THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 274, No. 53, Issue of December 31, pp. 37511–37516, 1999

Printed in U.S.A.

The New Kallikrein-like Gene, *KLK-L2*

MOLECULAR CHARACTERIZATION, MAPPING, TISSUE EXPRESSION, AND HORMONAL REGULATION*

(Received for publication, July 22, 1999, and in revised form, September 24, 1999)

George M. Yousef and Eleftherios P. Diamandis‡

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

Since in rodents the kallikreins are represented by a large multi-gene family, the restriction of this family in humans to three genes is somewhat surprising. In an effort to identify new human kallikrein genes, we examined a genomic area of about 300 kilobases on chromosome 19q13.3-q13.4, a region that contains most of the currently known kallikreins. By using the positional candidate approach, we were able to identify a new gene named *KLK-L2* (for kallikrein- like gene 2). Screening of human EST libraries allowed us to delineate the full genomic and cDNA structure of the new gene. *KLK-L2* consists of 5 coding exons and 4 introns and has significant similarities to other members of the kallikrein multi-gene family. Homology studies suggest that the protein is likely secreted. *KLK-L2* is expressed mainly in breast, brain, and testis and to a lesser extent in many other tissues. *KLK-L2* is up-regulated by estrogens and progestins in the breast cancer cell line BT-474.

Strategies for new gene discovery are undergoing rapid evolution. Traditional procedures for new gene identification, such as CpG island mapping and cross-species hybridization will soon be replaced by methods that are based on accumulating knowledge of chromosomal localization of genes and increasing amounts of nucleotide sequence data (1). Positional candidate cloning is a new approach for gene discovery that combines the knowledge of map position with the increasingly dense human transcript maps and the available expressed sequence tags (ESTs) (2).

The kallikreins are a group of serine proteases involved in the post-translational processing of polypeptide precursors (3). This enzyme family primarily consists of plasma kallikrein and tissue or glandular kallikreins. Plasma kallikrein is encoded by a single gene that is structurally different from the genes encoding tissue kallikreins. The tissue kallikreins comprise a large multigene family of enzymes in rodents, with a highly conserved sequence and tertiary structures. In the mouse genome, at least 24 genes have been identified (4). A similar family of 15–20 kallikreins has been found in the rat genome (5). The structural organization of the kallikrein genes includes five coding exons and is highly conserved in all species studied thus far (6).

In humans, the kallikrein gene family locus is on chromosome 19q13 (7–9), a region that is syntenic to the mouse tissue kallikrein gene family locus on chromosome 7 (10). However, in marked contrast to the large rodent kallikrein gene family, the human kallikrein gene family was, until recently, composed of three genes. From Southern blot analysis, the size of this family has been suggested to vary from just 3–4 genes (11–12) to as many as 19 genes (13). Recently, a few new putative members of the human kallikrein gene family have been discovered such as the zyme (14) (also called protease M (15) or neurosin (16)), the normal epithelial cell-specific-1 gene (*NES1*) (17), and neurosin (18).

In our efforts to study the role of the kallikrein gene family in the initiation and/or progression of cancer, we have analyzed a 300-kb[1] genomic region on chromosome 19q13.3 -q13.4 and identified a new gene named *KLK-L2* (for kallikrein-like gene 2). In this study, we describe the identification of the new gene, its genomic and mRNA structure, its precise location in relation to other known kallikreins, and its tissue expression pattern.

MATERIALS AND METHODS

DNA Sequence on Chromosome 19—We have obtained sequencing data of approximately 300 kb of nucleotides on chromosome 19q13.3-q13.4 from the web site of the Lawrence Livermore National Laboratory. This sequence was in the form of nine contigs of different lengths. We performed a restriction analysis study of the available sequences using the WebCutter computer program, and with the aid of the *Eco*RI restriction map of this area (also available from Lawrence Livermore National Laboratory) we constructed an almost contiguous stretch of genomic sequences. We identified the relative positions of the known kallikrein genes *PSA* (GenBank™ accession number X14810), *KLK2* (GenBank™ accession number M18157), and zyme (GenBank™ accession number U60801) by using the alignment program BLAST 2 (19).

New Gene Identification—We used a number of computer programs to predict the presence of putative new genes in the genomic area of interest. We initially tested these programs using the known genomic sequences of the *PSA*, protease M, and *NES1* genes. The most reliable computer programs GeneBuilder, Grail 2, and GENEID-3 were selected for further use.

EST Searching—The predicted exons of the putative new gene were subjected to homology search using the BLASTN algorithm (19) on the National Center for Biotechnology Information web server against the human EST data base (dbEST). Clones with >95% homology were obtained from the I.M.A.G.E. consortium (20) through Research Genetics Inc, Huntsville, AL (Table I). The clones were propagated, purified, and sequenced from both directions with an automated sequencer using insert-flanking vector primers.

Tissue Expression—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

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF135028.

