Vol. 159, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
February 28, 1989 Pages 95-102

CHARACTERIZATION OF THE PROSTATE-SPECIFIC ANTIGEN GENE: A NOVEL HUMAN KALLIKREIN-LIKE GENE

P.H.J. Riegman[1], R.J. Vlietstra[1], J.A.G.M. van der Korput[1], J.C. Romijn[2] and J. Trapman[1]

[1]Department of Pathology and [2]Department of Urology, Erasmus University, P.O.Box 1738, 3000 DR Rotterdam,The Netherlands

Received December 15, 1988

Using Prostate-specific Antigen cDNA fragments as hybridization probes a clone containing the information for the gene encoding Prostate-specific Antigen was isolated from a human genomic DNA library. The complete gene (about 6 kb) was sequenced and shown to be composed of four introns and five exons. Two major transcription initiation sites were found. The sequence of the promoter region revealed the presence of various well known transcription regulatory elements including a TATA box. A high percentage of homology was found between the Prostate-specific Antigen gene and the hGK-1 gene (82%). This homology extended into the promoter region. Two previously described variant Prostate-specific Antigen cDNAs can now be explained by intron retention and alternative splicing of the primary transcript.

Tissue kallikreins and kallikrein-like proteins are a subgroup of closely related serine proteases (see for recent reviews 1 and 2). They are able to cleave specific precursor proteins at highly selective sites, generating in this way the mature, biologically active proteins. Kallikrein-like proteins are produced by many different tissues, including salivary glands, kidney, prostate and pancreas (1,2). In rat and mouse kallikrein-like proteases are encoded by a multigene family (3-7). The total number of kallikrein-like genes in these species is between 15 and 25. In man the gene family seems to be much smaller (8-10). So far, the existence of three different members has been described. One gene (hGK-1) was completely sequenced (10), of two others (Prostate-specific Antigen (PA) (11,12) and kidney/pancreas kallikrein (8,9)) the cDNA structure has been established.

PA is a 35 kD glycoprotein, that is exclusively synthesized by the epithelial cells of the prostate gland (13-16). Its natural substrates probably are proteins, secreted by seminal vesicles, which cause the gel like structure of the semen (17,18). Recently, PA has attracted much attention as a reliable marker for human prostate cancer (19-22). Normally, PA is not detectable in the serum. In case of invasive growth and/or metastasis of prostate tumors significantly elevated PA levels can be measured. Because of its clinical importance, detailed knowledge of regulation of PA expression is of high interest.

0006-291X/89 $1.50

Previously we reported the molecular cloning and characterization of three different PA cDNAs and the detection of seven different PA transcripts (12). In this study the isolation and structure of the PA gene including part of the promoter region is described. In addition the mutual relationship between various variant PA cDNAs is clarified.

Methods

A human genomic library (partially Mbo I digested DNA in lambda EMBL 3) was kindly provided by Dr. G. Grosveld (Rotterdam). The library was screened with a [^{32}P]-labelled (23) 380 bp Eco RI-Sac I fragment derived from the cDNA clone PA 525 (12). Duplicate nitrocellulose filters were hybridized overnight at 65°C in 6xSSC containing 10x Denhardt, 0.1% SDS, and 100μg/ml salmon sperm DNA. Filters were washed twice in 3x SSC for 20 minutes at 65°C and twice in 1x SSC for 20 minutes at 65°C and once at 0.3x SSC for 10 minutes at 65°C. Filters were exposed to X-ray films for 18 hours at -70°C using intensifier screens. Positive phages were purified by two isolation cycles.

Phages were propagated, DNA isolated and fragments were subcloned into pUC9 using standard procedures (24,25). Nucleotide sequences of appropriate fragments subcloned into M13mp18/19 were determined by the dideoxy chain termination method (26) using sequenase (USB, Cleveland) and dGTP or dITP.

Primer extention experiments were performed using a method essentially identical as described in ref. 27, but omitting the dideoxy nucleotides in the reverse transcriptase mix. As primer a 22-mer oligonucleotide with the sequence 5'-TGCAGCACCAATCCACGTCACG-3' was used. This oligonucleotide is PA-specific (compare Fig. 2 position 74 to 87 and 1323 to 1330 with hGK-1 (ref. 10)). PolyA+ RNA was isolated from the human prostate tumors PC EW and PC 82, using standard procedures. Annealing of primer (5 ng) with polyA+ RNA (3μg) was for 6 hours at 65°C.

