Genomics **65,** 184–194 (2000)
doi:10.1006/geno.2000.6159, available online at http://www.idealibrary.com on IDEAL®

The Expanded Human Kallikrein Gene Family: Locus Characterization and Molecular Cloning of a New Member, KLK-L3 (KLK9)

George M. Yousef 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 December 23, 1999; accepted February 11, 2000

In rodents, kallikreins are encoded by a large multigene family but in humans, only three kallikrein genes were thought to exist. Based on the homology between the human and the rodent kallikrein loci, we defined a 300-kb human kallikrein gene region on chromosome 19q13.3–q13.4. By using linear sequence information, restriction analysis, PCR, and blotting techniques, we were able to construct the first detailed map of the human kallikrein gene locus. Comparative analysis of genes located in this area enabled us to expand the human kallikrein multigene family with some recently identified serine proteases and establish common structural features. We further identified a new kallikrein-like gene, named kallikrein-like gene 3 (KLK-L3; HGMW-approved symbol KLK9). We describe the structural characterization of the KLK-L3 gene, together with its precise chromosomal localization in relation to other kallikreins and its tissue expression pattern and hormonal regulation.

INTRODUCTION

The progressive development of genetic tools, together with the availability of powerful computer programs and the large amount of sequence information available from the Human Genome Project, has enabled the mapping of several loci responsible for genetic disease or the identification of areas where certain gene families are clustered.

Positional candidate cloning is a new approach for gene discovery that combines the knowledge of map position with the increasingly dense human transcript maps, the available expressed sequence tags (ESTs),[2] and the availability of candidate genes mapped to the same region (Ballabio, 1993). This method greatly expedites the search process and is already one of the predominant methods of gene discovery (Collins, 1995).

Serine proteases are a group of protein cleaving enzymes that contain a serine residue in their active site. Kallikreins are a subfamily of serine proteases, originally defined as enzymes cleaving vasoactive peptides (kinins) from kininogen (Schachter, 1980). 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 genes encoding tissue kallikreins (Clements, 1997). The tissue or glandular kallikreins are involved in posttranslational modification of many polypeptides and are central to many biological processes (Rittenhouse *et al.*, 1998).

In mouse and rat, the kallikreins are encoded by large multigene families. In the mouse genome, at least 24 genes have been identified (Evans *et al.*, 1987). The rat kallikrein gene family has been estimated to contain 20 members (Ashley and MacDonald, 1985). In humans, the kallikrein gene family locus is on chromosome 19q13.3–q13.4 (Riegman *et al.*, 1989a, 1992; Richards *et al.*, 1991), a region that is syntenic to the mouse kallikrein gene family locus on chromosome 7 (Nadeau *et al.*, 1991). The human kallikrein gene family is thought to have only three members: the tissue (pancreatic–renal) kallikrein (KLK1) (Evans *et al.*, 1988), the human glandular kallikrein (KLK2) (Schedlich *et al.*, 1987), and prostate-specific antigen (PSA or KLK3) (Riegman *et al.*, 1989b; Diamandis, 1998). Several early estimations of the size of this

[1] To whom correspondence and reprint requests 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: KLK, kallikrein; KLK-L, kallikrein-like; PCR, polymerase chain reaction; PSA, prostate-specific antigen; dNTPS, deoxynucleoside triphosphates; LLNL, Lawrence Livermore National Laboratory; RT, reverse transcription; EST, expressed sequence tag; TLSP, trypsin-like serine protease; HSCCE, human stratum corneum chymotryptic enzyme; NES1, normal epithelial cell-specific 1 gene; BAC, bacterial artificial chromosome; PAC, P1-derived artificial chromosome; RACE, rapid amplification of cDNA ends; BCM, Baylor College of Medicine; FISH, fluorescence *in situ* hybridization.

0888-7543/00 $35.00

family, using Southern blot analysis, predicted just 3–4 genes (Riegman *et al.*, 1989a; Schedlich *et al.*, 1987; Fukushima *et al.*, 1985; Baker and Shine, 1985). However, a Southern blot analysis using monkey cDNA suggested that it might contain as many as 19 genes (Murray *et al.*, 1990). Recently, a number of novel serine proteases were discovered and found to have significant similarities with the classical kallikrein genes. Also, a fragment of a putative human kallikrein gene adjacent to the KLK2 gene was identified on a human genomic cosmid clone (Stephenson *et al.*, 1999). This information suggested that the human kallikrein gene family might be larger than originally expected.

Based on mapping of the rodent kallikrein genes and the documented strong conservation between human chromosome 19q13.3–q13.4 and the 17 loci in a 20-cM proximal part of mouse chromosome 7 (Nadeau *et al.*, 1991; Saunders and Seldin, 1990), we identified a candidate genomic region for further analysis of the human kallikrein locus. In this paper, we describe the construction of the first detailed map of the human kallikrein gene locus and provide strong evidence that kallikreins comprise a large multigene family in humans. We further suggest that some of the recently cloned serine proteases can be considered members of this family. Also, we describe the cloning and characterization of a new kallikrein-like gene named kallikrein-like gene 3 (KLK-L3)[3] and present its precise chromosomal localization, tissue expression pattern, and hormonal regulation.

MATERIALS AND METHODS

Strategy for new gene discovery. We have obtained sequencing data for approximately 300 kb around chromosome 19q13.3–q13.4, from the Web site of the Lawrence Livermore National Laboratory (LLNL) (http://www-bio.llnl.gov/genome/genome.html). Different computer programs were used for putative new gene prediction, as previously described (Yousef *et al.*, 1999a).

Reverse transcriptase-polymerase chain reaction (RT-PCR) for KLK-L3 cDNA. Total RNA isolated from 26 different human tissues was purchased from Clontech (Palo Alto, CA). cDNA was prepared as described below and used for PCR amplification. A primer set (L3-F1 and L3-R1) was used to identify the presence of the gene in tissues, and the reverse primer (L3-R1) was used with another primer (L3-F2) to amplify and clone the full cDNA of the gene. These primer sequences are shown in Table 1. Tissue cDNAs were amplified at various dilutions so that different levels of expression of various tissues could be assessed semiquantitatively.

Two micrograms of total RNA was reverse-transcribed into first-strand cDNA using the Superscript preamplification system (Gibco BRL, Gaithersburg, MD). The final volume was 20 μl. Based on the combined information obtained from the predicted genomic structure of the new gene and the EST sequence (please see Results), two gene-specific primers (L3-F1 and L3-R1) were designed (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, 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.

TABLE 1

Primers Used for RT-PCR Analysis

Gene	Primer name	Sequence[a]
KLK-L3	L3-F1	CATGCAGTGTCTCATCTCAG
	L3-F2	CATGGAGGAGGAAGGAGATG
	L3-R1	CTTCGGCCTCTCTTGGTCTT
PSA	PSAS	TGCGCAAGTTCACCCTCA
	PSAAS	CCCTCTCCTTACTTCATCC
Actin	ACTINS	ACAATGAGCTGCGTGTGGCT
	ACTINAS	TCTCCTTAATGTCACGCACGA

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

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 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 100% ethanol were added into the culture medium, 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 total RNA extraction by the Trizol method (Gibco BRL). cDNA was prepared and amplified as described above. Control genes (PSA, pS2, and actin) were amplified as previously described (Yousef *et al.*, 1999c).

