Nucleotide Sequence of Cloned cDNA for Human Pancreatic Kallikrein†

Daikichi Fukushima, Naomi Kitamura, and Shigetada Nakanishi*

Institute for Immunology, Kyoto University Faculty of Medicine, Kyoto 606, Japan

Received April 15, 1985

ABSTRACT: Cloned cDNA sequences for human pancreatic kallikrein have been isolated and determined by molecular cloning and sequence analysis. The identity between human pancreatic and urinary kallikreins is indicated by the complete coincidence between the amino acid sequence deduced from the cloned cDNA sequence and that reported partially for urinary kallikrein. The active enzyme form of the human pancreatic kallikrein consists of 238 amino acids and is preceded by a signal peptide and a profragment of 24 amino acids. A sequence comparison of this with other mammalian kallikreins indicates that key amino acid residues required for both serine protease activity and kallikrein-like cleavage specificity are retained in the human sequence, and residues corresponding to some external loops of the kallikrein diverge from other kallikreins. Analyses by RNA blot hybridization, primer extension, and S1 nuclease mapping indicate that the pancreatic kallikrein mRNA is also expressed in the kidney and sublingual gland, suggesting the active synthesis of urinary kallikrein in these tissues. Furthermore, the tissue-specific regulation of the expression of the members of the human kallikrein gene family has been discussed.

Glandular kallikreins (EC 3.4.21.8) are members of a closely related subfamily of serine proteases that exhibit a strict specificity in their proteolytic actions [see review from Schachter (1980)]. They are characterized by their ability to release vasoactive peptides, kinins, from kininogen in vitro, though the physiological significance of proteolytic actions of these enzymes seems to be unrelated to the release of kinins at least in some instances, and certain kallikreins are thought to be involved in the specific processing for the generation of biologically active peptides as well as factors from their precursors. Recently, the mRNA sequences for mouse submaxillary gland kallikrein and for rat pancreatic kallikrein have been elucidated by molecular cloning and sequence analysis of their cloned cDNAs (Swift et al., 1982; Richards et al., 1982), and the structure of the multigene family for mouse kallikreins has been reported by molecular analysis of their cloned genomic DNAs (Mason et al., 1983).

Human kallikreins in various tissues have been purified and characterized (Amouric & Figarella, 1979; Geiger et al., 1980; Lottspeich et al., 1979), and the limited amino acid sequence in the amino-terminal portion of urinary kallikrein has been reported (Lottspeich et al., 1979). However, the complete amino acid sequences of human kallikreins remain to be determined. Furthermore, although studies have been extensively made about the physiological or pathophysiological roles of kallikreins in blood pressure and blood flow regulation, in hypertension, in kidney function or disfunction, etc. [see reviews from Mills (1979) and Colman & Wong (1979)], there is little information about the sites and the regulation of biosynthesis of human kallikreins.

To study the regulatory mechanisms involved in the biosynthesis of human kallikreins and to reveal their primary structures, we have undertaken the construction and sequence analysis of cDNAs for human pancreatic kallikrein. We here report the complete amino acid sequence of human pancreatic kallikrein deduced from the nucleotide sequence and its identity with urinary kallikrein. We also describe the nature of the preproenzyme and the expression of the mRNA in tissues other than the pancreas.

Experimental Procedures

Materials. Reagents were obtained as follows: [α-^{32}P]-dCTP (3000 Ci/mmol) and [γ-^{32}P]ATP (5000 Ci/mmol) from Amersham Co.; oligo(dT)–cellulose (type 7) and *Escherichia coli* ribonuclease H from P-L Biochemicals; *E. coli* DNA polymerase I and *E. coli* DNA ligase from New England Biolabs; S1 nuclease from Sankyo Co.; avian myeloblastosis virus reverse transcriptase from Life Science Inc.; aminobenzyloxymethyl–paper from Schleicher & Schuell; terminal deoxyribonucleotidyl transferase, bacterial alkaline phosphatase, and T4 polynucleotide kinase from Takara Shuzo Co.; restriction endonucleases from Takara Shuzo Co., Toyobo Co., Bethesda Research Laboratories, and New England Biolabs. The oligodeoxyribonucleotide was synthesized by the triester method with the aid of the full-automated solid-phase synthesizer, Model 25A, Genetic Design Inc.

Cloning Procedures. Isolation of cloned cDNAs for both rat and human kallikreins was performed according to the procedures described previously (Nawa et al., 1983; Kageyama et al., 1984). In brief, total RNA was extracted from a rat or human pancreas as described (Chirgwin et al., 1979), and poly(A) RNA was isolated by subjecting the total RNA extracted to oligo(dT)–cellulose chromatography (Aviv & Leder, 1972). A rat and a human pancreatic cDNA library was constructed by the method of Okayama & Berg (1982) from 2.2 μg of poly(A) RNA and 2.8 μg of the vector-primer DNA in both cases. *E. coli* HB101 was transformed and selected for ampicillin resistance (Morrison, 1979). About 10 000 transformants derived from the rat cDNA library were screened by hybridization at 38 °C with the oligodeoxyribonucleotide probe specified in the text; the probe was labeled at the 5′ end with [γ-^{32}P]ATP. One hybridization-positive clone prKK1 was isolated, and the 447-bp *Pvu*II–*Hin*fI and 149-bp *Hin*fI-*Dde*I fragments derived from this cDNA insert were used for subsequent cloning of cDNAs for human kallikrein. About 80 000 transformants derived from the human

† This investigation was supported in part by research grants from the Ministry of Education, Science and Culture of Japan and the Science and Technology Agency of Japan.

