Whitham, S. E., Murphy, G., Angel, P., Rahmsdorf, H. J., Smith, B. J., Lyons, A., Harris, T. J. R., Reynolds, J. J., Herrlich, P., & Docherty, A. J. P. (1986) *Biochem. J. 240*, 913–916.
Wigler, M., Sweet, R., Sim, G. K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S., & Axel, R. (1979) *Cell 16*, 777–785.
Wilhelm, S. M., Collier, I. E., Kronberger, A., Eisen, A. Z., Marmer, B. L., Grant, G. A., Bauer, E. A., & Goldberg, G. I. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6725–6729.
Woessner, J. F. (1962) *Biochem. J. 83*, 304–314.
Woessner, J. F. (1980) in *Collagenase in Normal and Pathological Connective Tissues* (Woolley, D. E., & Evanson, J. M., Eds.) pp 223–239, Wiley, New York.
Woessner, J. F., & Taplin, C. J. (1988) *J. Biol. Chem. 263*, 16918–16925.

Characterization of Genes Encoding Rat Tonin and a Kallikrein-like Serine Protease[†,‡]

Shaw-Yung Shai,[§] Cheryl Woodley-Miller,[§] Julie Chao,[∥] and Lee Chao*[,§]
Departments of Biochemistry and Pharmacology, Medical University of South Carolina, Charleston, South Carolina 29425
Received August 25, 1988; Revised Manuscript Received March 21, 1989

ABSTRACT: Tissue kallikreins are a group of serine proteases which may function as peptide hormone processing enzymes. Two rat kallikrein genomic clones (RSKG-5 and RSKG-50) were sequenced and characterized. The rat tonin gene and a kallikrein-like gene were found in clones RSKG-5 and RSKG-50, respectively. The tonin gene is 4146 base pairs in length, with both the variant CCAAA and TTTAAA boxes in the 5′-end region and an AATAAA polyadenylation signal at the 3′ end of the gene. It has five exons which are separated by four introns. Sequence analysis of 3.7-kb 5′ upstream and 7.5-kb 3′ downstream of the tonin gene failed to reveal a second kallikrein gene. Sequence comparisons of the RSKG-5 exons with tonin cDNA revealed that only one base in the 3′-noncoding region was different from that in the previously reported rat tonin cDNA. Characteristic TC- and TG-repeated sequences were also found in the first and second introns of the tonin gene. The tonin gene encodes a preprotonin of 259 amino acids (aa). The active enzyme consists of 235 aa and is preceded by a deduced signal peptide of 17 aa and a profragment of 7 aa. Northern blot analysis indicates that RSKG-5 is expressed in a sex-dependent manner in rat submandibular gland, with a higher level expressed in males. The RSKG-50 gene was truncated at an *Eco*RI site in the second intron, excluding its 5′ end. Compared to the coding sequence of pancreatic kallikrein, 12 nucleotides have been deleted in exon 3 of the RSKG-50 gene. The nucleotide sequences of the third, fourth, and fifth exons of the RSKG-50 gene encode a polypeptide of 188 aa residues. The translated peptide is 80% homologous to rat pancreatic kallikrein and 75% homologous to rat tonin in the corresponding regions. Key residues in the RSKG-50 gene product indicate a serine protease with kallikrein-like cleavage specificity at basic amino acids.

Tissue kallikrein belongs to a multigene family coding for a subgroup of serine proteases which are involved in the processing of bioactive peptides (Schachter, 1980; Mason et al., 1983). These enzymes including pancreatic kallikrein, tonin, arginine esterase A, γ subunit of nerve growth factor, and epidermal growth factor binding protein (Bothwell et al., 1979; Chao, 1983; Chao et al., 1984; Swift et al., 1982; Mason et al., 1983; Ashley & MacDonald, 1985a,b; Gerald et al., 1986a) show a high degree of amino acid and nucleotide sequence homology. The structural and functional similarities of the kallikreins have hindered the identification and definition of their physiological roles. Therefore, molecular biology approaches have been implemented to analyze kallikrein gene structure, organization, function, and expression in mammalian systems (Mason et al., 1983; Gerald et al., 1986a,b).

