

Identification and Characterization of *KLK-L4*, a New Kallikrein-like Gene That Appears to be Down-regulated in Breast Cancer Tissues*

(Received for publication, December 3, 1999, and in revised form, January 21, 2000)

George M. Yousef^{‡§}, Albert Chang^{‡§}, and Eleftherios P. Diamandis^{‡§¶}

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

Kallikreins are a subgroup of serine proteases and these proteolytic enzymes have diverse physiological functions in many tissues. Growing evidence suggests that many kallikreins are implicated in carcinogenesis. In rodents, kallikreins constitute a large multigene family, but in humans, only three genes were identified. By using the positional candidate gene approach, we were able to identify a new kallikrein-like gene, tentatively named *KLK-L4* (for kallikrein-like gene 4). This new gene maps to chromosome 19q13.3-q13.4, is formed of five coding exons and four introns, and shows structural similarity to other kallikreins and kallikrein-like genes. *KLK-L4* is expressed in a variety of tissues including prostate, salivary gland, breast, and testis. Our preliminary results show that *KLK-L4* is down-regulated, at the mRNA level, in breast cancer tissues and breast cancer cell lines. Its expression is regulated by steroid hormones in the breast cancer cell line BT-474. This gene may be involved in the pathogenesis and/or progression of breast cancer and may find applicability as a novel cancer biomarker.

Prostate-specific antigen (PSA)¹ testing has revolutionized the management of patients with prostate cancer (1). The PSA gene (*KLK3*) is a member of the human tissue kallikrein gene family, which is also comprised of human glandular kallikrein 2 (*KLK2*) (2) and pancreatic/renal kallikrein (*KLK1*) (3) genes. More recently, new serine proteases with a high degree of homology to the kallikrein genes were cloned (4–9). The successful diagnostic use of PSA in prostate cancer suggests that other related or unrelated molecules might be discovered and serve as diagnostic tests for breast, ovarian, and other cancers. In addition to PSA, human glandular kallikrein 2 (encoded by the *KLK2* gene) may be useful as an adjuvant diagnostic marker for prostate cancer (10). Accumulating evidence indicates that some members of the kallikrein gene family are implicated in carcinogenesis. The normal epithelial cell-specific

1 gene (*NES1*) was found to be a tumor suppressor (11) that is down-regulated during breast cancer progression. The zyme/protease M/neurosin gene is expressed in primary breast cancers but is down-regulated at metastatic sites (4).

The large size of the kallikrein gene family in other species, such as rat and mouse, where kallikreins are encoded by 13–24 genes (12–13), and the recent identification of new kallikrein-like genes suggested that the human kallikrein gene family may be larger than previously thought. The rodent kallikrein genes are located in clusters on chromosome 7, and the region between two mouse kallikrein genes in a cluster can be as small as 3–7 kb (14).

In our efforts to identify new kallikrein-like genes that might be useful as diagnostic and/or prognostic markers for cancer, we studied a genomic area of ~300 kb around chromosome 19q13.3-q13.4, where the known human kallikrein genes are localized. We were able to identify three new kallikrein-like genes; *KLK-L1* (for kallikrein-like gene 1) (15) *KLK-L2* (for kallikrein-like gene 2) (16, 17), and *KLK-L3*.² Here, we describe the cloning of a new kallikrein-like gene, named kallikrein-like gene 4 (*KLK-L4*), together with its precise chromosomal localization in relation to other kallikreins and its tissue expression pattern. We further describe identification of alternatively spliced forms of this gene in some tissues. We also provide preliminary evidence indicating that *KLK-L4* is down-regulated in breast cancer tissues and breast cancer cell lines and that it is hormonally regulated in the breast cancer cell line BT-474.

MATERIALS AND METHODS

DNA Sequence on Chromosome 19 and Prediction of New Genes—We have obtained sequencing data of ~300 kb of nucleotides, around chromosome 19q13.3-q13.4, from the web site of the Lawrence Livermore National Laboratory and constructed an almost contiguous stretch of genomic sequences. A number of computer programs were used to predict the presence of putative new genes in this genomic area (16).

Expressed Sequence Tag (EST) Searching—The predicted exons of the putative new gene were subjected to homology search using the BLASTN algorithm (18) 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 (19) through Research Genetics Inc., Huntsville, AL. The clones were propagated, purified, and sequenced from both directions with an automated sequencer using insert-flanking vector primers.

Rapid Amplification of cDNA Ends (3'-RACE)—According to the EST sequence data and the predicted structure of the gene, two gene-specific primers were designed, and two rounds of RACE reactions (nested PCR) were performed with 5 μl of Marathon Ready 228 cDNA of human testis (CLONTECH) as a template. The reaction mix and PCR conditions used were according to the manufacturer's recommendations.

* 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) AF135024.

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

¹ The abbreviations used are: PSA, prostate-specific antigen; KLK, kallikrein; kb, kilobase; KLK-L, kallikrein-like; EST, expressed sequence tag; RACE, rapid amplification of cDNA ends; PCR, polymerase chain reaction; bp, base pair(s); RT, reverse transcription; TLSP, trypsin-like serine protease.

² G. M. Yousef, A. Chang, and E. P. Diamandis, unpublished data.

Tissue Expression—Total RNA isolated from 26 different human tissues was purchased from CLONTECH. We prepared cDNA as described below and used it for PCR reactions with different sets of primers (Table I). 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), Rockville, MD. 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 and then harvested for mRNA extraction.

Reverse Transcriptase Polymerase Chain Reaction—Total RNA was extracted from the breast cancer tissues and cell lines 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 (L4-F1 and L4-R1, see Table I) 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 MgCl₂, 200 μM dNTPs, 150 ng of primers, and 2.5 units of AmpliTaq Gold DNA polymerase (Roche Molecular Systems) 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

2 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.

Normal and Malignant Breast Tissues—Normal breast tissues were obtained from women undergoing reduction mammoplasties. Breast tumor tissues were obtained from female patients at participating hospitals of the Ontario Provincial Steroid Hormone Receptor Program. The normal and tumor tissues were immediately frozen in liquid nitrogen after surgical resection and stored in this manner until extracted. The tissues were pulverized with a hammer at dry ice temperature, and RNA was extracted as described above, using Trizol reagent.

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. The 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 the "SAPS" (structural analysis of protein sequence) program.

RESULTS

Cloning of the *KLK-L4* Gene—Computer analysis of the genomic sequence around chromosome 19q13.3-q13.4 predicted a putative new gene formed of at least 3 exons. To experimentally verify the existence of this gene, the putative exons were subjected to sequence homology search against the human EST data base (dbEST), and four EST clones with >97% homology were identified (Table II). All ESTs were cloned from testicular tissue. These clones were obtained and inserts were sequenced from both directions. Sequences were then compared with the computer-predicted structure, and final selection of the intron/exon splice sites was made according to the EST sequences.

As shown in Fig. 1, all ESTs match almost perfectly with the predicted 3 exons (exons 3, 4, and 5) of the gene. However, each of the ESTs extends further upstream with different exonic patterns, suggesting the presence of different splice variants. Attempts to translate these clone sequences demonstrated the

TABLE I
Primers used for reverse RT-PCR analysis

Gene	Primer name	Sequence ^a
<i>KLK-L4</i>	L4-F1	AACTCTACAATGTGCCAACA
	L4-R1	TTATTGTGGGCCCTTCAACC
	L4-R3	GGATGGTCCATTATAGGAC
	L4-A	AGGCTGCCCTACTAGTGCAA
	L4-B	ATATTGCCCTAGGTGGATGTG
	L4-D	AAGACTTCAAGGAGCCAAGC
	L4-E	GACCCCTCACCTCCCAAAT
	L4-X1	CTAGTGATCGCCTCCCTGAC
pS2	PS2S	GGTGATCTGCGCCTGGTCTCT
	PS2AS	AGGTGTCCGGTGGAGGTGGCA
PSA	PSAS	TGCGCAAGTTCACCCTCA
	PSAAS	CCCTCTCCTTACTTCATCC
Actin	ACTINS	ACAATGAGCTGCGTGTGGCT
	ACTINAS	TCTCCTTAATGTACGCACGA

^a All nucleotide sequences are given in the 5'→3' orientation.

