Gene Structure and Chromosomal Localization of Plasma Kallikrein[†,‡]

G. Beaubien,[§] I. Rosinski-Chupin,[∥] M. G. Mattei,[⊥] M. Mbikay,[#] M. Chrétien,[#] and N. G. Seidah*[,§]

J. A. DeSève Laboratories of Biochemical and Molecular Neuroendocrinology, Clinical Research Institute of Montreal, 110 Pine Avenue West, Montreal, Quebec H2W 1R7, Canada, Hôpital d'enfants, Groupe hospitalier de la Timone, 13385 Marseille Cédex 5, France, and Unité de Génétique et du Développement, Institut Pasteur, 75914 Paris, France

Received August 13, 1990; Revised Manuscript Received October 29, 1990

ABSTRACT: Plasma kallikrein (Fletcher factor) is a hepatic serine proteinase that participates in the early phase of blood coagulation. From two genomic libraries, we succeeded to isolate four overlapping clones representing the entire rat plasma kallikrein gene. Using selective DNA sequencing, polymerase chain reactions, and restriction mapping, we demonstrated that the gene for rat plasma kallikrein was 22 kb in length. Similar to human factor XI [Asakai et al. (1987) *Biochemistry 26*, 7221-7228], we also found that the plasma kallikrein gene is composed of 15 exons and 14 introns. A potential transcription initiation step was determined by a novel application of the polymerase chain reaction technique. Computer analysis of the 5'-promoter region of this gene revealed some putative control elements that might regulate the rat plasma kallikrein gene expression. These data and the results of chromosomal localization reported in the present study for mouse (chromosome 8) and human (chromosome 4) plasma kallikrein genes strongly corroborate a genic duplication event from a common ancestor to both plasma kallikrein and factor XI.

Plasma prekallikrein (M_r 88 000) and factor XI (M_r 160 000) are zymogens of two distinct serine proteinases present in normal plasma. The cDNA sequences of rat (Seidah et al., 1989), mouse (Seidah et al., 1990), and human plasma kallikrein (PK) and factor XI (Chung et al., 1986; Fujikawa et al., 1986) are now known. However, at the genomic level, only the human factor XI sequence organization was published (Asakai et al., 1987). This gene was shown to be 23 kb in length, comprising 15 exons (I-XV) and 14 introns (A-N), and was localized on the most distal part (q35) of the long arm of human chromosome 4 (Kato et al., 1989).

Our search of prohormone processing enzymes led us to demonstrate that PK exhibits preferential cleavage at the C-terminal of pairs of basic residues, similar to sites utilized in pro-hormone processing (Cromlish et al., 1986; Metters et al., 1988; Seidah et al., 1986, 1988). Furthermore, PK is shown to participate in the early stage of the intrinsic pathway of blood coagulation, activating in a reciprocal fashion factor XII, and is capable to release the vasoactive peptide bradykinin from the high molecular weight kininogen (HK) (Takagaki et al., 1985). Plasma kallikrein was also reported to interact with human leukocytes in a variety of ways. It is a chemotactic factor for neutrophils and monocytes (Kaplan et al., 1972; Gallin et al., 1974), causing aggregation of neutrophils (Schapira et al., 1982) and release of elastase (Wachtfogel et al., 1983). It was also proposed that PK can participate in the activation of prourokinase into urokinase during the intrinsic fibrinolysis (Hauert, 1985) and can convert prorenin to renin in vitro (Sealy et al., 1979). The protein substrates for factor XIa include, among others, factor IX (Schiffman et al., 1963), factor XII (Cochrane et al., 1973), and plasminogen (Mandle et al., 1977).

The major source of PK is the liver, and previous studies have shown a higher level of rat PK mRNA in female than in male (Seidah et al., 1989, 1990), suggesting that PK might be regulated by estrogen or other related factors. Sequence analysis demonstrated an overall identity of 75% between human and rat or mouse PK (Seidah et al., 1989, 1990), suggesting a low rate of variation between species. Comparison of human factor XI and human PK cDNA sequences has shown a 58% identity. This high degree of sequence identity between PK and factor XI suggests a recent evolutionary divergence of these genes from a common ancestral gene. On the basis of the partial NH_2-terminal sequences of these proteinases, a tentative evolutionary tree was suggested (Veloso et al., 1986). To corroborate this hypothesis, we undertook the isolation and determination of the structural organization of the rat PK gene and the chromosomal localization of the genes coding for the human and mouse homologues. These data also demonstrate that the gene organization of the catalytic subunit of PK is quite similar to that of trypsin and other related serine proteases.

Experimental Procedures

Screening of Genomic Libraries. Two rat genomic libraries were screened by using either 5'-end (1500 bp) or 3'-end (940 bp) *Eco*RI fragments of the rat PK cDNA (Seidah et al., 1989) radioactively labeled with ribonucleoside [α-^{32}P]triphosphate (Maniatis et al., 1982). From a Wistar genomic library in λEMBL-3 (Rosinski-Chupin & Rougeon, 1990), 2 million recombinant phages were screened by using the hybridization technique of Benton and Davis (1977). Plaque were lifted on nylon membranes and subjected to in situ alkaline denaturation and neutralization (Hanahan & Meselson, 1983) before UV irradiation to fix DNA to the membrane (Reed & Mann, 1985). This permitted the characterization of most of the gene coding for rat PK except for the 5' end of the gene. In an attempt to isolate this segment of the gene, a second genomic library (Sprague-Dawley) obtained from Clontech Laboratories (Palo Alto, CA) and containing 2.3 × 10^6 independent clones in λEMBL-3 SP6/T7 was screened by using the same methodology as described above. DNA

[†] This work was supported by grants from the Medical Research Council of Canada and from the J. A. DeSève succession.

[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J05315.

* To whom correspondence should be addressed.

[§] J. A. DeSève Laboratory of Biochemical Neuroendocrinology, Clinical Research Institute of Montreal.

[∥] Institut Pasteur.

[⊥] Groupe hospitalier de la Timone.

[#] J. A. DeSève Laboratory of Molecular Neuroendocrinology, Clinical Research Institute of Montreal.

0006-2960/91/0430-1628$02.50/0

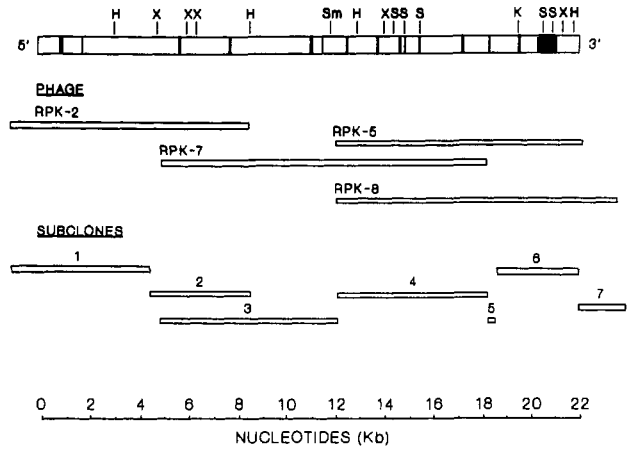

FIGURE 1: Recombinant phage, subclones, and a partial restriction map for the rat PK gene. On the top line is a representation of the gene coding for rat PK. The 15 exons are indicated by solid bars and are separated by 14 open boxes representing intervening sequences. The 5′ to 3′ orientation of transcription is shown. Sites for several restriction enzymes are shown: H, *Hin*dIII; X, *Xho*I; Sm, *Sma*I; S, *Sst*I; K, *Kpn*I. Four overlapping recombinant phage containing rat DNA coding for PK are shown. Fragments isolated from each phage insert and cloned into Bluescript (Stratagene) are shown under the subclone heading.

fragments containing regions of the rat PK gene were excised from recombinant phages and subcloned into appropriate restriction sites (*Eco*RI or *Xho*I) in a Bluescript plasmid (Stratagene).

Restriction Mapping and Southern Blotting. Restriction maps of the inserts subcloned (Figure 1) were generated by single and double digestions with the enzymes *Eco*RI, *Sst*I, *Sst*II, *Xba*I, *Bam*HI, *Sma*I, *Hin*dIII, *Acc*I, *Xho*I, *Kpn*I, and *Eco*RV. The DNA fragments were resolved by electrophoresis through 1% agarose gels (Maniatis et al., 1975). To identify fragments encoding PK sequences, the DNA was transferred to Hybond-N membrane (Amersham) by the technique of Southern (1975) and hybridized with specific oligonucleotides 5′-end-labeled with γ-^{32}P.

Oligonucleotide Synthesis. Assuming that the intron positions of factor XI and PK gene could be comparable, we synthesized 15 sense and 15 antisense oligonucleotides upstream and downstream of each of the hypothetical intron sites localized on the rat PK cDNA sequence. Most of them were 21 nucleotides in length (see Figure 2) and were used for sequencing and polymerase chain reactions (PCR). Seven other oligonucleotides were synthesized for the analysis of the promoter region of the gene. All these primers were synthesized on a 381A DNA synthesizer (Applied Biosystems).

