Genomic Organization of the Human Kallikrein Gene Family on Chromosome 19q13.3–q13.4

George M. Yousef,*† Albert Chang,*† Andreas Scorilas,*† and Eleftherios P. Diamandis*††

*Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada; and†Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

Received August 11, 2000

Kallikreins are a subgroup of serine proteases with diverse physiological functions. Recently, growing evidence indicates that many kallikrein genes are involved in malignancy. In rodents, kallikreins are encoded by a large multigene family, but in humans only three kallikreins were thought to exist. Based on the homology between the human and rodent kallikrein loci, we studied a 300 kb region of genomic sequences around the putative KLK1 gene locus on chromosome 19q13.3–q13.4. By using linear sequence information, restriction analysis, end sequencing, PCR and blotting techniques, as well as bioinformatic approaches, we were able to construct the first detailed map of the human kallikrein gene family. Comparative analysis of genes located in this area, provides strong evidence that the human kallikrein gene family locus on chromosome 19 is considerably larger than previously thought, containing at least fifteen genes. We have established, for the first time, the common structural features that apply to all members of the expanded kallikrein multigene family. Our map specifies the distance between genes to one base pair accuracy, the relative location, and the direction of transcription of all 15 genes. Determination of the true size of the kallikrein family in humans is important for our understanding of the contribution of the kallikreins to human biology and pathophysiology.

Key Words: kallikreins; kallikrein-like genes; serine proteases; human kallikrein locus; gene mapping; KLK3; chromosome 19.

Kallikreins are a subgroup of serine proteases found in glandular cells, neutrophils, and biological fluids (1). Kallikreins are divided into two main groups: tissue and plasma kallikreins (2). The two groups differ in their molecular weight, substrate specificity, and immunological characteristics. A single gene encodes for the plasma kallikrein enzyme (3), whereas tissue kallikreins belong to a multigene family which shows different patterns of tissue specific expression (4). Recently, growing evidence suggests that many kallikreins and kallikrein-like genes are associated with cancer (5).

In mouse and rat, kallikreins are encoded by large multigene families. In the mouse genome, at least 24 genes have been identified (6). Expression of 11 of these genes has been confirmed; the rest are presumed to be pseudogenes (7). A similar family of 15–20 kallikreins has been found in the rat genome (8) where at least 4 of these are known to be expressed (9). The human kallikrein gene family was thought to have only three members: the tissue (pancreatic-renal) kallikrein (KLK1, encoding for hK1) (10), the human glandular kallikrein (KLK2, encoding for hK2) (11), and prostate specific antigen (KLK3, encoding for PSA or hK3) (12, 13). Several early estimations of the size of this family in humans, using Southern blot analysis, predicted just 3–4 genes (11, 14–16). However, a Southern blot analysis using monkey cDNA suggested that it might contain as many as 19 genes (17). Recently, new kallikreins and kallikrein-like genes have been discovered (18–25). In addition, a number of newly discovered serine proteases were found to have significant homologies with the classical kallikrein genes. Also, a recent report indicated the presence of two putative...
kallikrein-like gene fragments in the region 19q13.3–q13.4 (26). Based on these findings, we hypothesized that the human kallikrein multigene family might be larger than previously thought.

The mouse kallikrein genes are clustered in groups of up to 11 genes on chromosome 7 and the distance between the genes in the various clusters can be as small as 3–7 kb (6). All three established human kallikrein genes have been assigned to chromosome 19q13.2–19q13.4 and the distance between KLK3 and KLK2 has been estimated to be 12 Kb (27).

Based on mapping of the rodent kallikrein genes and the documented strong conservation between human chromosome 19q13.3–q13.4 and the seventeen loci in a 20-cM proximal part of mouse chromosome 7 (28, 29), we identified a candidate genomic region (flanking the presumed KLK1 locus) 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 family, defining the order of genes along the chromosomal region, the direction of transcription, and estimating the distances between genes with one base pair accuracy. We provide strong evidence that kallikreins comprise a large multigene family in humans, and elucidate some common structural features that are conserved in all kallikrein genes.

**MATERIALS AND METHODS**

DNA sequences on chromosome 19. We obtained genomic sequences from the web site of the Lawrence Livermore National Laboratory (LLNL) (http://www-bio.llnl.gov/genome/genome.html). These sequences encompass a region flanking the putative KLK1 gene locus on chromosome 19q13.3–q13.4. The data were in the form of separate contigs constructed from BAC or cosmID clones. These contig sequences were subjected to an EcoR1 restriction analysis using the “webcutter” computer program (available at: http://www.firstmarket.com/cutter/cut2.html), and the fragment lengths were calculated.