TABLE II
EST clones with >95% homology to exons of *KLK-L4*

GenBank™	Tissue of origin	I.M.A.G.E. ID
AA399955	Testis	743113
AA401397	Testis	743113
AA846771	Testis	1392889
AI002101	Testis	1619045
AI032327	Testis	1644236

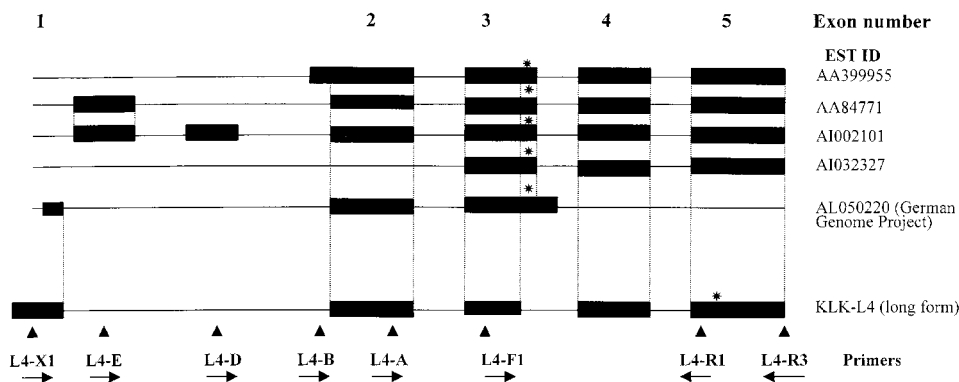


FIG. 1. Comparative genomic structure of the ESTs (Table II), the clone from The German Genome Project, and the long form of *KLK-L4*. Exons are represented by solid bars, and introns are represented by the connecting lines. Exon numbers on top of solid bars refer to our GenBank™ submission AF135024. The EST IDs represent GenBank™ accession numbers. Asterisks represent the positions of stop codons. Horizontal arrows indicate the direction of the PCR primers (described in Table I), and arrowheads indicate their position along the exons. Vertical dotted lines show alignment of identical fragments. For more details, see text.

presence, in some ESTs, of interrupting stop codons in all three possible reading frames. Homology search of the three common exons against the GenBank™ data base revealed a cDNA sequence from the German Human Genome Project. This clone has an identical exon 2 as the long form of *KLK-L4* gene (this form will be described below) but has an extended exon 3 that ends with a stop codon (Fig. 1). This clone was isolated from uterine tissue and is translated by software into a truncated protein product of 196 amino acids, which is followed by a

3'-untranslated region (GenBank™ accession no. AL050220).

Screening of cDNAs from 26 different tissues by RT-PCR, using gene-specific primers for exons 3 and 5 (L4-F1 and L4-R1) (Table I and Fig. 1), revealed that this gene is expressed in many tissues. Four tissues that show the highest level of expression (salivary gland, mammary gland, prostate, and testis (Fig. 2)) and uterus (the EST clone AL050220 was isolated from this tissue) were selected for identification of the full structure of the gene. Different PCR reactions were performed using one reverse primer (L4-R1) together with each of the forward primers located in upstream exons that were found in the different EST clones (primers L4-B, L4-D, L4-E) (Table I and Fig. 1). The PCR reactions were performed under different experimental conditions, using the EST clones as positive controls, and the PCR products were sequenced. None of these forms were found in any of the tissues, except in testis where all three forms were found (data not shown).

By RT-PCR of the *KLK-L4* gene using primers L4-R1 and L4-F1, it was found that the gene is expressed in a wide variety of tissues (Fig. 2). To obtain the structural forms that exist in these tissues, a homology study was performed. Aligning the predicted polypeptide of the *KLK-L4* gene with all other kallikreins and kallikrein-like genes suggested, by homology, that at least two more exons should be present upstream of the predicted three exons. The genomic fragment upstream of the third exon was subjected to further computer analysis for gene prediction, and exon 2 was identified based on: (a) a consensus exon/intron splice site; (b) preservation intron phase II after this exon, in agreement to intron phases of all other known kallikreins (see "Discussion" for details); (c) presence of the histidine residue of the catalytic triad (His⁷⁶) surrounded by a well conserved peptide motif (see below) just before the end of this exon; and (d) comparable exon length to other kallikrein genes. A potential first exon was also predicted from the upstream genomic sequence based on the preserved intron phase (phase I), and the existence of an in-frame start codon that is located at a comparable distance (in relation to other kallikreins) from the end of this exon. To verify this predicted

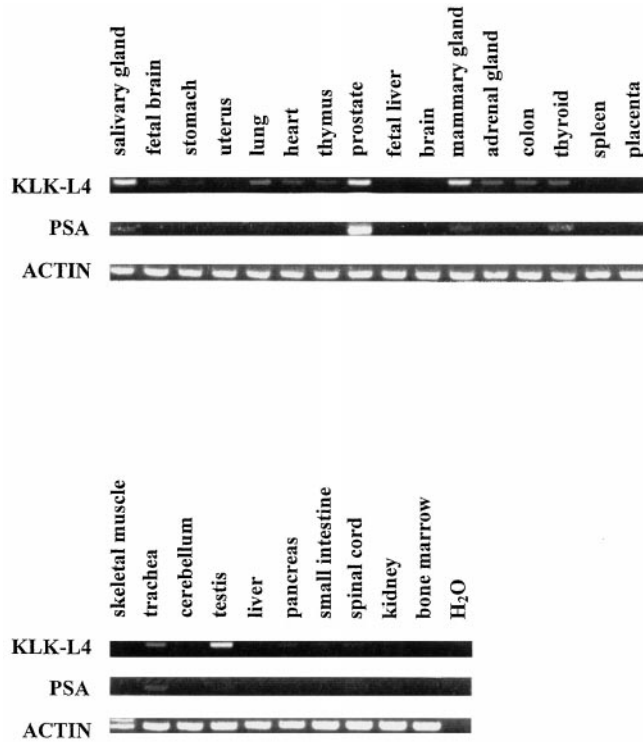


FIG. 2. Tissue expression of the *KLK-L4* gene as determined by RT-PCR. Actin and PSA are control genes. *KLK-L4* is highly expressed in breast, prostate, salivary gland, and testis.