0006-2960/85/0424-8037$01.50/0

cDNA library were screened by hybridization at 50 °C with the cDNA probe described; the probe was labeled by nick translation (Rigby et al., 1977) with [α-^{32}P]dCTP. The human genomic DNA library was kindly provided by Dr. T. Maniatis (Harvard University, Boston), and was a collection of recombinant phage that carried human fetal liver DNA fragments generated by partial digestion with *Hae*III and *Alu*I and joined to the λ Charon 4A arms (Lawn et al., 1978). The phage were screened by hybridization in situ with the cDNA fragments described above. Hybridization-positive phage clones were isolated by repeated plaque purification (Maniatis et al., 1982). Subcloning of genomic DNA fragments in plasmid pBR322 was performed as described (Maniatis et al., 1982). All of the cloning procedures were conducted in accordance with the guidelines for research involving recombinant DNA molecules issued by the Ministry of Education, Science, and Culture of Japan.

RNA Blot-Hybridization Analysis. Total RNA and poly(A) RNA were isolated from various tissues as described above. Poly(A) RNA (10 μg except for 3 μg of sublingual RNA) was denatured with 1 M glyoxal/50% (v/v) dimethyl sulfoxide, electrophoresed on a 1% agarose gel and transfered to diazobenzyloxymethyl-paper (McMaster & Carmichael, 1977). Hybridization and washing was carried out according to the procedures described (Alwine et al., 1977). The 562-bp *Ava*II fragment and the 233-bp *Ava*II–*Sau*3AI fragment were excised from clone phKK25 and were labeled by nick translation. A mixture of these two fragments was used as a probe. The size markers used were human rRNA.

Primer-Extension Analysis. The 77-bp *Bst*NI fragment was isolated from clone phKK25, labeled at the 5′ ends with [γ-^{32}P]ATP, and digested with *Hin*fI. The 51-bp *Hin*fI–*Bst*NI fragment corresponding to the 5′-terminal portion of the protein-coding region of the kallikrein mRNA was isolated, denatured at 80 °C for 5 min, and hybridized to poly(A) RNA (10 μg except for 1 μg of sublingual RNA) in a solution containing 80% (v/v) formamide, 40 mM 1,4-piperazinediethanesulfonic acid buffer, pH 6.4, 0.4 M NaCl, and 1 mM EDTA at 42 °C for 3 h. The RNA–DNA hybrids were precipitated by ethanol and then subjected to reverse transcription reaction (Ohkubo et al., 1983). The primer-extended cDNA transcripts were extracted with phenol, precipitated by ethanol, and electrophoresed on a 7 M urea/8% polyacrylamide gel. The sizes of the primer-extended cDNAs were estimated by comparison with chemically degraded products of the 340-bp *Rsa*I–*Hpa*II fragment isolated from clone phKK25, which were run on a DNA sequencing gel along with the primer-extended cDNAs.

S1 Nuclease Mapping Analysis. The 2129-bp *Ban*I fragment was isolated from clone phKK25, labeled at the 5′ ends, and digested with *Dra*I. The 1047-bp *Ban*I–*Dra*I fragment containing the 742-bp cDNA sequence and its flanking sequence composed of the 35-bp oligo(dC-dG) tail and the 270-bp vector DNA was isolated, denatured at 80 °C for 5 min, and hybridized to poly(A) RNA (2 μg of pancreatic RNA and 10 μg of kidney and liver RNAs) at 53 °C for 3 h in the solution described for the primer-extension analysis. The RNA–DNA hybrids were digested with S1 nuclease (10–500 units/μg of pancreatic RNA and 10–100 units/μg of kidney and liver RNAs) at 40 °C for 30 min (Berk & Sharp, 1977; Weaver & Weissmann, 1979). The S1 digestion products were precipitated by ethanol and electrophoresed on a 7 M urea/5% polyacrylamide gel. The size markers used were the *Pst*I-digested fragment (781 nucleotides) and the *Hha*I-digested fragment (658 nucleotides) of the *Ban*I–*Dra*I fragment described above as well as the *Hin*fI–digested pBR322 DNA.

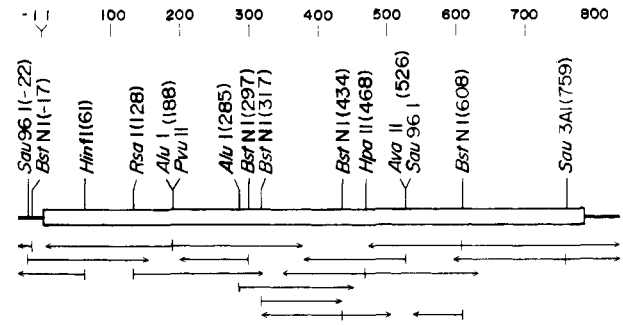

FIGURE 1: Strategy for determining the sequence of the cDNA insert in clone phKK25. The map displays only the relevant restriction endonuclease sites, which are identified by numbers indicating the 5′-terminal nucleotide generated by cleavage (for the nucleotide numbers, see Figure 2). The sequence corresponding to the coding region is indicated by the outlined box. The direction and extent of sequence determinations are shown by horizontal arrows; the sites of 5′-end labeling are indicated by short vertical lines on the arrows.

Other Analytical Procedures. Procedures for restriction endonuclease digestion and 5′-end labeling of DNA were as described (Kageyama et al., 1984). DNA sequence analysis was carried out by the procedure of Maxam & Gilbert (1980). DNA blot-hybridization analysis was conducted as described by Southern (1975); the 562-bp *Ava*II and 233-bp *Ava*II–*Sau*3AI fragments derived from clone phKK25 were labeled by nick translation and used as a hybridization probe.

RESULTS

Isolation of cDNA Clones. As an initial step for the isolation of cDNA clones for human pancreatic kallikrein, we cloned a DNA sequence complementary to the rat pancreatic kallikrein mRNA, the sequence of which was reported by molecular cloning and sequence analysis (Swift et al., 1982). The tetradecamer oligodeoxyribonucleotide 5′-AGGTGGAGCAGCAT-3′ was chemically synthesized according to the complementary sequence reported for the pentapeptide Met-Leu-Leu-His-Leu [residues 98–102 in Figure 2 of Swift et al., (1982)] of rat pancreatic kallikrein and was used for a hybridization probe to select rat kallikrein cDNA clones from a rat pancreatic cDNA library. This sequence was chosen because it contained a high GC content and exhibited no homology with the Okayama–Berg vector. One hybridization-positive clone prKK1 was isolated from about 10000 transformants, and this cDNA clone was verified to contain the cDNA sequence for rat pancreatic kallikrein by partial nucleotide sequence analysis (data not shown).