Results

Using PA cDNA fragments as hybridization probes a genomic DNA fragment (designated 5P1) containing the complete PA gene was isolated. The restriction map of clone 5P1 is depicted in Fig. 1A. A more detailed map of the region at which the PA gene is situated is presented in Fig. 1B. Fig. 2 shows the complete sequence of the PA gene including parts of the 5'- and 3'-flanking regions. Comparison of the genomic sequence with that of the most abundant PA cDNA (PA 75, corresponding to a 1.5 kb mRNA; Fig. 1C and ref. 12) revealed that the PA gene is composed of five exons and four introns. The total size of the gene is approximately 6 kb. As determined by primer extention experiments, using polyA+ RNA from two different human prostate tumors (PC 82, PC EW), two transcriptional initiation sites were found at 42, respectively 35 bp upstream from the first ATG of the open reading frame (see Fig. 3). The exons have a size of 81 or 87 bp (exon I), 160 bp (exon II), 287 bp (exon III), 137 bp (exon IV) and 793 bp (exon V). The respective introns have a size of 1235, 1628, 145 and 1373 bp. Exon I contains in addition to the leader sequence the information for the major part of the signal peptide. Exon II encodes the last two amino acids of the signal peptide, the seven amino acids of the pro-fragment and the amino-terminal region of the mature protein. Exon V

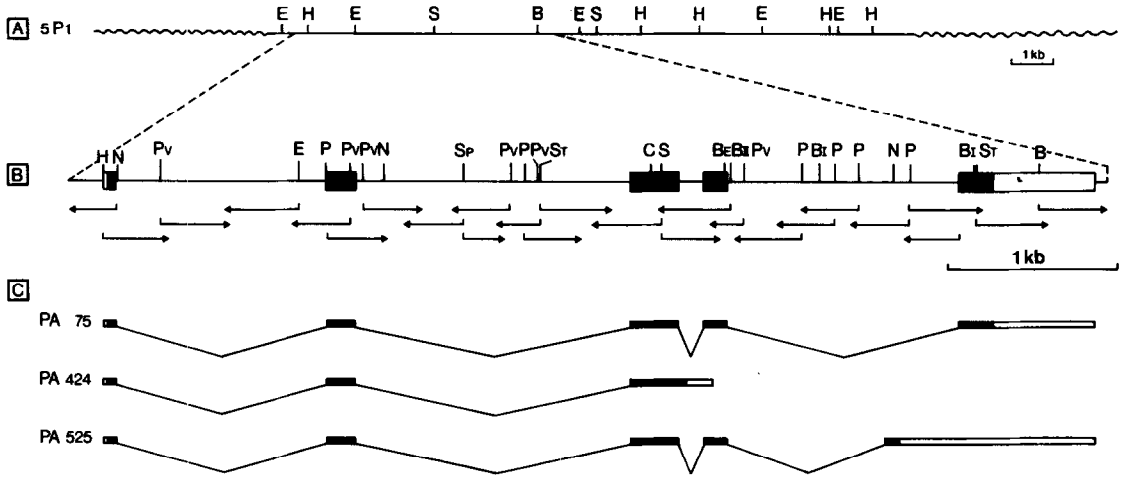
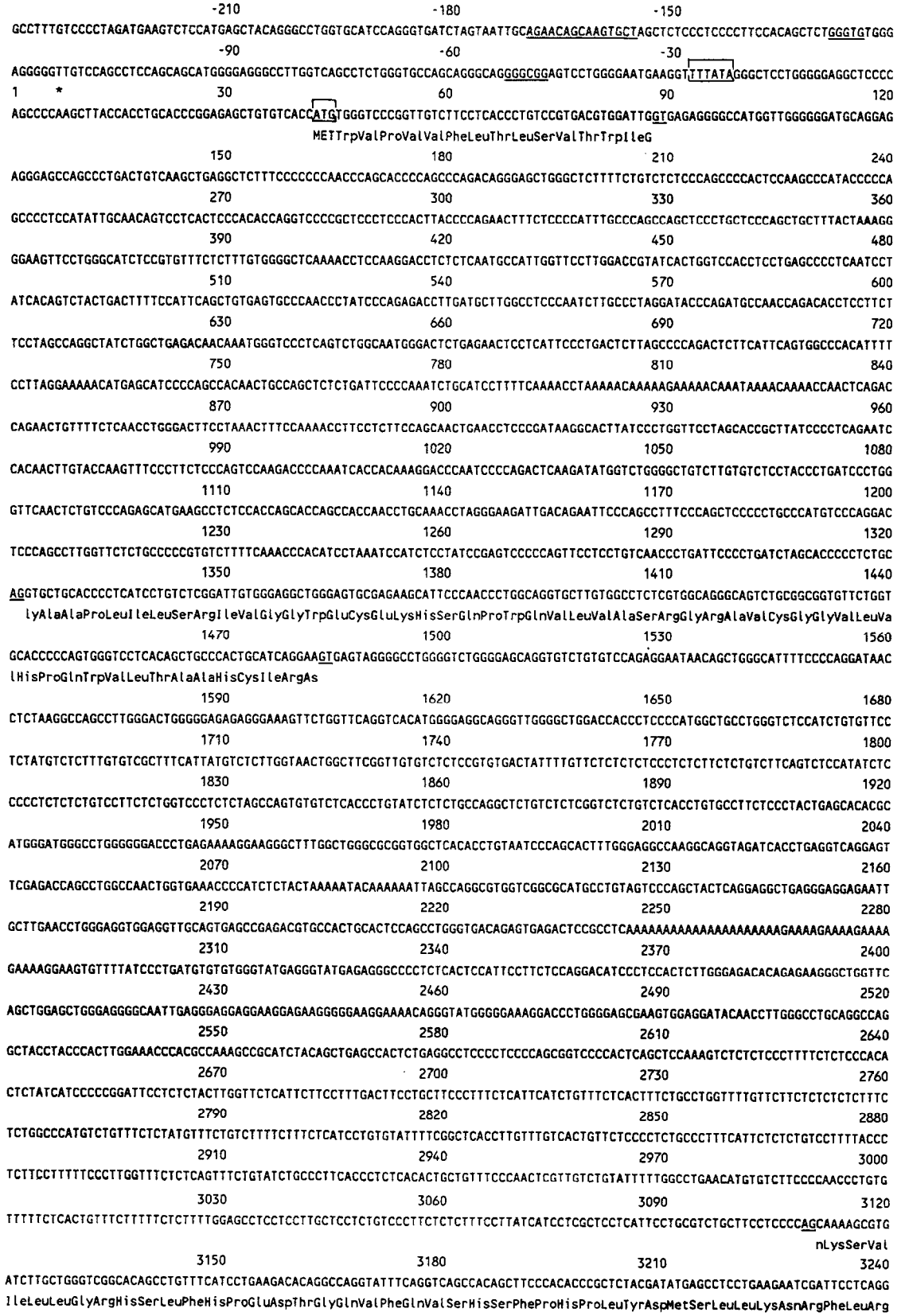


