Molecular Characterization of Zyme/Protease M/Neurosin (PRSS9), A Hormonally Regulated Kallikrein-like Serine Protease

George M. Yousef,* † Liu-Ying Luo,* † Stephen W. Scherer, ‡ Georgia Sotiropoulou, § and Eleftherios P. Diamandis* † 1

*Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada; † Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada; ‡ Department of Genetics, Hospital for Sick Children, Toronto, Ontario, Canada; and § Department of Pharmacy, University of Patras, Patras, Greece

Received July 14, 1999; accepted September 28, 1999

The cDNA for the zyme/protease M/neurosin gene (HGMW-approved symbol PRSS9) has recently been identified. Zyme appears to play a role in Alzheimer disease as well as in breast cancer. In this paper, we describe the complete genomic organization of the zyme gene. Zyme spans 10.5 kb of genomic sequence on chromosome 19q13.3-q13.4. The gene consists of seven exons, the first two of which are untranslated. All splice junctions follow the GT/AG rule, and the intron phases are identical to those of many other genes belonging to the same family, i.e., the kallikreins, NES1, and neuropsin. Fine-mapping of the genomic locus indicates that zyme lies upstream of the NES1 gene and downstream from the PSA and KLK2 genes. Tissue expression studies indicate that zyme is expressed mainly in brain tissue, including spinal cord and cerebellum, in mammary gland, and in kidney and uterus. Zyme is regulated by steroid hormones in the breast carcinoma cell line BT-474. Estrogens and progestins, and to a lesser extent androgens, up-regulate the zyme gene in a dose-dependent manner. © 1999 Academic Press

INTRODUCTION

Serine proteases are a group of protein-cleaving enzymes that contain a serine residue in their active sites. These enzymes play important roles in diverse physiological processes. The kallikreins are a subfamily of serine proteases, originally defined as enzymes cleaving vasoactive peptides (kinins) from kinogens (Schachter, 1980). Currently, the kallikreins comprise a large gene family in rodents (Evans et al., 1987; Gerald et al., 1986; Clements, 1997). In humans, the kallikrein (KLK)2 gene family was until recently known to include only three members: the tissue kallikrein (KLK1) (Evans et al., 1988), the glandular kallikrein (KLK2) (Schedlich et al., 1987), and prostatic-specific antigen (KLK3, PSA) (Riegman et al., 1989). More recently, newly discovered genes are also thought to be members of the kallikrein family, including the normal epithelial cell-specific 1 gene (NES1) (Liu et al., 1996), zyme/protease M/neurosin1 (Little et al., 1997; Anisowicz et al., 1996; Yamashiro et al., 1997), and other kallikrein-like genes (Yousef et al., 1999).

The zyme/protease M/neurosin gene (hereafter referred to as zyme) is a new member of the serine protease family and is most homologous to trypsin and other members of the kallikrein gene family (Little et al., 1997; Anisowicz et al., 1996; Yamashiro et al., 1997). The cDNA of this gene was first isolated using a differential display technique from primary and metastatic breast cancer cell lines, and it was named protease M (Anisowicz et al., 1996). Protease M was shown to be down-regulated in metastatic breast cancer cell lines but is strongly expressed, at the messenger RNA level, in some primary breast cancer cell lines and in ovarian cancer tissues and tumor cell lines (Anisowicz et al., 1996). The same gene was also cloned from a cDNA library prepared from a human colon adenocarcinoma cell line (COLO 201), and it was named neurosin (Yamashiro et al., 1997). Neurosin was found to be highly expressed in the brain, followed by a lower, but significant, level of expression in the spleen. Finally, the same cDNA was cloned from Alzheimer disease (AD) brain by polymerase chain reaction (PCR) amplification, with primer sequences representing ac-

2 Abbreviations used: AD, Alzheimer disease; KLK, kallikrein; KLK-L, kallikrein-like; PCR, polymerase chain reaction; PSA, prostate-specific antigen; dNTPs, deoxynucleoside triphosphates; BAC, bacterial artificial chromosome; PAC, P1-derived artificial chromosome; SDS, sodium dodecyl sulfate; BSA, bovine serum albumin; SSC, saline-sodium citrate; LLNL, Lawrence Livermore National Laboratory; RT, reverse transcription.