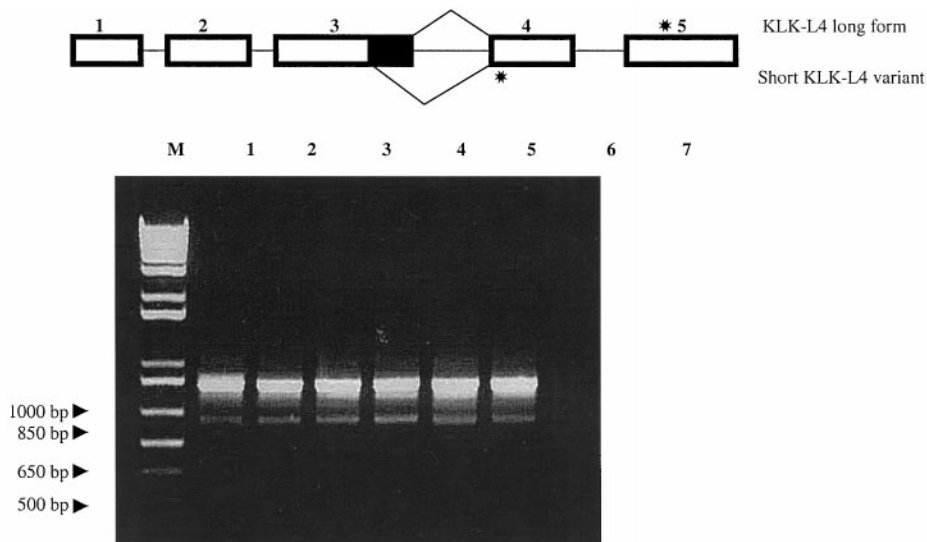


FIG. 3. Upper panel, diagram showing the comparative genomic structure of the long *KLK-L4* form and the short *KLK-L4* variant. Exons are represented by boxes, and introns are represented by the connecting lines. Exon numbers refer to our GenBank™ submission AF135024. The black region indicates the extra fragment (214 bp) that is found in the long, but not in the short form of the gene (see text for details). The positions of the stop codons of the two forms are marked with asterisks. Frame shifting occurs as a result of utilization of an alternative splice site, and a stop codon is generated at the beginning of exon 4 in the short form. Lower panel, PCR products of the amplification of the *KLK-L4* gene using L4-R1 and L4-X1 primers (Fig. 1 and Table I). Note the predominant long form and a minor band representing the short form of *KLK-L4* mRNA. M, markers with sizes in bp shown on the left. Tissues used: Lane 1, salivary gland; lane 2, mammary gland; lane 3, prostate; lane 4, testis; lane 5, uterus; lane 6, breast cancer tissue; lane 7, negative control.

TCAGGCCCGCCCGCCCTGCCCTCCCTCCCGATCCCGGAGCC ATC TGG CCC CTG GCC
 CTA GTG ATC GCC TCC CTG ACC TTG GCC TTG TCA GGA G...gtaaga...intron 1.....ttaccag
 L V I A G A L A L A L S G
 GT GTC TCC CAG GAG TCT TCC AAG GTT CTC AAC ACC AAT GGG ACC AGT GGG TTT
 G V S Q E S S K V L N T N G T S G F
 CTC CCA GGT GGC TAC ACC TGC TTC CCC CAC TCT CAG CCC TGG CAG GCT GCC
 L P G Y T A T C F P H S Q P W Q A A
 CTA CTA GTG CAA GGG CGG CTA CTC TGT GGG GGA GTC CTG GTC CAC CCC AAA
 L L V Q G R L L C G G V L V H P K
 TGG GTC CTC ACT GCC GCA CAC TGT CTA AAG GA gtagtg...intron 2.....cacag G GGG
 W V L T A A H C L K
 CTC AAA GTT TAC CTA GGC AAG CAC GCC CTA GGG CGT GTG GAA GCT GGT GAG
 L K V Y L G K H A L G R V E A G E
 CAG GTG AGG GAA GTT GTC CAC TCT ATC CCC CAC CCT GAA TAC CGG AGA AGC
 Q V R E V R V C L Y I P H P E Y R R S
 CCC ACC CAC CTG AAC CAC GAC CAT GAC ATC ATG CTT CTG GAG CTG CAG TCC
 P T H L N H D H D I M L L E L Q S
 CCG GTC CAG CTC ACA GGC TAC ATC CAA ACC CTG CCC CTT TCC CAC AAC AAC CGG
 P V Q L T G T Q T L Q T L P L S H N R
 CTA ACC CCT GGC ACC ACC TGT CGG GTG TCT GGC TGG GGC ACC ACC ACC AGC
 L T P G T T C T R V S G W G T T T S
 CCC CAG G gtagca...intron 3.....tcacc ag TG AAT TAC CCC AAA ACT CTA CAA TGT GCC
 P Q V N Y P K T L C A A
 AAC ATC CAA CTT CGC TCA GAT GAG GAG TGT CGT CAA GTC TAC CCA GGA AAG
 N I Q L R S D E E C R Q V Y P G K
 ATC ACT GAC AAC ATG TTG TGT GCC GGC ACA AAA GAG GGT GGC AAA GAC TCC
 I T D N M L C A T T K E G G K D S
 TGT GAG gtagca...intron 4.....aactgc GGT GAC TCT GGG GGC CCC CTG GTC TGT AAC
 C E G D S G G P L V C N
 AGA ACA CTG TAT GGC ATC GTC TCC TGG GGA GAC TTC CCA TGT GGG CAA CCT
 R T L Y G I V S W G D F P C G Q P
 GAG CGG CCT GGT GTC TAC ACC CGT GTC TCA AGA TAC GTC CTG TGG ATC CGT
 D R P G V Y T R V S R Y V L W I R
 GAA ACA ATC CGA AAA TAT GAA ACC CAG CAA AAA TGG TTG AAG GGC CCA
 E T I R K Y E T Q Q Q K W L K G P
 CAA TAA AAGTTGAGAAATGACCCGGCITCCATCCTGTGACCATGACTTCCTCAC
 Q
 ATGGTCTGCTTAGCCCTCTCTGCTCCTTATCCCAAGTGTTCATTTGAACCAAGTATCCATGTC
 CTGAAAAATGCTCAATCTCAGCTAACATTCCATGTTTCAGAAGCAATTCAGGCACTGCCAGGCT
 TGCAGTCTCCAGATGTTGATCCCTGAAACATCTCAACAACCTGAATGTCCTCAACCCAGACA
 ATGGCCAGGTCTCTCAACTTCATCAGTGTGGCTTCTATGAGCCAGATCACCACCTGAACGT
 TCTGTCTGGCCACATCTTAAATATTTCCATCAGCCCATCTCAACAATATATGTCTCTATAAAT
 GGACCATCCTTGACA

FIG. 4. Genomic organization and partial genomic sequence of the *KLK-L4* gene. Intronic sequences are not shown except for the splice junction areas. Introns are shown with lowercase letters, and exons are shown with capital letters. For the full sequence, see GenBank™ accession AF135024. The start and stop codons are circled, and the exon-intron junctions are underlined. The translated amino acids of the coding region are shown underneath by a single letter abbreviation. The catalytic residues are boxed. The putative polyadenylation signal is underlined.

structure, a PCR reaction was performed using one reverse primer (L4-R1) together with another forward primer that is located in the predicted first exon (primer L4-X1) (Table I and Fig. 1). Two main PCR bands were obtained from the tissues examined, the expected 819-bp band (predominant) and an additional minor band of about 650 bp (Fig. 3). Cloning and sequencing of these two bands revealed that the gene exists in two main forms in these tissues; the long form (our GenBank™ submission AF135024) and another form (referred to as the short *KLK-L4* variant) that utilizes an upstream alternative splice donor site located inside exon 3 thus creating an mRNA product that is 214 bp shorter. This alternative splice site causes frame shifting of the coding region that will generate a predicted stop codon at the beginning of exon 4 giving rise to a truncated protein product that does not contain the serine residue of the catalytic triad (Figs. 3 and 4).