DNA Sequence Analysis. DNA sequencing was done by the dideoxynucleotide chain-termination procedure (Sanger et al., 1980) using a Sequenase kit. The SK and KS primers (Stratagene) as well as rat PK cDNA sequence-specific oligonucleotides (see above and Figure 2) were used as primers. The reaction mixtures were subjected to electrophoresis in a 6% denaturing field gradient polyacrylamide gel (Ansorge & Labeit, 1984).

Polymerase Chain Reaction. The length of the introns within the rat PK gene was evaluated by using PCR amplification (Saiki et al., 1988) with the exonic primers just described. The amplifications were performed on a DNA thermal cycler (Perkin-Elmer, Cetus) using the PCR kit provided by the manufacturer. For each intron, the 100-μL mixture reaction contained 1 μg of the appropriate insert subclone (1 μL), 100 pmol (5 μL) of the corresponding primers (Figure 2), 10 μL of the 10× PCR buffer, 16 μL of a solution containing 1.25 mM each dNTP, 2.5 units (0.5 μL) of *Thermus aquaticus* DNA polymerase (AmpliTaq, Perkin-Elmer, Cetus), and 62.5 μL of sterile water. To prevent evaporation during the evaporation reactions, 50 μL of *n*-dodecane was added. Each cycle consisted of heating 1 min at 94 °C, followed by annealing for 1 min at the appropriate temperature, i.e., 5 °C below the predicted melting temperature of the oligonucleotides (estimated from the formula: T_m (°C) = 4[C+G] + 2[A+T]). Elongation was performed at 72 °C for 1 min. This was repeated for 30 cycles. The last cycle was followed by an additional incubation period of 10 min at 72 °C. Electrophoresis was performed on a 10-mL aliquot loaded onto a 1% agarose gel. The different DNA fragments obtained were visualized by ethidium bromide coloration (Figure 3).

Mouse and Human Gene Mapping by in Situ Hybridization. In situ hybridization experiments were carried out on a chromosome preparation obtained from WMP male mouse concanavalin A stimulated lymphocytes or from phytohemagglutinin-stimulated human lymphocytes cultured for 72 h. 5-Bromodeoxyuridine (60 μg/mL of medium) was added for the final six (mouse) or seven (human) hours to ensure a posthybridization chromosomal banding of good quality. The rat PK cDNA (Seidah et al., 1989) and the human PK cDNA (Chung et al., 1986) both cloned as 2200 bp inserts in Bluescript were tritium-labeled by nick-translation to a specific activity of 1.7×10^8 (rat PK) or 1.9×10^8 (human PK) dpm/μg. As previously described by Mattei et al. (1985), the radiolabeled probes were hybridized to metaphase spreads at a final concentration of 25 ng/mL of hybridization solution. Hybridization was performed for 16–18 h at 42 °C in a 50% formamide/2 × SSC (1 × SSC = 150 mM NaCl + 15 mM sodium citrate, pH 7.0) saturated environment. The slides were then dipped in 50% formamide/2 × SSC at 39 °C to remove the coverslip, and then rinsed 3 times in 50% formamide/2 × SSC at 39 °C and 3 times, for 10 min, in 2 × SSC at room temperature. Following this, the final washing was done for 1 h in 0.1 × SSC at room temperature and at 4 °C, prior to being dehydrated through an ethanol series and air-dried. After being coated with nuclear track emulsion (KODAK NTB_2), the slides were exposed for 17 (mouse) or 21 (human) days at 4 °C and then developed. To avoid any slipping of silver grains during the banding procedure, chromosome spreads were first stained with buffered Giemsa solution and metaphase-photographed. R banding was then performed by the Fluorochrome–Photolysis–Giemsa (F.P.G.) method (Camargo & Cervenka, 1982), and the metaphase were rephotographed before analysis.