Identification of positive BAC genomic clones from a human genomic DNA library. Many kallikrein genes were discovered independently and given various names. Recently, an international group of investigators established official nomenclature for these genes, as described in detail elsewhere (5). In Table 1, we present the conventional and official symbols of these genes. In the text, we use only the official nomenclature.

Gene-specific primers were used to amplify unique DNA sequences for the KLK1, KLK6, KLK13, KLK14 and the Siglec 9 gene (please see below). The primers used and the PCR reaction conditions are summarized in Table 2. The PCR products were purified and then labeled with 32P by the random primer method (30), and used as probes to screen the PCRI-11 human male BAC library, constructed by de Jong and associates (31). The library was spotted in duplicate on nylon membranes, for identification of positive clones. The filters were hybridized in 15% formamide, 500 mM Na2HPO4, 7% SDS, 1% BSA (w/v) at 65°C overnight, then washed sequentially with 2× SSC, 0.1% SDS at 65°C, and then exposed to X-ray film as described (30). Positive clones were obtained, plated on selective LB medium, and then a single colony was transferred into LB broth for overnight cultures. Three PAC clones positive for KLK10 were identified by a similar methodology as described elsewhere (20).

End sequencing of BAC clones. Purification of BAC and PAC DNA was done 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 (30). The ends of these clones were sequenced using vector-specific primers with an automated DNA sequencer.

Gene-specific amplification of other genes from genomic DNA. According to the published sequence of the 15 kallikrein genes identified in this region, we designed gene-specific primers for each of these genes and developed PCR-based amplification protocols and used them to screen the BAC and PAC clones described above. PCR was carried out in a reaction mixture containing 1 μl of DNA (0.1–1.0 μg), 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2 mM MgCl2, 200 μM dNTPs (deoxynucleoside triphosphates), 100 ng of primers and 2.5 units of

<table>
<thead>
<tr>
<th>New gene symbol1,2</th>
<th>Previous gene symbol(s)</th>
<th>New protein symbol</th>
<th>Other protein names/symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>KLK1</td>
<td>KLK1</td>
<td>hK1</td>
<td>Pancreatic/renal kallikrein, hPRK</td>
</tr>
<tr>
<td>KLK15</td>
<td>—</td>
<td>hK15</td>
<td>—</td>
</tr>
<tr>
<td>KLK3</td>
<td>KLK3</td>
<td>hK3</td>
<td>Prostate specific antigen, PSA</td>
</tr>
<tr>
<td>KLK2</td>
<td>KLK2</td>
<td>hK2</td>
<td>Human glandular kallikrein 1, hGK-1</td>
</tr>
<tr>
<td>KLK4</td>
<td>PRSS17, KLK-L1, KLK4</td>
<td>hK4</td>
<td>Prostase, KLK-L1 protein, EMSP1</td>
</tr>
<tr>
<td>KLK5</td>
<td>KLK-L2</td>
<td>hK5</td>
<td>KLK-L2 protein; HSCTE</td>
</tr>
<tr>
<td>KLK6</td>
<td>PRSS9</td>
<td>hK6</td>
<td>Zyme; Protease M, Neurosin</td>
</tr>
<tr>
<td>KLK7</td>
<td>PRSS6</td>
<td>hK7</td>
<td>HSCE</td>
</tr>
<tr>
<td>KLK8</td>
<td>PRSS19</td>
<td>hK8</td>
<td>Neuropsin; Ovasin; TADG-14</td>
</tr>
<tr>
<td>KLK9</td>
<td>KLK-L3</td>
<td>hK9</td>
<td>KLK-L3 protein</td>
</tr>
<tr>
<td>KLK10</td>
<td>PRSSL1, NES1</td>
<td>hK10</td>
<td>NES1 protein</td>
</tr>
<tr>
<td>KLK11</td>
<td>PRSS20</td>
<td>hK11</td>
<td>TLSP; Hippostasin</td>
</tr>
<tr>
<td>KLK12</td>
<td>KLK-L5</td>
<td>hK12</td>
<td>KLK-L5 protein</td>
</tr>
<tr>
<td>KLK13</td>
<td>KLK-L4</td>
<td>hK13</td>
<td>KLK-L4 protein</td>
</tr>
<tr>
<td>KLK14</td>
<td>KLK-L6</td>
<td>hK14</td>
<td>KLK-L6 protein</td>
</tr>
</tbody>
</table>

1 Genes are presented in their order on chromosome 19q13.3–q13.4 from centromere to telomere.
2 Full gene names are given under non-standard abbreviations.