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.

Identification of positive PAC and BAC genomic clones from human genomic DNA libraries. The PCR product generated with primer set Z1S and Z1AS (Table 2) was purified and then labeled with ^{32}P by the random primer method (Sambrook *et al.*, 1989) and used as a probe to screen a human genomic DNA BAC library, spotted in duplicate on nylon membranes, for identification of positive clones. The filters were hybridized in 15% formamide, 500 mM Na_2HPO_4, 7% SDS, 1% BSA (w/v) at 65°C overnight, washed sequentially with 2× SSC, 1× SSC, 0.2× SSC, containing 0.1% SDS at 65°C, and then exposed to X-ray film as described (Sambrook *et al.*, 1989). Positive clones were obtained and plated on selective LB medium, and then a single colony was transferred into LB broth for overnight cultures. A PAC clone positive for NES1 was identified by a similar methodology as described elsewhere (Luo *et al.*, 1998). PAC and BAC libraries were constructed by de Jong and associates (see Osoegawa *et al.*, 1999). Purification of BAC and PAC DNA was performed by a rapid alkaline lysis miniprep method, which is a modification of the standard Qiagen-Tip method. Positive clones were further confirmed by Southern blot analysis as described (Sambrook *et al.*, 1989).

5′ Rapid amplification of cDNA ends (5′ RACE). According to the EST sequences and the computer-predicted structure of the

[3] The HGMW-approved symbol for the gene described in this paper is KLK9.

TABLE 2

Primers Used for Gene-Specific PCR Amplification of the Kallikrein Genes Using DNA as a Template

Primer name	Sequence[a]	Coordinates	GenBank Accession No.	Gene name
Z1S	GACCCTGACATGTGACATCTA	979–999	U62801	Zyme
Z1AS	GCCACTGCCTGATGGAGACTG	1,422–1,402		
GL3-F1	AACATCAGCATCCTGGAGAA	7,324–7,343	AF135026	KLK-L3
LL3-R1	CTTCGGCCTCTCTTGGTCTT	8,051–8,060		
L2-1	GGGTCAGAGCTGCAGAGAAG	11,104–11,123	AF135028	KLK-L2
L2-2	GGGCCTGTCGTCTGCAATGG	11,522–11,541		
KLK-L1	ATGGCCACAGCAGGAAATCC	1,411–1,430	AF113141	KLK-L1
	GGTCACTTGTCTGCGCAGAC	1,990–2,019		
PS	CCCAACCCTGTGTTTTTCTC	3,634–3,653	M33105	PSA[b]
PAS	GGCCCTCCTCCCTCAGA	4,143–4,118		
K1S	ATCCCTCCATTCCCATCTTT	2–22	M18157	KLK1[c]
K1AS	CACATACAATTCTCTGGTTC	324–305		
K2S	AGTGACACTGTCTCAGAATT	131–150	AF024605	KLK2[d]
K2AS	CCCCAATCTCACCAGTGCAC	580–561		
NS	GCTTCCCTACCGCTGTGCT	552–570	AF055481	NES1[e]
NAS	CACTCTGGCAAGGGTCCTG	763–744		

[a] All nucleotide sequences are given in the 5′ → 3′ orientation.
[b] Prostate-specific antigen.
[c] Human renal kallikrein.
[d] Human glandular kallikrein.
[e] Normal epithelia cell-specific 1 gene.

KLK-L3 gene, two gene-specific primers were designed. Two rounds of RACE reactions (nested PCR) were performed with 5 μl Marathon Ready cDNA of human testis (Clontech) as a template. The reaction mix and PCR conditions were selected according to the manufacturer's recommendations. Positive bands were gel-purified using Qiagen Gel Purification kits according to the manufacturer's recommendations.

Gene-specific amplification of other genes from genomic DNA. According to the published sequence of PSA, KLK1, KLK2, normal epithelial cell-specific 1gene (NES1), KLK-L1, KLK-L2 and zyme genes, we designed gene-specific primers for each of these genes (Table 2) and developed polymerase chain reaction (PCR)-based amplification protocols that allowed us to generate specific PCR products with genomic DNA as a template. The PCRs were carried out as described above but using an annealing/extension temperature of 65°C.

Structure analysis studies. Multiple alignment was performed using the Clustal X software package available at 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). Phylogenetic studies were performed using the Phylip software package available from 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 Prompters program. The hydrophobicity study was performed using the BCM search launcher programs (http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html). Signal peptide was predicted using the SignalP WWW server (http://www.cbs.dtu.dk/services/signal). Protein structure analysis was performed with the SAPS (structural analysis of protein sequence) program (http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html).

RESULTS

Construction of a Contiguous Map of the Human Kallikrein Locus on Chromosome 19q13.3–q13.4

Sequence information around the human chromosome 19q13.3–q13.4 locus (the proposed kallikrein locus) is available at the Lawrence Livermore National Laboratory Web site. We have obtained sequences of approximately 300 kb in length. These sequences were in the form of contigs of different lengths. A restriction analysis study of the contigs was performed using various computer programs. With the aid of the *Eco*RI restriction map of this area, which is also available at the LLNL Web site, we were able to define the relative positions of these contigs in relation to one another. Some contigs were overlapping, enabling us to construct a contiguous segment; however, three gaps were present. BLAST analysis of these segments against the GenBank database (Altschul *et al.*, 1997) enabled us to define the precise location of two classical kallikreins, namely PSA and KLK2. We also localized other newly discovered serine proteases that are homologous with the kallikrein genes, namely protease M/zyme/neurosin (Anisowicz *et al.*, 1996; Little *et al.*, 1997; Yamashiro *et al.*, 1997), human stratum corneum chymotryptic enzyme (HSSCE) (Hansson *et al.*, 1994), neuropsin (Yoshida *et al.*, 1998), normal epithelial cell-specific 1 gene (NES1) (Liu *et al.*, 1996), trypsin-like serine protease (TLSP) (GenBank Accession No.

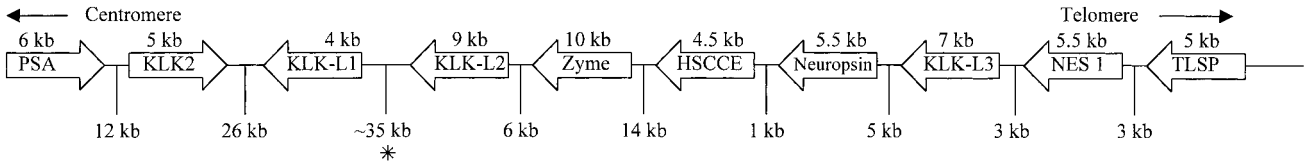

FIG. 1. An approximately 300-kb region of almost contiguous genomic sequence around chromosome 19q13.3–q13.4. Genes are represented by horizontal arrows denoting the direction of the coding sequence. Gene lengths and distances between genes are rounded to the nearest 0.5 kb. The site of the gap is marked with an asterisk. Telomeric to TLSP, there are likely another three kallikrein-like genes (Yousef *et al.*, 1999a; and GenBank Accession No. AF161221) but these are not as yet well characterized. For gene names and more details, see text. Figure is not drawn to scale.