Results and Discussion

Screening for the Rat PK Gene. From a rat (Wistar) genomic library, we isolated three positive clones (RPK-5, -7, and -8) that hybridized with the rat PK cDNA (Seidah et al., 1989). These clones were plaque-purified, and DNA sequence analysis of a series of *Eco*RI subclones revealed that none of these included the 5′ end of the gene. From the screening of a second rat (Sprague-Dawley) genomic library, 18 positive clones were isolated and 3 of them plaque-purified. One of these (RPK-2) was found by specific oligonucleotide hybridization to overlap with the initially isolated phage and to extend in the 5′ direction. The rat PK gene was entirely represented in a minimum of three overlapping inserts (RPK-2, -7, and -8), where RPK-5 was found to be entirely comprised within RPK-8 (see Figure 1). Six *Eco*RI restriction fragments ranging in size from 0.4 to 8.5 kb and encompassing the full length of the rat PK gene were obtained from the recombinant

```
atggattctg ggttagccaa ctgtgtagtt agagttggat atgcaaggct aggtttgaat
cttgggtcca cagaaaagag catttgcatca gggtccaccg tgattacatt gatgactgaa
ttcctagagc ccatgcttgg atcctcagat cagtgggcac gatcttatga ttgaggactg
ccttaacatg ttgatattga aatgagcctg aaaactgagc ctccagtggc ca gttaacg
aactcaagcc cgtgagcatg ggtttgttcc tgaatctaca atcgctgttc tgtaaatgag
ccctggaaat gggacaatct ggtggctggc ctgaggtatc atcctggagc ctgggtctgc
agggtcagcc tgataactgg catttcctgg aatggacttg tttacatttg tctttattca
tagggaacac atcgccattc tgtgctgccc aggcttagta gcagggtttt atggttaaag
tgaaggatcg aacactcttc aatgcctatt cactatgtcc acagtttcac cgggggctgt
aacctatcat ctggctttct taacacttgg aaaggcatgg ggaattgttc ccatggattc
ttctgcaaag ggatgcgtgg tggagattCC AATtttgtca tggtgtttac atgacttttt
ggaatcggta acaTATTAAg tacttgaacc tggactggaa ggtccATGGA CTGTATTGAC
AGGTCAAACA GAAGACACTG ATGCCAGAAG CCCAGTGTCA ACACTGGAGC CAAGCAGAGA
CCAACCTCAG TGCCATATTC GGAGAGCTTG AAGACTAGCT TCATGTGAAG ACTCCTTCTC
CTCCAGCAGC ACAAAGCAAC CATCCTTCCA G [[gtagctactctccaccattctcattgta
gcctatgtcttaggaatttatttttttaaaaaagtataactatatctgca...Intron A...c
ttgggagcatgataaatattttaatttttctatatccaaaatgtcccctcctccaaatcctctttc
tcctggtgattatgtgttgtgatttgcaatttagacatttaccaaaaaggtgaagtgtacattaa
attaaattgtctattaaatacatggtaaacatgatctcaacctagtagtagcgtaaatttatttt
                    M  I  L  F  K  Q  V  G  Y  F  V  S
tttctttcaaag ]]G ATG ATT TTA  TTC AAA CAA GTG GGT TAT TTT GTT TCC
 L  F  A  T  V  S  C
TTG TTC GCT ACA GTT TCC TGT G [[gtaagtattagcctgggagttcaaattaagtta
gtgtgtgaaactataccactcctggagaaagccttaaaagcttgggggaagaaggcagctgctg
atagtccaaggcaagaa...Intron B...tcttctatagtgtcacctaaatggccatttaggc
ctagagtcgacccagatccactcgttattctcggacgagtgttcagtaatgaacctctggagaga
accatgtatatgatcgttatctgggttggacttctgctttaagcccagataactggcctgaata
tgttaatgagagaatcggtattcctca...Intron B...tggtaattggtttaaaacagttgt
catatgtttcattctacctgtgtggttgtatatatctttgggtagtcattgggataaataaaaag
gaaaaacaacatagatgtatttctactaaaatggttacacttttctccggaatcaacttttatatc
        G  C  L  S  Q  L  Y  A  N  T  F  F  R  G  G
acag ]]GG TGT CTG TCA CAA CTG TAT GCA AAT ACC TTC TTC AGA GGT GGG
  D  L  A  A  I  Y  T  P  D  A  Q  H  C  Q  K
 GAT CTG  GCT GCC ATC TAC  ACC CCG GAT GCC CAG  CAC TGT CAG  AAG
 M  C  T  F  H  P  R  C  L  L  F  S  F  L  A  V
ATG TGC ACG TTT CAC CCT  AGG TGC CTG CTC TTC AGC TTC CTT GCC GTG
 S  P  T  K  E  T  D  K  R
AGT CCA ACC AAG GAG ACA GAT AAA AG [[ gtaagatgtccgtgagtccaaccaagg
agacagataaaaggtaagatgtgctgttttatcactgccgggtgagacgtttctcaaatagcttc
ctgt ... Intron C ... cgatgtagatagtatgaacttgtcttaagaatgtccttccagc
acgtgtcgagaaatgtaattagtttcctgaccttatttttatttattaacatatccaaaatgctgt
ataatttcttaatgtgtatttgtcgttttattccccttttaaaatgtgatttttttaaaccag
        F  G  C  F  M  K  Q  S  I  T  G  T  L  P  R
 ]] G TTT GGG TGC TTC ATG AAA GAG AGC ATT ACA GGG ACT TTG CCA AGA
  I  H  R  T  G  A  I  S  G  H  S  L  K  Q  C  G
ATA CAC CGG ACA GGG GCC ATT TCT GGT CAT TCT TTA AAA CAG  TGT GGC
 H  Q  L  S
CAT CAA TTA AGT G [[ gtaagatgtcggatttttttccaactaaatttatgttactagca
ctcaaactcagagtagtttttgctgaaagtctatactatgatggacttttagaagagaacaccga
gaaagataacagacggagcagcatcatggacctaaggaggggtctttcttcctgataatgttcat
gtctttatcagttgtaatgcacggccttaaggcttcc ... Intron D...aaaatgactccg
agggctgagcagtttatctgtgttcaatatacatatattataatacatatatacatatatgta
                                                A  C  H  Q  D
tgtatatatatatggtcgaatgtctctctgtgtttcatttcag ]]CT TGC CAC CAA GAC

 I  Y  E  G  L  D  M  R  G  S  N  F  N  I  S  K
ATA TAC GAA  GGA CTG GAT ATG AGA GGG TCC  AAC TTT AAT ATA TCT AAG
 T  D  S  I  E  E  C  Q  K  L  C  T  N  N  I  H
ACT GAC  AGT ATT GAA GAA TGC CAG AAA CTG  TGC ACA AAT AAT ATT CAC
 C  Q  F  F  T  Y  A  T  K  A  F  H  R  P  E  Y
TGC CAA TTT TTC ACA TAT GCT ACA AAA GCA TTT CAC AGA CCA  GAG TAC
 R
AG [[ gtgagtgagtcatggttcccctgagaggacattccaggatgcatcagcatccgcagtgaa
aaagacaagtctgcttcccctccagtcaaagacaaggggtgggcagtcctgag ... Intron E
... ttagtgactggtaagctatgtatgggcataatgtcagacaatgatgtccgccactccatt
tgcagccgcctgtccccttgcagaagggaaaagtgtctattgatcactaatggcccaactgtgc
ccctttgtggtaacaaatattatctgcacacctttgcacacctcaaggaaggaagcataacatct
                        K  S  C  L  L  K  R  S  S  S  G
tttgttgacactcag ]] G AAG AGT TGC CTG CTG AAG CGC AGT TCA AGT GGA
 T  P  T  S  I  K  P  V  D  N  L  V  S  G  F  S
ACG CCC ACC AGT ATA AAG CCA GTG GAC AAC CTG GTG TCT GGA TTC TCA
 L  K  S  C  A  L  S  E  I
CTG AAG TCC TGT GCT CTC TCA GAG ATC G [[ gtaattagatgacgattattcctc
tgtgtctgtgatgtcctaaggaagggttcccagaaacatccacatgctgttcccgggaacccagg
gaatatgatgccttttatagatgtgattacgtggaaatagtttgcaatcagagttatttgggat
taccagaatatgtgcctatccactctcaggggttcttagatgataaagcgggtaacagagaggtc
agtgtcagaggtagatagacttgaaaggtggaacaagagtc ... Intron F...tatattaa
ctatttaatttaatattccccttttctagtttttggaaacttgtctataggagataggctctt
tattcaacagtgttcccttttgtatactgcgcttatcttagactgcaccctgcagctatagccag
atgccagaggactgggaaacaggcaatttaaaactcaggatgctgactgctgcaaatacctaattg
tcataatcaccttagtttctatgttgaaaagcaagcaaactccacatcgtgtcactgagaaatgt
                            G  C  P  M  D  I  F  Q  H  F  A
tcaccactgagttcacag ]] GT TGC CCC ATG GAT ATT TTC CAG CAC TTT GCC
 F  A  D  L  N  V  S  Q  V  V  T  P  D  A  F  V
TTT GCA GAC CTG AAT GTA AGC CAG GTC GTC ACC  CCC GAT GCC TTC GTG
 C  R  T  V  C  T  F  H  P  N  C  L  F  F  T  F
TGT CGC ACC GTT TGC ACC TTC CAT CCC AAC TGC CTC TTC TTC ACA  TTC
 Y  T  N  E  W  E  T  E  S  Q  R
TAC ACG AAT GAG TGG GAG ACG GAA TCA CAG AG [[ gtgagtgtcagcagcatat
gctccctcttagcctttccggtcctagcagaggcttgcaagggtcttgctggtttccctaaca
gcgctggagcctgtggttaacaggatgcttttcagacgggtgcagg ... Intron G ...
accacccacaatgggctggtcctttcccccttgatcactggttgagaaaatgcctacagctggg
tctcatggaggcatttcctcaaggagtctccttttctctgtgataactccagcttgtgtcaagct
gacacacaaaaccaatctgtacaggaaggaaggaaggaaagaaagaatatgtctagtgcttctaa
                                                N  V  C  F  L  K
tgctgtcttctttctctctttcttgtttatttttaatag ]]G AAT GTT TGT TTT CTT AAG