Tonin (EC 3.4.99.–), a member of the rat tissue kallikrein family, produces angiotensin II from angiotensinogen, angiotensin I, or the tetradecapeptide renin substrate (Boucher et al., 1972). Angiotensin II is the most potent vasoconstrictive peptide known. It maintains arterial pressure in peripheral arterioles, modulates renin release in the kidney, and stimulates secretion of aldosterone in the adrenal cortex. Tonin has been shown to cleave proopiomelanocortin to produce adrenocorticotropin hormone (ACTH)[1] and to yield opiate-like peptides from β-lipotropic hormone (β-LPH) (Seidah et al., 1979a,b) and to degrade substance P (Chretien et al., 1980). Moreover, tonin has been reported to have kininogenase activity and can produce bradykinin from both high and low molecular weight kininogens (Ikada & Arakawa, 1984). In addition, tonin was shown to have the capability to activate renin from prorenin (Gutkowska et al., 1982). In these proteolytic activities, tonin prefers to cleave the substrate selectively at Phe– or Arg–peptide bonds (Seidah et al., 1979a,b; Chretien et al., 1980), giving tonin both trypsin- and chy-

[†] This work was supported by NIH Grant HL33552.
[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02860.
* Address correspondence to this author at the Department of Biochemistry, Medical University of South Carolina, 171 Ashley Ave., Charleston, SC 29425.
[§] Department of Biochemistry.
[∥] Department of Pharmacology.

[1] Abbreviations: ACTH, adrenocorticotropin hormone; β-LPH, β-lipotropic hormone; aa, amino acid(s).