Aligning the long *KLK-L4* form with the ESTs (Fig. 1) demonstrated that all ESTs utilize a different splice donor site located 80 bp downstream from the end of exon 3. These additional 80 bp contain an in-frame stop codon at nucleotide position 5505, which will lead to the formation of a shorter polypeptide product. They also utilize an alternative polyadenylation signal located at position 8706 (numbers refer to our Gen-

Bank™ submission AF135024). The clone from the German Genome Project utilizes another splice donor site that is located further downstream, inside intron 3 and ends up with a poly(A) tail without having a fourth or fifth exon. The same stop codon (position 5505) will be in-frame, and therefore, a truncated protein product is predicted to be formed (Fig. 1).

To obtain the 3'-end of the gene, a 3'-RACE reaction was performed, and an additional 375-bp fragment of 3'-untranslated region, downstream from PCR primer L4-R1, was obtained. This fragment was further confirmed to be present in all tissues tested by performing a PCR reaction using primers L4-F1 and L4-R3 (Table I and Fig. 1). This fragment ends with a putative polyadenylation signal variant (TATAAA).

Structural Characterization of the *KLK-L4* Gene and Its Protein Product—The long form of the *KLK-L4* gene is presented in Fig. 4. *KLK-L4* is formed of five coding exons and four intervening introns, spanning an area of 8905 bp of genomic sequence on chromosome 19q13.3-q13.4. The lengths of the coding regions are 52, 187, 269, 137, and 189 bp, respectively. The predicted protein coding region of the gene is formed of 831 bp, encoding a deduced 277-amino acid protein with a predicted molecular mass of 30.6 kDa (Fig. 4). The intron/exon splice sites (*mGT*. . . *AGm*, where *m* is any base) and their flanking sequences are in agreement with the consensus splice site sequence (20). A potential translation initiation codon is present at position 45 of the predicted first exon (numbers refer to our GenBank™ submission AF135024). The cDNA extends at least 382 bp further downstream from the stop codon, and a putative polyadenylation signal (TATAAA) is present at the end of this region (Fig. 4).

Hydrophobicity analysis revealed that the amino-terminal region is quite hydrophobic (Fig. 5), consistent with the possibility that this region may harbor a signal sequence, analogous to other serine proteases. Fig. 5 also shows the presence of several evenly distributed hydrophobic regions throughout the *KLK-L4* polypeptide, which are consistent with a globular protein, similar to other serine proteases (6). Computer analysis of the amino acid sequence of *KLK-L4* predicted a cleavage site between amino acids 20 and 21 (GVS-QE). Sequence homology with other serine proteases (Fig. 6) predicted another potential cleavage site (Lys²⁵) in close proximity. Most other kallikreins are activated by cleavage after arginine or lysine (data not shown). Thus, although the protein product has not as yet been directly characterized, it is very likely to be a secreted protein. The dotted region in Fig. 6 indicates an 11-amino acid loop characteristic of the classical kallikreins (PSA, *KLK1*, and *KLK2*) that is not found in *KLK-L4* or other members of the kallikrein multigene family (5–7, 15, 16, 21).

Sequence analysis of eukaryotic serine proteases indicates the presence of 29 invariant amino acids (22). Twenty eight of them are conserved in the *KLK-L4* protein and the remaining amino acid (Gln¹⁸² instead of Pro) is not conserved among all other kallikreins (Fig. 6). Ten cysteine residues are present in the putative mature *KLK-L4* protein. These are conserved in all the serine proteases that are aligned in Fig. 6 and would be expected to form disulphide bridges.

The presence of aspartate in position 239 suggests that *KLK-L4* will possess a trypsin-like cleavage pattern, similarly to most of the other kallikreins (e.g. *KLK1*, *KLK2*, TLSP, neuropsin, zyme, prostate, and enamel matrix serine proteinase) but different from PSA, which has a serine residue in the corresponding position and is known to have chymotrypsin-like activity (Fig. 6) (23, 24).

Mapping and Chromosomal Localization of the *KLK-L4* Gene—Alignment of the *KLK-L4* gene and the sequences of other known kallikrein genes within the 300-kb area of interest

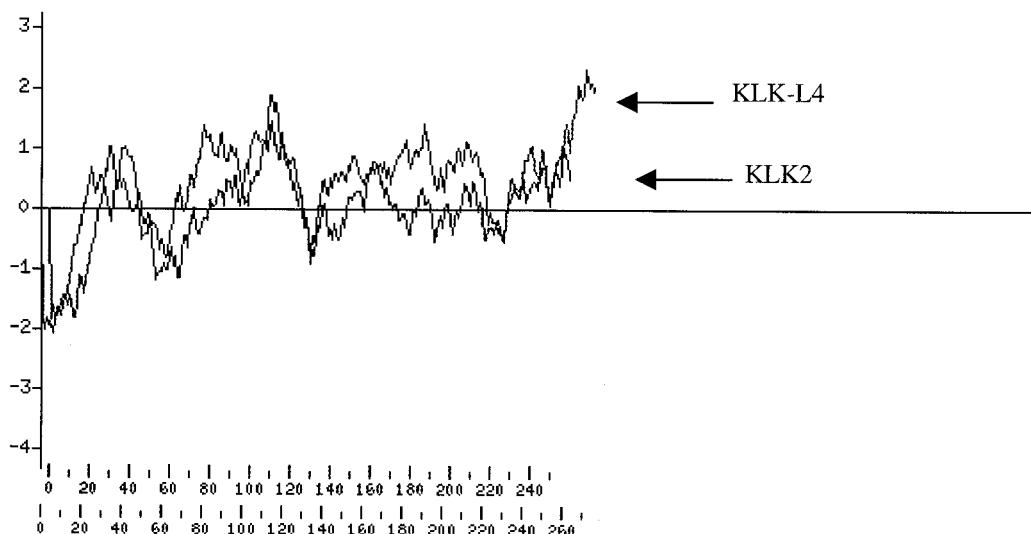


FIG. 5. Plot of hydrophobicity and hydrophilicity of the *KLK-L4* protein, as compared with the glandular kallikrein gene 2 (*KLK2*). Note the hydrophobic region at the amino terminus, suggesting presence of a signal peptide. For details, see text.

(the human kallikrein gene family locus) enabled us to precisely localize all known genes and to determine the direction of transcription, as shown by the arrows in Fig. 7. The PSA gene lies between *KLK1* and *KLK2* genes and is separated by 13,319 bp from *KLK2*, and both genes are transcribed in the same direction (centromere to telomere). All other kallikrein-like genes are transcribed in the opposite direction. *KLK-L4* is 13 kb centromeric from *KLK-L6* (GenBankTM accession AF161221), and 21 kb more telomeric to *KLK-L5* (GenBankTM accession AF135025).²

Homology with the Kallikrein Multigene Family—Alignment of the amino acid sequence of the *KLK-L4* protein (long form) against the GenBankTM data base and the known kallikreins, using the BLAST algorithm (18), indicated that *KLK-L4* has 51% amino acid sequence identity with the TLSP and zyme genes, 49% identity with *KLK-L2*, and 47 and 45% identity with PSA and *KLK2* genes, respectively. A multiple alignment study shows that the typical catalytic triad of serine proteases is conserved in the *KLK-L4* gene (His¹⁰⁸, Asp¹⁵³, and Ser²⁴⁵), 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 (WLLTAAHC), serine (GDSGGP), and aspartate (DLMLI)) (Fig. 6) (2, 3, 5–7, 15–16). In addition, several other residues were found to be fully or partially conserved among the human kallikrein gene family, as further shown in Fig. 6. To predict the phylogenetic relatedness of the *KLK-L4* 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. 8). Phylogenetic analysis separated the classical kallikreins (*KLK1*, *KLK2*, and PSA) and grouped *KLK-L4* with zyme, TLSP, *KLK-L3*, neuropsin, and NES1 genes, consistent with previously published studies (8, 24, 25) and indicating that this group of genes probably arose from a common ancestral gene by duplication.