 T  S  K  S  G  R  P  S  P  P  I  I  Q  E  N  A
ACA TCT AAA AGT GGA AGA CCA AGT CCC CCT ATT ATT  CAA GAA AAT GCT
 V  S  G  Y  S  L  F  T  C  R  K  A  R  P
GTA TCT GGA TAC AGT CTC TTC ACC TGC AGA AAA GCT CGC CCT G [[ gtaa
tgtcactcaacctcgatggcgtgtgacttgtgtcctggtttcgttcttggttgctgtgataaata
```

```
ccatgaccaaaagcaagttggggagga  aggtgttttgttcacatgtccctgttatagccgat
tcctgagggaagtcaggctaggaactcaagcagaggcaaagacaggaactgagcatagacctcag
agaaatgctgcgtttttgacttgctcctgtgattgatcatatagtgtatgtgtatgtatgtgtga
tgtgtgtgttgtgtgtgtatgtgtatatgatgtgtagat ... Intron H ...ggatgttcc
atcttttcaggtgactcttttatgtcaagttgacaaaacactaggtgtaccacatcttgggagcct
ctaagagttaatgaaatgggtctcaagtcccctagaaacgttgccaatatgccttatgtcctgact
                                        E  P  C  H  F  K  I
tgtcatatctcctttggtgtttgtatgtggtgcag ]] AA CCC TGC CAT TTC AAG ATT
 Y  S  G  V  A  F  E  G  E  E  L  N  A  T  F  V
TAC TCT GGA GTT GCC TTC GAA GGG GAA GAA CTG AAC GCG ACC  TTC GTG
 Q  G  A  D  A  C  Q  E  T  C  T  K  T  I  R  C
CAG GGA GCA GAT GCG TGC CAA  GAG ACC TGT ACA AAG ACC ATC CGC TGT
 Q  F  F  T  Y  S  L  L  P  Q  D  C  K  A  E  G
CAG TTT TTT ACT TAC TCA TTG CTT CCC CAA GAC TGC AAG GCA GAG GG [[
gtaagggaaccttttctttaacgataatcagcaaggtaatttttttctagttttctctctctgtgg
                                        C  K  C  S  L  R  L
gtttaattttgtactgttcatttttttccag ]]G TGT AAA TGT TCC TTA AGG TTA
 S  T  D  G  S  P  T  R  I  T  Y  E  A  Q  G  S
TCC ACG GAT GGC TCT CCA ACT AGG ATC ACC TAT GAG GCA CAG GGG AGC
 S  G  Y  S  L  R  L  C  K  V  V  E  S  S
TCT GGT TAT TCT CTG AGA CTG TGT AAA GTT GTG GAG AGC TCT G [[ gtga
gtgaagctcacttgttatggaactgcatggctggcttgactgttgtgatgaatgtcttggtctta
tataatgggattgtttatatggtttctgttatggtttgtatgtg ... Intron J ... ctt
ggtgataatggactgaacctctaaacctctaaaccagctccaattaaatgttgtctttataagag
ttgccttggtcatggtgtctgttcgcacagtaaaaaccctaactaagacagtttccttggatgga
                                                D  C
agggagggaatgctgatctgtgtctaattctgccacacatctcaacacttttcag ]] AC TGT
 T  T  K  I  N  A  R  I  V  G  G  T  N  S  S  L
ACG ACA AAA ATA AAT GCA CGT ATT GTG GGA GGA ACA AAC TCT  TCT TTA
 G  E  W  P  W  Q  V  S  L  Q  V  K  L  V  S  Q
GGA GAG TGG CCA TGG CAG GTC AGC CTG CAA GTG  AAG TTG GTT TCT CAG
 N  H  M  C  G  G  S  I  I  G  R  Q  W  I  L  T
AAC CAT ATG TGT GGA GGG TCC ATC ATT GGA CGC CAA TGG ATA CTG ACG
 A  A  H  C  F  D  G
GCT GCC CAT TGC TTT GAT GG [[gtaagttttcgatccatcttatttttaccaatgtgtg
tgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgagcatgcacgtgtatggaccctacctgcacat
gtgtatataaagaggttagcctcaggtatcttcgtgtactgcttgcttatctttatttgttgagac
agggttgttcactgagtctagtgctaaccctttgact ... Intron K ...ttttattataaa
ctaaatcttcgaagtccaatactgctgctgtaaacaacaataggtggccttgccagaaaaagtaa
ctaggtcatatggaatgttttcgttcttctccctcctcccctcaaagaaatgtgctaagaaaaa
                                                I  P
ggagcaaggaaggatgccaacgttactttgttctctgcttctctcctag ]] G ATT CCC
 Y  P  D  V  W  R  I  Y  G  G  I  L  N  L  S  E
TAT CCA GAC GTG TGG CGT ATA  TAT GGC GGG ATT CTT AAT CTG TCA GAG
 I  T  N  K  T  P  F  S  S  I  K  E  L  I  I  H
ATT ACA AAC AAA ACG CCT TTC TCA AGT ATA AAG GAG CTT ATT ATT  CAT
 Q  K  Y  K  M  S  E  G  S  Y  D  I  A  L  I  K
CAG AAA TAC AAA ATG  TCA GAA GGC AGT TAC GAT ATT GCC TTA ATA AAG
 L  Q  T  P  L  N  Y  T
CTT CAG ACA CCG TTG AAT TAT ACT G [[gtatgcagcgtatttaaggggaaactgta
caattcacttgtattggcgttagcttccagttcaatttgagcgagaagacagatctgcagcaatt
gtctttgtgtcctggttggtggggctgagggagagatgaggctgcctgggaagtaaaaaccccac
gacctaccatgcatttgctgctactt..Intron L..gaattctgttacagtgtaggtttacca
aagcataaggatgtttaccaatacatttttggttgaagagtcccctcttccccatggaaatttcatt
ctgaaatgccattcaaactgtgagaatttccttgaatcatgggctttagaaatggctggagggcc
attctttatactctttagtttcagcaatggtttcattatatctcatacatgggcatggtatgttt
tgattgcatgcacccctcaccctctcttgccctcccactcctgttgttcgccctactcctcc
cgtgtagtcccacatggctcattcttgctctgttgcttttgtagaaaaagaataaggcaaaagata
gatattaatttttcagatccctatctatggttttaattagcctaaatctattcatgaaatttct

                E  F  Q  K  P  I  C  L  P  S  K  A  D  T
ttcag ]] AA TTC CAA AAA CCA ATA TGC CTG CCT TCC  AAA GCT GAC ACA
 N  T  I  Y  T  N  C  W  V  T  G  W  G  Y  T  K
AAT ACA ATT  TAT ACC AAC TGC TGG GTG ACT GGA TGG GGC TAC ACA AAG
 E  R
GAA CGA G [[gtaagcctggcatgtggaatgttcctctcttatgtgctaaggtacatggagta
gagccactcaacaacccgtagtcatgatcctttctgtagatccattcaacagcccttactcatggc
ctttctgttcattccaaaatgatcccaagaatatgacttttttgcattaaaaactgatgagcta
atcattaaacatattactcataatagaga ... Intron M ...catagaaaccaagtctgtg
gtagcttgggtgataggggagcttcagtcaacacagatcattgcttcttcatctccatctcctc
ccctgctctccctccacttcttgcttccttttctccctccctcttactcccactattgtattt
                                        G  E  T  Q  N  I  L
cctctctttaatgctacttcttgattgcttcatag ]] GT GAG ACC CAA AAT ATT CTA
 Q  K  A  T  I  P  L  V  P  N  E  E  C  Q  K  K
CAA AAG GCA ACT ATT CCC TTG GTA CCA AAT GAA GAA TGC CAG AAA  AAA
 Y  R  D  Y  V  I  T  K  Q  M  I  C  A  G  Y  K
TAT AGA GAT  TAT GTT ATA ACC AAG CAG ATG ATC TGT GCT GGC TAC AAA
 E  G  G  I  D  A  C  K
GAA GGT GGA ATA GAT GCT TGT AAG [[gtaactctgggaaaaatacgtaatagttcac
tggagactgctcatccaaaagcaaatttcccaagttacatttcattagtttcagttttgaagttg
cagagttaagtgacatggagacttttgtttttaaagatttatttttatttatgtgagtgtacact
gtcactgttttcaccagaagagggcatcggattccattacagatgg ...Intron N... aaa
tatgatagcaaatgttaaagacatctatctttgaacaaggggctcctaatttgaatcatcactgt
                                                G  D  S  G
tatagctctctatcatttttttttctactcactgtgatccag ]] GGA GAT TCC  GGT
 G  P  L  V  C  K  H  S  G  R  W  Q  L  V  G  I
GGC CCC  TTA GTT TGC AAA CAT AGT GGA AGG TGG CAG TTG GTG GGT ATC
 T  S  W  G  E  G  C  A  R  K  E  Q  P  G  V  Y
ACC AGC TGG GGT GAA GGC TGT GCC CGC AAG GAG CAA  CCA GGA GTC TAC
 T  K  V  A  E  Y  I  D  W  I  L  E  K  I  Q  S
ACC AAA  GTT GCT GAG TAC ATT GAC TGG ATA TTG GAG AAG ATA CAG AGC
 S  K  E  R  A  L  E  T  S  P  A  stop
AGC AAG GAA AGA GCT  CTG GAG ACA TCT CCA GCA TGA GGA GGC TGG GTA
CTG ACG GGG AAG AGC  CCA GCT GGC ACC AGC TTT ACC ACC TGC CCT CAA
GTC CTA CTA GAG CTC  CAG AGT TCT CTT CTG CAA AAT GTC GAT AGT GGT
GTC TAC CTC GCA TCC  TTA CCA TAG GAT TAA AAG TCC AAA TGT AGA CAC
AGT TGC TAA AGA CAG  CGC CAT GCT CAA GCG TGC TTC CTG CCT TGA GCA
ACA GGA ACG CCA ATG  AGA ACT ATC CAA AGA TTA CCA AGC CTG TTT GGA
```

```
AAT AAA ATG GTC AAG  GGA TTT TAT TAG GTA GTG AAA TTA GGT AGT TGT
 CCT TGG AAC CAT CCT  CAT GTA ACT GTT GAC TCT GGA CCT CAG CAG ATC
 ACA GTT ACC TTC TGT  CCA CTT TTG ACA TTT GTG TAC TGG AAC CTG ATG
 CTG TTC TTC CAC TTG  GAG CAA AGA ACT GAG AAA CCT GGT TCT ATC CAT
 TGG GAA AAA GAG ATC  TTT GTA ACA TTT CCT TTA CAA TAA AAA GAT GTT
```

```
CTA CTT GAA ⇓ CTTGATGGGACAGAGCACCTTTTCATATAAACCACCTTTCTGGCCATTGC
TGCTGAGTCCTGCATGGTCATGAGTAAGGAGCTCGTGACAGGGTTTCCCACCCCTGGAGCAAAA
CCAGCTAGATGTTAGGCTTGTTTGTATAAATTTCAAGTTCCTCCTTTGGGAAGTATCTCTCTGC
CTAGATTATTGGGGGAATATCCTCTCTGTCAGGGCAGCCCGGAAGATTGTGG
```