0006-2960/89/0428-5334$01.50/0

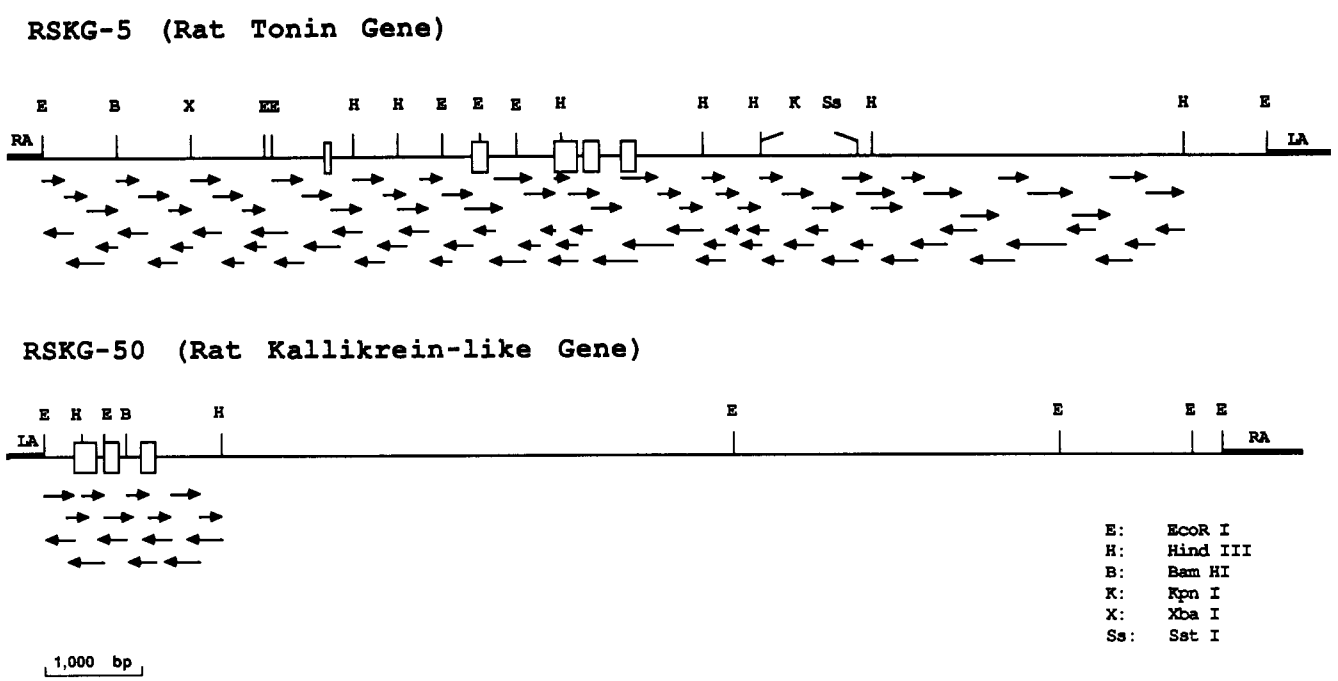


FIGURE 1: Restriction maps and sequencing strategies of RSKG-5 and RSKG-50. Exons are indicated by open boxes. Introns and the 5′- and 3′-flanking regions are indicated by thin lines. The letters represent restriction enzyme sites utilized for gene mapping and subcloning. The arrow indicates the direction and the length of sequences obtained in each sequencing reaction. RA, right arm; LA, left arm.

motrypsin-like activities. These findings suggest that tonin may play a general role in the cleavage of prohormones.

In order to understand the structure and regulation of kallikrein-like or tonin-like genes, we have screened a group of 34 rat kallikrein genomic clones (RSKG) using a rat tissue kallikrein cDNA probe (Gerald et al., 1986a). In this paper, we present the nucleotide sequences and the deduced amino acid sequences of two rat tissue kallikrein genes found in clones RSKG-5 and RSKG-50. The gene contained in RSKG-5 is the complete rat tonin gene. The gene in RSKG-50 is truncated at an *Eco*RI site in the second intron, excluding the 5′ end of the gene. The deduced amino acid sequences showed that the RSKG-50 gene encodes a serine protease with kallikrein-like substrate specificity.

Experimental Procedures

Nucleotide Sequence Analysis. DNA from the rat kallikrein genomic clones was prepared as described by Maniatis et al. (1982). The restriction map of each clone was determined, and the restriction fragments were then subcloned into M13 mp18 and mp19 (Messing, 1983). Single-stranded DNA from the M13 recombinant phage was prepared according to Schreier and Cortese (1979). The nucleotide sequence of the inserts was determined by Sanger's chain termination method (Sanger et al., 1977). All of the oligodeoxynucleotide primers used in the DNA sequencing reactions except the M13 universal primer were synthesized by a DNA synthesizer (ABI Model 380B).