Tissue Expression and Hormonal Regulation of the *KLK-L4* Gene—As shown in Fig. 2, the *KLK-L4* gene is primarily expressed in mammary gland, prostate, salivary gland, and testis, but as is the case with other kallikreins, lower levels of expression are found in many other tissues. To verify the RT-PCR specificity, the PCR products were cloned and sequenced.

A steroid hormone receptor-positive breast cancer cell line (BT-474) was used as a model to verify whether the *KLK-L4*

gene is under steroid hormone regulation. PSA was used as a control gene, known to be up-regulated by androgens and progestins, and pS2 was used as an estrogen up-regulated control gene in the same cell line. Our preliminary results indicate that *KLK-L4* is up-regulated by progestins and androgens and to a lower extent by estrogens (Fig. 9).

Expression of *KLK-L4* in Breast Cancer Tissues and Cell Lines—To characterize the extent and frequency of expression of the *KLK-L4* gene in breast tumors, we used cDNA derived from 3 normal and 19 malignant breast tissues and 3 breast cancer cell lines. The data were interpreted by comparison of band intensities. Out of the 19 tumors, *KLK-L4* gene expression was undetectable in 7, lower than normal tissues in 9, comparable to the normal tissues in 1, and higher than normal tissues in 2 tumors. Without hormonal stimulation, the BT-474 and T-47D cell lines had no detectable *KLK-L4* mRNA, whereas the MCF-7 cell line was positive (data not shown). These preliminary results suggest that this gene is down-regulated in the majority (16/19) of breast tumors.

DISCUSSION

Kallikreins are a subgroup of serine proteases traditionally defined by their ability to release vasoactive peptides (kinins) from kininogens (23). However, it is now recognized that in both humans and rodents, kallikreins exhibit a variety of functions in different tissues. *KLK-L4* is defined as a kallikrein-like gene based on the criteria of structural homology and chromosomal localization (25). Irwin *et al.* (26) proposed that the serine protease genes could be classified into five different groups according to intron position. The established kallikreins (*KLK1*, *KLK2*, and PSA), trypsinogen, and chymotrypsinogen belong to a group that has: (1) an intron just downstream from the codon for the active site histidine residue, (2) a second intron downstream from the exon containing the codon for the active site aspartic acid residue, and (3) a third intron just upstream from the exon containing the codon for the active site serine residue. Fig. 10 shows that *KLK-L4* meets the above mentioned criteria; moreover, it is located in close proximity to other kallikrein genes on the chromosomal locus 19q13.3-q13.4 (Fig. 7).

Our preliminary finding, supporting that the *KLK-L4* gene may be down-regulated in a subset of breast cancers, is not surprising. There is now growing evidence that many of the kallikreins and kallikrein-like genes that are clustered in the