Figure 1
A. Partial restriction map of the genomic clone 5P1.
B. Restriction map of the genomic region containing the PA gene. Arrows indicate the sequence strategy. C. Organization of the different PA cDNAs described so far (see refs 11,12) Closed boxes correspond to the open reading frame. Open boxes represent the remaining regions of the various cDNAs. The boxes in B. correspond to the major PA transcript (see also cDNA clone PA75). B=Bam HI, BI=Bgl I, BII=Bgl II, C=Cla I, E=Eco RI, H=Hind III, N=Nco I, P=Pst I, Pv=Pvu II, S=Sac I, Sl=Sal I, Sp=Sph I and St=Stu I.

provides information for the carboxy-terminal part (51 amino acids) of the protein and the complete 3'-non-coding region.

Seven distinct PA transcripts have been detected (12). In addition to cDNA corresponding to the 1.5 kb mRNA (PA 75), two cDNAs have been characterized in detail: PA 424 and PA 525. Both contain an extra internal fragment as compared to PA 75 cDNA (Fig. 1c and ref.12). In addition, PA 424 is truncated at the 3'-terminus because of the use of an alternative polyadenylation signal sequence (12). From the organization of the PA gene it can now clearly be established that PA 424 contains as extra information the complete retained small intron III (Fig. 1c). PA 525 results from the use of an alternative splice acceptor site upstream from position 4608 instead of the site at position 5051 used in PA 75 (see Fig. 2).

In general, elements important for regulation of transcription are located within a region of 200 bp upstream from the transcription start site. This region of the PA gene is characterized by the presence of several well known regulatory elements. A variant TATA-box (TTTATA) is found at position -28 to -23, a GC-box at -53 to -48, a CACCC-box at -129 to -125. Interestingly, starting at position -170 the imperfect palindromic sequence AGAACAGCAAGTGCT is found, which is closely related to reverse complement of the consensus sequence for steroid (glucocorticoid, progestin, androgen (28)) hormone receptor binding (TGTACANNNTGTC/TCT).