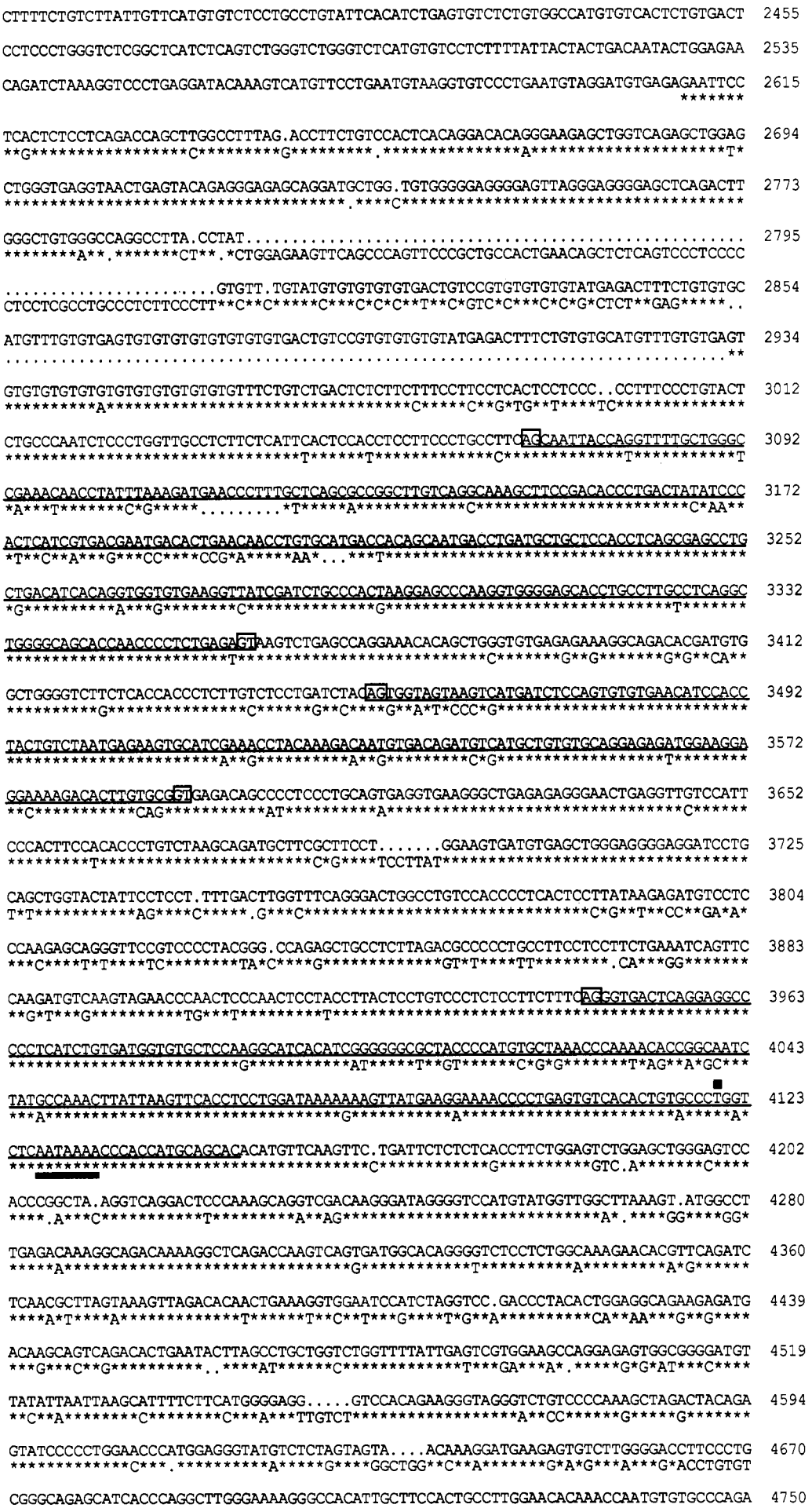
RNA Preparation and Northern Blot Hybridization. Male and female Sprague-Dawley rats weighing 250 g were anesthetized and perfused with phosphate-buffered saline. Submandibular glands were removed and homogenized in a solution of 4 M guanidine isothiocyanate, 0.5% sodium laurylsarcosine, 25 mM sodium citrate, and 0.1 M β-mercaptoethanol. Total RNAs were then extracted by centrifuging the tissue homogenates through a 5.7 M CsCl gradient (Davis et al., 1986). Fifty micrograms of total RNA was denatured in formaldehyde, separated by electrophoresis through a 1.5% agarose–formaldehyde gel, and then transferred to nitrocellulose by capillary blotting. The filter was air-dried and baked under vacuum at 80 °C for 2 h. Prehybridization was performed for 4 h at 55 °C in a solution of 5× SSC (1× SSC: 0.15 M NaCl/0.015 M sodium citrate), 50 mM sodium phosphate, pH 6.8, 1 mM sodium pyrophosphate, 100 μg/mL sonicated herring sperm DNA, 0.02% bovine serum albumin, 0.02% Ficoll, and 0.02% poly(vinylpyrrolidone). Hybridization to the ^{32}P-labeled oligodeoxynucleotide probe (5×10^5 cpm/mL of hybridization solution) was performed at 55 °C overnight, and the filter was washed 4 times in 6× SSC and 0.5% sarcosyl at room temperature for 15 min each and in 2× SSC at 30 °C twice. The filter was then air-dried and exposed to Kodak X-AR film at −80 °C with intensifying screens.