1 The HGMW-approved symbol for the gene described in this paper is PRSS9.
TABLE 1

Primer name  Sequence*  Coordinates  GenBank Accession No.  Gene name
Z1S  GACCCTGACATGTGACATCTA  979–999  U62801  Zyme
Z1AS  GCCACCTCCTGATGGAGACTG  1422–1402  Zyme
Z2S  TACCAAGCTGCCCTCATCAC  348–367  U62801  Zyme
Z2AS  ACAAGGCTGGAAGGGGTGTC  877–858  Zyme
Z3S  CGGACCAAAGGCCGATTGTTC  4–23  U62801  Zyme
Z3AS  GCCATGCACCAAATTTATCT  301–320  Zyme
PS  CCCAACCTGTTTCTCTCTCTCT  3634–3653  M27274  PSA
PAS  GCCCTCCTCCTCCCATGACAGA  4143–4118  PSA
K1S  ATCCCTTCATCCCATCTTT  2–2 exon 3  M33105  KLK1
K1AS  CACATAAAATTCTTCTGGTTC  324–305 exon3  KLK1
K2S  AGTGACACTGGTCCTCAGAATT  131–150  M18157  KLK2
K2AS  CCCAATTCACAGTGCACTGCAGGAC  580–561  KLK2
NS  GTCTCCCTACCGCTGCAGTCTCTCTCTG  552–570  AF024605  NES1
NAS  CACTCTGGCAAGGGTCCTG  763–744  NES1

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

Materials and Methods

Amplification of the zyme gene by PCR. Two primers (Z1S/Z1AS) were designed in the 3'- untranslated region of zyme cDNA based on the published sequence (Little et al., 1997) (see Table 1). PCR was performed with human genomic DNA from normal individuals in a 50-μl reaction mixture containing 100 ng DNA, 10 mM Tris–HCl (pH 8.3), 50 mM KC1, 2.5 mM MgCl2, 200 μM dNTPs (deoxynucleoside triphosphates), 150 ng primers, and 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 10 min, followed by 43 cycles of 94°C for 30 s, 64°C for 1 min, and a final extension step at 64°C for 9 min. The PCR products were electrophoresed on a 2% agarose gel.

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.

Long PCR amplification of the zyme gene. Two pairs of zyme-specific primers were designed, based on the published zyme cDNA sequence (Z2S/Z2AS and Z3S/Z3AS). PCR was performed using the Expand Long Template PCR System (Boehringer Mannheim GmbH, Mannheim, Germany) with human BAC clones that were positive for the zyme gene, as a template. We used a 50-μl reaction mixture containing 100 ng DNA, 20 mM Tris–HCl (pH 7.5), 100 mM KCl, 2.25 mM MgCl2, 1 mM dithiothreitol, 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, 65°C for 30 s, 68°C for 6 min, 20 cycles of 94°C for 10 s, 65°C for 30 s, 68°C for 6 min (increasing by 20 s per subsequent cycle) (4 min of extension only for the second set of primers), and a final extension
Purification and sequencing of zyme long PCR products. PCR products were purified with exonuclease I plus shrimp alkaline phosphatase treatment as described elsewhere (Birren et al., 1989). Briefly, 1 µl of exonuclease I and 1 µl of shrimp alkaline phosphatase were added to the PCR products, and then the mixture was incubated at 37°C for 15 min and then at 85°C for 15 min. Excess oligonucleotides and dNTPs were removed by a spin dialysis method using Centricon 30 microconcentrators (Amicon, Beverly, MA). Four sequential cycles of centrifugation and dilution were performed as described (Sambrook et al., 1989). After purification, PCR products were directly sequenced using an automated DNA sequencer.