-210 -180 -150
GCCTTTGTCCCCTAGATGAAGTCTCCATGAGCTACAGGGCCTGGTGCATCCAGGGTGATCTAGTAATTGCAGAACAGCAAGTGCTAGCTCTCCCTCCCCTTCCACAGCTCTGGGTGTGGG
-90 -60 -30
AGGGGGTTGTCCAGCCTCCAGCAGCATGGGGAGGGCCTTGGTCAGCCTCTGGGTGCCAGCAGGGCAGGGGCGGAGTCCTGGGGAATGAAGGTTTTATAGGGCTCCTGGGGGAGGCTCCCC
1 * 30 60 90 120
AGCCCCAAGCTTACCACCTGCACCCGGAGAGCTGTGTCACCATGTGGGTCCCGGTTGTCTTCCTCACCCTGTCCGTGACGTGGATTGGTGAGAGGGGCCATGGTTGGGGGGATGCAGGAG
METTrpValProValValPheLeuThrLeuSerValThrTrpIleG
150 180 210 240
AGGGAGCCAGCCCTGACTGTCAAGCTGAGGCTCTTTCCCCCCCAACCCAGCACCCCAGCCCAGACAGGGAGCTGGGCTCTTTTCTGTCTCTCCCAGCCCCACTCCAAGCCCATACCCCCA
270 300 330 360
GCCCCTCCATATTGCAACAGTCCTCACTCCCACACCAGGTCCCCGCTCCCTCCCACTTACCCCAGAACTTTCTCCCCATTTGCCCAGCCAGCTCCCTGCTCCCAGCTGCTTTACTAAAGG
390 420 450 480
GGAAGTTCCTGGGCATCTCCGTGTTTCTCTTTGTGGGGCTCAAAACCTCCAAGGACCTCTCTCAATGCCATTGGTTCCTTGGACCGTATCACTGGTCCACCTCCTGAGCCCCTCAATCCT
510 540 570 600
ATCACAGTCTACTGACTTTTCCATTCAGCTGTGAGTGCCCAACCCTATCCCAGAGACCTTGATGCTTGGCCTCCCAATCTTGCCCTAGGATACCCAGATGCCAACCAGACACCTCCTTCT
630 660 690 720
TCCTAGCCAGGCTATCTGGCTGAGACAACAAATGGGTCCCTCAGTCTGGCAATGGGACTCTGAGAACTCCTCATTCCCTGACTCTTAGCCCCAGACTCTTCATTCAGTGGCCCACATTTT
750 780 810 840
CCTTAGGAAAAACATGAGCATCCCCAGCCACAACTGCCAGCTCTCTGATTCCCCAAATCTGCATCCTTTTCAAAACCTAAAAACAAAAAGAAAAACAAATAAAACAAAACCAACTCAGAC
870 900 930 960
CAGAACTGTTTTCTCAACCTGGGACTTCCTAAACTTTCCAAAACCTTCCTCTTCCAGCAACTGAACCTCCCGATAAGGCACTTATCCCTGGTTCCTAGCACCGCTTATCCCCTCAGAATC
990 1020 1050 1080
CACAACTTGTACCAAGTTTCCCTTCTCCCAGTCCAAGACCCCAAATCACCACAAAGGACCCAATCCCCAGACTCAAGATATGGTCTGGGGCTGTCTTGTGTCTCCTACCCTGATCCCTGG
1110 1140 1170 1200
GTTCAACTCTGTCCCAGAGCATGAAGCCTCTCCACCAGCACCAGCCACCAACCTGCAAACCTAGGGAAGATTGACAGAATTCCCAGCCTTTCCCAGCTCCCCCTGCCCATGTCCCAGGAC
1230 1260 1290 1320
TCCCAGCCTTGGTTCTCTGCCCCCGTGTCTTTTCAAACCCACATCCTAAATCCATCTCCTATCCGAGTCCCCCAGTTCCTCCTGTCAACCCTGATTCCCCTGATCTAGCACCCCCTCTGC
1350 1380 1410 1440
AGGTGCTGCACCCCTCATCCTGTCTCGGATTGTGGGAGGCTGGGAGTGCGAGAAGCATTCCCAACCCTGGCAGGTGCTTGTGGCCTCTCGTGGCAGGGCAGTCTGCGGCGGTGTTCTGGT
lyAlaAlaProLeuIleLeuSerArgIleValGlyGlyTrpGluCysGluLysHisSerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCysGlyGlyValLeuVa
1470 1500 1530 1560
GCACCCCCAGTGGGTCCTCACAGCTGCCCACTGCATCAGGAAGTGAGTAGGGGCCTGGGGTCTGGGGAGCAGGTGTCTGTGTCCAGAGGAATAACAGCTGGGCATTTTCCCCAGGATAAC
lHisProGlnTrpValLeuThrAlaAlaHisCysIleArgAs
1590 1620 1650 1680
CTCTAAGGCCAGCCTTGGGACTGGGGGAGAGAGGGAAAGTTCTGGTTCAGGTCACATGGGGAGGCAGGGTTGGGGCTGGACCACCCTCCCCATGGCTGCCTGGGTCTCCATCTGTGTTCC
1710 1740 1770 1800
TCTATGTCTCTTTGTGTCGCTTTCATTATGTCTCTTGGTAACTGGCTTCGGTTGTGTCTCTCCGTGTGACTATTTTGTTCTCTCTCTCCCTCTCTTCTCTGTCTTCAGTCTCCATATCTC
1830 1860 1890 1920
CCCCTCTCTCTGTCCTTCTCTGGTCCCTCTCTAGCCAGTGTGTCTCACCCTGTATCTCTCTGCCAGGCTCTGTCTCTCGGTCTCTGTCTCACCTGTGCCTTCTCCCTACTGAGCACACGC
1950 1980 2010 2040
ATGGGATGGGCCTGGGGGGACCCTGAGAAAAGGAAGGGCTTTGGCTGGGCGCGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCAAGGCAGGTAGATCACCTGAGGTCAGGAGT
2070 2100 2130 2160
TCGAGACCAGCCTGGCCAACTGGTGAAACCCCATCTCTACTAAAAATACAAAAAATTAGCCAGGCGTGGTCGGCGCATGCCTGTAGTCCCAGCTACTCAGGAGGCTGAGGGAGGAGAATT
2190 2220 2250 2280
GCTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCGAGACGTGCCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCGCCTCAAAAAAAAAAAAAAAAAAAAAAGAAAAGAAAAGAAAA
2310 2340 2370 2400
GAAAAGGAAGTGTTTTATCCCTGATGTGTGTGGGTATGAGGGTATGAGAGGGCCCCTCTCACTCCATTCCTTCTCCAGGACATCCCTCCACTCTTGGGAGACACAGAGAAGGGCTGGTTC