‡ To whom correspondence should be addressed: Mount Sinai Hospital, Department of Pathology and Laboratory Medicine, 600 University Ave., Toronto, Ontario M5G 1X5, Canada. Tel.: 416-586-8443; Fax: 416-586-8628; E-mail: ediamandis@mtsinai.on.ca

[1] The abbreviations used are: kb, kilobase(s); bp, base pair(s); KLK, kallikrein; KLK-L, kallikrein-like; RT-PCR, reverse transcriptase-polymerase chain reaction; PSA, prostate-specific antigen; dNTPs, deoxynucleoside triphosphates; EST, expressed sequence tag; EMSP1, enamel matrix serine proteinase 1; TLSP, trypsin-like serine protease; HSCCE, human stratum corneum chymotryptic enzyme; *NES1*, normal epithelial cell-specific 1 gene; contig, group of overlapping clones.

TABLE I
EST clones with >95% homology to exons of KLK-L2

GenBank™ No.	Tissue of origin	I.M.A.G.E. ID	Homologous exons
W73140	Fetal heart	344588	4,5
W73168	Fetal heart	344588	3,4,5
AA862032	Squamous cell carcinoma	1485736	4,5
AI002163	Testis	1619481	3,4,5
N80762	Fetal lung	300611	5
W68361	Fetal heart	342591	5
W68496	Fetal heart	342591	5
AA292366	Ovarian tumor	725905	1,2
AA394040	Ovarian tumor	726001	5

for PCR reactions with the primers described in Table II. Tissue cDNAs were amplified at various dilutions.

Breast Cancer Cell Line and Hormonal Stimulation Experiments—The breast cancer cell line BT-474 was purchased from the American Type Culture Collection (ATCC), Manassas, VA. Cells were cultured in RPMI media (Life Technologies, Inc.) supplemented with glutamine (200 mmol/liter), bovine insulin (10 mg/liter), 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. 24 h before the experiments, the culture media were changed into phenol red-free media containing 10% charcoal-stripped fetal bovine serum. For stimulation experiments, various steroid hormones dissolved in 100% ethanol were added into the culture media, at a final concentration of 10^{-8} M. Cells stimulated with 100% ethanol were included as controls. The cells were cultured for 24 h, then harvested for mRNA extraction.

Reverse Transcriptase-Polymerase Chain Reaction—Total RNA was extracted from the breast cancer cells using Trizol reagent (Life Technologies, Inc.) following the manufacturer's instructions. RNA concentration was determined spectrophotometrically. 2 μg of total RNA was reverse-transcribed into first strand cDNA using the Superscript™ preamplification system (Life Technologies, Inc.). The final volume was 20 μl. Based on the combined information obtained from the predicted genomic structure of the new gene and the EST sequences, two gene-specific primers were designed (Table II), 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 Biochemicals) 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, 63 °C for 1 min, and a final extension at 63 °C for 10 min. Equal amounts of PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. All primers for RT-PCR spanned at least two exons to avoid contamination by genomic DNA.

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.

Structure Analysis—Multiple alignment was performed using the Clustal X software package and the multiple alignment program available from the Baylor College of Medicine, Houston, TX. Phylogenetic studies were performed using the Phylip software package. Distance matrix analysis was performed using the Neighbor-Joining/UPGMA program, and parsimony analysis was done using the Protpars program. A hydrophobicity study was performed using the Baylor College of Medicine search launcher programs. Signal peptide was predicted using the SignalP server. Protein structure analysis was performed by SAPS (structural analysis of protein sequence) program.

RESULTS

Identification of the KLK-L2 Gene—Computer analysis of the genomic sequence predicted a putative new gene consisting of four exons. This gene was detected by all programs used, and all exons had high prediction scores. EST sequence homology search of the putative exons against the human EST data base (dbEST) revealed nine EST clones from different tissues with >95% identity to the putative exons of our gene (Table I). Positive clones were obtained, and the inserts were sequenced from both directions. The Blast 2 sequences program was used to compare the EST sequences with the predicted exons, and final selection of the exon-intron splice sites was done according to the EST sequences. The presence of many areas of overlap between the various EST sequences allowed us to further verify the structure of the new gene. The coding and genomic sequence of the gene has been deposited in GenBank™ (accession number AF135028). The 3′ end of the gene was verified by the presence of poly(A) stretches that are not present in the genomic sequence at the end of two of the sequenced ESTs. One of the sequenced ESTs revealed the presence of an additional exon at the 5′ end. The nucleotide sequence of this exon matches exactly with the genomic sequence. To further identify the 5′ end of the gene, 5′-Rapid amplification of cDNA ends was performed, but no additional sequence could be obtained. However, as is the case with other kallikreins, the presence of further up-stream untranslated exon(s) could not be excluded.

Mapping and Chromosomal Localization of the KLK-L2 Gene—Alignment of *KLK-L2* gene and the sequences of other known kallikrein genes within the 300-kb area of interest enabled us to precisely localize all genes and to determine the direction of transcription, as shown by the arrows in Fig. 1. The *PSA* gene was found to be the most centromeric, separated by 12,508 base pairs (bp) from KLK2, and both genes are transcribed in the same direction (centromere to telomere). The prostase/*KLK-L1* gene (GenBank™ accession number AF135023) is 26,229 bp more telomeric and transcribes in the opposite direction, followed by *KLK-L2*. The distance between *KLK-L1* and *KLK-L2* is about 35 kb; however, we could not establish it more precisely due to the presence of a gap in the genomic sequence. The zyme gene is 5,981 bp more telomeric, and the latter 3 genes are all transcribed in the same direction (Fig. 1).