	1	15	16	30	31	45	46	60	61	75	76	90			
1	KLK-L1/prostase	---MATAGN	PWGWFLG---	YLIL	GVAG-----	---	SLVSG-----	---	SCSQ	IINGEDCS	PHSQPWQ	45			
2	KLK-L2	---MATARP	PWMWVLCALITALLL	GVTEHVLANN	NDVSCD	HPSENTVP	SGSNQDLG	AGAGEDARS	DDSSSR	IINGSDCD	DMHTQPWQ	81			
3	PSA	---	MWVPVVF---	TLSP	---	---	AAPLI-----	---	L---	SR	IVGWECEKHSQPWQ	39			
4	KLK2	---	MWDLVLS---	IALLSV	GCTG-----	---	AVPLI-----	---	Q---	SR	IVGGWECEKHSQPWQ	39			
5	KLK1	---	MWFLVLC---	LALLSL	GGTG-----	---	AAPPI-----	---	Q---	SR	IVGWECEQHSQPWQ	39			
6	trypsinogen	---	MNPLLI---	LTFAVA	AALA-----	---	APFDD-----	---	D---	DK	IVGGYNCENSVYQ	38			
7	zyme/protease M	---	MKKLMV---	VLSL	IAAA-----	---	WAEQNK-----	---	---	---	LVHGGPCDKTSHPYQ	36			
8	KLK-L4	---	MW	PLALVIA---	SLTL	ALSG-----	---	GVSQESSKV---	---	LNTMGTSGF	LPGGYTCFPHSQPWQ	50			
9	TLSP	---	---	ILQLILLALATGLVG	GETR-----	---	I-----	---	---	---	IKGFCEKPHSQPWQ	36			
10	neuropsin	---	MGRPRPRA	AKTWMFLLLLGGAWA	GHSR-----	---	AQEDK-----	---	---	---	VLGGHECQPHSQPWQ	47			
11	NES1	---	MRAPHLHLSAASGAR	ALAKLLPLLMAQLWA	AEAA-----	---	LLPQN-----	---	---	DTRLDP	EAYGAPCARGSQPWQ	60			
							▲	▲							
		91	105	106	120	121	135	136	150	151	165	166	180		
1	prostase	AALVM-ENELFC	SGV	LVHPQWVLSA	AHCFCQ	NSYTTIGLGLH	SLSEAD	QEPGSMQVEAS	LSVR	HPEYNRPLLAN	---	---	DLMLIKLD	123	
2	KLK-L2	AALLLRPNQ	LYCGAV	LVHPQWVLSA	AHCRCR	KVFRVRLGHYS	LSPLV	YESGQMFQGV	KSIP	HPGYSHPGHSN	---	---	DLMLIKLN	160	
3	PSA	VLVAS-RGRA	VC	LVHPQWVLSA	AHCIR	NKSVILLGRHS	LFP	-EDTGQVFPV	SHSFP	HPLYDMSLLK	NRFLR	PGDSSSHD	MLLRLS	127	
4	KLK2	VAVYS-HGWA	HCGGV	LVHPQWVLSA	AHCCLK	KNSQWVLRG	RNLFP	-EDTGRVPSV	SHSFP	HPLYNMSLLK	HQSLR	PDEDSSH	DLMLLRLS	127	
5	KLK1	AALYH-FSTF	QCGGI	LVHPQWVLSA	AHCIS	DNYQLWLGRH	NLFDD	-ENTAQFVHV	SEFP	HPGFNMSLLEN	HTRO	ADEYSHD	MLLRLT	127	
6	trypsinogen	VSLNS--GYH	FCGGS	LINEQWVVS	AHCYK	SRIQVRLGEH	NIEVL	-EGNEQFINA	AKIIR	HPQYDRKTLN	N	---	DLMLIKLS	114	
7	zyme	AALYT-SGHL	CCGV	LHPLWVLSA	AHCCK	PNLQVFLGKH	NLRQR	-ESSQESSV	VRAVI	HPDYDA	---	---	SHDQD	IMLLRLA	113
8	KLK-L4	AALLV-QGRLL	CCGV	LVHPQWVLSA	AHCCLK	EGLKVYLK	HALGRV	-EAGEQVRE	VHSIP	HPEYRRSP	THL	---	NHDH	IMLLELQ	131
9	TLSP	AALFE-KTRLL	CCGAT	LIAPRWVLSA	AHCCLK	PRYIVHLG	QHNLQKE	-EGCEQTRT	ATESFP	HPGFNNSL	PNK	---	DHRN	DMVMA	117
10	neuropsin	AALFQ-GQQL	CCGV	LVGGNWLTA	AHCCK	PKYTVRLGD	HSLQNK	-DGPEQEI	PVVQSIP	HPCYNSSD	VE	---	DHNH	DMMLQLR	127
11	NES1	VSLFN-GLSF	HCAVG	LVDQSWVLSA	AHCCKG	KPLWARVGD	DHLLLL	Q--GEQLRR	TRTSV	HPKYHQGS	GPILPR	R--TDEH	DMMLKL	144	
		181	195	196	210	211	225	226	240	241	255	256	270		
1	prostase	ESVS-ESDT	TIRSISI	ASQCPTAG	NSCLVS	GWGLLANG--	RMPTV	LQCVNVSVV	SEEVCS	KLVDPLYH	PSMFCAG	GGHDQK	DSCNGD	SGG	209
2	KLK-L2	RRIR-PTK	DVRPINV	SSHCP	SAG-TKCLVS	GWGTTKSP	QVHFPSK	LQCLNLSVLS	QKRCE	DAYPRQI	DDTMFCAG	DKAG-RD	SCGDSGG	247	
3	PSA	EPAE-LTD	AVKMDL	PTQEPALG	-TTCYAS	GWGSI	EEFLTPPK	LQCVLDLHV	ISNDVCA	QVHPQKV	TKFMLCAG	RWTGGK	STCSGD	SGG	215
4	KLK2	EPAK-ITD	VVKVGLG	PTQEPALG	-TTCYAS	GWGSI	EEFLRPRS	LQCVSLHLL	SNDMCA	RAYSEK	VTEFMLCAG	LWTGGK	DTCCGDSGG	215	
5	KLK1	EPADTITD	AVKVEL	PTEEP	EVG-STCLAS	GWGSI	EPENFSFDD	LQCVLDL	KILPNDECK	KAHVQKV	TFDMLCVG	HLEGGK	DCVGD	SGG	216
6	trypsinogen	SRAV-INAR	VSTISL	PTAPPATG	-TKCLIS	GWGNT	ASSGADYPDE	LQCLDPA	VLVSQAKCE	ASYPGK	ITSNMFCVG	FLEGGK	DCSCGDSGG	202	
7	zyme	RPAK-LSE	LIQPLPL	ERDCSANT	-TSCHIL	GWGKTADG	--DPDPT	IQCAYIHL	VSRREECE	HAYPGQ	ITQNMMLCAG	DEKYGK	DCSCGDSGG	199	
8	KLK-L4	SPVQ-LT	GIQTLPL	SHNRLT	PGTTCRVS	GWGTT	SPQVNYPKT	LQCANI	QLRSDEECR	QVYPGK	ITDMLCAG	TEKEGK	DCSCGDSGG	220	
9	TLSP	SPVS-IT	WAVRPLTL	SSRCVTAG	-TSCLIS	GWGST	SSPQLRLPHT	LRCANIT	IEHQKCE	NAYPGN	ITDMLCAG	VQEGGK	DCSCGDSGG	205	
10	neuropsin	DQAS-LG	SKVKPISL	ADHCTQPG	-QKCTVS	GWGTV	TSPRENFPDT	LNC	AEVKIPQK	KCE	DAYPGQ	ITDGMV	CAGSSKG-ADTC	CGDSGG	214
11	NES1	RPVVP-GPR	VRALQYL	PYRCAQPG	-DQCQA	GWGTTA	ARRVKYKNG	LTCS	SITILSPKECE	VFYPGV	VVTTNMLCAG	LDRG-QDPC	QSDSGG	231	
		271	285	286	300	301	315	316	330	331					
1	prostase	PLVCG	NGYLOGLV	SFG	KAPCGQV	GVPGVYTN	LCKFT	EWIEKTVQAS	-----	254					
2	KLK-L2	FVVCN	GSLOGLV	SWG	DYPCARP	NRPGVYTN	LCKFT	KWIKETIQAN	S-----	293					
3	PSA	PLVCG	NGVLOGIT	SWG	SEPCAL	PERPSLYTK	VVHYR	KWIKDTIVAN	P-----	261					
4	KLK2	PLVCG	NGVLOGIT	SWG	PEPCAL	PEKPAVYTK	VVHYR	KWIKDTIAAN	P-----	261					
5	KLK1	PLMCD	GVLOQV	TSWG	YVPCGT	PNKPSVAVR	VLSYV	KWIEDTIAEN	S-----	262					
6	trypsinogen	FVVCN	NGLOQV	VSWG	-DGA	CAKPKGVYTK	VVNYV	KWIKNTIAAN	S-----	247					
7	zyme	PLVCG	DHLRGLV	SWG	NIPCGS	KEKPGVYTN	VCRYT	NWIKTIQAK	-----	244					
8	KLK-L4	PLVCG	NRTLYGLV	SWG	DFPCQ	QDRPGVYTR	VSRYL	VWIRETIRKY	ETQQQ	KWLK	RGPO	277			
9	TLSP	PLVCG	NQSLQGI	ISWG	QDPCA	ITRPGVYTK	VCKYV	DWIQETMKNN	-----	250					
10	neuropsin	PLVCG	DGALQGIT	SWG	SDPCGR	SDKPGVYTN	ICRYL	DWIKKIIGSK	G-----	260					
11	NES1	PLVCG	DETLOGLV	SWG	VYPCG	SAQHPAVYTK	ICKYMS	WINKVKRSR	-----	276					

FIG. 6. Alignment of the deduced amino acid sequence of *KLK-L4* with members of the kallikrein multigene family. Genes are (from top to bottom and the GenBank™ accession numbers are in parentheses): *KLK-L1/prostase* (AAD21581), enamel matrix serine proteinase 1 (*EMSP*) (NP 004908), *KLK-L2* (AF135028), *PSA* (P07288), *KLK2* (P20151), *KLK1* (NP 002248), trypsinogen (P07477), *zyme* (Q92876), *KLK-L4* (AF135024), *TLSP* (BAA33404), *KLK-L3* (AF135026), neuropsin (BAA28673), and *NES1* (O43240). Dashes represent gaps to bring the sequences to better alignment. The residues of the catalytic triad are typed in bold, and conserved motifs around them are highlighted in gray. The 29 invariant serine protease residues are denoted by ●, and the cysteine residues are denoted by ◆. The predicted cleavage sites are indicated by ▲. The dotted area represents the kallikrein loop sequence. The trypsin-like cleavage pattern of *KLK-L4* with the Asp residue, is indicated by a star.

same chromosomal region (Fig. 7) are related to malignancy. PSA is the best marker for prostate cancer so far (1). A recent report provided evidence that PSA has antiangiogenic activity, and that this activity may be related to its function as a serine protease (27). This study suggested that other serine proteases, including new members of the kallikrein multigene family of enzymes, should also be evaluated for potential antiangiogenic activity (27). Recent reports suggest that human glandular kallikrein 2 (encoded by the *KLK2* gene) could be another useful diagnostic marker for prostate cancer (10, 28). *NES1* appears to be a tumor suppressor gene (11). The protease M gene was shown to be differentially expressed in primary breast and ovarian tumors (4), and the human stratum corneum chymotryptic enzyme has been shown to be expressed at abnormally high levels in ovarian cancer (29). Another recently identified kallikrein-like gene, located close to *KLK-L4* and tentatively named tumor associated, differentially expressed

gene-14 (*TADG14*) (an alternatively spliced form of neuropsin, see Fig. 7) and was found to be overexpressed in about 60% of ovarian cancer tissues (29). Also, prostase/*KLK-L1*, another newly discovered kallikrein-like gene, is speculated to be linked to prostate cancer (8). Thus, extensive new literature suggests multiple connections of many kallikrein genes to various forms of human cancer. This subject has recently been reviewed (30). The removal of intervening RNA sequences (introns) from the premessenger RNA in eukaryotic nuclei is a major step in the regulation of gene expression (31). RNA splicing provides a mechanism whereby protein isoform diversity can be generated, and the expression of particular proteins with specialized functions can be restricted to certain cell or tissue types during development (31). The sequence elements in the pre-mRNA at the 5'- and 3'-splice sites in metazoans have very loose consensus sequence; only the first and the last two bases (GT..AG) of the introns are highly conserved (32). These sequences cannot