2430 2460 2490 2520
AGCTGGAGCTGGGAGGGGCAATTGAGGGAGGAGGAAGGAGAAGGGGGAAGGAAAACAGGGTATGGGGGAAAGGACCCTGGGGAGCGAAGTGGAGGATACAACCTTGGGCCTGCAGGCCAG
2550 2580 2610 2640
GCTACCTACCCACTTGGAAACCCACGCCAAAGCCGCATCTACAGCTGAGCCACTCTGAGGCCTCCCCTCCCCAGCGGTCCCCACTCAGCTCCAAAGTCTCTCTCCCTTTTCTCTCCCACA
2670 2700 2730 2760
CTCTATCATCCCCCGGATTCCTCTCTACTTGGTTCTCATTCTTCCTTTGACTTCCTGCTTCCCTTTCTCATTCATCTGTTTCTCACTTTCTGCCTGGTTTTGTTCTTCTCTCTCTCTTTC
2790 2820 2850 2880
TCTGGCCCATGTCTGTTTCTCTATGTTTCTGTCTTTTCTTTCTCATCCTGTGTATTTTCGGCTCACCTTGTTTGTCACTGTTCTCCCCTCTGCCCTTTCATTCTCTCTGTCCTTTTACCC
2910 2940 2970 3000
TCTTCCTTTTTCCCTTGGTTTCTCTCAGTTTCTGTATCTGCCCTTCACCCTCTCACACTGCTGTTTCCCAACTCGTTGTCTGTATTTTTGGCCTGAACATGTGTCTTCCCCAACCCTGTG
3030 3060 3090 3120
TTTTTCTCACTGTTTCTTTTTCTCTTTTGGAGCCTCCTCCTTGCTCCTCTGTCCCTTCTCTCTTTCCTTATCATCCTCGCTCCTCATTCCTGCGTCTGCTTCCTCCCCAGCAAAAGCGTG
nLysSerVal
3150 3180 3210 3240
ATCTTGCTGGGTCGGCACAGCCTGTTTCATCCTGAAGACACAGGCCAGGTATTTCAGGTCAGCCACAGCTTCCCACACCCGCTCTACGATATGAGCCTCCTGAAGAATCGATTCCTCAGG
IleLeuLeuGlyArgHisSerLeuPheHisProGluAspThrGlyGlnValPheGlnValSerHisSerPheProHisProLeuTyrAspMetSerLeuLeuLysAsnArgPheLeuArg

```
                    3270                    3300                    3330                    3360
CCAGGTGATGACTCCAGCCACGACCTCATGCTGCTCCGCCTGTCAGAGCCTGCCGAGCTCACGGATGCTGTGAAGGTCATGGACCTGCCCACCCAGGAGCCAGCACTGGGGACCACCTGC
ProGlyAspAspSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeuThrAspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuGlyThrThrCys
                    3390                    3420                    3450                    3480
TACGCCTCAGGCTGGGGCAGCATTGAACCAGAGGAGTGTACGCCTGGGCCAGATGGTGCAGCCGGGAGCCCAGATGCCTGGGTCTGAGGGAGGAGGGGACAGGACTCCTAGGTCTGAGGG
TyrAlaSerGlyTrpGlySerIleGluProGluGluP
                    3510                    3540                    3570                    3600
AGGAGGGCCAAGGAACCAGGTGGGGTCCAGCCCACAACAGTGTTTTTTGCCTGGCCCGTAGTCTTGACCCCAAAGAAACTTCAGTGTGTGGACCTCCATGTTATTTCCAATGACGTGTGT
                                                        heLeuThrProLysLysLeuGlnCysValAspLeuHisValIleSerAsnAspValCys
                    3630                    3660                    3690                    3720
GCGCAAGTTCACCCTCAGAAGGTGACCAAGTTCATGCTGTGTGCTGGACGCTGGACAGGGGGCAAAAGCACCTGCTCGGTGAGTCATCCCTACTCCCAAGATCTTGAGGGGAAAGGTGAG
AlaGlnValHisProGlnLysValThrLysPheMetLeuCysAlaGlyArgTrpThrGlyGlyLysSerThrCysSer
                    3750                    3780                    3810                    3840
TGGGGACCTTAATTCTGGGCTGGGGTCTAGAAGCCAACAAGCATCTGCCTCCCCTGCTCCCCAGCTGTAGCCATGCCACCTCCCCGTGTCTCATCTCATTCCCTCCTTCCCTCTTCTTTG
                    3870                    3900                    3930                    3960
ACTCCCTCAAGGCAATAGGTTATTCTTACAGCACAACTCATCTGTTCCTGCGTTCAGCACACGGTTACTAGGCACCTGCTATGCACCCAGCACTGCCCTAGAGCCTGGACATAGCAGTGA
                    3990                    4020                    4050                    4080
ACAGACAGAGAGCAGCCCCTCCCTTCTGTAGCCCCCAAGCCAGTGAGGGGCACAGGCAGGAACAGGGACCACAACACAGAAAAGCTGGAGGGTGTCAGGAGGTGATCAGGCTCTCGGGGA
                    4110                    4140                    4170                    4200
GGGAGAAGGGGTGGGGAGTGTGACTGGGAGGAGACATCCTGCAGAAGGCGGGAGTGAGCAAACACCTGCCGCAGGGGAGGGGAGGGCCTGCGGCACCTGGGGGAGCAGAGGGAACAGCAT