Chromosomal localization of the zyme gene by somatic cell hybrid and radiation hybrid mapping. A panel of human–rodent somatic cell hybrids each containing a single human chromosome (NIGMS human–rodent, somatic cell hybrid, mapping panel 2) and a whole-genome radiation hybrid panel (GeneBridge 4, Research Genetics, Huntsville, AL) were PCR-screened using the zyme-specific primers Z1S and Z1AS. PCR was carried out as described above. The PCR results of the radiation hybrid panel were submitted to the Whitehead Institute/MIT Center for Genome Research (http://www.genome.wi.mit.edu) for statistical localization of the gene.

Identification of positive PAC and BAC genomic clones from human genomic DNA libraries. The PCR product generated with primer set Z1S/Z1AS was purified, labeled with 32P by the random primer method (Sambrook et al., 1989), and used as a probe to screen a human genomic DNA BAC library (Roswell Park Cancer Institute, RPCI-11). The filters were hybridized in 15% formamide, 500 mM Na2HPO4, 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). A PAC clone positive for NES1 was identified by a similar methodology as described elsewhere (Luo et al., 1998). Purification of BAC and PAC DNA was achieved by a rapid alkaline lysis miniprep method, which is a modification of the standard Qiagen-Tip method (Vogelstein and Gillespie, 1979). Positive clones were further confirmed by Southern blot analysis as described (Sambrook et al., 1989).

Gene-specific amplification of other genes from genomic DNA. According to the published sequence of PSA, human renal kallikrein (KLK2), NES1, and zyme gene, we designed gene-specific primers for each of these genes (see Table 1) and developed PCR-based amplification protocols, which allowed us to generate specific PCR products with genomic DNA as a template. PCR were carried out as described above but using an annealing/extension temperature of 65°C. The PCR products were purified using a QIAquick PCR purification kit (Qiagen Inc., Valencia, CA) according to the manufacturer's instructions and then cloned and sequenced to verify their identity, as described above. These PCR products were used only with PAC and BAC clones as templates; to assist with the mapping.

DNA sequences on chromosome 19. Sequencing information for this chromosome is available at the Web site of the Lawrence Livermore National Laboratory (LLNL; http://www-bio.llnl.gov/genome/). We have obtained approximately 300 kb of genomic sequence from that Web site, encompassing a region on chromosome 19q13.3–q13.4, where the known kallikrein genes are localized. This 300 kb of sequence is represented by nine contigs of variable lengths. By using a number of different computer programs, and with the aid of an EcoRI restriction map of that area (also available from the LLNL Web site) and the computer program WebCutter (http://www.firstmarket.com/cutter/cut2.html), we performed a restriction study analysis of the available sequence and were able to construct an almost contiguous sequence map of the region. By using the published cDNA sequences of PSA, KLK2, NES1, and protease M and the computer software BLAST 2 (Altschuel et al., 1997), we were able, using alignment strategies, to identify the relative positions of these genes on the contiguous map.

### Table 2

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer name</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zyme</td>
<td>Z2S, Z2AS</td>
<td>531</td>
</tr>
<tr>
<td></td>
<td>pS2, pS2AS</td>
<td>328</td>
</tr>
<tr>
<td>PSA</td>
<td>PSAAS, PSAAS</td>
<td>754</td>
</tr>
</tbody>
</table>

*All nucleotide sequences are given in the 5’ → 3’ orientation.

Stimulation experiments with the breast cancer cell line BT-474. The breast cancer cell line BT-474 was purchased from the American Type Culture Collection (ATCC; Rockville, MD). BT-474 cells were cultured in RPMI medium (Gibco BRL, Gaithersburg, MD) supplemented with glutamine (200 mmol/L), bovine insulin (10 mg/L), fetal bovine serum (10%), antibiotics, and antimycotics, in plastic flasks, to near confluency. The cells were then aliquoted into 24-well tissue culture plates and cultured to 50% confluence. 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 to the culture medium, at a final concentration of 10⁻⁹ M. Cells stimulated with 100% ethanol were included as controls. The final ethanol concentration was less than 1%. The cells were cultured for 24 h and then harvested for mRNA extraction.