AF164623), KLK-L1 (Yousef *et al.*, 1999c; Nelson *et al.*, 1999) (GenBank Accession No. AF135023), and KLK-L2 (Yousef and Diamandis, 1999) (GenBank Accession No. AF135028). The gaps in the 300-kb genomic sequence were partially filled as follows:

(a) The margins of the first gap were found to contain the 5′ and 3′ ends of the KLK2 gene; this enabled us to fill this gap with the published genomic structure of the KLK2 gene (GenBank Accession No. M18157).

(b) The margins of the third gap (gaps are numbered from centromere to telomere) were found to have the 3′ and 5′ ends of the zyme gene mRNA sequence; thus, a radiolabeled probe specific for the zyme gene was used to screen a human BAC library, and two positive clones were obtained. Restriction analysis was performed, followed by Southern blotting, and a fragment containing the zyme gene was obtained and sequenced, thus filling this gap (Yousef *et al.*, 1999b).

(c) The second gap (between KLK-L1 and KLK-L2 genes) still exists, and we used the *Eco*RI restriction map of this area to define its length approximately (Fig. 1).

Further support for the relative locations of these genes was obtained by performing PCRs with gene-specific primers to screen genomic DNA clones. The most centromeric group of genes (PSA, KLK2, KLK-L1, KLK-L2, and zyme) were found to be clustered in one genomic BAC clone, and the next group (HSCCE, neuropsin, KLK-L3, and NES1) were found to be clustered together in another clone, as expected from the data of Fig. 1.

Cloning of the KLK-L3 Gene

A putative new gene, formed by three exons, was predicted by computer analysis of the genomic sequence. The predicted exons were subjected to sequence homology search against the human EST database (dbEST) and revealed an EST clone (GenBank Accession No. AA583908) that exhibited 99% identity with our putative gene. This EST was obtained, purified, and sequenced, and the sequence was aligned by BLAST software (Altschul *et al.*, 1997) against the genomic area that contains the putative gene. An additional exon, downstream of our predicted structure, was identified. The 3′ end of the gene was verified by (a) the presence of the serine residue (S) of the catalytic triad in a well-conserved region; this highly conserved motif (GDSGGP) always occurs at the beginning of the last exon in all known kallikreins; (b) the presence of a stop codon that is in-frame with the predicted amino acid sequence; and (c) the presence of a 19-poly(A) stretch at the end of the EST that was not found in the genomic sequence.

To verify the accuracy of the cDNA sequence of the gene, PCRs were performed using gene-specific primers for the first and last exons of the predicted structure of the gene (L3-F2 and L3-R1) with cDNA isolated from different human tissues as putative templates. A positive band of the expected size was isolated from testis cDNA and fully sequenced. Its sequence was aligned by BLAST against the genomic sequence to define the exon/intron boundaries unequivocally. For further characterization of the 5′ end of the gene, 5′RACE reaction was performed using Marathon Ready cDNA from testis as a template. This allowed us to identify an additional exon that contains the start codon and 5′ untranslated region. The full sequence of the gene has now been deposited with GenBank (Accession No. AF135026).

Structural Characterization of the KLK-L3 Gene

As shown in Fig. 2, the KLK-L3 gene is formed of five coding exons and four intervening introns, although, as with other kallikreins, the presence of a further upstream untranslated exon(s) could not be ruled out (Luo *et al.*, 1998; Yoshida *et al.*, 1998; Yousef *et al.*, 1999b). All of the exon/intron splice sites conform to the consensus sequence for eukaryotic splice sites (Iida, 1990). The gene further follows strictly the common structural features of the human kallikrein multigene family, as described below.

The predicted protein-coding region of the gene is formed of 753 bp, encoding a deduced polypeptide with a predicted molecular mass of 27.5 kDa. A potential translation initiation codon is found at position 28 of the predicted first exon (numbers of nucleotides refer to GenBank Accession No. AF135026). This codon does not match well with the consensus Kozak (1987) sequence; however, it has a purine at position (−3), which occurs in 97% of vertebrate mRNAs (Kozak, 1987), and it is almost identical to the sequence of the zyme gene flanking the start codon (Yousef *et al.*,