                    4230                    4260                    4290                    4320
CTGGCCAGGCCTGGGAGGAGGGGCCTAGAGGGCGTCAGGAGCAGAGAGGAGGTTGCCTGGCTGGAGTGAAGGATCGGGGCAGGGTGCGAGAGGGAAGAAGGACCCCTCCTGCAGGGCCTC
                    4350                    4380                    4410                    4440
ACCTGGGCCACAGGAGGACACTGCTTTTCCTCTGAGGAGTCAGGAACTGTGGATGGTGCTGGACAGAAGCAGGACAGGGCCTGGCTCAGGTGTCCAGAGGCTGCCGCTGGCCTCCCTATG
                    4470                    4500                    4530                    4560
GGATCAGACTGCAGGGAGGGAGGGCAGCAGGGATGTGGAGGGAGTGATGATGGGGCTGACCTGGGGGTGGCTCCAGGCATTGTCCCCACCTGGGCCCTTACCCAGCCTCCCTCACAGGCT
                    4590                    4620                    4650                    4680
CCTGGCCCTCAGTCTCTCCCCTCCACTCCATTCTCCACCTACCCACAGTGGGTCATTCTGATCACCGAACTGACCATGCCAGCCCTGCCGATGGTCCTCCATGGCTCCCTAGTGCCCTGG
                    4710                    4740                    4770                    4800
AGAGGAGGTGTCTAGTCAGAGAGTAGTCCTGGAAGGTGGCCTCTGTGAGGAGCCACGGGGACAGCATCCTGCAGATGGTCCTGGCCCTTGTCCCACCGACCTGTCTACAAGGACTGTCCT
                    4830                    4860                    4890                    4920
CGTGGACCCTCCCCTCTGCACAGGAGCTGGACCCTGAAGTCCCTTCCCCACCGGCCAGGACTGGAGCCCCTACCCCTCTGTTGGAATCCCTGCCCACCTTCTTCTGGAAGTCGGCTCTGG
                    4950                    4980                    5010                    5040
AGACATTTCTCTCTTCTTCCAAAGCTGGGAACTGCTATCTGTTATCTGCCTGTCCAGGTCTGAAAGATAGGATTGCCCAGGCAGAAACTGGGACTGACCTATCTCACTCTCTCCCTGCTT
                    5070                    5100                    5130                    5160
TTACCCTTAGGGTGATTCTGGGGGCCCACTTGTCTGTAATGGTGTGCTTCAAGGTATCACGTCATGGGGCAGTGAACCATGTGCCCTGCCCGAAAGGCCTTCCCTGTACACCAAGGTGGT
          GlyAspSerGlyGlyProLeuValCysAsnGlyValLeuGlnGlyIleThrSerTrpGlySerGluProCysAlaLeuProGluArgProSerLeuTyrThrLysValVa
                    5190                    5220                    5250                    5280
GCATTACCGGAAGTGGATCAAGGACACCATCGTGGCCAACCCCTGAGCACCCCTATCAACTCCCTATTGTAGTAAACTTGGAACCTTGGAAATGACCAGGCCAAGACTCAAGCCTCCCCA
lHisTyrArgLysTrpIleLysAspThrIleValAlaAsnPro *
                    5310                    5340                    5370                    5400
GTTCTACTGACCTTTGTCCTTAGGTGTGAGGTCCAGGGTTGCTAGGAAAAGAAATCAGCAGACACAGGTGTAGACCAGAGTGTTTCTTAAATGGTGTAATTTTGTCCTCTCTGTGTCCTG
                    5430                    5460                    5490                    5520
GGGAATACTGGCCATGCCTGGAGACATATCACTCAATTTCTCTGAGGACACAGATAGGATGGGTTGTCTGTGTTATTTGTGGGGTACAGAGATGAAAGAGGGGTGGGGATCCACACTGAG
                    5550                    5580                    5610                    5640
AGAGTGGAGAGTGACATGTGCTGGACACTGTCCATGAAGCACTGAGCAGAAGCTGGAGGCACAACGCACCAGACACTCACAGCAAGGATGGAGCTGAAAACATAACCCACTCTGTCCTGG
                    5670                    5700                    5730                    5760
AGGCACTGGGAAGCCTAGAGAAGGCTGTGAGCCAAGGAGGGAGGGTCTTCCTTTGGCATGGGATGGGGATGAAGTAAGGAGAGGGACTGGACCCCCTGGAAGCTGATTCACTATGGGGGG
                    5790                    5820                    5850                    5880
AGGTGTATTGAAGTCCTCCAGACAACCCTCAGATTTGATGATTTCCTAGTAGAACTCACAGAAATAAAGAGCTGTTATACTGTGGTTTATTCTGGTTTGTTACATTGACAGGAGACACAC
                    5910
TGAAATCAGCAAAGGAAACAGGCATCTAAGTGG
```

