biochim. Biophys. Acta* **1350:** 11–14.

Yoshida, S., and Shiosaka, S. (1999). Plasticity-related serine proteases in the brain (review). *Int. J. Mol. Med.* **3:** 405–409.

Yoshida, S., Taniguchi, M., Hirata, A., and Shiosaka, S. (1998a). Sequence analysis and expression of human neuropsin cDNA and gene. *Gene* **213:** 9–16.

Yoshida, S., Taniguchi, M., Suemoto, T., Oka, T., He, X., and Shiosaka, S. (1998b). cDNA cloning and expression of a novel serine protease, TLSP. *Biochim. Biophys. Acta* **1399:** 225–228.

Yousef, G. M., and Diamandis, E. P. The new kallikrein-like gene KLK-L2: Molecular characterization, mapping, tissue expression and hormonal regulation. *J. Biol. Chem.* **274:** 37511–37516.

Yousef, G. M., Luo, L. Y., and Diamandis, E. P. (1999a). Identification of novel human kallikrein-like genes on chromosome 19q13.3–q13.4. *Anticancer Res.* **19:** 2843–2852.

Yousef, G. M., Luo, L. Y., Scherer, S. W., Sotiropoulou, G., and Diamandis, E. P. Molecular characterization of zyme/ protease M/neurosin (PRSS9), a hormonally-regulated kallikrein-like serine protease. *Genomics,* in press.

Yousef, G. M., Obiezu, C. V., Luo, L. Y., Black, M. H., and Diamandis, E. P. (1999b). Prostase/KLK-L1 is a new member of the human kallikrein gene family, is expressed in the prostate and breast tissues and is hormonally regulated. *Cancer Res.* **59:** 4252–4256.