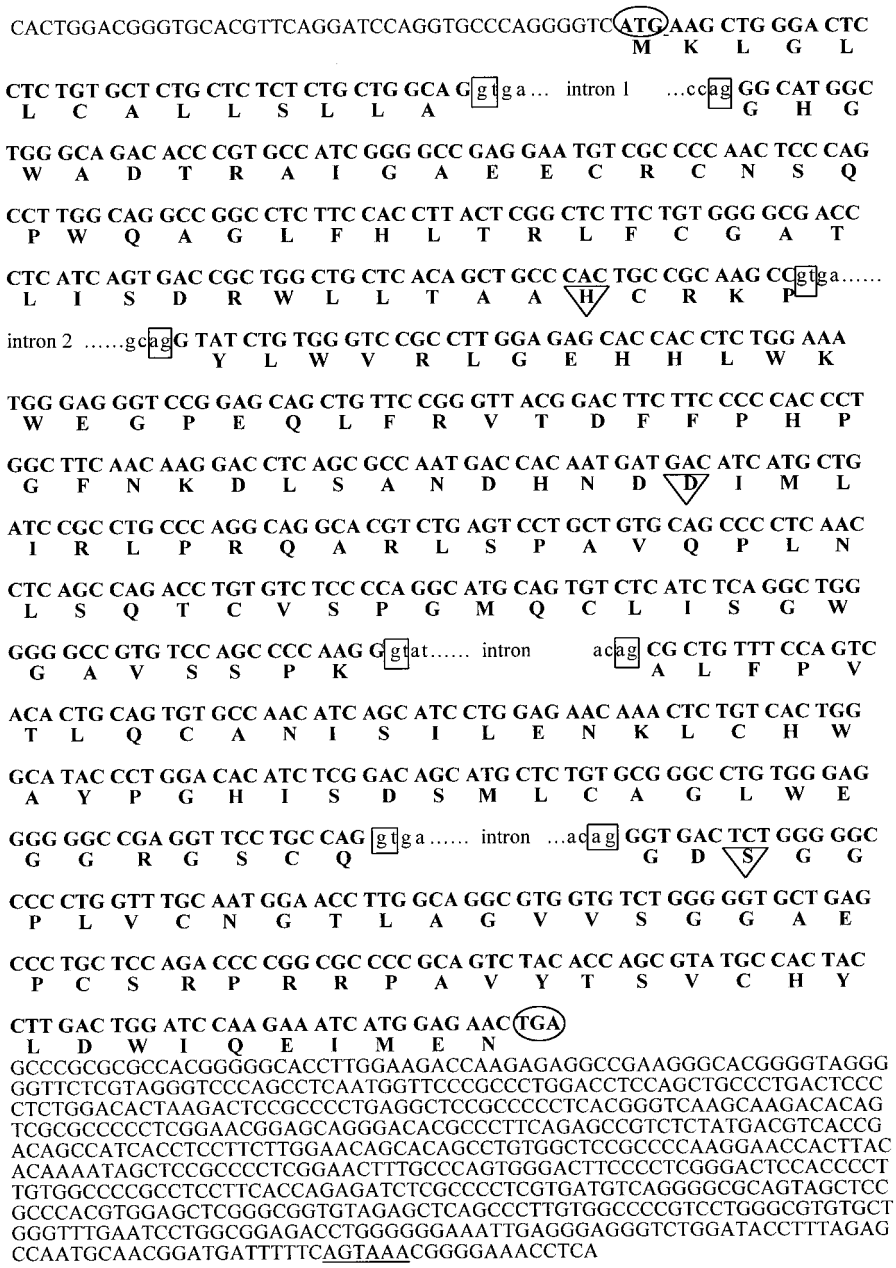

FIG. 2. Genomic organization and partial genomic sequence of the KLK-L3 gene. Intronic sequences are not shown except for the splice junctions. Introns are shown with lowercase letters and exons with uppercase letters. For the full sequence, see GenBank Accession No. AF135026. The start and stop codons 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.

1999b). It should also be noted that most kallikreins do not have the consensus G nucleotide at position (+4).

Nucleotides 6803–6808 (AGTAAA) closely resemble a consensus polyadenylation signal (Proudfood and Brownlee, 1976) and are followed by a stretch of 19 poly(A) nucleotides not found in genomic DNA, after a space of 14 nucleotides. No other potential polyadenylation signals were discernible in the 3′ untranslated region, suggesting that the above motif is indeed the polyadenylation signal. The same polyadenylation signal motif was predicted for the KLK1 and KLK2 genes (Evans *et al.*, 1988; Schedlich *et al.*, 1987).

Although the KLK-L3 protein sequence is unique, comparative analysis revealed that it is homologous to other members of the kallikrein multigene family. KLK-L3 shows 40% protein identity with the TLSP gene product and 38 and 33% identity with the KLK-L2 and KLK1 proteins, respectively. Hydrophobicity analysis revealed that the amino-terminal region is quite hydrophobic (Fig. 3), consistent with the possibility that this region may harbor a signal sequence, analogous to other serine proteases. Computer analysis of the amino acid sequence of KLK-L3 predicted a cleavage site between amino acids 19 and 20 (GWA-DT). Sequence alignment (Fig. 4) also revealed a potential cleavage site (Arg^{22}), at a site homologous to other

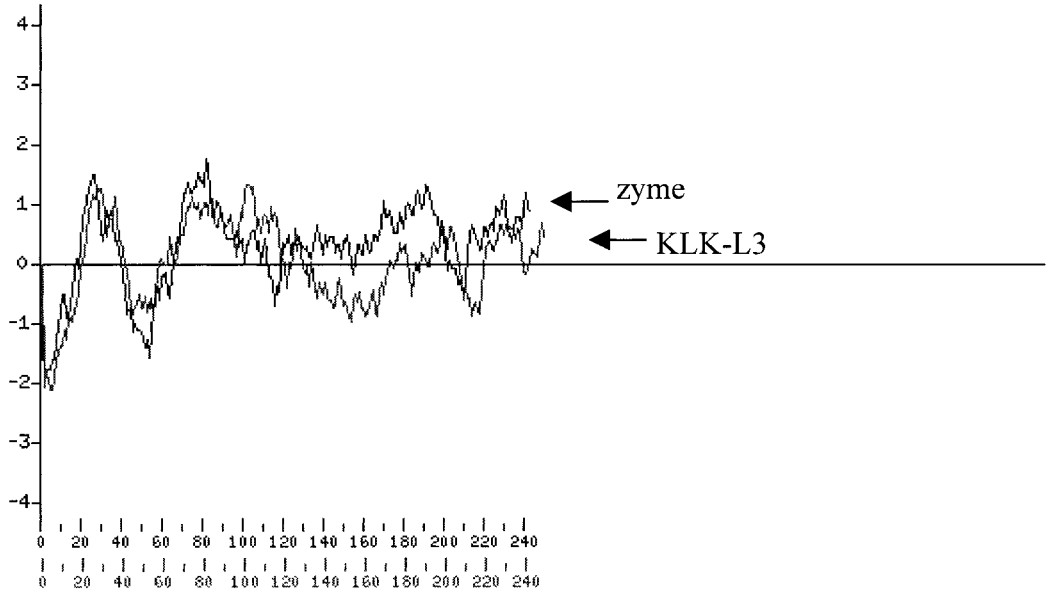

FIG. 3. Plot of hydrophobicity and hydrophilicity, comparing the pattern of the KLK-L3 with that of the zyme gene. Note the hydrophobic region around the first 20 amino acids, likely representing the signal peptide.

serine proteases (lysine (K) or arginine (R) is present in most cases) (Keil, 1971). Several evenly distributed hydrophobic regions throughout the KLK-L3 polypeptide are consistent with a globular protein, similar to other kallikreins and serine proteases. The dotted region in Fig. 4 indicates an 11-amino-acid loop characteristic of the classical kallikreins (PSA, KLK1, and KLK2) but not found in KLK-L3 or other members of the kallikrein multigene family (Yousef *et al.*, 1999b,c; Little *et al.*, 1997; Nelson *et al.*, 1999; Yousef and Diamandis, 1999).

Twenty-nine "invariant" amino acids surrounding the active site of serine proteases have been described (Dayhoff, 1978). Of these, 26 are conserved in KLK-L3. One of the unconserved amino acids (Ser^{168} instead of Pro) is also found in prostase, KLK-L2, and enamel matrix serine proteinase (EMSP1) genes. The second (Leu^{58} instead of Val) is also found in TLSP and KLK-L2 genes, and the third is Ala^{26} instead of Gly. 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). Twelve cysteine residues are present in the putative mature KLK-L3 protein; 10 of them are conserved in all the serine proteases that are aligned in Fig. 4 and would be expected to form disulfide bridges. The other two (C^{136} and C^{238}) are not found in PSA, KLK1, KLK2, or trypsinogen; however, they are found at similar positions in prostase, HSCCE, zyme neuropsin, and TLSP genes and are expected to form an additional disulfide bond.

To predict the phylogenetic relatedness of the KLK-L3 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. 5). Phylogenetic analysis separated the classical kallikreins (KLK1, KLK2, and PSA) and grouped KLK-L3 with TLSP, neuropsin, zyme, HSCCE, and prostase/KLK-L1, consistent with previously published studies (Little *et al.*, 1997; Nelson *et al.*, 1999).