FIGURE 2: DNA sequence of the rat PK gene including its 5′- and 3′-flanking regions, exons, and exon–intron boundaries. The promoter region is shown in small letters with the potential CAAT box and TATA box in evidence. Intervening sequences are also in small letters with the GT-AG consensus sequences in boldface and underlined. The proposed site of initiation of transcription is indicated by a capital letter in boldface. All the oligonucleotides used for the sequence and PCR on genomic clones are shown. Sense primers are underlined once and antisense primers are doubly underlined. Exon sequences are indicated in capital letters as nucleotide triplets, with the predicted amino acids in the single-letter code shown above each triplet. The polyadenylation signal AATAAA is also indicated, and an arrow shows the poly(A) addition site at the 3′ end of the gene.

Table I: Rat Plasma Kallikrein Gene Exon–Intron Junctions[a]

INTRON Type*	EXON	JUNCTION (donor)	JUNCTION (acceptor)	EXON	INTRON (bp) PK.	FXI.
A (nc)	AG	GTAGCTACTC.....TCTTTCAAAG		GA	700	900
B (I)	TG	GTAAGTATTA.....TATATCACAG		GG	4000	4500
C (II)	AG	GTAAGATGTC.....TTTAAACCAG		GT	2000	1400
D (I)	TG	GTAAGATGTC.....TTCATTTCAG		CT	3000	1200
E (II)	AG	GTGAGTGAGT.....TGACACTGAG		GA	415	1400
F (I)	CG	GTAATTAGAT.....GAGTTCACAG		GT	1100	600
G (II)	AG	GTGAGTGTCA.....TTTTTAATAG		GA	1350	3400
H (I)	TG	GTAATGTCAC.....TGTGGTGCAG		AA	860	100
I (II)	GG	GTAAGGGAAC.....TTTTTTCCAG		GT	96	91
J (I)	TG	GTGAGTGAAG.....CACTTTTCAG		AC	600	3700
K (II)	GG	GTAAGTTTTC.....TCTCTCCTAG		GA	1800	1400
L (I)	TG	GTATGCAGCG.....TTTCTTTCAG		AA	1100	850
M (I)	AG	GTAAGCCTGG.....TGCTTCATAG		GT	1600	1200
N (0)	AG	GTAACTCTGG.....TGTGATCCAG		GG	770	700
Consensus:	G	GTUAG.....YAG (U=Purine, Y=Pyrimidine)		G		
			cDNA:		2462	2097
			GENE LENGTH:		22000	23000

[a] The asterisk indicates intron type relative to the reading frame of the genes.

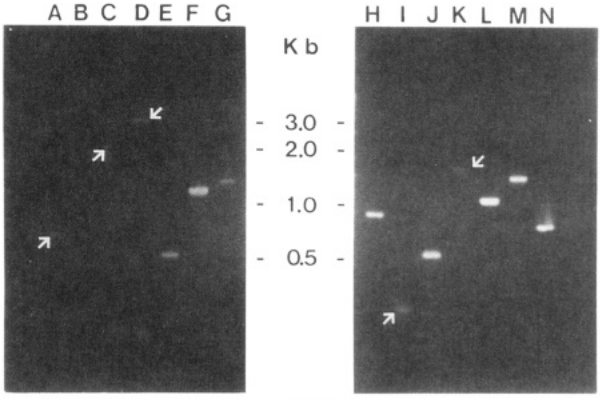

FIGURE 3: PCR amplification of 13 introns of the rat PK gene. The sense and antisense primers used are predicted in Figure 2. The products of the amplification reactions were analyzed by electrophoresis on 1.5% agarose gels. The estimated intron lengths are indicated in Table I. Since some of the bands are less visible than others, white arrows emphasize the positions of the PCR-amplified products of introns A, C, D, I, and K. Intron B cannot be evaluated by PCR, but restriction enzyme studies indicated a length of 4 kb.

phage (see Figure 1). The rat PK partial genomic sequence deduced from these subclones is shown in Figure 2.

Localization of Intron and Exon Junctions. Considering the intron sites of human factor XI gene (Asakai et al., 1987) and using synthetic oligonucleotides based on rat PK cDNA sequence, we precisely localized 14 intron–exon junctions (Table I). It is noteworthy that all the introns of the rat PK gene were of the same type, relative to the reading frame of the gene, as the equivalent ones found in the human factor XI gene (Asakai et al., 1987). The consensus proposed by Mount (1982) and the GT-AG rule (Breathnach et al., 1978) are obeyed at the 5′ and 3′ termini of each intervening sequence (see Table I). Furthermore, Figure 2 shows that the 5′ end of each intron contains a complementary sequence (Konarska et al., 1985) to the consensus 3′-end structure (Ruskin et al., 1984; Keller et al., 1984) which is considered to be important in splicing.

Intron Length and Gene Organization. The length of each intron was determined by PCR using as primers the same synthetic oligonucleotides as those used for the genomic sequence. The smallest fragment amplified is intron I (96 bp), and the largest one is intron D (3 kb) (see Figure 3 and Table I). Since the length of intron B could not be obtained by PCR, it was evaluated by restriction enzyme analyses and estimated to be about 4 kb in length. For each genomic fragment subcloned, the sum of the estimated exon and intron sizes corresponded to that of the entire subclone, confirming that all intronic sequences were identified. The estimated length of the rat PK gene is approximately 22 kb (starting at the transcription initiation site), a very similar value to the one reported for the human factor XI gene (23 kb) (Asakai et al., 1987). The size range of the exons is 59–763 bp; the entire length of the 3′-untranslated region is part of the largest exon (Figure 1). Excluding exons 2 (59 bp) and 15 (763 bp), the remaining exons showed an average of 141 bp in length, close to the mean value of 150 bp reported (Naora & Deacon, 1982; Blake, 1983) for exons in higher eukaryotes.

Intron sequence analysis also revealed at least two stretches of alternating purines and pyrimidines $(TA)_{26}$ and $(TG)_{19}$ in introns D and K, respectively. These sequences are frequent in mammalian genomes and were also reported in the case of globin (Miesfeld et al., 1981), ANF (Argentin et al., 1985), and immunoglobulin superfamily (Hamada et al., 1982). These types of sequences were also identified in human factor XI gene (introns B and M). The exact function of these structures is not yet known, but they have been proposed to exhibit enhancer properties playing a role in the regulation of transcription (Cereghini et al., 1983) and to promote recombinational events (Norheim & Rich, 1983; Wells et al., 1982).

A 40-nucleotide sequence spanning the donor splicing site of intron C (31 nucleotides on the coding site and 9 of the noncoding one) is immediately repeated inside the noncoding sequence of intron C. This might be a possible region where alternative splicing can occur. This putative site of alternative splicing might be important in the regulation of rat PK expression. Between the first donor splicing site and the second one (40 nucleotides downstream), we found a stop codon, "TAA". Therefore, should the second donor splicing site be used in the immature RNA processing reaction, no functional PK translation product would be produced.

Promoter Sequence Analysis. Even though PK and factor XI share similar gene organization, except for putative TATA and CAAT boxes little was known about the promoter sequence of factor XI (Asakai et al., 1987). Furthermore, the 5′ end of all cDNA sequences reported for PK (Chung et al., 1986; Seidah et al., 1989) or factor XI (Fujikawa et al., 1986) did not include the site of transcription initiation. Accordingly, we have extended our sequence of the rat PK gene to include 911 nucleotides at the 5′ end of the gene, hoping to cover most

```
-745 TTGCTTCACT GGGCAAACCT AGTAGTCATG TATCTCATCA ATGGATTCTG
                                                 +++++++++
-695 GGTTAGCCAA CTGTGTAGTT AGAGTTGGAT ATGCAAGGCT AGGTTTGAAT
                                                 .........
-645 CTTGGGTCCA CAGAAAAGAG CATTGCATCA GGGTCCACCG TGATTACATT
     .......
-595 GATGACTGAA TTCCTAGAGC CCATGCTTGG ATCCTCAGAT CAGTGGGCAC
-545 GATCTTATGA TTGAGGACTG CCTTAACATG TTGATATTGA AATGAGCCTG
                          *******
-495 AAAACTGAGC CTCCAGTGGC CATGTTAACG AACTCAAGCC CGTGAGCATG
-445 GGTTTGTTCC TGAATCTACA ATCGCTGTTC TGGAAATGAG CCCTGGAAAT
-395 GGGACAATCT GGTGGCTGGC CTGAGGTATC ATCCTGGAGC CTGGGTCTGC
-345 AGGGTCAGCC TGATAACTGG CATTTCCTGG AATGGACTTG TTTACATTTG
-295 TCTTTATTCA TAGGGAACAC ATCGCCATTC TGTGCTGCCC AGGCTTAGTA
-245 GCAGGGTTTT ATGGTTAAAG TGAAGGATCG AACACTCTTC AATGCCTATT
-195 CACTATGTCC ACAGTTTCAC CGGGGGCTGT AACCTATCAT CTGGCTTTCT
-145 TAACACTTGG AAAGGCATGG GGAATTGTTC CCATGGATTC TTCTGCAAAG
                                                    +1
- 95 GGATGCGTGG TGGAGATTCC AATTTTGTCA TGGTGTTTAC ATGACTTTCT
- 45 GGAATCGGTA ACATATTAAG TACTTGAACC TGGACTGGAA GGTCCATGGA
   6 CTGTATTGAC AGGTCAAACA GAAGACACTG ATGCCAGAAG CCCAGTGTCA
  56 ACACTGGAGC CAAGCAGAGA CCAACCTCAG TGCCATATTC GGAGAGCTTG
 106 AAGACTAGCT TCATGTGAAG ACTCCTTCTC CTCCAGCAGC ACAAAGCAAC
 156 CATCCTTCCA GG ATG
```

FIGURE 4: Sequence of the 5′-upstream region of the rat PK gene. The site of transcriptional initiation is indicated by the number +1. The arrow represents the position of intron A, and the site of initiation of translation is represented by boldface and the underlined ATG codon. The symbolized putative elements, the activating factor, and the reported consensus sequences are designated as follows: octamer, OCT-1, ATTTGCAT; thyroid responsive element, thyroid hormone, GATCANNNNNNTGACC; glucocorticoid responsive element, glucocorticoid, TGT(C/T)CT; estrogen responsive element, estrogen, GGTCANNNTGACC; AP-2, AP-2/AMPc/phorbol ester, CCCCAGGC. The TATA and CAAT box elements are also indicated.

of the promoter region.