**TABLE 1**

<table>
<thead>
<tr>
<th>New gene symbol1,2</th>
<th>Previous gene symbol(s)</th>
<th>New protein symbol</th>
<th>Other protein names/symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>KLK1</td>
<td>KLK1</td>
<td>hK1</td>
<td>Pancreatic/renal kallikrein, hPRK</td>
</tr>
<tr>
<td>KLK15</td>
<td>—</td>
<td>hK15</td>
<td>—</td>
</tr>
<tr>
<td>KLK3</td>
<td>KLK3</td>
<td>hK3</td>
<td>Prostate specific antigen, PSA</td>
</tr>
<tr>
<td>KLK2</td>
<td>KLK2</td>
<td>hK2</td>
<td>Human glandular kallikrein 1, hGK-1</td>
</tr>
<tr>
<td>KLK4</td>
<td>PRSS17, KLK-L1, KLK4</td>
<td>hK4</td>
<td>Prostase, KLK-L1 protein, EMSP1</td>
</tr>
<tr>
<td>KLK5</td>
<td>KLK-L2</td>
<td>hK5</td>
<td>KLK-L2 protein; HSCTE</td>
</tr>
<tr>
<td>KLK6</td>
<td>PRSS9</td>
<td>hK6</td>
<td>Zyme; Protease M, Neurosin</td>
</tr>
<tr>
<td>KLK7</td>
<td>PRSS6</td>
<td>hK7</td>
<td>HSCE</td>
</tr>
<tr>
<td>KLK8</td>
<td>PRSS19</td>
<td>hK8</td>
<td>Neuropsin; Ovasin; TADG-14</td>
</tr>
<tr>
<td>KLK9</td>
<td>KLK-L3</td>
<td>hK9</td>
<td>KLK-L3 protein</td>
</tr>
<tr>
<td>KLK10</td>
<td>PRSSL1, NES1</td>
<td>hK10</td>
<td>NES1 protein</td>
</tr>
<tr>
<td>KLK11</td>
<td>PRSS20</td>
<td>hK11</td>
<td>TLSP; Hippostasin</td>
</tr>
<tr>
<td>KLK12</td>
<td>KLK-L5</td>
<td>hK12</td>
<td>KLK-L5 protein</td>
</tr>
<tr>
<td>KLK13</td>
<td>KLK-L4</td>
<td>hK13</td>
<td>KLK-L4 protein</td>
</tr>
<tr>
<td>KLK14</td>
<td>KLK-L6</td>
<td>hK14</td>
<td>KLK-L6 protein</td>
</tr>
</tbody>
</table>

1 Genes are presented in their order on chromosome 19q13.3–q13.4 from centromere to telomere.
2 Full gene names are given under non-standard abbreviations.
HotStar Taq polymerase (Qiagen, Valencia, CA) on an Eppendorf thermal cycler. The cycling conditions were 95°C for 15 min to activate the HotStar Taq polymerase, followed by 40 cycles of 94°C denaturation for 30 s, annealing for 30 s, and 72°C extension for 1 min, and a final extension at 72°C for 10 min. Equal amounts of PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. Primers and annealing temperatures used for the different reactions are summarized in Table 2.

Cloning and sequencing of the PCR products. Due to the high degree of homology between the genes in this genomic region, all primers were designed to be specific for each gene, annealing away from conserved regions. To further 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.

Long PCR amplification. A pair of primers connecting the first exon of KLK1 (CTC CGA TCC CCA GAG CTG TA) and the centromeric end of the KLK3 genomic sequence (CTG GTG GTG CCA AGG ATC CA), were used to amplify the gap between KLK1 and KLK3. PCR was performed using the Expand Long Template PCR System (Roche Molecular Systems, Indianapolis, IN) with human genomic DNA as a template. We used a 50 μl reaction mixture containing 300 ng DNA, 20 mM Tris-HCl (pH 7.5), 100 mM KC1, 2.25 mM MgCl2, 1 mM dithiothreitol (DTT), 0.1 mM EDTA, 0.5% (v/v) Tween 20, 0.5% (v/v) Nonidet P40, 500 μM dNTPs, 300 nM primers and 2.6 units of enzyme mix containing thermostable Taq and Pwo DNA polymerases, on a Perkin-Elmer 9600 thermal cycler. The cycling conditions were 94°C for 2 min, followed by 10 cycles of 94°C for 10 s, 63°C for 30 s, 68°C for 8 min, then 20 cycles of 94°C for 10 s, 63°C for 30 s, 68°C for 8 min (increasing by 20 s per subsequent cycle), in addition to a final extension step at 68°C for 7 min. The PCR products were electrophoresed on a 0.8% agarose gel.