Reverse transcriptase-polymerase chain reaction (RT-PCR). Total RNA was extracted from the breast cancer cells using Trizol reagent (Gibco BRL) following the manufacturer’s instructions. RNA concentration was determined spectrophotometrically. Two micrograms of total RNA was reverse-transcribed into first-strand cDNA using the Superscript preamplification system (Gibco BRL). The final volume was 20 µl. 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 (deoxynucleoside triphosphates), 150 ng of primers, and 2.5 units of AmpliTaq Gold DNA polymerase on a Perkin–Elmer 9600 thermal cycler. The cycling conditions were 94°C for 12 min to activate the Taq Gold DNA polymerase, followed by 43 cycles of 94°C for 30 s, 68°C for 1 min, and a final extension at 68°C for 10 min. Equal amounts of PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. The sequences of the amplification primers used for RT-PCR are described in Table 2. All primers for RT-PCR spanned at least two exons to avoid contamination by genomic DNA.

### Tissue expression of zyme

We purchased total RNA isolated from 26 different human tissues, from Clontech (Palo Alto, CA). Then, we prepared cDNA as described above for the tissue culture experiments and used it for PCR amplification with the primers described in Table 2. Tissue cDNAs were analyzed at different dilutions to obtain semiquantitative information on zyme expression.

### RESULTS

**Sequence Identification of the Zyme Gene**

According to the published cDNA sequence of the zyme gene (Anisowicz et al., 1996), we designed two PCR primers (Z2S/Z2AS), to amplify, by long PCR, an
area that covers most of the coding and noncoding regions of the gene. By using genomic DNA as a template, we obtained a fragment of ~10 kb in length. This fragment was purified and sequenced. The sequence of the PCR product was found to contain part of the corresponding sequence of the cDNA of zyme and, in addition, sequences of intervening introns, as expected. Another long PCR was performed using a different set of gene-specific primers (Z3S/Z3AS) to elucidate the 5′ region of the gene. The PCR products were sequenced and verified to contain parts of the zyme cDNA sequence plus two additional intervening introns. Further confirmation of the sequence was accomplished by BLAST 2 analysis (Altschul et al., 1997).
with sequences obtained from two contigs constructed at the LLNL (contigs 26 and 27). We found that these contigs were adjacent to each other, based on the EcoRI restriction map of chromosome 19.

Genomic Structure of the Zyme Gene

By comparing the genomic sequences identified above with the published cDNA sequences, we were able to determine the genomic organization of the zyme gene. Figure 1 demonstrates that the zyme gene is made up of seven exons separated by six introns. The gene covers 10.5 kb of genomic sequences on chromosome 19q13.3 and is processed to an mRNA that is 1526 nucleotides in length plus a poly(A) tail. All splice junctions agree with the conserved consensus sequence GT/AG (Lida, 1990). The first two exons are untranslated, and the start codon is on the third exon. The exact transcription initiation site is not known, and no TATA box or CAAT sequences were identified upstream of the longest published cDNA sequence of the gene.

Mapping and Chromosomal Localization of the Zyme Gene

We used PCR and two gene-specific primers (Z1S/ Z1AS) to screen a panel of somatic cell hybrids, each containing a single human chromosome. The zyme gene was found to reside on chromosome 19. To determine precisely the zyme locus, the same PCR was performed on a panel of GeneBridge 4 radiation hybrid DNA. The expected PCR bands were obtained from 29 of 93 DNAs from the hybrid cell lines. Computer analysis of these results assigned the zyme gene to chromosome 19 between markers NIB1805 and WI-5264. In addition, the PSA gene (WI-9055; GenBank Accession No. M21896) and the NES1 gene were mapped between markers NIB1805 and WI-5264 by the same method, indicating that they all reside on the same chromosomal locus. The PSA, KLK1, and KLK2 genes were previously assigned to be clustered in an area of ~60 kb in the same genomic region (Riegman et al., 1992). We thus decided to study the relative positions of these genes on chromosome 19q13.3–q13.4.