A human pancreatic cDNA bank constructed by using human pancreatic poly(A) RNA and the Okayama–Berg vector was then selected by hybridization with the cDNA fragments derived from clone prKK1. Eleven hybridization-positive clones were isolated from about 80000 transformants, and the largest cDNA in the 11 clones (clone phKK25) was subjected to nucleotide sequence analysis according to the strategy indicated in Figure 1.

Primary Structure of Human Pancreatic Kallikrein mRNA. The primary structure of the mRNA was deduced from the 871-nucleotide cDNA sequence determined (Figure 2). The translational initiation site was assigned to the methionine codon AUG located nearest the 5′ end of the mRNA, because the primer-extension analysis described below indicated that the cDNA insert in clone phKK25 covered almost the full length of the mRNA sequence. The open-reading frame of the mRNA consists of 786 nucleotides encoding 262 amino

```
                                                                         UCCUCCACCUGCUGGCCCCUGGACACCUCUGUCACC   -1

-24             -20                                     -10                                 -1   1
Met Trp Phe Leu Val Leu Cys Leu Ala Leu Ser Leu Gly Gly Thr Gly↓Ala↓Ala↓Pro Pro Ile Gln Ser Arg Ile Val Gly Gly Trp Glu
AUG UGG UUC CUG GUU CUG UGC CUC GCC CUG UCC CUG GGG GGG ACU GGU GCU GCG CCC CCG AUU CAG UCC CGG AUU GUG GGA GGC UGG GAG   90

            10                                  20                                  30
Cys Glu Gln His Ser Gln Pro Trp Gln Ala Ala Leu Tyr His Phe Ser Thr Phe Gln Cys Gly Gly Ile Leu Val His Arg Gln Trp Val
UGU GAG CAG CAU UCC CAG CCC UGG CAG GCG GCU CUG UAC CAU UUC AGC ACU UUC CAG UGU GGG GGC AUC CUG GUG CAC CGC CAG UGG GUG  180

            40                                  50                                  60
Leu Thr Ala Ala[His]Cys Ile Ser Asp Asn Tyr Gln Leu Trp Leu Gly Arg His Asn Leu Phe Asp Asp Glu Asn Thr Ala Gln Phe Val
CUC ACA GCU GCU[CAU]UGC AUC AGC GAC AAU UAC CAG CUC UGG CUG GGU CGC CAC AAC UUG UUU GAC GAC GAA AAC ACA GCC CAG UUU GUU  270

            70                                  80                                  90
His Val Ser Glu Ser Phe Pro His Pro Gly Phe Asn Met Ser Leu Leu Glu Asn His Thr Arg Gln Ala Asp Glu Asp Tyr Ser His[Asp]
CAU GUC AGU GAG AGC UUC CCA CAC CCU GGC UUC AAC AUG AGC CUC CUG GAG AAC CAC ACC CGC CAA GCA GAC GAG GAC UAC AGC CAC[GAC] 360

            100                                 110                                 120
Leu Met Leu Leu Arg Leu Thr Glu Pro Ala Asp Thr Ile Thr Asp Ala Val Lys Val Val Glu Leu Pro Thr Gln Glu Pro Glu Val Gly
CUC AUG CUG CUC CGC CUG ACA GAG CCU GCU GAU ACC AUC ACA GAC GCU GUG AAG GUC GUG GAG UUG CCC ACC CAG GAA CCC GAA GUG GGG  450

            130                                 140                                 150
Ser Thr Cys Leu Ala Ser Gly Trp Gly Ser Ile Glu Pro Glu Asn Phe Ser Phe Pro Asp Asp Leu Gln Cys Val Asp Leu Lys Ile Leu
AGC ACC UGU UUG GCU UCC GGC UGG GGC AGC AUC GAA CCA GAG AAU UUC UCA UUU CCA GAU GAU CUC CAG UGU GUG GAC CUC AAA AUC CUG  540

            160                                 170                                 180
Pro Asn Asp Glu Cys Glu Lys Ala His Val Gln Lys Val Thr Asp Phe Met Leu Cys Val Gly His Leu Glu Gly Gly Lys[Asp]Thr Cys
CCU AAU GAU GAG UGC GAA AAA GCC CAC GUC CAG AAG GUG ACA GAC UUC AUG CUG UGU GUC GGA CAC CUG GAA GGU GGC AAA[GAC]ACC UGU  630

            190                                 200                                 210
Val Gly Asp[Ser]Gly Gly Pro Leu Met Cys Asp Gly Val Leu Gln Gly Val Thr Ser Trp Gly Tyr Val Pro Cys Gly Thr Pro Asn Lys
GUG GGU GAU[UCA]GGG GGC CCG CUG AUG UGU GAU GGU GUG CUC CAA GGU GUC ACA UCA UGG GGC UAC GUC CCU UGU GGC ACC CCC AAU AAG  720

            220                                 230                     238
Pro Ser Val Ala Val Arg Val Leu Ser Tyr Val Lys Trp Ile Glu Asp Thr Ile Ala Glu Asn Ser
CCU UCU GUC GCC GUC AGA GUG CUG UCU UAU GUG AAG UGG AUC GAG GAC ACC AUA GCG GAG AAC UCC UGA ACGCCCAGCCCUGUCCCCUACCCCCAG  816

UAAAAUCAAAUGUGCAUCC                                                                                                    835
```