Tissue Expression and Hormonal Regulation of the KLK-L3 Gene

As shown in Fig. 6, the KLK-L3 gene is primarily expressed in thymus, testis, spinal cord, cerebellum, trachea, mammary gland, prostate, brain, salivary gland, ovary, and skin (the last two tissues are not shown in the figure). Lower levels of expression are seen in fetal brain, stomach, lung, thyroid, placenta, liver, small intestine, and bone marrow. No expression was seen in uterus, heart, fetal liver, adrenal gland, colon, spleen, skeletal muscle, pancreas, or kidney. To verify the RT-PCR specificity, representative PCR products were cloned and sequenced. Figure 7 shows that KLK-L3 gene is regulated by steroid hormones in the human breast cancer cell line BT-474.

DISCUSSION

Kallikreins are a subgroup of serine proteases, and these enzymes play important roles in diverse physiological processes (Clements, 1997). Recent evidence suggests that at least some members of the kallikrein family are implicated in breast, prostate, and other human cancers (Diamandis, 1998; Anisowicz *et al.*, 1996; Liu *et al.*, 1996; Diamandis and Yu, 1995; Goyal *et al.*, 1998; Stenman, 1999). Since in rodents there are many kallikrein genes, the restriction of this family to only three members in humans was somewhat surprising. Determination of the true size and cloning of all members of the human kallikrein gene family are im-

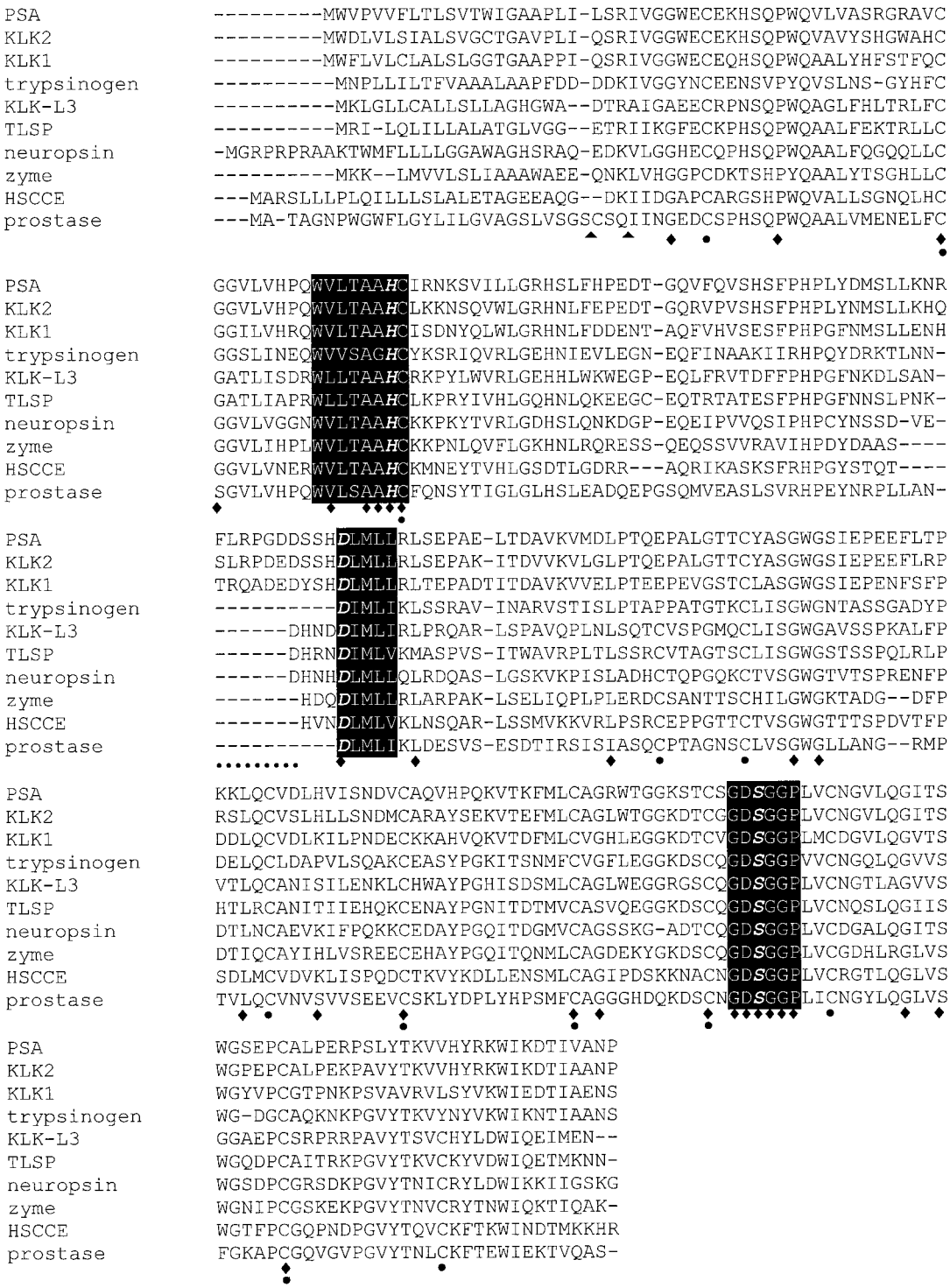

FIG. 4. Alignment of the deduced amino acid sequence of KLK-L3 with members of the kallikrein multigene family. Genes are (from top to bottom; the GenBank Accession number is given in parentheses): PSA (P07288), KLK2 (P20151), KLK1 (NP_002248), trypsinogen (P07477), KLK-L3 (AF135026), trypsin-like serine protease (TLSP) (BAA33404), neuropsin (BAA28673), zyme (Q92876), human stratum corneum chymotryptic enzyme (HSCCE) (AAD49718), and/prostase/KLK-L1 (AAD21581). Dashes represent gaps to bring the sequences to better alignment. The residues of the catalytic triad are in boldface italics, and the 29 invariant serine protease residues are denoted by (◆). Cysteine residues are marked by (●). Conserved areas around the catalytic triad are highlighted in black. The arrowheads (▲) represent the potential cleavage sites. The dotted area represents the kallikrein loop sequence.

portant to our understanding of the contribution of the kallikreins to human biology.