Structural Characterization of the KLK-L2 Gene and Its Protein Product—The *KLK-L2* gene, as presented in Fig. 2, is formed of 5 coding exons and 4 intervening introns, spanning an area of 9,349 bp of genomic sequence on chromosome 19q13.3-q13.4. The lengths of the exons are 73, 262, 257, 134, and 156 bp. The intron/exon splice sites (mGT . . . AGm) and their flanking sequences are closely related to the consensus splicing sites (mGTAAGT . . . CAGm, where m is any nucleotide) (21). The presumptive protein-coding region of the *KLK-L2* gene is formed of a 879-bp nucleotide sequence encoding a deduced 293-amino acid polypeptide with a predicted molecular mass of 32 kDa. There are two potential translation initiation codons (ATG) at positions 1 and 25 of the predicted first exon (numbers refer to GenBank™ accession number AF135028). We assume that the first ATG will be the initiation codon, since 1) the flanking sequence of that codon (GCGGCCATGG) matches closely with the Kozak consensus sequence for initiation of translation (GCC(A/G)CCATGG) (22) and is exactly the same as that of the homologous zyme gene. (15) At this initiation codon, the putative signal sequence at the N terminus is similar to other trypsin-like serine proteases (prostase and enamel matrix serine proteinase (EMSP) (Fig. 3). The cDNA ends with a 328-bp 3′-untranslated region containing a conserved polyadenylation signal (AATAAA) which was located 11 bp up-stream of the poly(A) tail (at a position exactly the same as that of the zyme poly(A) tail) (14).

A hydrophobicity study of the *KLK-L2* gene shows a hydrophobic region in the N-terminal region of the protein (Fig. 4), suggesting that a presumed signal peptide is present. By computer analysis, a 29-amino acid signal peptide is predicted with a cleavage site at the carboxyl end of Ala^{29}. For better characterization of the predicted structural motif of the KLK-L2 protein, it was aligned with other members of the kallikrein multigene family, (Fig. 3), and the predicted signal peptide cleavage

TABLE II
Primers used for RT-PCR analysis

Gene	Primer name	Sequence[a]	Product size
			base pairs
KLK-L2	KS	GGATGCTTACCCGAGACAGA	342
	KAS	GCTGGAGAGATGAACATTCT	
pS2	PS2S	GGTGATCTGCGCCCTGGTCCT	328
	PS2AS	AGGTGTCCGGTGGAGGTGGCA	
PSA	PSAS	TGCGCAAGTTCACCCTCA	754
	PSAAS	CCCTCTCCTTACTTCATCC	
Actin	ACTINS	ACAATGAGCTGCGTGTGGCT	372
	ACTINAS	TCTCCTTAATGTCACGCACGA	
KLK-L2	R1	CCGAGACGGACTCTGAAAACTTTCTTCC	
	R2	TGAAAACTTTCTTCCTGCAGTGGGCGGC	

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

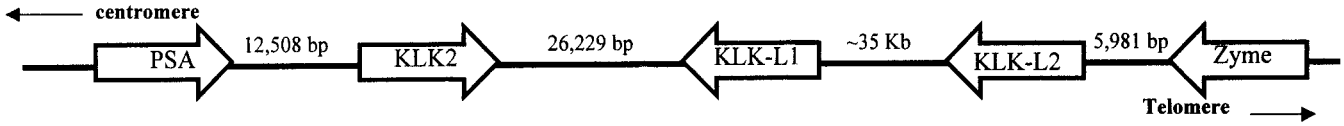

FIG. 1. **An approximate 300-kb region of almost contiguous genomic sequence around chromosome 19q13.3- q13.4.** Genes are represented by *horizontal arrows* denoting the direction of the coding sequence. Distances between genes are mentioned in base pairs. The figure is not drawn to scale.