Relative Position of PSA, KLK2, Zyme, and NES1 on Chromosome 19q13.3–q13.4

Screening of the human BAC library identified two clones that were positive for the zyme gene (clones BAC H-N0288H01 and BAC H-BH0076F07). These BACs were further analyzed by PCR and primers specific for PSA, NES1, KLK1, and KLK2 (the primers are shown in Table 1). These analyses indicated that both BACs were positive for zyme, PSA, and KLK2 and negative for the KLK1 and NES1 genes. Screening of a human PAC genomic library identified a PAC clone that was positive for NES1 (PAC H-DJ0043B01) (Luo et al., 1998). Further PCR analysis indicated that this PAC clone was positive for NES1 and KLK1 genes and negative for PSA, KLK2, and zyme. Combination of this information with the EcoRI restriction map of the region allowed us to establish the relative positions of these four genes. PSA is the most centromeric, followed by KLK2, Zyme, and NES1, respectively. Further alignment of the known sequences of these genes with the 300-kb area of the contigs constructed at the LLNL (Table 3) enabled us to localize precisely all four genes and determine the coding strand for each (Fig. 2). Contigs 5 and 7 were found to be adjacent to each other with a 2.2-kb overlap between them. By BLAST 2 sequence analysis, the PSA gene was assigned to contig 5 and KLK2 to contigs 7 and 8; the distance between the two genes was calculated to be 12,508 bp.

The 5’ end of the zyme gene was mapped close to the

---

**TABLE 3**

<table>
<thead>
<tr>
<th>Clone ID</th>
<th>Contig number</th>
<th>Contig length (kb)</th>
<th>Genomic content</th>
</tr>
</thead>
<tbody>
<tr>
<td>F 25479</td>
<td>5</td>
<td>35.6</td>
<td>PSA</td>
</tr>
<tr>
<td>4</td>
<td>4.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R33359</td>
<td>8</td>
<td>35.8</td>
<td>KLK2</td>
</tr>
<tr>
<td>7</td>
<td>5.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BC 85745</td>
<td>27</td>
<td>56.0</td>
<td>Zyme</td>
</tr>
<tr>
<td>26</td>
<td>32.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BC 349142</td>
<td>37</td>
<td>80.0</td>
<td>NES1</td>
</tr>
<tr>
<td>36</td>
<td>51.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Contigs and clone IDs are named all according to the Lawrence Livermore National Laboratory database (see also Fig. 2).

---

**FIG. 2.** Approximately 300 kb of contiguous genomic sequence around chromosome 19q13.3–q13.4 represented by seven contigs; each contig is shown with its approximate length in kilobases. The contig numbers refer to those reported at the Lawrence Livermore National Laboratory Web site. Vertical arrows denote contig boundaries. Genes are represented by horizontal arrows denoting the direction of the gene coding sequence. See text for accurate distances between genes and for full names of genes. The figure is not drawn to scale.
end of contig 26, and the rest of the gene matched the reverse complementary strand of contig 27. According to the EcoRI restriction map, the reverse complementary strand of contig 27 and contig 8 are adjacent, and the zyme gene is about 80 kb away from the KLK2 gene, in the complementary DNA strand. There is an area of overlap of 1.7 kb between contig 26 and the reverse complementary strand of contig 37, and thus, the NES1 gene is 43,468 bp telomeric to the zyme and is transcribed in the same direction (Fig. 2). The complete intronic sequences of zyme have been deposited with GenBank.

Tissue Expression of Zyme

We have examined by RT-PCR which human tissues express zyme. The experiments were performed at various dilutions of the cDNAs. RT-PCR for actin was used as a control, and RT-PCR for PSA was used as another control with known tissue-restricted specificity. The PSA gene was found to be highly expressed in the prostate, as expected, and to a lesser extent in mammary gland and salivary gland, as also expected from recent literature reports (Diamandis et al., 1994; James et al., 1996; Zarghami and Diamandis, 1996). We further found a very low level of PSA expression in the thyroid gland, trachea, and testis, a finding that, to our knowledge, has not been reported earlier.

The tissue expression of zyme is summarized in Fig. 3. This protease is primarily expressed in the brain, spinal cord, kidney, and mammary tissues and to a lesser extent in many other tissues (Table 4).