Purification and sequencing of long PCR products. The PCR products were purified with exonuclease I plus shrimp alkaline phosphatase treatment as described elsewhere (32). Briefly, 1 μl of exonuclease I and 1 μl of shrimp alkaline phosphatase were added to the PCR products, then the mixture was incubated at 37°C for 15 min, then at 85°C for 15 min. Excess oligonucleotides and dNTPs were removed by a spin dialysis method using Centricon 30 microconcentrators (Amicon, (30)). After purification, PCR products were directly sequenced using an automated DNA sequencer.

RESULTS

Construction of a Continuous Chromosomal Region Encompassing the Human Kallikrein Gene Locus

We obtained genomic sequences around the human kallikrein gene locus on chromosome 19q13.3–13.4 from the Lawrence Livermore National Laboratory. These sequences were in the form of separate contigs
Comparison of the EcoR1 restriction maps of the contigs with the restriction maps of their parent clones (available from the LLNL web site) enabled us to align the contigs along the chromosomal region. Some contigs were overlapping, enabling us to construct a contiguous segment; however, few gaps were present.

A Blast search was done for the genomic sequence against the non-redundant GenBank database (33). This allowed us to define the precise location of two classical kallikrein genes; PSA (KLK3), KLK1 and KLK2. We also localized other newly discovered serine proteases which are homologous with the kallikrein genes, KLK6 (19), human stratum corneum chymotryptic enzyme (KLK7) (34), KLK8 (23), normal epithelial cell-specific 1 gene (KLK10) (20, 35), and trypsin-like serine protease (KLK11) (36, 37).

Three gaps were present in our genomic sequence. These gaps were closed as follows: (1) The margins of the first gap (G1) were found to contain the 5' and 3' ends of the KLK2 gene; this enabled us to fill the gap with the published genomic structure of the KLK2 gene (GenBank Accession No. M18157). (2) The margins of the third gap (G3) were found to contain the 3' and 5' ends of the KLK6 gene mRNA sequence; thus, a radiolabeled probe specific for the KLK6 gene was used to screen a human BAC library (see Materials and Methods) and two positive clones were obtained. Restriction digestion was performed, followed by Southern blotting and a fragment containing the KLK6 gene was obtained and sequenced, thus filling this gap (25). (3) The second gap (G2) was closed by performing a long PCR reaction using primers located at the margins of a BAC clone that covers this area and the BAC clone as a template (see below).

Identification of New Kallikrein Genes in the Human Kallikrein Locus

A number of computer programs were used to predict the presence of putative new genes within the contiguous genomic region and six new putative genes were predicted, as described elsewhere (38). The full structure of these genes was verified by using various methods, including EST sequencing and homology search, PCR, 3' and 5' RACE, as described elsewhere (24, 25, 39). In addition, a novel member of the Siglec gene family was identified at the telomeric end of the genomic sequence (GenBank Accession No. AF135027).

Relative Localization of the Kallikrein Genes

Verification of the accuracy of the constructed human kallikrein gene locus and determination of the relative localization of the kallikrein genes were performed as follows:

(A) Screening of human BAC and PAC libraries. A human BAC library was screened using radiolabeled probes specific to KLK1, KLK6, KLK13, KLK14, and the Siglec 9 gene. Six BAC clones which were positive to any of the probes were selected for further analysis. Three PAC clones, known to be positive for the KLK10 gene (20), were also included.

(B) PCR screening of the BAC and PAC clones. The above-mentioned clones were screened by PCR using gene-specific primers for each of the 15 kallikrein genes that were identified in this region (Table 2). The results of this screening are summarized in Table 3 and a representative gel is shown in Fig. 1. These data are consistent with our predicted order for the 15 genes along the physical map.

(C) End sequencing of BAC and PAC clones. Further confirmation of the relative location of these genes
in relation to each other was obtained by end sequencing of the BAC and PAC clones. Each clone was sequenced from both ends, and the sequences obtained were aligned with the genomic sequence to identify the ends of each clone. Figure 2 shows a diagrammatic representation of genomic areas covered by each clone (horizontal lines), and the ends of each clone as determined by end sequencing (vertical lines).