In this report, we have precisely defined the human kallikrein gene locus and constructed the first detailed map describing the relative positions of the kallikreins and other kallikrein-like genes (Fig. 1). This map is consistent with previous reports on the localization of the classical kallikreins and the approximate mapping

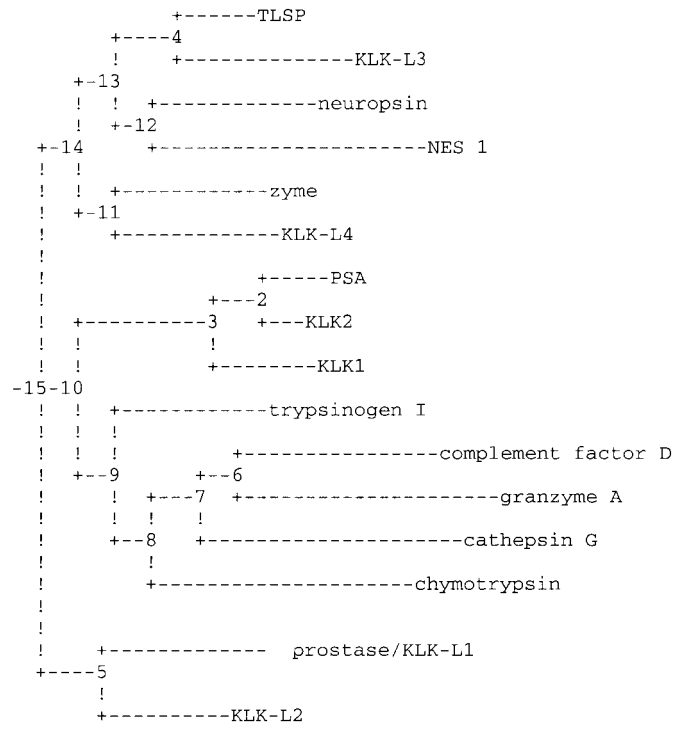

FIG. 5. Dendrogram of the predicted phylogenetic tree for some serine proteases and kallikrein genes. Neighbor-joining/UPGMA method was used to align KLK-L3 with other members of the kallikrein gene family. Gene names and accession numbers are listed in the legend to Fig. 4. The tree grouped the classical kallikreins (KLK1, KLK2, and PSA) together and aligned the KLK-L3 gene in one group with TLSP, neuropsin, and NES 1 genes. KLK-L4 (GenBank Accession No. AF135024) lies further telomeric to TLSP (Yousef *et al.,* 1999a).

of some new kallikreins by radiation hybrid and FISH techniques (Riegman *et al.,* 1992; Luo *et al.,* 1998; Anisowicz *et al.,* 1996; McCormack *et al.,* 1995). It should be noted, however, that the lengths of certain segments of this map (as depicted in Fig. 1) are dependent on the EcoRI restriction map of the area and are measured in terms of approximate kilobase units. Also, the measure of intervals between genes may change slightly in the future, since some kallikreins may have an extra 5′exon(s) that has not as yet been identified. Kallikreins with verified 5′ untranslated exons include NES1 (Luo *et al.,* 1998), zyme (Yousef *et al.,* 1999b), and neuropsin (Yoshida *et al.,* 1998). This map is also directional; it indicates that PSA and KLK2 genes are transcribed in the same direction (centromere to telomere) and that the rest of the kallikrein-like genes are transcribed in the reverse direction (Fig. 1).

An early report indicated that KLK1 is located approximately 31 kb centromeric to PSA (Riegman *et al.,* 1992). Our map extends only 24 kb centromeric to PSA, and for this reason, KLK1 was not precisely localized. Also, a recent report failed to define precisely the location of KLK1 in relation to the other kallikreins (Stephenson *et al.,* 1999). Thus, the exact location of the KLK1 gene is still to be defined from linear chromosome 19 sequencing data. The possibility still exists that this locus is extended further and that other kallikrein-like genes may be located upstream of PSA or downstream from TLSP. Evidence for this has been described elsewhere (Yousef *et al.,* 1999a).

Traditionally, kallikreins are characterized by their ability to liberate lysyl-bradykinin (kallidin) from kininogen (Clements, 1997). In humans, only KLK1 meets this "functional" definition. KLK2 and KLK3 are assigned to the same family based on the strong structural similarities of the genes and proteins and the close localization of these genes in the same chromosomal region (Rittenhouse *et al.,* 1998; Richards *et al.,* 1982). More recently, a new structural concept has emerged to describe kallikreins. Richards and co-workers introduced the concept of a "kallikrein multigene family" in mice, to refer to these genes (Richards *et al.,* 1982; Van Leeuwen *et al.,* 1986). This definition is not based much on the specific enzymatic function of the gene product, but more on its sequence homology and its close linkage on mouse chromosome 7.

Irwin *et al.* (1988) proposed that the serine protease genes can 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. Our results indicate the presence of some more common structural features that are found in all kallikreins (including the newly identified KLK-L3 gene): (1) All genes are formed of five coding exons and four intervening introns (with the possibility that some genes may have an extra 5′ untranslated exon(s) (Luo

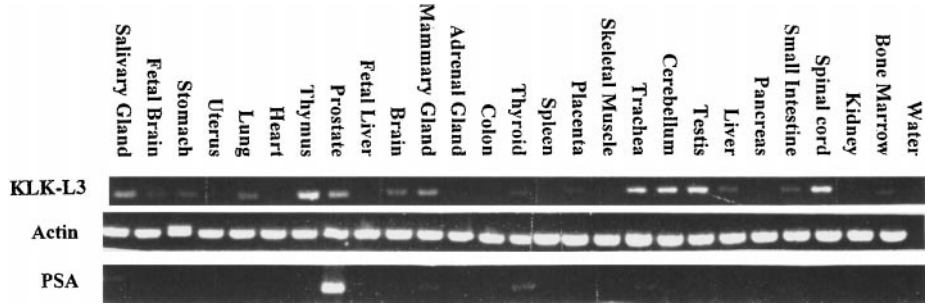

FIG. 6. Tissue expression of the KLK-L3 gene as determined by RT-PCR. Actin and PSA are control genes. For discussion, see text.

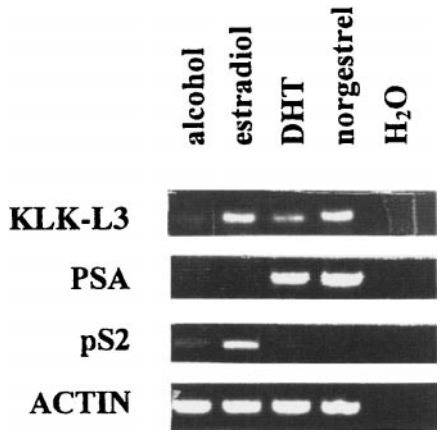

FIG. 7. Hormonal regulation of the KLK-L3 gene in the BT-474 breast carcinoma cell line. DHT, dihydrotestosterone. Steroids were at 10^{-8} M final concentrations. Actin (not regulated by steroid hormones), pS2 (up-regulated by estrogens), and PSA (upregulated by androgens and progestins) are control genes. KLK-L3 is upregulated by progestins, estrogens, and androgens, in that order. For more details, see text.

et al., 1998; Yoshida *et al.*, 1998; Yousef *et al.*, 1999b) (Fig. 8). (2) The exon lengths are usually comparable. (3) The intron phases are always conserved (I-II-I-0) (see the legend to Fig. 8 for a description of intron phases). (4) These genes are clustered in the same chromosomal region, apparently without any intervening non-kallikrein-like genes (Fig. 1). Thus, we can conclude that all the recently identified serine proteases that are present in this region (zyme, HSCCE, neuropsin, NES1, prostase/KLK-L1, KLK-L2, and TLSP), together with the newly identified kallikrein-like gene (KLK-L3), could be considered members of the expanded human kallikrein multigene family.