Hormonal Regulation of Zyme

We have used the steroid hormone receptor-positive breast carcinoma cell line BT-474 to investigate the hormonal regulation of zyme by RT-PCR. As shown in Fig. 4, the controls performed as expected, i.e., actin positivity without hormonal regulation in all cDNAs, only estrogen up-regulation of the pS2 gene, and up-regulation of the PSA gene by androgens and progestins. Zyme is up-regulated primarily by estrogens and progestins and to a lesser extent by androgens. This up-regulation was dose responsive and easily recognized at steroid hormone levels $>10^{-10}$ M (Fig. 5).

**DISCUSSION**

Kallikrein genes are a subfamily of serine proteases, traditionally characterized by their ability to liberate lysyl-bradykinin (kallidin) from kininogen (Schachter, 1980). More recently, however, a new, structural concept has emerged to describe kallikreins. From accumulated DNA sequence data, it is now clear that the mouse has many genes with a high degree of homology.
to kallikrein coding sequences (Clements, 1997; Morris et al., 1981; Richards et al., 1982). Richards and co-workers have contributed to the concept of a “kallikrein multigene family” to refer to these genes (Van Leeuwen et al., 1986; Evans and Richards, 1985). 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, the functional definition of a kallikrein is met only by KLK1. KLK2 has relatively potent trypsin-like enzymatic activity, and KLK3 (PSA) has relatively weak chymotrypsin-like enzymatic activity. The known enzymatic action and function of KLK2 and KLK3 do not liberate biologically active peptides from precursors. Based on the structural definition, members of the kallikrein family should include not only the gene for KLK1, but also genes encoding other homologous proteases that reside in the same chromosomal region and show some degree of homology. Therefore, it is important to underscore the clear distinction between the enzyme kallikrein and a kallikrein or a kallikrein-like gene. Based on the structural definition of kallikreins, zyme, NES1, and a number of other human genes (Yousef et al., 1999) qualify as members of this family.

The genomic organization of the zyme gene, presented in Fig. 1, demonstrates that it is made up of seven exons and six intervening introns. The gene spans 10.5 kb on chromosome 19q13.3 and is processed to yield an mRNA that is ~1526 nucleotides in length plus a poly(A) tail. All splice junctions agree with the consensus sequence GT/AG (Lida, 1990).

When the protein coding sequence of zyme was compared with the four most closely related proteins, it was found that glandular kallikrein (KLK2) has 44% exact matches and 48% matches with conserved changes. Trypsin I has 43% identity, and both KLK1 and PSA contain 39% exact matches and 44% matches with conservative changes (Anisowicz et al., 1996). The catalytic triad of serine proteases is conserved in zyme (i.e., histidine-62, aspartate-106, and serine-197; amino acid numbers refer to the published protease M sequence, GenBank Accession No. U62801). The presence of aspartate at position 191 indicates that the protein is a trypsin-like enzyme. Zyme contains 12 cysteine residues. Ten of these are conserved in KLK1, KLK2, PSA, and human trypsin. Twenty-seven of 29 “invariant” amino acids surrounding the active site of serine proteases (Dayhoff, 1978) are conserved in zyme, and one of the nonconserved amino acids (His-161 instead of Pro) is also found in glandular kallikrein and PSA.

Irwin et al. (1988) have proposed that the serine protease genes can be classified into five different groups according to intron position. The kallikreins 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. As seen in Fig. 6, the genomic organization of zyme is very similar to that of this group of genes. The lengths of the coding parts of exons 3–7 of zyme are 48, 157, 248, 137, and 153 bp, respectively, which are almost identical to the lengths of the exons of the kallikrein genes and also quite similar to those of many newly discovered genes.
in the same chromosomal region, including the NES1 (Liu et al., 1996; Luo et al., 1998) and neuoropsin (Yoshida et al., 1998) genes. One unique feature of zyme is the presence of two untranslated exons in the 5’ region, since no such exons have been reported in this position except for one exon in both the NES1 and the neuoropsin genes (Luo et al., 1998; Yoshida et al., 1998). Another highly conserved feature of all these genes is the intron phase, as shown in Fig. 6 and as explained in the legend to Fig. 6.