The data of Fig. 2 and Table 3 are consistent with the predicted physical map of this region. The relative locations of individual genes in relation to each other were verified as follows (see Fig. 2 and Table 3): (1) The position of KLK1 in relation to KLK3, and the direction of transcription of both genes was verified by performing a long PCR reaction, with a forward primer from the reverse complementary sequence of the first exon of KLK1 and a reverse primer from the sequence adjacent to the KLK15 gene, using human genomic DNA as a template. The reverse primer was located 2 kb from the centromeric end of the contig, containing the KLK3 gene and the PCR product was a band of ~9 kb in length (data not shown). The PCR product was cloned and sequenced to verify the specificity of the reaction. This data confirms both the tandem localization of the KLK1 and KLK3 genes and the distance between them, as reported by others (~31 kb). (2) Using bioinformatics, we have further predicted the existence of a novel serine protease gene, mapping between KLK1 and KLK3. This gene was then cloned and sequenced and was found to transcribe in the same direction as KLK1. The gene is known as KLK15 (GenBank Accession No. AF242195). (3) BAC 288H1 is positive for both KLK3 and KLK2, and PAC 121B1 is positive for KLK2, but not KLK3 (Fig. 2), thus, KLK3 must be centromeric to KLK2, also in accordance with previous reports (26). (4) KLK4 and KLK2 are present in two different contigs, and the alignment of these contigs according to the EcoR1 restriction map indicates that KLK4 is telomeric to KLK2 (data not shown). (5) KLK5 must be telomeric to KLK4, because PAC clone 86023 is positive for KLK5 but not for KLK4. (6) The KLK6 gene sequence bridges the gap between two adjacent contigs, 26 and 27 of clone 85745 (numbers refer to contigs of the LLNL) so, KLK6 must be at the telomeric end of contig 26, and thus further downstream from KLK5. (7) BAC 10111 is positive for KLK7 but not KLK6, and in addition, its end sequence falls within the first part of the KLK6 gene; thus, KLK6 must be centromeric to KLK7. (8) The KLK8 gene is telomeric to KLK7, as confirmed by the fact that both genes are located on one contiguous fragment that fits the EcoR1 map of this region. (9) By comparing telomeric sequences of clones 288H1, 76F7, 121B1 and 86023, we verified the order of genes telomeric to KLK8 as follows: KLK9, KLK10, KLK11, KLK-L5 and KLK13. (10) By obtaining additional sequences from the centromeric end of BAC 433D04 by primer walking, we confirmed that KLK13 is centromeric to KLK14. (11) Finally, we observed that BAC10111 contains the telomeric group of this gene family, except for the novel Siglec 9 gene, indicating that this gene is the most telomeric. Siglec 9 is not a kallikrein or a serine protease and its location indicates the end of the kallikrein gene family and the start of another gene family (Siegles).

In Fig. 3 we present the physical location of the 15 members of the human kallikrein gene family, around the chromosomal locus 19q13.3–q13.4. The lengths of the genes and between genes are given with 1 bp accuracy. These distances may change slightly in the future if more untranslated exons are found in the 3' or 5' ends of these genes. Additionally, the transcription initiation site of each gene has not as yet been accurately defined.

In Fig. 4 we present the genomic organization of the 15 genes, describing intron and exon sizes, start and stop codons, untranslated exons, intron phases and location of the amino acids of the catalytic triad.
DISCUSSION

Kallikreins are involved in diverse physiological processes through their serine protease enzymatic activity (2). Determination of the gene family size and cloning of all its members is important 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

![Diagram](image)

**FIG. 2.** Schematic presentation of the results of PCR and end-sequencing for the different BAC and PAC clones. The genes are shown, in order, in open boxes along a horizontal line and BAC and PAC clone IDs are inside solid bars. Horizontal dotted lines represent areas covered by the corresponding clones, and vertical dotted lines the locations of both ends of each clone, as determined by end-sequencing. Figure is not drawn to scale. For complete gene names and details, see Tables 1 and 3 and text.