The chromosomal band 19q13 is nonrandomly rearranged in a variety of human solid tumors including ovarian cancers (Mitelman, 1994), and the currently available data indicate that the kallikrein gene locus is related to many malignancies. At least three kallikrein genes (PSA, zyme, and NES1) are down-regulated in breast cancer (Anisowicz *et al.*, 1996; Liu *et al.*, 1996; Yu *et al.*, 1998; Sauter *et al.*, 1996), and NES1 appears to be a novel tumor suppressor gene (Goyal *et al.*, 1998). Furthermore, PSA exhibits potent antiangiogenic activity (Fortier *et al.*, 1999). It is possible that some of these kallikreins are involved in a cascade pathway, similar to the coagulation or apoptotic process, whereby proforms of proteolytic enzymes are activated and then act upon downstream substrates. Such activity was found for the KLK2 gene product, which acts upon and activates pro-PSA (Kumar *et al.*, 1997; Lovgren *et al.*, 1997).

We have shown here that the expanded human kallikrein gene family has a similar number of members as the rodent family of genes. Although the precise localization of this family of genes has now been achieved, its functional role and connection to human disease are still incompletely understood. Some new compelling data have raised the possibility that at least some of these genes behave as tumor suppressors (Goyal *et al.*, 1998), as negative regulators of cell growth (Lai *et al.*, 1996), as antiangiogenic molecules (Fortier *et al.*, 1999), and as apoptotic molecules (Balbay *et al.*, 1999). The paramount diagnostic value of some members is also well established (Rittenhouse *et*

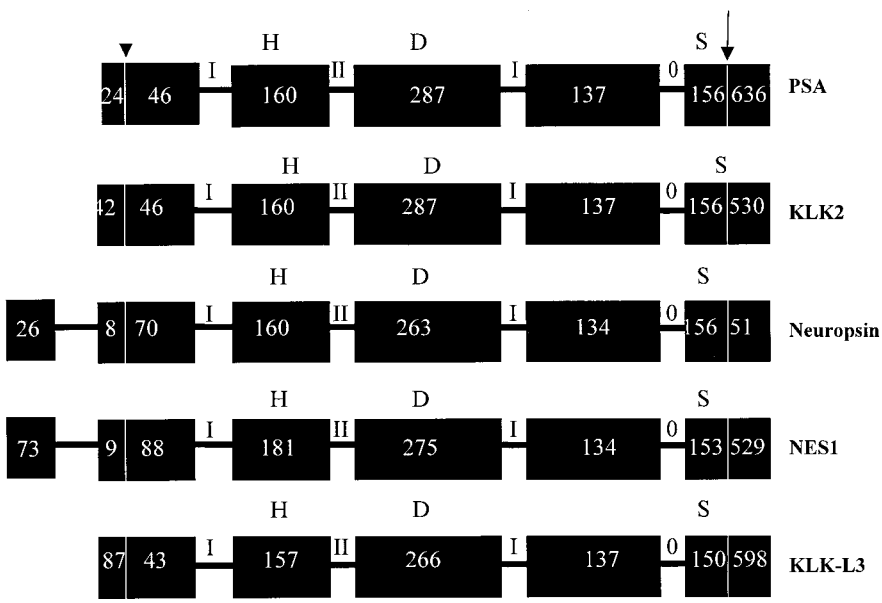

FIG. 8. Schematic diagram showing the comparison of the genomic structure of PSA, KLK2, neuropsin, NES1, and KLK-L3 genes. Exons are shown by black boxes and introns by the connecting lines. Arrowheads show the start codon, and arrows show the stop codon. Letters above boxes indicate relative positions 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 indicates 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.

al., 1998; Diamandis, 1998; Stenman, 1999). For these reasons, it is important to check all members of this family of genes as potential diagnostic or prognostic markers or as candidate therapeutic targets.

The newly identified KLK-L3 gene is expressed in many tissues, including skin, thymus, central nervous system, breast, prostate, and testis. The wide range of tissue expression of KLK-L3 should not be surprising since, by using the more sensitive RT-PCR technique, many kallikrein genes were found to be expressed in a wide variety of tissues. For example, PSA, KLK2, prostase/KLK-L1, and KLK-L2 are now known to be expressed in breast and many other tissues (Rittenhouse *et al.*, 1998; Yousef *et al.*, 1999b,c; Diamandis and Yu, 1995).

Like many other kallikreins, KLK-L3 is regulated by steroid hormones but in a more complex fashion than PSA and KLK2, which are up-regulated by androgens and progestins (Zarghami *et al.*, 1997). In the breast carcinoma cell line studied, KLK-L3 appears to be up-regulated by progestins > estrogens > androgens (Fig. 7). Some other kallikrein-like genes behave similarly (Yousef *et al.*, 1999b). The functional characterization of the KLK-L3 promoter will further clarify the precise mechanism of KLK-L3 regulation by steroid hormones. The possible biological function of KLK-L3 is under investigation.

ACKNOWLEDGMENTS

We thank Linda Ashworth for her assistance in using the sequencing data from the Lawrence Livermore National Laboratory and Dr. S. Scherer for advice and help in the screening of the genomic libraries.

REFERENCES

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

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.

Ashley, P. L., and MacDonald, R. J. (1985). Tissue-specific expression of kallikrein-related genes in the rat. *Biochemistry* **24:** 4520–4525.

Baker, A. R., and Shine, J. (1985). Human kidney kallikrein: cDNA cloning and sequence analysis. *DNA* **4:** 445–450.

Balbay, M. D., Juang, P., Llansa, N., Williams, S., McConkey, D., Fidler, I. J., and Pettaway, C. A. (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.

Ballabio, A. (1993). The rise and fall of positional cloning? *Nat. Genet.* **3:** 277–279.

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

Collins, F. S. (1995). Positional cloning moves from perditional to traditional. *Nat. Genet.* **9:** 347–350.

Dayhoff, M. O. (1978). Atlas of protein sequence and structure. *Natl. Biomed. Res. Found.* **5**(Suppl. 3): 79–81.

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

Diamandis, E. P., and Yu, H. (1995). New biological functions of prostate-specific antigen? *J. Clin. Endocrinol. Metab.* **80:** 1515–1517.

Evans, B. A., Drinkwater, C. C., and Richards, R. I. (1987). Mouse glandular kallikrein genes: Structure and partial sequence analysis of the kallikrein gene locus. *J. Biol. Chem.* **262:** 8027–8034.

Evans, B. A., Yun, Z. X., Close, J. A., Tregear, G. W., Kitamura, N., Nakanishi, S., Callen, D. F., Baker, K., Hyland, V. J., Sutherland, G. R., *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.