FIGURE 2: Primary structure of human pancreatic preprokallikrein mRNA. The nucleotide sequence of the mRNA was deduced from that of the cDNA insert of clone phKK25. Nucleotide residues are numbered in the 5′ to 3′ direction, beginning with the first residue of the initiation methionine codon, and the nucleotides on the 5′ side of residue 1 are indicated by negative numbers. The amino acid sequence is numbered from the amino terminus of the predicted active form of the kallikrein, and the amino acid residues on the amino-terminal side of residue 1 are indicated by negative numbers. The predicted cleavage points between the signal peptide and the profragment are indicated by arrows. The amino acid residues of the charge-relay system and the recognition site for substrates are boxed. The variant-sequence AGUAAA as a possible polyadenylation and processing signal is underlined. The sequence complementary to the conserved 3′-terminal sequence of eukaryotic 18S rRNA has a line above it.

acid residues with a calculated M_r of 28 889. The deduced amino acid sequence of 31 residues starting with Ile (residue 1 in Figure 2) precisely corresponded to the amino-terminal sequence partially reported for human urinary kallikrein (Lottspeich et al., 1979). Thus, the amino-terminal sequence of 24 amino acid residues most likely represents the signal peptide and the proenzyme fragment. When compared with the sequences of members of the kallikrein-like subfamily of serine proteases, the key amino acid residues that determine the serine protease activity as well as the cleavage preference of kallikrein-like enzymes were retained in the deduced amino acid sequence (see Discussion). Furthermore, all but one cysteine residue in the signal peptide region were observed at positions equivalent to those reported for rat and mouse kallikreins. Thus, we concluded that the sequence encoded by clone phKK25 represents a human preprokallikrein.

The 5′-untranslated region of the human kallikrein mRNA contained a sequence complementary to the conserved 3′-terminal sequence of eukaryotic 18S rRNA as observed for the rat kallikrein mRNA (Swift et al., 1982). The 3′-untranslated region of the human mRNA consists of 49 nucleotides. Instead of the canonical hexanucleotide AAUAAA observed as the polyadenylation signal in most eukaryotic mRNAs (Proudfoot & Brownlee, 1976), the sequence AGUAAA was found 16 nucleotides upstream from the polyadenylation site of the mRNA. This sequence was also identified in the two other cDNA clones isolated independently as well as one genomic clone determined (data not shown).

Expression of the Kallikrein mRNA in Other Tissues. In order to investigate the expression of the pancreatic kallikrein

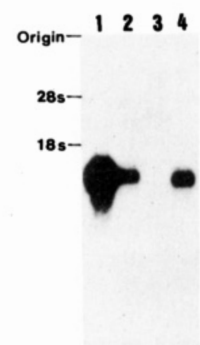

FIGURE 3: Blot-hybridization analysis of kallikrein mRNA in various tissues. Poly(A) RNA samples analyzed were derived from the following tissues: (lane 1) pancreas; (lane 2) kidney; (lane 3) liver; (lane 4) sublingual gland. Experimental details were described under Experimental Procedures.

mRNA or its related mRNAs in other tissues, RNAs isolated from various tissues were analyzed by blot-hybridization techniques with use of a cDNA fragment derived from clone phKK25 as a probe. The results of such an analysis are presented in Figure 3. The poly(A) RNA preparations isolated from the pancreas, kidney, and sublingual gland showed a hybridization-positive band with an identical mobility corresponding to a size of approximately 1050 nucleotides. No hybridization-positive band was detected with the liver poly(A) RNA. Among the RNA preparations analyzed, the pancreatic poly(A) RNA exhibited the most intense hybridization signal.

The above result indicates that the kidney as well as the sublingual gland expresses a mRNA sequence identical with

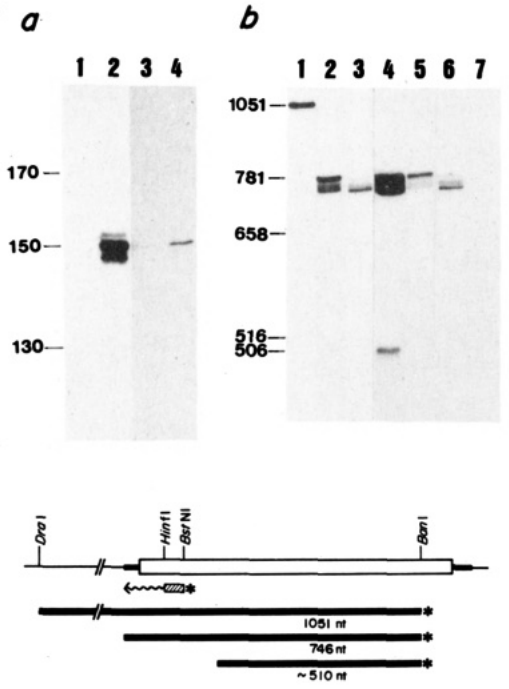

FIGURE 4: Primer-extension and S1 nuclease mapping analyses of kallikrein mRNA in various tissues. In the primer extension analysis shown in (a), RNA preparations used are as follows: (lane 1) yeast tRNA; (lanes 2–4) poly(A) RNA preparations derived from pancreas, sublingual gland, and kidney, respectively. In the S1 nuclease mapping analysis displayed in (b), lane 1 shows the cDNA probe without S1 nuclease digestion, and lanes 2–4, lanes 5 and 6, and lane 7 show the S1 digestion products with the poly(A) RNA preparations derived from the pancreas, kidney, and liver, respectively. The amounts of S1 nuclease used were 10 units/μg of RNA (lanes 5 and 7), 100 units/μg of RNA (lanes 2, 4, and 6), and 500 units/μg of RNA (lane 3). Lane 4 was obtained by exposing lane 2 for 3 times its exposure time. Experimental details were described under Experimental Procedures. The restriction maps relevant to the primer-extension and the S1 nuclease mapping analyses are shown under the autoradiographs; the open box, the solid line, and the thin line stand for the protein-coding region, the 5′- and 3′-untranslated regions, and the vector sequence, respectively. The shaded box and the wavy line under the restriction map indicate a DNA fragment used for the primer-extension analysis and a primer-extended cDNA transcript, respectively. The thick lines show S1 nuclease resistant fragments. The asterisks denote ^{32}P-labeled sites.