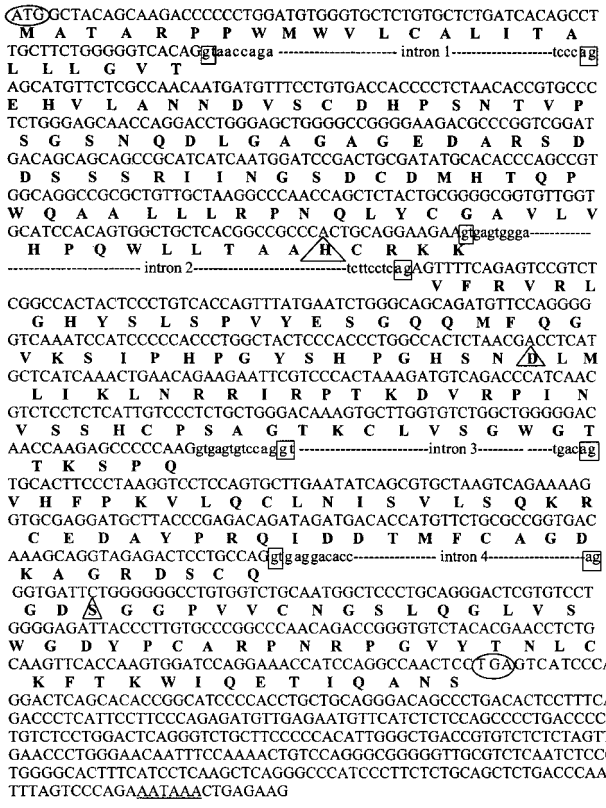

FIG. 2. **Genomic organization and partial genomic sequence of the *KLK-L2* gene.** Intronic sequences are not shown except for the splice junctions. Introns are shown with *lowercase letters*, and exons are shown with *capital letters*. For the full sequence, see GenBank™ accession number AF135028. The start and stop codons are *circled*, and the exon-intron junctions are *boxed*. The translated amino acids of the coding region are shown underneath by a single0letter abbreviation. The catalytic residues are *inside triangles*. The putative polyadenylation signal is *underlined*.

site was found to match with the predicted signal cleavage sites of zyme (14), *KLK1* (23), *KLK2* (24), and prostase/*KLK-L1* (25) genes. Also, sequence alignment supports by analogy the presence of a cleavage site at the carboxyl end of Ser^{66}, which is the exact site predicted for cleavage of the activation peptide of all the other kallikreins aligned in Fig. 3. Interestingly, the starting amino acid sequence of the mature protein IING(E/S)DC is conserved in the prostase and enamel matrix serine proteinase 1 (*EMSP*) genes. Thus, like other kallikreins, KLK-L2 is likely also synthesized as a pre-proenzyme that contains an N-terminal signal peptide (prezymogen) followed by an activation peptide and the enzymatic domain.

The presence of aspartate (D) in position 239 suggests that KLK-L2 will possess a trypsin-like cleavage pattern like most of the other kallikreins (*e.g.* KLK1, KLK2, trypsin-like serine protease (TLSP), neuropsin, zyme, prostase, and EMSP) but different from PSA, which has a serine (S) residue in the corresponding position and is known to have a chymotrypsin-like activity (Fig. 3) (3). The dotted region in Fig. 3 indicates an 11-amino acid loop characteristic of the classical kallikreins (*PSA*, *KLK1*, and *KLK2*) but not found in *KLK-L2* or other members of the kallikrein-like gene family (14).

Homology with the Kallikrein Multi-gene Family—We aligned the mature 227-amino acid sequence of the predicted protein against the GenBank™ data base and the known kallikreins using the BLASTP and BLAST 2 sequence programs. KLK-L2 is found to have 54% amino acid sequence identity and 68% similarity with the *EMSP1* gene, 50% identity with both TLSP and neuropsin genes, and 47, 46, and 42% identity with trypsinogen, zyme, and *PSA* genes, respectively. Multiple alignment study shows that the typical catalytic triad of serine proteases is conserved in the *KLK-L2* gene (His^{108}, Asp^{153}, and Ser^{245}), and 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), serine-(GD*S*GGP), and aspartate(*D*LMLI)) (14, 16).

Twelve cysteine residues are present in the putative mature KLK-L2 protein; 10 of them are conserved in all the serine proteases that are aligned in Fig. 3 and would be expected to form disulfide bridges. The other two cysteines (Cys^{178} and Cys^{279}) are not found in PSA, KLK1, KLK2 or trypsinogen; however, they are found in similar positions in prostase, *EMSP1*, zyme, neuropsin, and *TLSP* genes and are expected to form an additional disulfide bond. Twenty-nine "invariant" amino acids surrounding the active site of serine proteases