Fukushima, D., Kitamura, N., and Nakanishi, S. (1985). Nucleotide sequence of cloned cDNA for human pancreatic kallikrein. *Biochemistry* **24:** 8037–8043.

Goyal, J., Smith, K. M., Cowan, J. M., Wazer, D. E., Lee, S. W., and Band, V. (1998). The role of NES1 serine protease as a novel tumor suppressor. *Cancer Res.* **58:** 4782–4786.

Hansson, L., Stomqvist, M., Backman, A., Wallbrandt, P., Carlstein, A., and Egelrud, T. (1994). Cloning, expression and characterization of stratum corneum chymotryptic enzyme. A skin-specific human serine proteinase. *J. Biol. Chem.* **269:** 19420–19426.

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

Irwin, D. M., Robertson, K. A., and MacGillivary, R. T. (1988). Structure and evolution of the bovine prothrombin gene. *J. Mol. Biol.* **212:** 31–45.

Keil, B. (1971). Trypsin. *In* "The Enzymes" (P. D. Boyer, Ed.), 3rd ed., Vol. 3, pp. 249–275, Academic Press, New York.

Kozak, M. (1987). An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15:** 8125–8148.

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

Kumar, A., Mikolajczyk, S. D., Goel, A. S., Millar, L. S., and Saedi, M. S. (1997). Expression of proform of prostate-specific antigen by mammalian cells and its conversion to mature, active form by human kallikrein 2. *Cancer Res.* **57:** 3111–3114.

Lai, L. C., Erbas, H., Lennard, T. W., 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., Dobbins, J. R., Wyrick, T., Miller, J. R., MacKellar, W., *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.

Lovgren, J., Rajakoski, K., Krap, M., Lundwall, A., and Lilja, H. (1997). Activation of the zymogen form of prostate-specific antigen by human glandular kallikrein 2. *Biochem. Biophys. Res. Commun.* **238:** 549–555.

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.

McCormack, R. T., Rittenhouse, H. G., Finlay, J. A., Sokoloff, R. L., Wang, T. J., Wolfert, R. L., Lilja, H., and Oesterling, J. E. (1995). Molecular forms of prostate-specific antigen and the human kallikrein gene family: A new era. *Urology* **45:** 729–744.

Mitelman, F. (1994). "Catalog of Chromosome Aberrations in Cancer," 5th ed., pp. 3067–3198, Wiley–Liss, New York.

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

Murray, S. R., Chao, J., Lin, F., and Chao, L. (1990). Kallikrein multigene families and the regulation of their expression. *J. Cardiovasc. Pharmacol.* **15**(Suppl.): S7–S15.

Nadeau, J., Grant, P., and Kosowsky, M. (1991). Mouse and human homology map. *Mouse Genome* **89:** 31–36.

Nelson, P. S., Gan, L., Ferguson, C., Moss, P., Gelinas, R., Hood, L., and Wang, K. (1999). Molecular cloning and characterization of prostase, an androgen-regulated serine protease with prostate-restricted expression. *Proc. Natl. Acad. Sci. USA* **96:** 3114–3119.

Osoegawa, K., Woon, P. Y., Zhau, B., Frengen, E., Tateno, M., Catanese, J. J., and de Jong, P. J. (1999). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* **52:** 1–8.

Proudfoot, N. J., and Brownlee, G. G. (1976). 3′Non-coding region sequences in eukaryotic messenger RNA. *Nature* **263:** 211–214.

Richards, R. I., Catanzaro, D. F., Mason, A. J., Morris, B. J., Baxter, J. D., and Shine, J. (1982). Mouse glandular kallikrein genes. Nucleotide sequence of cloned cDNA coding for a member of the kallikrein arginyl estero-peptidase group of serine proteases. *J. Biol. Chem.* **257:** 2758–2761.

Richards, R. I., Holman, K., Shen, Y., Kozman, H., Harley, H., Brook, D., and Shaw, D. (1991). Human glandular kallikrein genes: Genetic and physical mapping of the KLKl locus using a highly polymorphic microsatellite PCR marker. *Genomics* **11:** 77–82.

Riegman, P. H., Vlietstra, R. J., Klaassen, P., van der Korput, J. A., Van Kessel, G. A., Romijn, J. C., and Trapman, J. (1989a). The prostate-specific antigen gene and the human glandular kallikrein-1 gene are tandemly located on chromosome 19. *FEBS Lett.* **247:** 123–126.

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

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

Rittenhouse, H. G., Finlay, J. A., Mikolajczyk, S. D., and Partin, W. (1998). Human kallikrein 2 (hk2) and prostate-specific antigen(PSA): Two closely related, but distinct, kallikreins in prostate. *Crit. Rev. Clin. Lab. Sci.* **35:** 275–368.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Saunders, A. M., and Seldin, M. F. (1990). A molecular genetic linkage map of mouse chromosome 7. *Genomics* **8:** 525–535.

Sauter, E. R., Daly, M., Linahan, K., Ehya, H., Engstrom, P. F., Bonney, G., Ross, E. A., Yu, H., and Diamandis, E. P. (1996). Prostate specific antigen levels in nipple aspirate fluid correlate with breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.* **5:** 967–970.

Schachter, M. (1980). Kallikreins (kininogenases)—A group of serine proteases with bioregulatory actions. *Pharmac. Rev.* **31:** 1–17.

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

Stenman, U. H. (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., Verity, K., Ashworth, L. K., and Clements, J. A. (1999). Localization of a new prostate-specific antigen-related serine protease gene KLK4, is evidence of an expanded human kallikrein gene family cluster on chromosome 19q13.3–q13.4. *J. Biol. Chem.* **27:** 23210–23214.

Van Leeuwen, B. H., Evans, B. A., Tregear, G. W., and Richards, R. I. (1986). Mouse glandular with breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.* **5:** 967–970.

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

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

Yousef, G., and Diamandis, E. P. (1999). The new kallikrein-like gene, KLK-L2: Molecular characterization, mapping, tissue expression, and hormonal regulation. *J. Biol. Chem.* **274:** 37511–37516.

Yousef, G., Luo, L., and Diamandis, E. P. (1999a). Identification of novel human kallikrein-like genes on chromosome 19q13.3–q13.4. *Anticancer Res.* **79:** 2843–2852.

Yousef, G., Luo, L., Scherer, S., Sotiropoulou, G., and Diamandis, E. P. (1999b). Molecular characterization of Zyme/Protease M/Neurosin (PRSS9), A hormonally regulated kallikrein-like serine protease. *Genomics* **62:** 251–259.

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

Yu, H., Levesque, M. A., Clark, G. M., and Diamandis, E. P. (1998). Prognostic value of prostate specific antigen for women with breast cancer. *Clin. Cancer Res.* **4:** 1489–1497.

Zarghami, N., Grass, L., and Diamandis, E. P. (1997). Steroid hormone regulation of prostate specific antigen gene expression in breast cancer. *Br. J. Cancer* **75:** 579–588.