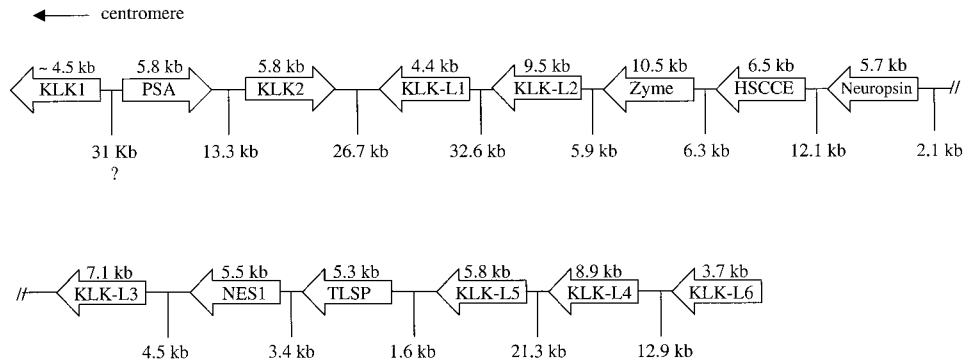


FIG. 7. An approximate 300-kb region of almost contiguous genomic sequence around chromosome 19q13.3-q13.4. Genes are represented by horizontal arrows denoting the direction of the coding sequence. Their lengths are shown on top of each arrow. Distances between genes are mentioned in base pairs below the arrows. The distance between *KLK1* and *PSA* is not accurately known. For gene names, see abbreviations. This figure is not drawn to scale.

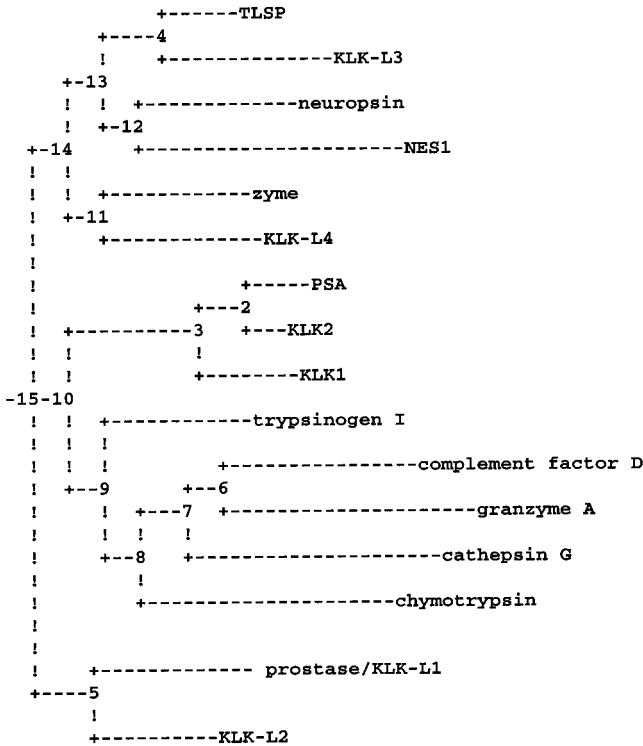


FIG. 8. Dendrogram of the predicted phylogenetic tree for some kallikrein and serine protease genes. The neighbor-joining/UPGMA method was used to align *KLK-L4* with other serine proteases and members of the kallikrein gene family. The tree grouped the classical kallikreins (*KLK1*, *KLK2*, and *PSA*) together and aligned the *KLK-L4* gene in one group with *zyme*, *NES1*, *neuropsin*, *KLK-L3*, and *TLSP*. Other serine proteases were aligned in different groups, as shown.

be the sole determinants of splice site selection, because identical, but not ordinarily active, consensus sequences can be found within both exons and introns of many eukaryotic genes. Other protein factors and sequences downstream of the splice sites are also involved.

The existence of multiple splice forms is frequent among kallikreins. Distinct RNA species are transcribed from the *PSA* gene, in addition to the major 1.6-kb transcript (33). Several distinct *PSA* transcripts have been described by Reigman *et al.* (34, 35). Interestingly, one of these clones lacks the 3'-untranslated region, the first 373 nucleotides of the open reading frame, and has an extended exon that contains a stop codon, a pattern that is comparable with some alternative forms of the *KLK-L4* cDNA, as described here (Fig. 1). Heuze *et al.* (33) reported the cloning of a full-length cDNA corresponding to a

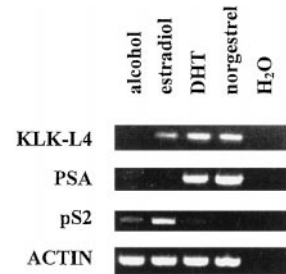


FIG. 9. Hormonal regulation of the *KLK-L4* gene in the BT-474 breast carcinoma cell line. Steroids were added at 10⁻⁸ M final concentrations. Actin (not regulated by steroid hormones), *pS2* (up-regulated by estrogens), and *PSA* (up-regulated by androgens and progestins) are control genes. *KLK-L4* is up-regulated by androgens and progestins and to a lesser extent by estrogens. H₂O was used to check for PCR specificity in all PCR reactions. For more details, see text. *DHT*, dihydrotestosterone.

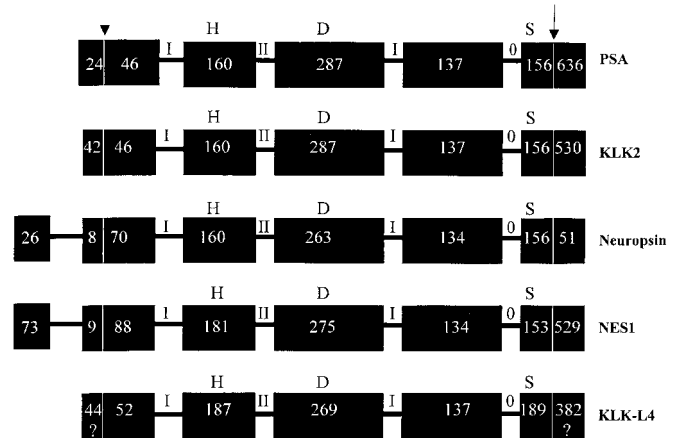


FIG. 10. Schematic diagram showing the comparison of the genomic structure of *PSA*, *KLK2*, *neuropsin*, *NES1*, and *KLK-L4* genes. Exons are shown by black boxes, and introns are shown by the connecting lines. The arrowhead shows the start codons, and the arrow shows the stop codons. Letters above boxes indicate the relative positions of the amino acids of the catalytic triad; H denotes histidine, D denotes aspartic acid, and S denotes serine. Roman numerals indicate intron phases. The intron phase refers to the location of the intron within the codon; I, the intron occurs after the first nucleotide of the codon; II, the intron occurs after the second nucleotide; 0, the intron occurring between codons. Numbers inside boxes indicate exon lengths in base pairs. The question mark indicates the possibility of more untranslated bases. This figure is not drawn to scale.