that of the pancreatic kallikrein mRNA or its closely related sequence. In order to further characterize the mRNA sequence detected in the kidney and sublingual gland, the RNA preparations isolated from these tissues were analyzed in parallel with that from the pancreas by primer extension and S1 nuclease mapping methods. As shown in Figure 4a, multiple but identical bands were observed among the three RNA preparations upon the primer-extension analysis of cDNA transcripts formed by elongation of the *Hin*fI–*Bst*NI fragment corresponding to the 5′ terminus of the protein-coding region of the mRNA. This result indicates that the mRNA for the pancreatic kallikrein is also expressed in the kidney as well as in the sublingual gland and that the mRNAs expressed in these three tissues share an identical sequence with one or more 5′ termini starting with 33–40 nucleotides upstream from the translational initiation site; it remains to be determined whether the observed multiple bands represent a heterogeneity of the 5′ ends of the kallikrein mRNA or result from heterogeneous terminations of the reverse transcription reaction at the 5′ end of the template. The sequence identity between the kallikrein mRNA expressed in the pancreas and kidney was further supported by S1 nuclease mapping analysis. In the experiment shown in Figure 4b, the cDNA probe representing the pancreatic mRNA was hybridized to the RNA preparation derived from either the pancreas or the kidney, and the hybrids formed were then subjected to S1 nuclease digestion. As can be seen in the autoradiograph presented in Figure 4b, the kidney mRNA, like the pancreatic mRNA, gave rise to an S1 nuclease resistant band of 746 nucleotides, which corresponded to the size of the whole cDNA sequence used; several extra bands observed above the band of the 746 nucleotides probably represent artifactual S1 nuclease resistant products that may result from the hybrid formation of the large dG·dC tail of the cDNA probe used. This result thus indicates that the kidney kallikrein mRNA shares an identical sequence with the pancreatic kallikrein mRNA. The pancreatic RNA preparation revealed one more minor band with a mobility corresponding to the size of approximately 510 nucleotides (Figure 4b, lanes 2 and 4). However, characterization for this band was not further proceeded.

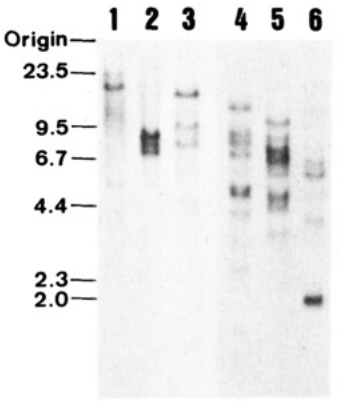

FIGURE 5: Blot-hybridization analysis of human and rat cellular DNAs. Total human placental cellular DNA (lanes 1–3) and rat liver cellular DNA (lanes 4–6) were digested with *Eco*RI (lanes 1 and 4), *Bam*HI (lanes 2 and 5), and *Hin*dIII (lanes 3 and 6). Other details were described under Experimental Procedures. The size markers used were the *Hin*dIII-digested λ DNA.

Human Kallikrein Gene Family. It has previously been reported that kallikreins of mouse and rat are encoded by a large gene family of 25–30 members (Mason et al., 1983; Howles et al., 1984). In order to characterize the human kallikrein genes, total cellular DNAs isolated from human placenta as well as from rat livers were digested with several restriction enzymes, and these DNA preparations were analyzed by blot-hybridization techniques with a cDNA fragment isolated from clone phKK25 as a probe. As shown in Figure 5, the human DNA revealed multiple hybridization-positive bands, but the number of bands exhibiting intense signals was found to be less in the human DNA than in the rat DNA. The characterization of the human kallikrein genes was also conducted by isolating genomic clones for human kallikreins by hybridization in situ with the cDNA probe used for the isolation of cDNA clones for human pancreatic kallikrein. Fifty four hybridization-positive clones were isolated from about 300000 plaques derived from a human genomic library. When one of these clones that exhibited a strong hybridization signal (λhKK3) was subjected to partial sequence analysis (data not shown), the amino acid sequence encoded by two exons was found to be highly homologous to the corresponding sequence of amino acid residues 92–141 and 188–238 presented in Figure 2 (see Figure 6). However, these two amino acid sequences differ in 32 out of 101 positions, corresponding to 68% homology. Thus, although the DNA blot-hybridization analysis suggests that the sequences closely related to the human pancreatic kallikrein mRNA are limited in the human genome, the large number of genomic clones isolated together