**FIG. 3.** An approximate 300 kb region of contiguous genomic sequence around chromosome 19q13.3–q13.4. The direction of transcription of each gene is illustrated by arrows. Boxes represent genes and contain the gene name and its genomic length in base pairs. Distances between genes are described in base pairs between boxes. Figure is not drawn to scale. For full gene names see abbreviations footnote and Table 1.
map describing the relative positions of the kallikreins and other kallikrein-like genes (Fig. 3). This map is consistent with previous reports on the localization of the classical kallikreins (26, 27), and the approximate mapping of some new kallikreins by radiation hybrid and FISH techniques (20, 21, 38). It should be noted, however, that the estimation of intervals between genes depends on the published sequences of the genes, and might change slightly in the future, since some kallikreins may have extra 5’ exon(s) that have not as yet been identified. This map is also directional; it indicates that KLK3 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. The possibility still exists that this locus is extended further, and that other kallikrein-like genes may be located upstream of KLK1, downstream from KLK14 or in between. The second possibility is more unlikely since the novel Siglec 9 gene identified, belongs to a multigene family which is localized telomeric to the kallikrein locus (our unpublished data).

In our study, all results of BAC and PAC end sequencing and screening were consistent with each other as well as with the EcoR1 restriction map of the region, with the exception of one PAC clone (43B1), which was positive for KLK10, KLK9, KLK8 and KLK1 genes and negative for all other genes. These results...
were not matching with data from all other PAC and BAC clones, and this clone was excluded from the analysis.

Kallikreins were traditionally characterized by their ability to liberate lysyl-bradykinin (kallidin) from kininogen (2). More recently, however, a new, structural concept has emerged to describe kallikreins. From accumulated sequencing data, it is now clear that the mouse has many genes with high homology to kallikrein coding sequences (40, 41). The concept of a "kallikrein multigene family" proposed by Richards and co-workers, refers to these genes (42, 43). This definition is not based much on specific enzymatic function of the gene product, but more on its sequence homology and their close linkage on mouse chromosome 7. In humans, only KLK1 meets the functional definition of a kallikrein. Based on the newer definition, members of the kallikrein family include, not only the gene for the kallikrein enzyme, but also genes encoding other homologous proteases that are located in the same chromosomal region. Therefore, it is important to note the clear distinction between the enzyme kallikrein and a kallikrein or a kallikrein-like gene.

An early report proposed that the serine proteases could be classified into five different groups according to intron position (44). The established kallikreins (KLK1, KLK2, and KLK3) belong to a group in which: (1) The intron phases are always conserved (I-II-I-0) (see Fig. 4 for description of intron phases). (2) There is an intron just downstream from the codon for the active site histidine residue, (3) a second intron is present downstream from the exon containing the codon for the active site aspartic acid residue, and (4) a third intron is located just upstream from the exon containing the codon for the active site serine residue. Our results indicate the presence of some additional common structural features that are found in all kallikreins, as shown in Fig. 4: (1) All genes are formed of 5 coding exons and 4 intervening introns (with the possibility that some genes may have extra 5' untranslated exon(s) (20, 23, 38). (2) The exon lengths are usually comparable. (3) These genes are clustered in the same chromosomal region, apparently without any intervening non kallikrein-like genes (Fig. 3). Thus, we can conclude that all the recently identified serine proteases that are present in this region could be considered members of the expanded human kallikrein multigene family.

Evidence is accumulating that many of the kallikreins and kallikrein-like genes that are clustered in the same chromosomal region are related to malignancy. KLK3 is the best marker for prostate cancer so far (13). A recent report provided evidence that KLK3 has antiangiogenic activity, and that this activity may result from its action as a serine protease. This study also suggested that other serine proteases and members of the kallikrein multigene family of enzymes should be evaluated for potential antiangiogenic action (45). Recent literature suggests that KLK2 could be a useful diagnostic marker for certain subtypes of prostate cancer (46). KLK10 is a tumor suppressor gene (47). KLK6 is differentially expressed in primary breast and ovarian tumors (18), and the human stratum corneum chymotryptic enzyme (KLK7) has been shown to be expressed at abnormally high levels in ovarian cancer (48). Another recently identified kallikrein-like gene, the tumor-associated differentially expressed gene-14 (TADG-14/Neurospin; KLK8) was found to be overexpressed in a subset of ovarian cancers (49). Also, KLK4 is speculated to be linked to prostate cancer (21).

It is possible that some of these kallikreins participate in pathways which are involved in tumor initiation, progression, or suppression. Identifying the relative positions of these genes in relation to one another may be important for their regulation. Many of these genes have already been found to be regulated by steroid hormones, as summarized elsewhere (5). We anticipate that the detailed description of the genomic organization of the human kallikrein gene family will facilitate further progress in the field and accelerate discoveries which may have practical clinical value.

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

Genomics