Figure 2
Nucleotide sequence of the PA gene. Numbering starts at the first transcription initiation site; the second transcription initiation site is indicated by an asterix. Putative transcriptional regulatory elements, splice consensus sequences and polyadenylation signals are underlined. The TATA-box and the translational start codon are boxed. The stop codon is indicated by an asterix in the protein sequence.

Another, to our knowledge not yet described, repeated motif is located in the region -123 to -72. Both the sequences GGGAGGG and CAGCCTC are present in duplicate at a mutual distance of 7 nucleotides.

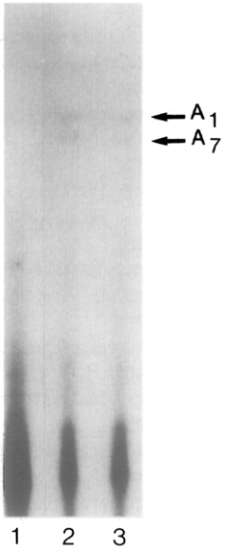


Figure 3
Primer extention analysis of PA mRNA. As size marker the sequence, performed with the 22-mer oligonucleotide as a primer, of the corresponding genomic region, as determined by the dideoxy chain termination method was used. Lane 1: control tRNA; lane 2: PC 82 polyA+ RNA; lane 3: PC EW polyA+ RNA. The two transcriptional initiation sites are indicated by arrows, the number indicates the corresponding position of the nucleotide in the genomic sequence (see Fig. 2).

Discussion

In this study we determined the structure of the PA gene. Comparison with other kallikrein-like genes (mouse, rat and the human hGK-1 gene) revealed a similar organization (3-6,10). Transcription in rat and mouse kallikrein-like genes initiates at positions comparable to the ones found here (3,7). All genes are composed of five exons and four introns; the splice junction sites are completely conserved. Rat and mouse kallikrein-like genes are somewhat smaller than PA and hGK-1, mainly because of the smaller size of the last intron. As expected, the highest percentage of homology is found with the human gene hGK-1 (see Fig. 4 and ref. 10). The

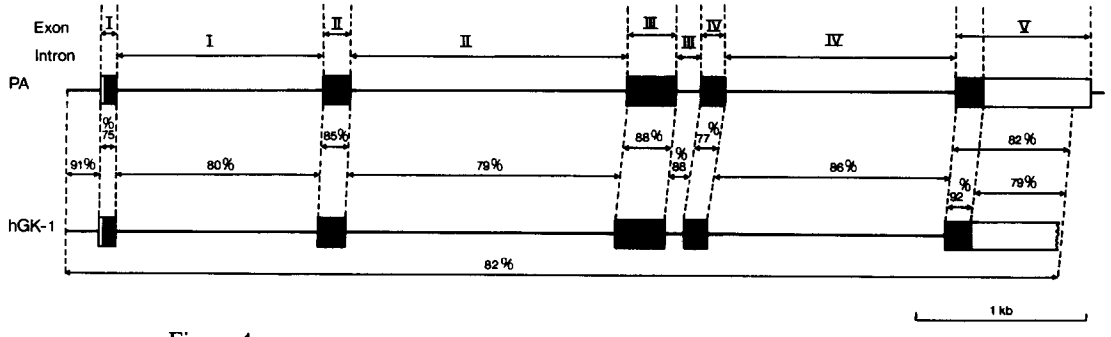


Figure 4
Comparison of the organization of the PA gene with that of the hGK-1 gene. The hGK-1 sequence data are from ref. 10.

```
PA        -190 CAGGGTGATCTAGTAATTGCAGAACAGCAAGTGCTAGCTCTCCCTCCCCTTCCA
hGK-1          ...AC.........TG..TG.............G.................

PA        -136 CAGCTCTGGGTGTGGGAGGGGGTTGTCCAGCCTCCAGCAGCATGGGGAGGGCCT
hGK-1          .......[.....].........................................

PA         -82 TGGTCAGCCTCTGGGTGCCAGCAGGGCAGGGGCGGAGTCCTGGGGAATGAAGGT
hGK-1          ........A...A.......A.......A[......]G.......A.........C

PA         -28 TTTATAGGGCTCCTGGGGGAGGCTCCCC -1
hGK-1          [......].........CA.......C....
```

Figure 5
Comparison of the sequence of the promoter region of the PA gene with that of the hGK-1 gene. Dots represent identical nucleotides. Putative transcriptional regulatory elements are boxed. The hGK-1 sequence data are from ref. 10.

various regions show a mutual homology ranging from 75 to 92 percent. The highest homology is found in the promoter region and in the protein coding region of exon V (over 90 %); the structural similarity is the smallest in the first exon, encoding the non-translated leader and the major part of the signal sequence, and in exon IV (75-77 %). Intron regions are as well conserved as the protein coding parts.