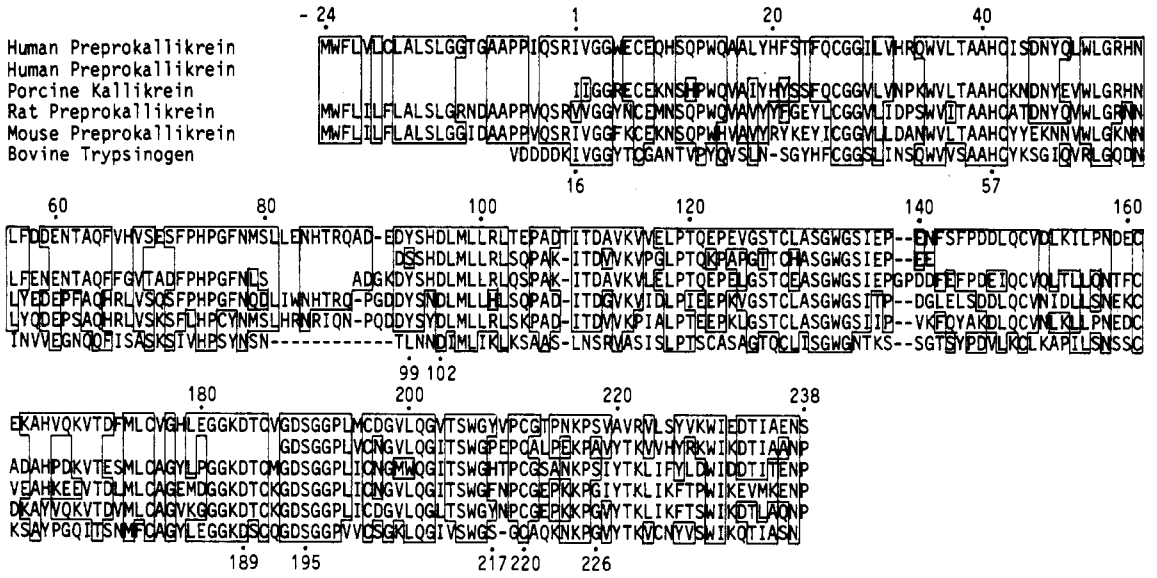

FIGURE 6: Amino acid sequence comparison of mammalian kallikreins. The numbering scheme starts at the amino terminus of the predicted active enzyme form of human kallikrein. The partial amino acid sequence of human kallikrein shown in the second line was deduced from the genomic DNA sequence of λhKK3. To facilitate comparison with other serine proteases, the sequence of bovine trypsinogen is shown on the bottom line and chymotrypsinogen numbering is included for key amino acid positions. The boxes delineate regions of amino acid sequence identical with that of human pancreatic preprokallikrein. The data of porcine kallikrein, rat preprokallikrein, mouse preprokallikrein, and bovine trypsinogen were taken from Bode et al. (1983), Swift et al. (1982), Mason et al. (1983), and Young et al. (1978), respectively.

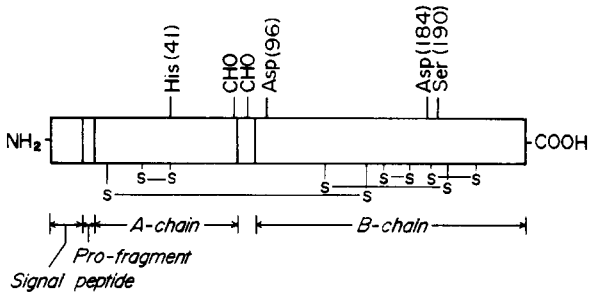

FIGURE 7: Schematic representation of human preprokallikrein. The cleavage site between the signal peptide and the profragment is not definitive. The positions for the amino acid residues of the charge-relay system and the recognition site for substrates are indicated. S–S and CHO indicate the disulfide linkage and the possible glycosylation site. For other explanations, see also the text.

with the partial nucleotide sequence determined for one of these clones support the view that they are members of the human kallikrein gene family.

Discussion

The comparison of the amino acid sequence of the human preprokallikrein revealed in this study with those reported for preprokallikreins or mature kallikreins of various animal species (Swift et al., 1982; Bode et al., 1983; Mason et al., 1983) is summarized in Figure 6. According to this comparison, the structure of the human preprokallikrein is diagrammatically illustrated in Figure 7. We have found that the sequence of residues 1–31 completely coincides with the partial sequence of 31 residues reported for the amino-terminal portion of human urinary kallikrein (Lottspeich et al., 1979). Furthermore, the amino-terminal sequences following the translational initiation site are highly conserved between the preprokallikrein of the human and those of the mouse and rat. Thus, it is reasonable to assume that the sequence preceding the mature human kallikrein is composed of the signal peptide and the profragment as observed in other serine proteases (MacDonald et al., 1982). However, the location of the cleavage site between the signal peptide and the profragment is not clear. The signal peptides usually terminate in amino acids with a small neutral side chain (Steiner et al., 1980), and the signal peptide sequences reported for other serine proteases consist of 15–18 amino acid residues (MacDonald et al., 1982). Thus, we have tentatively assigned the cleavage site of the signal peptide at a position after either Gly (−9), Ala (−8), or Ala (−7) in the human preprokallikrein sequence. It has been reported that the human pancreatic kallikrein purified from duodenal juice can be dissociated by reducing agents into two chains of estimated M_r 9000 and 14 000 and that equimolar isoleucine and alanine have been identified as amino-terminal residues for these two chains (Hofmann & Geiger, 1983). If a single-chain molecule of human kallikrein is split at positions corresponding to those observed in porcine kallikrein, the two chains can be formed by removing the central portion located between residues 81 and 88. Noteworthy is also that 10 cysteine residues in human kallikrein are all located at positions equivalent to those observed in porcine kallikrein. Thus, five disulfide linkages can be formed in human kallikrein as observed in the porcine counterpart (Bode et al., 1983), and one of these linkages between residues 7 and 150 should be involved in the connection of the two chains. There are two potential glycosylation sites conforming to the canonical Asn-X-Ser/Thr sequence at positions 78–80 and 84–86, one of which is observed at equivalent positions in the mouse and the rat kallikreins.

Amino acid residues required for either serine protease activity or kallikrein-like cleavage specificity are retained in the human pancreatic enzyme. His-57, Asp-102, Ser-195, and Asp-189 are present at positions equivalent to those observed in other serine proteases; to facilitate comparison with other serine proteases, the amino acid numbers described after the hyphen are of chymotrypsinogen numbering (see Figure 6). The former three residues constitute the charge-relay system for the proteolytic catalysis, while the latter one serves as a recognition site for the arginyl group of substrates (Young et al., 1978). The X-ray crystallographic analysis of porcine kallikrein has indicated that Tyr-99 protrudes into the binding site and interferes with the binding of peptide substrates (Chen & Bode, 1983; Bode et al., 1983). It has also been shown that the kallikrein specificity pocket is significantly enlarged due

to a longer peptide segment of residue 217 to residue 220 and due to the outward orientation of the carbonyl group of Pro-219 and that the side chain of Ser-226 partially covers Asp-189 at the bottom of the pocket. These characteristic features involved in the substrate cleavage specificity of kallikrein are preserved in human kallikrein. In contrast, clusters of small sections of divergent sequences occur at intervals along the whole sequences of human and porcine kallikreins, and these divergent sections (e.g., residues 32–34, 43–44, 58–59, 140–141, 171–172, and 219–225 in the human sequence) correspond to the external loops of porcine kallikrein. One amino acid addition observed at position 108 in the human sequence as compared with the porcine counterpart is also located at the region corresponding to the external loop of porcine kallikrein. Noteworthy also is that human kallikrein terminates with the serine residue instead of the proline residue, which is generally observed in other kallikreins, and the carboxyl terminus of kallikrein is again reported to belong to an external region in the kallikrein structure (Bode et al., 1983). It has thus been found that human and porcine kallikreins have an overall identity in 67.4% of the amino acid positions.