Through phylogenetic studies, the zyme gene was aligned with a group of proteases that includes chymotrypsin, trypsinogens I–IV, and the three kallikrein genes (Little et al., 1997). Taken together, these data indicate that these serine proteases are all derived from a common ancestral gene by the mechanism of exon shuffling.

No TATA box or CAAT sequences were found in the 5’ flanking region of the zyme gene. The same observation was made for other kallikreins like the neuorpsin gene (Yoshida et al., 1998) and the prothrombin gene (Irwin et al., 1988). However, using a promoter prediction program that is based on detection of multiple functional sites that are involved in the RNA polymerase-binding process, such as the TATA box and the transcription start site (“initiator”), a segment –281 to –320 bp from the mRNA start of Fig. 1 was detected as a putative promoter sequence. We should thus consider that the actual transcription start site is not yet fully ascertained and needs to be further studied.

A single PAC clone was positive for both NES1 and KLK1 genes and negative for zyme, PSA, and KLK2. Also, two BAC clones were positive for zyme, PSA, and KLK2 but negative for NES1 and KLK1. Moreover, the KLK1 gene sequence was not identified on any of the studied contigs. Taken together, these data suggest that the KLK1 gene is further telomeric to NES1, in contradiction to the previously published data suggesting that it is more centromeric to PSA (Riegman et al., 1992). We did not attempt to characterize more precisely the genomic position of the KLK1 gene.

Our tissue expression studies (Table 4 and Fig. 3) indicate, in agreement with previous reports (Little et al., 1997; Anisowicz et al., 1996; Yamashiro et al., 1997), that zyme is most highly expressed in brain tissue, including cerebellum and spinal cord. Significant expression was also seen in kidney and mammary tissues. A lower level of expression is notable in many other tissues. The tissue expression pattern of zyme is not consistent in all studies, presumably due to technical differences, i.e., RT-PCR (this study) versus Northern blotting, which is much less sensitive. In this paper, we provide the first indication that zyme is regulated by steroid hormones but in a more complex manner than PSA or pS2 (Figs. 4 and 5). However, the qualitative nature of RT-PCR does not allow drawing of definitive conclusions, and currently, we do not possess antibodies against zyme to verify the data by Western blot analysis. Up-regulation by estrogens, progestins, and androgens has also been observed for the homologous gene NES1 (Luo et al., submitted for publication) that was recently classified as a tumor suppressor (Goyal et al., 1998).

A more unified theme appears to be emerging regarding KLK2, PSA, zyme, and NES1, a group of homologous proteins that are all localized on chromosome 19q13.3–q13.4 (Fig. 2). All four genes are serine proteases, they are all regulated by steroid hormones, and they are all expressed in prostatic as well as mammary tissues. Strikingly, PSA was found to be down-regulated in breast cancer, and more recently, low levels of PSA were found to be associated with unfavorable outcome and more aggressive forms of breast cancer (Yu et al., 1995, 1996, 1998). Low levels of PSA in nipple aspirate fluids were associated with higher risk for breast cancer (Sauter et al., 1996). High levels of PSA decrease cell proliferation (Lai et al., 1996) and induce apoptosis (Balbay et al., 1999). Similarly, NES1 is down-regulated in breast cancer (Liu et al., 1996), and more recently, it was found to be a novel tumor suppressor (Goyal et al., 1998). KLK2 is an activating enzyme for PSA and may be part of a regulatory loop controlling PSA enzymatic activity (Lovgren et al., 1997; Magklara et al., 1999). Zyme is now also known to be significantly down-regulated in breast cancer in comparison to normal tissue (Anisowicz et al., 1996).

The role of these genes in breast cancer initiation and progression merits further investigation. It is clear from current knowledge that all of them are down-regulated in breast cancer in comparison to normal breast tissue. Furthermore, the literature suggests that 19q13 is rearranged in a variety of solid tumors (Mitelman, 1994). It will be interesting to examine whether any of these genes are disrupted by rearrangements in breast cancer and other cancers. Also, loss of heterozygosity studies on gliomas have shown a commonly deleted region (81%) at 19q13.2–q13.4 (Reifenberger et al., 1994). It should be investigated whether this group of genes is involved in these deletions.

REFERENCES