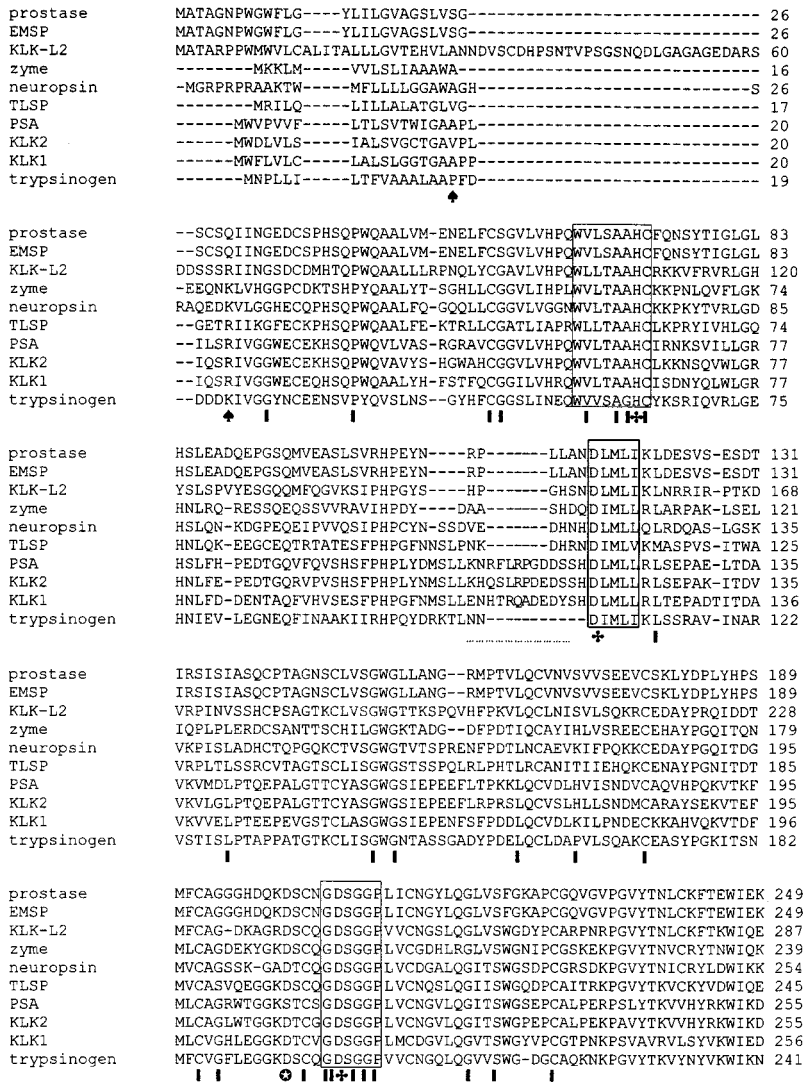

FIG. 3. **Alignment of the deduced amino acid sequence of *KLK-L2* with members of the kallikrein multi-gene family.** Genes are (from top to bottom) prostase/*KLK-L1* (GenBank™ accession number AAD21581), *EMSP1* (GenBank™ accession number NP number 04908), *KLK-L2* (GenBank™ accession number AF135028), zyme (GenBank™ accession number Q92876), neuropsin (GenBank™ accession number BAA28673), trypsin-like serine protease (*TLSP*) (GenBank™ accession number BAA33404), PSA (GenBank™ accession number P07288), KLK2 (GenBank™ accession number P20151), KLK1 (GenBank™ accession number NP number 002248), and trypsinogen (GenBank™ accession number P07477). *Dashes* represent gaps to bring the sequences to better alignment. The residues of the catalytic triad are represented by ⌘, and the 29 invariant serine protease residues are represented by *bold* ❙ or ⌘ signs. Conserved areas around the catalytic triad are *boxed*. The predicted cleavage sites are indicated by ♠ . The *dotted area* represents the kallikrein loop sequence. The trypsin-like cleavage pattern is indicated by ✪.

have been described (26). Of these, 26 are conserved in KLK-L2. One of the nonconserved amino acids (Ser^{210} instead of Pro) is also found in prostase and *EMSP1* genes, the second (Leu^{103} instead of Val) is also found in *TLSP* gene, and the third (Val^{174} instead of Leu) is also not conserved in prostase or *EMSP1* genes. According to protein evolution studies, each of these amino acid changes represents a conserved evolutionary substitution to a protein of the same group (26, 27).

Evolution of the KLK-L2 Gene—To predict the phylogenetic relatedness of the *KLK-L2* gene with other serine proteases, the amino acid sequences of the kallikrein genes were aligned together using the Clustal X multiple alignment program, and a distance matrix tree was predicted using the Neighbor-joining/UPGMA method (Fig. 4). Phylogenetic analysis separated the classical kallikreins (*KLK1*, *KLK2*, and *PSA*) and grouped the *KLK-L2* with prostase/*KLK-L1*, *EMSP1*, and *TLSP*, in consistence with previously published studies (25, 28).

Tissue Expression of the KLK-L2 Gene—As shown in Table III and Fig. 5, the *KLK-L2* gene is primarily expressed in the brain, mammary gland, and testis, but 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 KLK-L2 Gene—A steroid hormone receptor positive breast cancer cell line (BT-474) was used as a model to verify whether the *KLK-L2* gene is under steroid hormone regulation. PSA was used as a control known to be up-regulated by androgens and progestins, and pS2 was used as an estrogen up-regulated control. Our results indicate that *KLK-L2* is up-regulated by estrogens and progestins (Fig. 6).

DISCUSSION

Kallikreins are a subgroup of serine proteases that play important roles in diverse physiological processes (3). Recently, some of the kallikrein genes were found to be linked to the development and/or progression of different types of malignancies. Human kallikrein 3 (KLK3/PSA) is the best diagnostic and prognostic marker for prostate cancer to date (29, 30). Recombinant KLK2 has been shown to activate PSA *in vitro* (31), and the combination of KLK2 and free PSA has recently been found to increase the discrimination between prostate cancer and benign prostatic hyperplasia in patients with moderately elevated total PSA levels (32). *NES1* appears to be a tumor suppressor gene (33), and zyme (protease M/neurosin) may be important in the establishment of breast and ovarian tumors and may function later in progression as a potential metastatic inhibitor (15). At least three kallikrein genes (*PSA*, *NES1*, and zyme) are down-regulated in breast cancer.