PA is expressed at high level in the prostate (11,12). Recently data have been published, indicating that hGK-1 is also expressed in the prostate gland, although at a lower level (29). The third known human kallikrein-like gene is expressed in pancreas and kidney (9). Comparison of the promoter regions of the three human kallikrein-like genes can, therefore, provide important information about elements involved in tissue specific gene expression. The promoter region of the hGK-1 gene has been described (10). So far, however, no data are available about the structure of the kidney/pancreas kallikrein gene. As depicted in Fig. 5 the putative regulatory elements present in the PA promoter region are in general also present in the 5'-flanking region of hGK-1. A difference which can be of interest is the G at -170 in the SRE (steroid responsive element) of hGK-1 instead of A, which makes that the putative hGK-1 SRE deviates from the consensus sequence. At present, experiments are in progress to find out whether or not PA/hGK-1 transcription is steroid dependent and which sequences are involved in prostate specific gene expression.

Acknowledgments: We are indebted to Mrs. Paula Delfos for photography. This work has been supported by a grant from the Netherlands Cancer Society (KWF).

References

1. Drinkwater, C.C., Evans, B.A. & Richards, R.I. (1988) Trends Biochem. Sci. 13, 169-1722.
2. Macdonald, R.J., Margolius, H.S. & Erdös, E.G. (1988) Biochem. J. 253, 313-321.
3. Mason, A.J., Evans, B.A., Cox, D.R., Shine, J. & Richards, R.I. (1983) Nature 303, 300-307.
4. Ashley, P.L. & MacDonald, R.J. (1985) Biochem. 24, 4512-4520.

5. Gerald, W.L., Chao, J. and Chao, L. (1986) Biochim. Biophys. Acta 866, 1-14.
6. Evans, B.A., Drinkwater, C.C. and Richards, R.I. (1988) J. Biol. Chem. 262, 8027-8034.
7. Chen, Y.P., Chao, J. & Chao, L. (1988) Biochem. 27, 7189-7196.
8. Baker, A.R. & Shine, J. (1985) DNA 4, 445-450.
9. Fukushima, D., Kitamura, N. & Nakanishi, S. (1985) Biochem. 24, 8037-8043.
10. Schedlich, L.J., Bennetts, B.H. & Morris, B.J. (1987) DNA 6, 429-437.
11. Lundwall, A. & Lilja, H. (1987) FEBS Lett. 214, 317-322.
12. Riegman, P.H.J, Klaassen, P., Korput, J.A.G.M. van der, Romijn, J.C. & Trapman, J. (1988) Biochem. Biophys. Res. Commun. 155, 181-188.
13. Wang, M.C., Valenzuela, L.A., Murphy, G.P. & Chu, T.M. (1979) Invest. Urol. 17, 159-163.
14. Wang, M.C., Papsidero, L.D., Kuriyama, M., Valenzuela, L.A., Murphy, G.P. & Chu, T.M. (1981) The Prostate 2, 89-96.
15. Watt, K.W.K., Lee, P-J., M'Timkulu, T., Chan, W-P. & Loor, R. (1986) Proc. Natl. Acad. Sci. USA 83, 3166-3170.
16. Gallee, M.P.W., Vroonhoven, C.C.J. van, Korput, J.A.G.M. van der, Kwast, T.H. van der, Kate, F.J.W. ten, Romijn, J.C. & Trapman, J. (1986) The Prostate 9, 33-45.
17. Lilja, H. (1985) J. Clin. Invest. 76, 1899-1903.
18. McGee, R.S. & Herr, J.C. (1988) Biol. Reprod. 39, 499-510.
19. Kuriyama, M., Wang, M.C., Papsidero, L.D., Kilian, C.S., Shimano, T., Valenzuela, L., Nishiura, T., Murphy, G.P. & Chu, T.M. (1980) Cancer Res. 40, 4658-4662.
20. Killian, C.S., Yang, N., Emrich, L.J., Vargas, F.P., Kuriyama, M., Wang, M.C., Slack, N.H., Papsidero, L.D., Murphy, G.P. & Chu, T.M. (1985) Cancer Res. 45, 886-891.
21. Killian, C.S., Emrich, L.J., Vargas, F.P., Yang, N., Wang, M.C., Priore, R.L., Murphy, G.P. & Chu, T.M. (1986) J. Natl. Cancer Inst. 76, 179-185.
22. Stamey, T.A., Yang, N., Hay, A.R., McNeal, J.E., Freiha, F.S. & Redwine, E. (1987) New Engl. J. Med. 317, 909-916.
23. Feinberg, A.P. & Vogelstein, P. (1983) Anal. Biochem. 132, 6-13.
24. Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982) Molecular cloning: A Laboratory Manual, Cold Spring Harbor, NY.
25. Davis, L.G., Dibner, M.D. & Battey, J.F. (1986) Basic Methods in Molecular Biology, Elsevier, NY.
26. Sanger, F., Nicklen, S. & Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
27. Geliebter, J., Zeff, R.A., Melvold, R.W. & Nathenson, S.G. (1986) Proc. Natl. Acad. Sci. USA 83, 3371-3375.
28. Ham, J., Thomson, A., Needham, M., Webb, P. & Parker, M. (1988) Nucl. Acid Res. 16, 5263-5275.
29. Chapdelaine, P., Paradis, G., Tremblay, R.R. & Dube, J.Y. (1988) FEBS Lett. 236, 205-208.