The rat PK promoter region (see Figure 4) reveals a *TATTAA* sequence which corresponds to the consensus TATA box reported to bind the RNA polymerase II (Breathnach & Chambon, 1981). A total of 44 nucleotides upstream from this TATA box, we found a CCAAT box element (Benoist et al., 1980). Computer analysis of this promoter region revealed putative elements that could bind regulatory factors. At positions −343/−330, we found a possible estrogen responsive element (ERE), "GGTCAGCCTGATA", which is very similar to the consensus sequence reported as "GGTCANNNTGACC" (Maurer & Notides, 1987). Though not yet proven, this sequence might be responsible for the sexual dimorphism observed in rat PK hepatic mRNA (Seidah et al., 1989). The hexanucleotide sequence "TGTCCT", which is reported to act as a core sequence for binding the glucocorticoid receptor (Mitchell & Tjian, 1989), was localized at positions −402/−414. Other putative elements like the octomer for Oct-1 (Mitchell & Tjian, 1989), the thyroid responsive element (TRE) (Evans, 1988), and the AP-2 element (Greene & Struhl, 1988) were also identified in the upstream region of the rat PK gene. Interestingly, a 10-nucleotide palindromic structure (*GTCCA TGGAC*) was found exactly at the transcription start site proposed (see below). This structure might have some importance on the transcription regulation of the PK gene. The confirmation of these regulatory elements will require experimental proof, for example, by using transcriptional assays following transfection

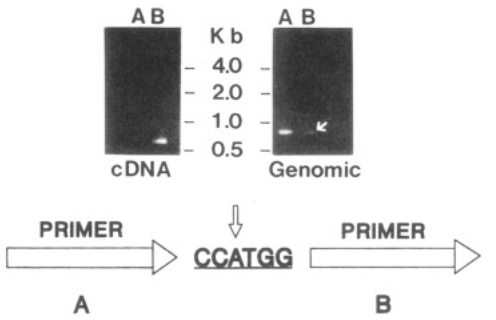

FIGURE 5: Determination of the initiation of transcription using specific PCR amplification. Primers A and B are described in the text. The left panel shows, on a 1.0% agarose gel, the amplification carried out on cDNA obtained after reverse transcription of rat liver total RNA. The right panel represents the control made on rat genomic DNA (a small arrow, on the gel, represents a less visible band). The vertical arrow indicates the proposed transcription initiation site.

of various 5′-end segments fused to a reporter gene (Maniatis et al., 1989).

Site of Initiation of Transcription. Attempts to determine the site of initiation and transcription using nuclease S1 (Greene & Struhl, 1988) or primer extension experiments (Kingston, 1987) on either liver total RNA or poly(A+) RNA [prepared as described by Seidah et al. (1989)] were so far inconclusive. This might be due to the very low level of mRNA coding for rat PK estimated to be less than 0.01% (Seidah et al., 1989). It is generally accepted that transcription begins 25–30 nucleotides downstream of the TATA box. Within the promoter region of rat PK (Figure 4), we found 2 pyrimidines preceding an adenine residue (5′-CCA-3′) at 32 nucleotides 3′ to the beginning of the TATA box. This adenine can serve as a transcription start site (Corden et al., 1980). To confirm that the initiation of transcription takes place at this position, we synthesized two sense oligonucleotides: one upstream (5′-ACCTGGACTGGAAGGTC-3′, primer A) and the other one downstream (5′-GGACTGTATTGACAGGTC-3′, primer B) of this sequence (see Figures 2 and Figures 5). In combination with an antisense oligonucleotide, localized 700 bp downstream on the cDNA sequence, we alternatively used primer A or B to amplify by PCR and cDNA prepared from rat liver total RNA, as described (Seidah et al., 1989). As shown in Figure 5 (left panel), no PCR-amplified product was obtained by using primer A. With primer B, we obtained a band migrating at the expected size (700 bp), supporting our hypothesis. As a control, we performed the PCR reaction on genomic DNA by using an antisense oligonuclotide localized at the 3′ end of intron A (Figure 2) together with either primer A or primer B. The two PCR-amplified segments migrate at the expected position as shown in Figure 5. It is emphasized that the proposed transcription initiation site is tentative, and the method used could have missed the presence of other initiation sites which may be identified by RNA protection assays (Maniatis et al., 1989). We cannot exclude the possibility of another upstream 5′-noncoding intron (containing the primer A sequence) in the PK gene, which would also explain the results of Figure 5. However, we feel that this interpretation of the PCR results is questionable, in view of the presence of both a TATA box and a CAAT box at the correct distance from each other, and from the proposed transcription initiation site (Figure 2).

Mouse and Human Gene Mapping and Evolutionary Relatedness. For the mouse PK, in the 100 metaphase cells examined after in situ hybridization, there were 135 silver grains associated with chromosomes, and 66 of these (48.8%) were located on chromosome 8; the distribution of grains on this chromosome was not random: 78.8% of them mapped to the B1–B3 region of chromosome 8. The chromosomal

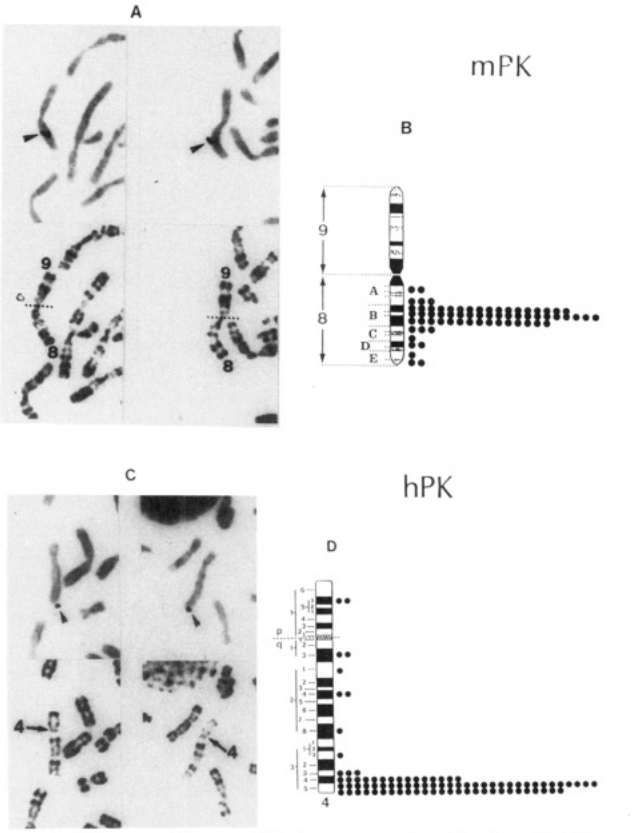

FIGURE 6: Mouse and human PK chromosomal localization. (A) Two partial WMP mouse metaphases, showing the specific site of hybridization to chromosome 8. (Top) Arrowheads indicate silver grains on Giemsa-stained chromosomes, after autoradiography. (Bottom) Chromosomes with silver grains were subsequently identified by R banding. (B) Diagram of WMP mouse Rb (8, 9) chromosome, indicating the distribution of labeled sites. (C) Two partial human metaphases showing the specific site of hybridization to chromosome 4. Top and bottom as in (A). (D) Idiogram of the human G-banded chromosome 4 illustrating the distribution of labeled sites for the human PK cDNA probe.

localization of mouse factor XI gene is not yet known. For the same number of metaphase cells examined in the human mapping, there were 234 silver grains associated with chromosomes, and 85 of these (36.3%) were located on chromosome 4; 85.8% of them mapped to the q34–q35 region of the long arm of chromosome 4, with a maximum in the q35 band (Figure 6). It could be interesting to look at this locus in Fletcher trait individuals who were described to lack prekallikrein) (Wuepper, 1972). The human factor XI gene was also localized to this region of the human chromosome 4 (Kato et al., 1989). The relatively close linkage of these two genes on the same chromosome, the number and position of introns within those genes, the similarity at the level of intron type [that appears to be a much more conservative trait in the evolution than the number and position of introns (Patthy, 1987)], and the 58% identity between human factor XI and PK cDNA strongly corroborate the hypothesis of a gene duplication event that was estimated to occur between 280 and 256 million years ago (Chung et al., 1986). Even though we used stringent hybridization and washing conditions (see Experimental Procedures), we cannot totally exclude the possibility of some cross-hybridization of the rPK or hPK probes with either mouse or human factor XI. However, it is emphasized that no other hybridizing band appears at any other chromosomal location within the mouse and human genome.