In search for new genes that may be involved in malignant diseases, we investigated the possible existence of new kallikreins. Based on mapping of the rodent kallikrein genes and the documented strong conservation between the human chro-

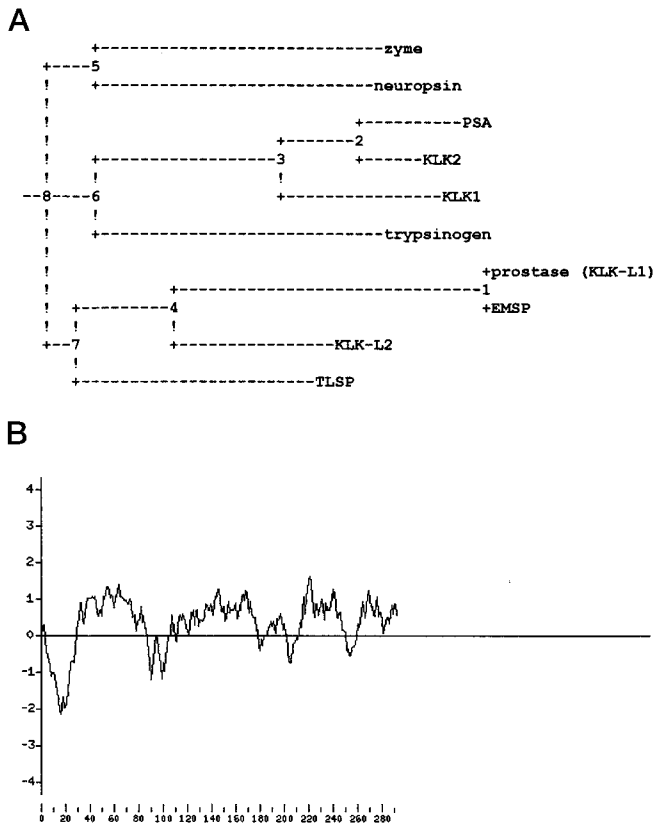

FIG. 4. *A*, dendrogram of the predicted phylogenetic tree for some kallikrein genes. Neighbor-joining/UPGMA method was used to align *KLK-L2* with other members of the kallikrein gene family. Gene names and accession numbers are listed in Fig. 3. The tree grouped the classical kallikreins (*KLK1*, *KLK2*, and *PSA*) together and aligned the *KLK-L2* gene in one group with *EMSP*, prostase, and *TLSP*. *B*, plot of hydrophobicity and hydrophilicity of *KLK-L2*.

TABLE III
Tissue expression of KLK-L2 by RT-PCR analysis

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

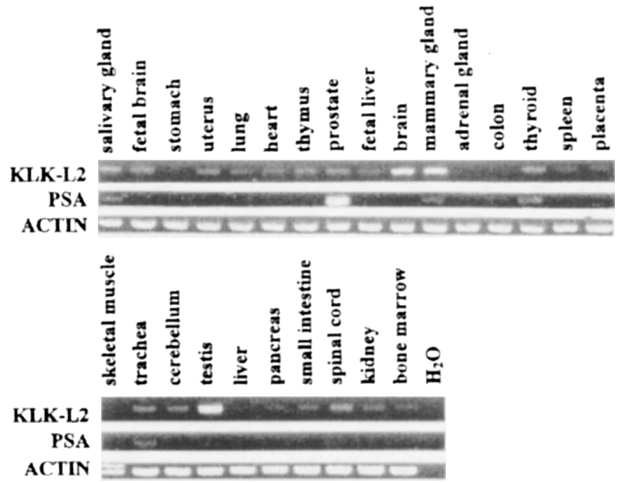

FIG. 5. **Tissue expression of the *KLK-L2* gene as determined by RT-PCR.** Actin and *PSA* are control genes. Interpretations are presented in Table III.

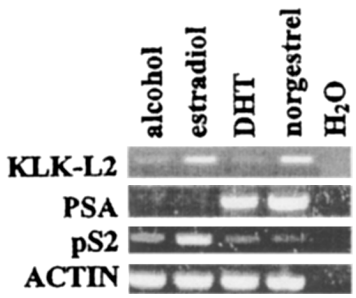

FIG. 6. **Hormonal regulation of the *KLK-L2* gene in BT-474 breast carcinoma cell lines.** Steroids were at 10^{-8} M final concentrations. Actin (not regulated by steroid hormones), *pS2* (up-regulated by estrogens), and *PSA* (up-regulated by androgens and progestins) are control genes. *KLK-L2* is up-regulated by estrogens and progestins. For more details see text. *DHT*, dihydrotestosterone.

mosome 19q13.1-q13.4 region and the 17 loci in a 20-centimorgan proximal part of mouse chromosome 7 (10, 34), we identified a putative "critical region." With the aid of computer programs for gene prediction and the available EST data base, we were able to identify a new gene, named *KLK-L2* (for kallikrein-like gene 2). The 3′ end of the gene was verified by the presence of poly(A) stretches in the sequenced ESTs that were not found in the genomic sequence, and the start of translation was identified by the presence of a start codon in a well conserved consensus Kozak sequence.