2.1-kb *PSA* mRNA. This form results from the alternative splicing of intron 4 and lacks the serine residue that is essential for catalytic activity. Also, Reigman *et al.* (36) reported the identification of two alternatively spliced forms of the human

glandular kallikrein 2 (*KLK2*) gene. A novel transcript of the tissue kallikrein gene (*KLK1*) was also isolated from the colon (37). Interestingly, this transcript lacks the first two exons of the tissue kallikrein gene, but the last three exons were fully conserved, a pattern that is similar to our findings with some ESTs containing parts of the *KLK-L4* gene (Fig. 1). Neuropsin, a recently identified kallikrein-like gene, was found to have two alternatively spliced forms, in addition to the major form (29, 38). Here, we describe the cloning of the *KLK-L4* gene and the identification of a number of alternative mRNA forms. These forms may result from alternative splicing (32), a retained intronic segment (34), or from the utilization of an alternative transcription initiation site (37). Because the long form of *KLK-L4* and the major alternative splice variant (short *KLK-L4* variant) (Fig. 3) have an identical 5'-sequence required for translation, secretion, and activation, it is possible to assume that both mRNAs encode for a secreted protein (33).

To investigate the relative predominance of the long *KLK-L4* and related forms, cDNA from various tissues was amplified by PCR. Although, in general, it is difficult to use PCR for quantitative comparisons between mRNA species, in this experiment (mRNAs of comparable sizes, using one set of primers under identical conditions), such a comparison is reasonable (36). In all five normal tissues examined (breast, prostate, testis, salivary gland, and uterus) the long form of *KLK-L4* was the predominant with minimal level of expression of the short form (Fig. 3).

The presence of alternatively spliced forms may be related to malignancy. Recent literature suggests that distinct molecular forms of PSA could be expressed differently by malignant versus benign prostate epithelium (39). Aberrant PSA mRNA splicing in benign prostatic hyperplasia, as opposed to prostate cancer, has been described by Henttu *et al.* (40). In addition, it has been recently postulated that different prostatic tissues potentially harboring occult cancer could account for the presence of various forms of PSA (39). Clearly, the alternatively spliced forms of *KLK-L4* should be examined and compared in detail, in various normal and malignant tissues.

REFERENCES

- Diamandis, E. P. (1998) *Trends Endocrinol. Metab.* **9**, 310–316
- Schedlich, L. J., Bennetts, B. H., and Morris, B. J. (1987) *DNA (N. Y.)* **6**, 429–437
- Evans, B. A., Yun, Z. X., Close, J. A., Tregear, G. W., Kitamura, N., Nakanishi, S., Callen, D. F., Baker, E., Hyland V. J., and Sutherland, G. R., *et al.* (1988) *Biochemistry* **27**, 3124–3129
- Anisowicz, A., Sotiropoulou, G., Stenman, G., Mok, S. C., and Sager, R. (1996) *Mol. Med.* **2**, 624–636
- Little, S. P., Dixon, E. P., 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
- Liu, X. L., Wazer, D. E., Watanabe, K., and Band, V. (1996) *Cancer Res.* **56**, 3371–3379
- Yoshida, S., Taniguchi, M., Hirata, A., and Shiosaka, S. (1998) *Gene (Amst.)* **213**, 9–16
- Nelson, P. S., Gan, L., Ferguson, C., Moss, P., Gelinis, R., Hood, L., and Wang, K. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 3114–3119
- Hansson, L., Stomqvist, M., Backman, A., Wallbrandt, P., Carlstein, A., and Egelrud, T. (1994) *J. Biol. Chem.* **269**, 19420–19426
- Stenman, U. H. (1999) *Clin. Chem.* **45**, 753–754
- Goyal, J., Smith, K. M., Cowan, J. M., Wazer, D. E., Lee, S. W., and Band, V. (1998) *Cancer Res.* **58**, 4782–4786
- Evans, B. A., Drinkwater, C. C., Richards, R. I. (1987) *J. Biol. Chem.* **262**, 8027–8034
- Ashley, P. L., MacDonald, R. J. (1985) *Biochemistry* **24**, 4520–4527
- 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
- Yousef, G., Obiezu, C., Luo, L., Black, M. H., and Diamandis, E. P. (1999) *Cancer Res.* **59**, 4252–4256
- Yousef, G. M., Luo, L. Y., and Diamandis, E. P. (1999) *Anticancer Res.* **19**, 2843–2852
- Yousef, G., and Diamandis, E. P. (1999) *J. Biol. Chem.* **274**, 37511–37516
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402
- Lennon, G., Auffray, C., Polymeropoulos, M., and Soares, M. B. (1996) *Genomics* **33**, 151–152
- Iida, Y. (1990) *J. Theor. Biol.* **145**, 523–533
- Yousef, G., Luo, L., Scherer, S., Sotiropoulou, G., and Diamandis, E. P. (1999) *Genomics* **62**, 251–259
- Dayhoff, M. O. (1978) *Natl. Biomed. Res. Found.* **5**, Suppl. 3, 79–81
- Clements, J. (1997) in *The Kinin System* (Farmer, S., ed) pp. 71–97, Academic Press, New York
- 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
- Mason, A. J., Evans, B. A., Cox, D. R., Shine, J., and Richards, R. I. (1983) *Nature* **303**, 300–307
- Irwin, D. M., Robertson, K. A., and MacGillivray, R. T. (1988) *J. Mol. Biol.* **212**, 31–45
- Fortier, A. H., Nelson, B. J., Grella, D. K., and Holaday, J. W. (1999) *J. Natl. Cancer Inst.* **91**, 1635–1640
- Black, M. H., Magklara, A., Obiezu, C. V., Melegos, D. N., and Diamandis, E. P. (1999) *Clin. Chem.* **45**, 790–799
- Underwood, L. J., Tanimoto, H., Wang, Y., Shigemasa, K., Parmley, T. H., and O'Brien, T. J. (1999) *Cancer Res.* **59**, 4435–4439
- Diamandis, E. P., Yousef, G. M., Luo, L. Y., Magklara, A., and Obiezu, C. V. (2000) *Trends Endocrinol. Metab.* **11**, 54–60
- Adams, M. D., Rudner, D. Z., and Rio, D. C. (1996) *Curr. Opin. Cell Biol.* **8**, 331–339
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Heuze, N., Olayat, S., Gutman, N., Zani, M. L., and Courty, Y. (1999) *Cancer Res.* **59**, 2820–2824
- Riegman, P. H., Vlietstra, R. J., van der Korput, J. A., Romijn, J. C., and Trapman, J. (1989) *Biochem. Biophys. Res. Commun.* **159**, 95–102
- Riegman, P. H., Klaassen, P., van der Korput, J. A., Romijn, J. C., and Trapman, J. (1988) *Biochem. Biophys. Res. Commun.* **155**, 181–188
- Riegman, P. H., Vlietstra, R. J., van der Korput, H. A., Romijn, J. C., and Trapman, J. (1991) *Mol. Cell. Endocrinol.* **76**, 181–190
- Chen, L. M., Murray, S. R., Chai, K. X., Chao, L., and Chao, J. (1994) *Braz. J. Med. Biol. Res.* **27**, 1829–1838
- Mitsui, S., Tsuruoka, N., Yamashiro, K., Nakazato, H., and Yamaguchi, N. (1999) *Eur. J. Biochem.* **260**, 627–634
- Baffa, R., Moreno, J. G., Monne, M., Veronese, M. L., and Gomella, L. G. (1996) *Urology* **47**, 795–800
- Henttu, P., Luukkainen, O., and Vihko, P. (1990) *Int. J. Cancer* **45**, 654–660