Comparison of the structural organization of the genes for several serine proteinases (Figure 7) reveals an extensive similarity between the catalytic portion of PK and factor XI and that of tissue plasminogen activator, urokinase, and tissue kallikrein genes. The distribution of the characteristic serine

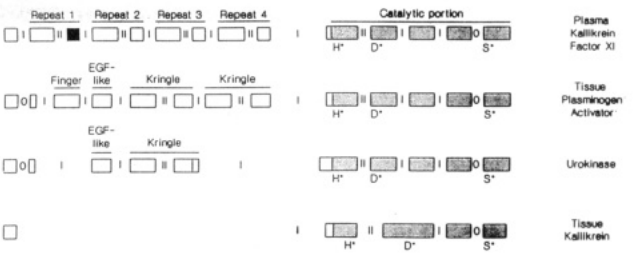

FIGURE 7: Comparison of the structural organization of the genes for several serine proteases. Exons in the catalytic portion are represented by dotted boxes and those in the heavy chain domain by open boxes. The types of introns, according to the reading frame, are indicated between each exon. The dark box represents the segment reported to bind HK in human factor XI (Baglia et al., 1990). [Modified from Rogers (1985).]

protease active sites (His*, Asp*, and Ser*) is the same in all these genes except for the tissue kallikrein (and trypsin) where one intron is missing and the active-site Asp* is consequently found in a larger exon. In this study, we demonstrated that the genic organizations of PK and the factor XI regulatory domain are similar. These types of tandem repeats have not been reported for other serine proteases.

Analysis of the gene chromosomal distribution within the serine protease family has shown an important difference between glandular kallikrein genes and those for coagulation or fibrinolytic enzymes. In the case of the mouse glandular kallikrein gene family (about 25 members), a close evolutionary relationship is reflected by their high degree of identity but also by their close linkage on mouse chromosome 7 (Mason et al., 1983). The human glandular kallikrein family seems to be much smaller (three genes), two of them [human glandular kallikrein 1 (hCK-1) and prostate-specific antigen (PA)] being tandemly located on chromosome 19 (Riegman et al., 1989). This type of organization may suggest that the ancestral gene duplications occurred relatively recently or that these are functional constraints requiring that the genes remain closely linked. In the case of coagulation, fibrinolytic, or pancreatic enzymes, the several chromosomal localizations reported do not describe such a cluster of genes. For example, coagulation factor IX was located on human chromosome X (Camerino et al., 1984); tissue plasminogen activator and urokinase genes were identified on human chromosomes 8 and 10 and mouse chromosomes 8 and 14, respectively (Rajput et al., 1985, 1987). Trypsin, chymotrypsin B, and elastase genes were identified on human chromosomes 7q22, 16, and 12, respectively (Honey et al., 1984). Despite the sequence similarity between these serine proteases, the presence of their genes on different chromosomes implies that there are no functional or regulatory constraints requiring linkage as reported by Honey et al. (1984). We report here the clustering of PK and factor XI genes, both mapping to the q34–q35 region of the long arm of human chromosome 4. The fact that these two genes are synthetic might be explained by their relatively recent divergence from each other, but this linkage may also reflect the functional constraints between these two molecules.

Interestingly, close to the mouse PK locus is found an "esterase cluster" suggested to have arisen from a common ancestral gene (Peters & Nash, 1978), with esterases 2, 5, and 11 having almost identical substrate specificities. Whether this region of the chromosome contains other clusters of closely related genes is yet to be answered. Finally, it has been recently reported (Baglia et al., 1990) that the 56–86 amino acid segment of factor XI, contained within the first tandem repeat (Figure 7, dark box), represents the binding site of factor XI to high molecular weight kininogen (HK). On the basis of sequence identity, it was possible that PK also binds HK through the equivalent region of the molecule. However, data

from Page and Coleman (1989) suggest that the binding segment of PK resides between the third and fourth repeats. Therefore, even though PK and factor XI have similar structures, gene organization, and chromosomal localization originating from a recent genic divergence, they developed specialized functions exhibited by different substrate specificities and binding sites to HK.

ACKNOWLEDGMENTS

We especially thank François Rougeon (Institut Pasteur) for allowing us to use the rat genomic library developed in his laboratory. Many thanks to Drs. E. W. Davie and D. Chung (University of Washington, Seattle) for their generous gift of the human plasma kallikrein's cDNA. We also thank Dr. Ben Liu for his help during the subcloning of the genomic fragments and Dr. Lazslô Gaspar for his comments on the manuscript. The secretarial help of Mrs. Sylvie Emond and Claudette Allard and the expert technical assistance of Josée Hamelin and Jim Rochemont are highly appreciated.

Registry No. Kallikreinogen, 9055-02-1; DNA (rat kallikreinogen gene coding region), 126700-10-5; kallikreinogen (rat reduced), 126701-15-3; blood coagulation factor XI, 9013-55-2; RNA (rat kallikreinogen messenger), 131321-99-8.

REFERENCES

Ansorge, W., & Labeit, S. (1984) *J. Biochem. Biophys. Methods 10*, 237–243.

Argentin, S., Nemer, M., Drouin, J., Scott, G. K., Kennedy, B. P., & Davies, P. L. (1985) *J. Biol. Chem. 260*, 4568–4571.

Asakai, R., Davie, E. W., & Chung, D. W.. (1987) *Biochemistry 26*, 7221–7228.

Benoist, C., O'Hare, K., Breathnach, R., & Chambon, P. (1980) *Nucleic Acids Res. 8*, 127–142.

Benton, W. D., & Davis, R. W. (1977) *Science 196*, 180–182.

Blake, C. (1983) *Nature 306*, 535–537.

Breathnach, R., & Chambon, P. (1981) *Annu. Rev. Biochem. 50*, 349–383.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., & Chambon, P. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4853–4857.

Camargo, M., & Cervenka, J. (1982) *Am. J. Hum. Genet. 34*, 757–780.

Camarino, G., Grzeschik, K. H., Jaye, M., De La Salle, H., Tolstoshev, P., Lecocq, J. P., Heilig, R., & Mandel, J. L. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 498–502.

Cereghini, S., Herbomel, P., Jounneau, J., Saragosti, S., & Katinka, M. (1983) *Cold Spring Harbor Symp. Quant. Biol. 47*, 935–944.

Chung, D. W., Fujikawa, K., McMullen, B. A., & Davie, E. W. (1986) *Biochemistry 25*, 2410–2417.

Cochrane, C. G., Revak, S. D., & Wuepper, K. D. (1973) *J. Exp. Med. 138*, 1564–1583.

Corden, J., Wasyluk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., & Chambon, P. (1980) *Science 209*, 1406–1414.

Cromlish, J. A., Seidah, N. G., & Chrétien, M. (1986) *J. Biol. Chem. 261*, 10858–10870.

Doolittle, R. F. (1983) *Ann. N.Y. Acad. Sci. 408*, 13–26.

Evans, R. E. (1988) *Science 240*, 889–895.

Fitch, W. M. (1976) in *Molecular Evolution* (Ayala, F. J., Ed.) p 160, Sinauer, Sunderland, MA.

Fujikawa, K., Chung, D. W., Hendrickson, L. E., & Davie, E. W. (1986) *Biochemistry 25*, 2417–2424.

Gallin, J. I. & Kaplan, A. P. (1974) *J. Immunol. 113*, 1928–1934.

Greene, J. M., & Struhl, K. (1988) in *Current Protocols in Molecular Biology* (Ausubel, F. M., et al., Eds.) pp 461–463, Wiley & Sons, New York.

Hamada, H., Petrino, M. G., & Kakunaga, T. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6465–6469.

Hanahan, D., & Meselson, M. (1983) *Methods Enzymol. 100*, 333–343.

Hauert, J., & Bachmann, F. (1985) *Thromb Haemostasis 54*, 122–130.

Honey, N. K., Sakaguchi, A. Y., Quinto, C., MacDonald, R. J., Bell, G. I., Craik, C., Rutter, W. J., & Naylor, S. L. (1984) *Somat. Cell. Mol. Genet. 10(4)*, 369–376.

Kaplan, A. P., Kay, A. B., & Austen, K. F. (1972) *J. Exp. Med. 135*, 81–97.

Kato, A., Asakai, R., Davie, E. W., & Aoki, N. (1989) *Cytogenet. Cell Genet. 52(1–2)*, 77–78.

Keller, E. B., & Noon, W. A. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7417–7420.

Kinston, R. E. (1987) in *Current Protocols in Molecular Biology* (Ausubel, F. M., et al., Eds.) pp 481–483, Wiley & Sons, New York.

Konarska, M. M., Grabowski, P. J., Padgett, R. A., & Sharp, P. A. (1985) *Nature 313*, 552–557.