As is the case with other kallikreins, the *KLK-L2* gene is composed of 5 coding exons and 4 intervening introns and, except for the second coding exon, the exon lengths are comparable to those of other members of the kallikrein gene family (Fig. 7). The exon-intron splice junctions were identified by comparing the genomic sequence with the EST sequence and were further confirmed by the conservation of the consensus splice sequence (mGT . . . AGm, where m is any nucleotide) (21) and the fully conserved intron phases (35), as shown in Fig. 7. Furthermore, the position of the catalytic triad residues in relation to the different exons is also conserved (Fig. 7). As is the case with most other kallikreins, except PSA and HSCCE, KLK-L2 is more functionally related to trypsin than to chymotrypsin (3). The wide range of tissue expression of KLK-L2 should not be surprising since by using the more sensitive RT-PCR technique instead of Northern blot analysis (36), many kallikrein genes were found to be expressed in a wide variety of tissues including salivary gland, kidney, pancreas, brain, and tissues of the reproductive system (uterus, mammary gland, ovary, and testis) (3). *KLK-L2* is highly expressed in the brain. Another kallikrein, neuropsin, was also found to be highly expressed in the brain and has been shown to have important roles in neural plasticity in mice (18). Also, the zyme gene is highly expressed in the brain and appears to have amyloido-

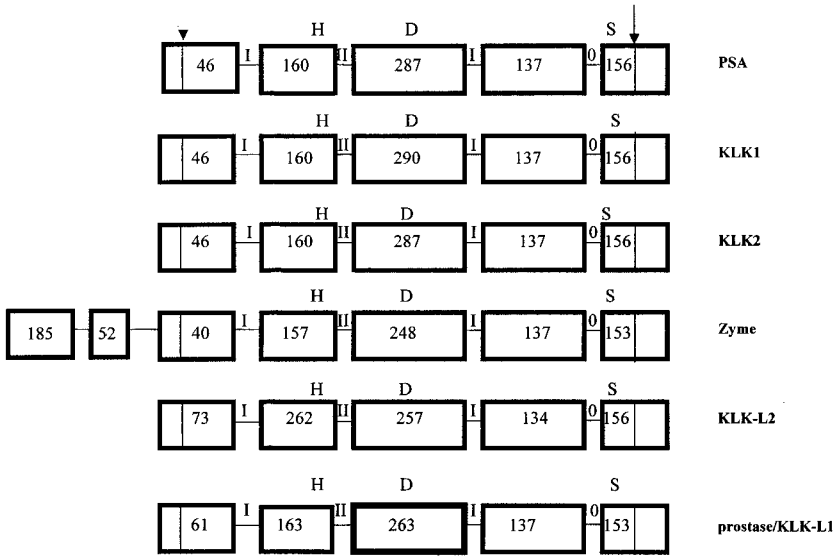

FIG. 7. **Schematic diagram showing the comparison of the genomic structure of *PSA*, *KLK1*, *KLK2*, zyme, *KLK-L2*, and *KLK-L1*/prostase genes.** Exons are shown by *open boxes*, and introns by the *connecting lines*. *Arrowheads* show the start codon, and *arrows* show the stop codon. *Letters* above the *boxes* indicate relative positions of the catalytic triad; *H* denotes histidine, *D* denotes aspartic acid, and *S* denotes 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* denotes the intron occurs after the second nucleotide, *0* denotes the intron occurs between codons. *Numbers* inside *boxes* indicate exon lengths in base pairs. The figure is not drawn to scale.

genic potential (14). Taken together, these data point out to a possible role of *KLK-L2* in the central nervous system.

It was initially thought that each kallikrein enzyme has one specific physiological substrate. However, the increasing number of substrates, which purified proteins can cleave *in vitro*, has led to the suggestion that they may perform a variety of functions in different tissues or physiological circumstances. The biological function of *KLK-L2* is not yet known. Serine proteases encode protein cleaving enzymes that are involved in digestion, tissue remodeling, blood clotting, etc., and many of the kallikrein genes are synthesized as precursor proteins that must be activated by cleavage of the pro-peptide. The predicted trypsin-like cleavage specificity of *KLK-L2* makes it a candidate activator of other kallikreins or it may be involved in a "cascade" of enzymatic reactions similar to those found in fibrinolysis and blood clotting (31, 37).

In conclusion, we characterized a new member of the human kallikrein gene family, *KLK-L2*. This gene is hormonally regulated, and it is mostly expressed in the brain, mammary gland, and testis. Based on our experience with other kallikreins that are already used as valuable tumor markers (*PSA*, *hk2*), we speculate that *KLK-L2* may also have utility in similar applications. This possibility, as well as the physiological function of the protein need further investigation.

Acknowledgments—We thank David Irwin and Irvin Bromberg for helpful suggestions and advice on use of some computer programs and Liu-Ying Luo and Margot Black for technical assistance.

Note Added in Proof—We have now identified an untranslated exon of the *KLK-L2* gene; this is described in our updated GenBank™ submission (AF135028).