It has been suggested that hexanucleotide AAUAAA found in 15–25 nucleotides upstream from the polyadenylation site in most eukaryotic mRNAs serves as at least part of a recognition signal required for proper polyadenylation and processing of eukaryotic mRNAs (Proudfoot & Brownlee, 1976; Proudfoot, 1984), and a number of lines of evidence has indeed supported this concept (Montell et al., 1983; Wickens & Stephenson, 1984). Instead of the common AAUAAA sequence, however, we have found that the human kallikrein mRNA sequences deduced from the three independent cDNA inserts as well as that inferred from one of the genomic DNAs contain the AGUAAA sequence at the corresponding region of the mRNA. Interestingly, the same AGUAAA sequence has been reported as the polyadenylation and processing signal for the c-Ha-ras-1 oncogene mRNA as well as for the long terminal repeat of the mouse mammary tumor virus (Donehower et al., 1981; Capon et al., 1983). However, the particular function of this sequence in the expression of these mRNAs and the efficiency as a signal for the polyadenylation and processing remain to be determined.

The primer-extension and the S1 nuclease mapping analyses have indicated that the same mRNA sequence for kallikrein is expressed in the pancreas, kidney, and sublingual gland. Furthermore, the identity between this kallikrein and urinary kallikrein has been indicated by their limited sequence comparison. Thus, the source of urinary kallikrein can be attributed to not only its synthesis by the kidney but also the circulating kallikrein from pancreatic and/or sublingual origins that is filtered from the kidney into the urine, although the sole involvement of the former or the latter kallikrein cannot be excluded.

Kallikrein mRNAs exhibit intriguing patterns of the tissue-specific regulation. The presence of the identical kallikrein mRNA sequence in the pancreas, kidney, and sublingual gland implies that the same structural gene is expressed in these tissues. Interestingly, this mRNA is expressed in relatively high but variable levels in the above tissues but not in the liver. We also noted that the mRNA encoded by one of the genomic clones, λhKK3, was not detected in the RNA preparations of pancreas, kidney, sublingual gland, or liver (data not shown). Thus, the expression of each kallikrein gene seems to be differently regulated, and the kallikrein gene family would provide an intriguing system for understanding the regulation of expression of eukaryotic genes.

Registry No. DNA (human pancreas kallikrein messenger RNA complementary), 99052-62-7; preprokallikrein (human pancreas reduced), 99052-64-9; kallikrein (human pancreas reduced), 99052-63-8; kallikrein, 9001-01-8.

REFERENCES

Alwine, J. C., Kemp, D. J., & Stark, G. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5350–5334.

Amouric, M., & Figarella, C. (1979) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 457–465.

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

Berk, A. J., & Sharp, P. A. (1977) *Cell (Cambridge, Mass.) 12*, 721–732.

Bode, W., Chen, Z., Bartels, K., Kutzbach, C., Schmidt-Kastner, G., & Bartunik, H. (1983) *J. Mol. Biol. 164*, 237–282.

Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H., & Goeddel, D. V. (1983) *Nature (London) 302*, 33–37.

Chen, Z., & Bode, W. (1983) *J. Mol. Biol. 164*, 283–311.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry 18*, 5294–5299.

Colman, R. W., & Wong, P. Y. (1979) *Handb. Exp. Pharmacol. 25 (Suppl.)*, 569–607.

Donehower, L., Huang, A., & Hager, G. (1981) *J. Virol. 37*, 226–238.

Geiger, R., Stuckstedte, U., & Fritz, H. (1980) *Hoppe-Seyler's Z. Physiol. Chem. 361*, 1003–1016.

Hofmann, W., & Geiger, R. (1983) *Hoppe-Seyler's Z. Physiol. Chem. 364*, 413–423.

Howles, P. N., Dickinson, D. P., DiCaprio, L. L., Woodworth-Gutai, M., & Gross, K. W. (1984) *Nucleic Acids Res. 12*, 2791–2805.

Kageyama, R., Ohkubo, H., & Nakanishi, S. (1984) *Biochemistry 23*, 3603–3609.

Lawn, R. M., Fritsch, E. F., Parker, R. C., Blobe, G., & Maniatis, T. (1978) *Cell (Cambridge, Mass.) 15*, 1157–1174.

Lottspeich, F., Geiger, R., Henshen, A., & Kutzbach, C. (1979) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 1947–1950.

MacDonald, R. J., Stary, S. J., & Swift, G. H. (1982) *J. Biol. Chem. 257*, 9724–9732.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *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 (London) 303*, 300–307.

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.

McMaster, G. K., & Carmichael, G. G. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4835–4838.

Mills, I. H. (1979) *Handb. Exp. Pharmacol. 25 (Suppl.)*, 549–567.

Montell, C., Fisher, E. F., Caruthers, M. H., & Berk, A. J. (1983) *Nature (London) 305*, 600–605.

Morrison, D. A. (1979) *Methods Enzymol. 68*, 326–331.

Nawa, H., Kitamura, N., Hirose, T., Asai, M., Inayama, S., & Nakanishi, S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 90–94.

Ohkubo, H., Kageyama, R., Ujihara, M., Hirose, T., Inayama, S., & Nakanishi, S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2196–2200.