Mandle, R., Jr., & Kaplan, A. P. (1977) *J. Biol. Chem. 252*, 6097–6104.

Maniatis, T., Jeffrey, A., & van deSande, H. (1975) *Biochemistry 14*, 3787–3794.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Mason, A. J., Evans, B. A., Cox, D. R., Shine, J., & Richards, R. I. (1983) *Nature 303*, 300–307.

Mattei, M. G., Philip, N., Passage, E., Moisan, J. P., Mandel, J. L., & Mattei, J. F. (1985) *Hum. Genet. 69*, 268–271.

Maurer, R. A., & Notides, A. C. (1987) *Mol. Cell. Biol. 7*, 4247–4254.

Metters, K. M., Rossier, J., Paquin, J., Chrêtien, M., & Seidah, N. G. (1988) *J. Biol. Chem., 263*, 12543–12553.

Miesfeld, R., Krystal, M., & Arnheim, N. (1981) *Nucleic Acids Res. 9*, 5931–5938.

Mitchell, P. J., & Tjian, R. (1989) *Science 245*, 371–378.

Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459–472.

Naora, H., & Deacon, N. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6196–6200.

Nordheim, A., Lafer, E. M., Peck, L. J., Wang, J. C., Stollar, B. P., & Rich, A. (1982) *Cell (Cambridge, Mass.) 31*, 309–318.

Page, J., & Colman, R. (1989) *Blood 74(7) (Suppl. 1)*, 291a.

Patthy, L. (1987) *FEBS Lett. 214(1)*, 1–7.

Peters, J., & Nash, H. R. (1978) *Biochem. Genet. 16(5/6)*, 553–569.

Rajput, B., Marshall, A., Killary, A. M., Lalley, P. A., Naylor, S. L., Belin, D., Rickles, R. J., & Strickland S. (1987) *Somat. Cell Mol. Genet. 13(5)*, 581–586.

Rujput, B., Degen, S. F., Reich, E., Waller, E. K., Axelrod, J., Eddy, R. L., & Shows, T. B. (1985) *Science 230*, 672–674.

Reed, K. C., & Mann, D. A. (1985) *Nucleic Acids Res. 13*, 7207–7221.

Riegman, P. H. J., Vlietstra, R. J., Klaassen, P., van der Korput, J. A. G. M., Geurts van Kessel, A., Romijn, J. C., & Trapman, J. (1989) *FEBS Lett. 247(1)*, 123–126.

Rogers, J. (1985) *Nature 315*, 458–459.

Rosinski-Chupin, I., & Rougeon, F. (1990) *J. Biol. Chem. 265*, 10709–10713.

Ruskin, B., Krainer, A. R., Maniatis, T., & Green, M. R.

(1984) *Cell 38*, 317–331.
Saiki, R., Gelffand, D. H., Stoffel, S., Scharf, S., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988) *Science 239*, 487–491.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Schapira, M., Despland, E., Scott, C. F., Boxer, L. A., & Colman, R. W. (1982) *J. Clin. Invest. 69*, 1199–1202.
Schiffman, S., Rapaport, S. I., & Patch, M. J. (1963) *Blood 22*, 733.
Sealy, J. E., Atlas, S. A., Laragh, J. H., Silvergerg, M., & Kaplan, A. P. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 5914–5918.
Seidah, N. G., Cromlish, J. A., Hamelin, J., Thibault, G., & Chrêtien, M. (1986) *Biosci. Rep. 6*, 835–842.
Seidah, N. G., Paquin, J., Hamelin, J., Benjannet, S., & Chrêtien, M. (1988) *Biochimie 70*, 33–46.
Seidah, N. G., Ladenheim, R., Mbikay, M., Hamelin, J., Lutfalla, G., Rougeon, F., Lazure, C., & Chrêtien, M. (1989) *DNA 8*, 563–574.
Seidah, N. G., Sawyer, N., Hamelin, J., Mion, P., Beaubien, G., Brachpapa, L., Rochemont, J., Mbikay, M., & Chrêtien, M. (1990) *DNA Cell Biol. 9*, 737–748.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–507.
Tait, J. F., & Fujikawa, K. (1987) *J. Biol. Chem. 262*, 11651–11656.
Takagaki, Y., Kitamura, N., & Nakanishi, S. (1985) *J. Biol. Chem. 260*, 8601–8609.
Veloso, D., Shilling, J., Shine, J., Fitch, W. M., & Colman, R. W. (1986) *Thromb. Res. 43*, 153–160.
Wachtfogel, Y. T., Kucich, U., James, H. L., Scott, C. F., Schapira, M., Zimmerman, M., Cohen, A. B., & Colman, R. W. (1983) *J. Clin. Invest. 72*, 1672–1677.
Wells, R. D., & Miglietta, J. J., Klysik, J., Larson, J. E., Stirdivant, S. M., & Zacharias, W. (1982) *J. Biol. Chem. 257*, 10166–10171.
Wilson, A. C., Carlson, S. S., & White, T. J. (1977) *Annu. Rev. Biochem. 46*, 573–639.
Wuepper, K. D. (1973) *J. Exp. Med. 138*, 1345–1355.

Nucleotide Sequence Analysis of the *Pseudomonas putida* PpG7 Salicylate Hydroxylase Gene (*nahG*) and Its 3′-Flanking Region[†,‡]

In-Soon You,[*,§] Debabrota Ghosal,[∥] and Irwin C. Gunsalus[§]

Department of Biochemistry, University of Illinois, Urbana, Illinois 61801, and DST Unit on Genetic Engineering, Center for Biotechnology, Jawaharlal Nehru University, New Delhi, India

Received August 8, 1990; Revised Manuscript Received November 2, 1990

ABSTRACT: Gene *nahG* of naphthalene/salicylate catabolic plasmid NAH7 encodes a protein of molecular weight 45 000, salicylate hydroxylase. This enzyme catalyzes the formation of catechol from salicylate, a key intermediate in naphthalene catabolism. DNA sequence analysis of the 3.1-kilobase *Hin*dIII fragment containing the *nahG* locus reveals an open reading frame (ORF) of 1305 base pairs that corresponds to a protein of 434 amino acid residues. The predicted amino acid sequence of salicylate hydroxylase is in agreement with the molecular weight, NH_2-terminal amino acid sequence, and total amino acid composition of the purified salicylate hydroxylase [You, I.-S., Murray, R. I., Jollie, D., & Gunsalus, I. C. (1990) *Biochem. Biophys. Res. Commun. 169*, 1049–1054]. The amino acid sequence between positions 8 and 37 of salicylate hydroxylase shows homology with known ADP binding sites of other FAD-containing oxidoreductases, thus confirming its biochemical function. The sequence of the *Pseudomonas putida* salicylate hydroxylase was compared with those of other similar flavoproteins. A small DNA segment (831 base pairs) disrupts the continuity of the known gene order *nahG* and *nahH*, the latter encoding catechol 2,3-dioxygenase. The complete nucleotide sequence of the intergenic region spanning genes *nahG* and *nahH* has been determined and its biological role proposed.

Pseudomonas putida PpG7 plasmid NAH7 carries all the genetic information necessary for the conversion of naphthalene to pyruvate and acetaldehyde. In plasmid NAH7, the naphthalene catabolic genes are physically and functionally organized in two operons, *nah* and *sal*, on a 25-kb *Eco*RI fragment (Yen & Gunsalus, 1982; Grund & Gunsalus, 1983). The positive regulatory gene *nahR*, which is divergently transcribed from the *sal* operon, is located between the *nah* and *sal* operons (Yen & Gunsalus, 1982, 1985; Schell, 1986). The *nah* operon (*nahABCDE*) encodes the enzymes responsible for the conversion of naphthalene to salicylate while the *sal* operon (*nahGHINLMJK*) specifies enzymes that convert salicylate to pyruvate and acetaldehyde via the meta-cleavage pathway. The gene order of the two operons follows the order of biochemical and enzymatic steps in the naphthalene catabolic pathway (Yen & Gunsalus, 1982; Harayama et al., 1987). The *nahG* gene, the most proximal gene of the *sal* operon, codes for salicylate hydroxylase.

Salicylate hydroxylase (E.C. 1.14.13.1), a flavoprotein, converts salicylate to catechol via decarboxylative hydroxylation (Figure 1). This enzyme was first purified from *P. putida* (Yamamoto et al., 1965) and later from *Pseudomonas cepacia* (Tu et al., 1981) and other soil microorganisms (White-Stevens & Kamin, 1972). Although the biochemistry of salicylate hydroxylase is known in detail (Wang & Tu, 1984;

[†] This work was supported in part by NIH Grant DK00562-33 and a grant in aid from Du Pont.
[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J05317.
[*] Correspondence should be addressed to this author. Present address: Department of Biology, California State University, Fresno, CA 93740-0073.
[§] University of Illinois.
[∥] Jawaharlal Nehru University.

0006-2960/91/0430-1635$02.50/0