REFERENCES

1. Boehm, T. (1998) *Methods* **14,** 152–158
2. Collins, F. S. (1995) *Nat. Genet.* **9,** 347–350
3. Clements, J. (1997) in *The Kinin System* (Farmer, S., ed) pp. 71–97, Academic Press, Inc., New York
4. Evans, B. A., Drinkwater, C. C., and Richards R. I. (1987) *J. Biol. Chem.* **262,** 8027–8034
5. Ashley, P. L., and MacDonald, R. J. (1985) *Biochemistry* **24,** 4520–4527
6. Lin, F. K., Lin, C. H., Chou, C. C., Chen, K., Lu, H. S., Bacheller, W., Herrera, C., Jones, T., Chao, J., and Chao, L. (1993) *Biochim. Biophys. Acta* **1173,** 325–328
7. Riegman, P. H., Vlietstra, R. J., Suurmeijer, L, Cleutjens, C. B., and Trapman, J. (1992) *Genomics* **14,** 6–11
8. Riegman, P. H., Vlietstra, R. J., Klaassen, P., van der Korput, J. A., Geurts van Kessel, A., Romijn, J. C., and Trapman, J. (1989) *FEBS Lett.* **247,** 123–126
9. Richards, R. I., Holman, K., Shen, Y. Kozman, H., Harley, H., Brook, D., and Shaw, D. (1991) *Genomics* **11,** 77–82
10. Nadeau, J. H., Davisson, M. T., Doolittle, D. P., Grant, P., Hillyard, A. L., Kosowsky, M. R., and Roderick, T. H. (1992) *Mamm. Genome* **3,** 480–536
11. Fukushima, D., Kitamura, N., and Nakanishi, S. (1985) *Biochemistry* **24,** 8037–8043
12. Baker, A. R., and Shine, J. (1985) *DNA* (*N. Y.*) **4,** 445–450
13. Murray, S. R., Chao, J., Lin, F., and Chao, L. (1990) *J. Cardiovasc. Pharmacol.* **15,** (suppl.) 7–15
14. Little, S. P., Dixon E. P., and Norris F., Buckley, W., Becker, G. W., Johnson, M., Dobbins, J. R., Wyrick, T., Miller, J. R., MacKellar, W., Hepburn, D., Corvalan, J., McClure, D., Liu, X., Stephenson, D., Clemens, J., and Johnstone, E. M. (1997) *J. Biol. Chem.* **272,** 25135–25142
15. Anisowicz, A., Sotiropoulou, G., Stenman, G., Mok, S. C., and Sager, R. (1996) *Mol. Med.* **2,** 624–636
16. Yamashiro, K., Tsuruoka, N., Kodama, S., Tsujimoto, M., Yamamura, Y., and Tanaka, T., Nakazato, H., and Yamaguchi, N. (1997) *Biochim. Biophys. Acta* **1350,** 11–14
17. Liu, X. L., Wazer, D. E., Watanabe, K., and Band, V. (1996) *Cancer Res.* **56,** 3371–3379
18. Yoshida, S., Taniguchi, M., Hirata, A., and Shiosaka, S. (1998) *Gene* **213,** 9–16
19. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402
20. Lennon, G., Auffray, C., Polymeropoulos, M., and Soares, M. B. (1996) *Genomics* **33,** 151–152
21. Iida, Y. (1990) *J. Theor. Biol.* **145,** 523–533
22. Kozak, M. (1991) *J. Cell Biol.* **115,** 887–892
23. Evans, B. A., Yun, Z. X., Close, J. A., Tregear, G. W., Kitamura, N., Nakanishi, S., Callen, D. F., Baker, E., Hyland, V. J., Sutherland, G. R., *et al.* (1988) *Biochemistry* **27,** 3124–3129
24. Schedlich, L. J., Bennetts, B. H., and Morris, B. J. (1987) *DNA* (*N. Y.*) **6,** 429–437
25. Nelson, P. S., Gan, L., Ferguson, C., Moss, P., Gelinas, R., Hood, L., and Wang, K. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 3114–3119
26. Dayhoff, M. O. (1976) *Fed. Proc.* **35,** 2132–2138
27. Miyata, T., Miyazawa, S., and Yasunaga, T. (1979) *J. Mol. Evol.* **12,** 219–236
28. Simmer, J. P., Fukae, M., Tanabe, T., Yamakoshi, Y., Uchida, T., Xue, J., Margolis, H. C., Shimizu, M., DeHart, B. C., Hu, C. C., and Bartlett, J. D. (1998) *J. Dent. Res.* **77,** 377–386
29. Diamandis, E. P. (1998) *Trends Endocrinol. Metab.* **9,** 310–316
30. Oesterling, J. E. (1991) *J. Urol.* **145,** 907–923
31. Takayama, T. K, Fujikawa, K., and Davie, E. W. (1997) *J. Biol. Chem.* **272,** 21582–21588
32. Stenman, U. (1999) *Clin. Chem.* **45,** 753–754
33. Goyal, J., Smith, K. M., Cowan, J. M., Wazer, D. E., Lee, S. W., and Band, V. (1998) *Cancer Res.* **58,** 4782–4786
34. Saunders, A. M., and Seldin, M. F. (1990) *Genomics* **8,** 525–535
35. Irwin, D. M., Robertson, K. A., and MacGillivary, R. T (1988) *J. Mol. Biol.* **212,** 31–45
36. Clements, J. A., Mukhtar, A., Verity, K., Pullar, M., McNeill P., Cummins, J., and Fuller, P. J. (1996) *Clin. Endocrinol.* **44,** 223–231
37. Davie, E. W., Fujikawa, K., and Kisiel, W. (1991) *Biochemistry* **3,** 10363–10370