Okayama, H., & Berg, P. (1982) *Mol. Cell. Biol. 2*, 161–170.

Proudfoot, N. J. (1984) *Nature (London) 307*, 412–413.

Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 252*, 359–362.

Richards, R. I., Catanzaro, D. F., Mason, A. J., Morris, B. J., Baxter, J. D., & Shine, J. (1982) *J. Biol. Chem. 257*, 2758–2761.
Rigby, P. W. J., Diechmann, M., Rhodes, C., & Berg, P. (1977) *J. Mol. Biol. 113*, 237–251.
Schachter, M. (1980) *Pharm. Rev. 31*, 1–17.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.
Steiner, D. F., Quinn, P. S., Chan, S. J., Marsh, J., & Tager, H. S. (1980) *Ann. N.Y. Acad. Sci. 343*, 1–16.
Swift, G. H., Dagorn, J.-C., Ashley, P. L., Cummings, S. W., & MacDonald, R. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 7263–7267.
Weaver, R. F., & Weissmann, C. (1979) *Nucleic Acids Res. 7*, 1175–1193.
Wickens, M., & Stephenson, P. (1984) *Science (Washington, D.C.) 226*, 1045–1051.
Young, C. L., Barker, W. C., Tomaselli, C. M., & Dayhoff, M. O. (1978) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, Suppl. 3, pp 73–93, National Biomedical Research Foundation, Silver Springs, MD.

DNA Filter Retention Assay for Exonuclease Activities. Application to the Analysis of Processivity of Phage T5 Induced 5′-Exonuclease

Martine Joannes, Jean-Marie Saucier,* and Alain Jacquemin-Sablon

Laboratoire de Biochimie et Enzymologie (INSERM Unité 140 and CNRS Laboratoire Associé 147), Institut Gustave Roussy, 94800 Villejuif, France

Received March 27, 1985

ABSTRACT: The 5′-exonuclease of phage T5 has been purified nearly to homogeneity by using a simple and fast procedure. The kinetic properties of the purified enzyme have been studied by using a new sensitive assay based upon retention by nitrocellulose filters of DNA with short protruding single-stranded ends. The enzyme is specifically stimulated by KCl. Its K_m is 2.2×10^{-7} M at 30 °C, and its turnover number is 0.33 DNA molecule transformed per minute. The filter retention assay shows that the T5 exonuclease acts by a semiprocessive mechanism, removing from DNA ends about 30 nucleotides on the average per cycle. The degree of enzyme processivity increases with increasing magnesium concentrations.

A 5′-exonuclease, coded for by gene D15 of bacteriophage T5, is required for the transcription of late genes (Chinnadurai & McCorquodale, 1973) and the replication of phage DNA (Frenkel & Richardson, 1971b). It has been suggested that these functions are carried out by a single multienzyme complex which has been isolated from page T5 infected *Escherichia coli* (Ficht & Moyer, 1980). In this complex, phage DNA is found associated with bacterial RNA polymerase and several phage-induced proteins including the 5′-exonuclease and DNA polymerase. The 5′-exonuclease of phage T5 has been purified, and the properties of the purified enzyme have been extensively studied (Paul & Lehman, 1966; Frenckel & Richardson, 1971a). Oligonucleotides of varying lengths (four residues on the average) are released from the 5′ ends of either double-stranded or single-stranded DNA at approximately the same rate. The enzyme can initiate hydrolysis at single-strand breaks in duplex molecules but exhibits no activity on circular single-stranded DNA or covalently closed double-stranded circles and is weakly active on RNA.

New methods of assaying exonucleases may provide meaningful information relevant to the mechanism of action and functions of these enzymes. This prompted us to develop an assay based upon the property of nitrocellulose filters to retain in appropriate conditions double-stranded DNA molecules with short protruding single-stranded ends, whereas intact double-stranded molecules pass through the filter. This specific and sensitive assay has been used to purify nearly to homogeneity the T5-induced 5′-exonuclease and to study the kinetics of hydrolysis of double-stranded DNA by this enzyme. Our results show that the T5-induced exonuclease hydrolyzes DNA through a partially processive mechanism, the number of nucleotides removed between each enzyme–DNA association and dissociation depending on the magnesium concentration.

Experimental Procedures

DNAs. λ DNA was prepared by thermal induction of *E. coli* 159T⁻ (λcI857 S7). ^{14}C-Labeled λ DNA was obtained by addition, after prophage induction, of [2-^{14}C]thymine of specific activity 47.5 mCi/mmol prepared by C.E.A. (France). The [^{14}C]DNA used in this work had a specific activity of 1.31×10^7 cpm/μmol. λ DNA *Sma*I fragments were obtained by two successive 1-h digestions at 25 °C with 30 units/mL restriction endonuclease *Sma*I (Boehringer Mannheim) followed by phenol extraction and dialysis. Before digestion, the DNA was cyclized for 2 h at 50 °C in 2 M NaCl and 10 mM trisodium ethylenediaminetetraacetate (Na_3EDTA) and dialyzed against the incubation medium containing 6 mM $MgCl_2$, 5 mM KCl, and 15 mM tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), pH 8.5. Cleavage of DNA by *Sma*I was checked by electrophoresis of the fragments on 1% agarose gels. [^{3}H]DNA from phage PM2, kindly provided by Dr. J. Pierre, had a specific activity of 1.07×10^7 cpm/μmol. The covalently closed circular form (I) and open circular form (II) were separated by equilibrium density centrifugation in a CsCl gradient containing 100 μg/mL ethidium bromide. DNA concentrations are expressed in moles of nucleotides per liter.

^{32}P 5′ End Labeling of DNA Fragments. The λ DNA fragments were labeled by the polynucleotide kinase exchange reaction (Berkner & Folk, 1980) using T4 polynucleotide kinase from NEN and [γ-^{32}P]ATP (specific activity 2.9 Ci/μmol from NEN. The excess of [γ-^{32}P]ATP was removed

0006-2960/85/0424-8043$01.50/0