Synthesis of the Oligodeoxynucleotide Probes. Primers used in the DNA sequencing reactions, oligodeoxynucleotides, 16–20 bases long, corresponding to the 3′ end of the previous reactions were synthesized by using the DNA synthesizer. Before use, the crude primers were dissolved in sterile water, and the unincorporated nucleotides were removed by a Sephadex G-25 spun column. The tonin-specific oligodeoxynucleotide probe used for Northern blot analysis was described by Ashley and MacDonald (1985b) as S2. The oligodeoxynucleotide was end labeled at the 5′ termini using T_4 polynucleotide kinase. Briefly, 100 pmol of oligonucleotide, 10 units of T_4 polynucleotide kinase (New England Biolab), 100 μCi of [γ-^{32}P]ATP (New England Nuclear, specific activity 6000 Ci/mmol), and 6 μL of 5× forward buffer (350 mM Tris-HCl, pH 7.6, 50 mM $MgCl_2$, and 25 mM dithiothreitol) were added to a total of 30 μL; this mixture was incubated at 37 °C for 1 h. The reaction was terminated by the addition of EDTA to a final concentration of 12.5 mM. Free [γ-^{32}P]ATP was removed by a Sephadex G-25 spun column.

Results

Primary Structures of RSKG-5 and RSKG-50 Genes. Rat kallikrein genomic clones were isolated from a rat λ Charon 4A genomic library using a kallikrein cDNA as probe (Gerald et al., 1986a,b). These clones have been systematically analyzed (Gerald et al., 1986b; Chen et al., 1988), and RSKG-5 and RSKG-50 represent further analysis of this group of clones. RSKG-5 and RSKG-50 were subcloned into M13

```
ACCTCAAAATATTTACTTACATATTTATTCCCTTCTGTCAGTCAAGGATGACCTTGAATTTCTGATCCTCCTGCCTCCAC -1224
ATCACACATGTTAGTTCTACAAGCATGCACCACCTCACCTGATTATACACAGCACTGTGGCTCAAACCCAAGGCTTCCTG -1144
CATTCTAGGCAATGAATCCACCACTTGGCCTGCATCCCAATCCCACCCATCCCCACACATACTTCCTTAAGAAGACAAGG -1064
TCTCATGTAGCCTGACCTAGAACTTGTTCTGTAGCTGAGCTGACCTTGAACTTCTGGTCCTCCTGCCTCCACTTCCCAAG  -984
TGCTGAGGTAACAGGTGTGACCACCATGCAGGCCCTGTCGGAAAATCTTGCCCAGGGAAGTCAATTAGTCCAGATCTTTG  -904
GGCTTTTGAAGGATGGGACCTTGAGTCCCCTTACATAGGAAACAAGAGAAAGTCATTTCCAGTCCAACACTGTTCAGGGC  -824
GTGTGGCACTCCACAGGAATAACCTGAATTCTCATATACTAGAACCAAAGAGTGATCAGGACATGATAATGTGCACCGTA  -744
ACTTCTATTTATCTTACTTTGTTTTATTGGTTTGTTTTGAGACAGCTTTATATATAACCCAGGCTGTCCTGGAATTCACT  -666
ATGTAGACCAGGTTGGTCTACATCACAGATTATTAGCCCCTAAAATCACAGAGATCTGCCTGCCTCTGGCTTCAAACTTC  -586
CTGGATTAAAGGTGTACACCATTACCCAGCTCCCAAACGTCTAATAGGAAAAATAAAATCAGTGTTCCATAGTCCAAGGA  -506
AGGCAAGATTTTTGGCATTCTCAAAAAAACAAAACTTTTATCTGTGGGGAGCTTGCAGACTTCTTGGCTGTCAATGGTGC  -426
AAAGAATTTGTTCTAATGGCTAATTTCTGAGGATCATCTTGGTCTTGAGGTACAGCATCTTCCTTAGAGCTGTGTTTTGG  -346
GGGTCACTTGTTCCTCCACGCACCTATCATACCCTTTCAATATCACTCCTGACCTTGCCCTCTGACCTTGGTCCACAGAA  -266
CATGAACAGGTCCAGGAGCAGTAGTCACTGCCCCACCCAATCTCACCCCTGGGAAGGTATCCAGGGAAACCTAGGATTTC  -186
CTGGACAGCAGGTGCAATAGGGCCCTCCCCTCCCCCAGCCATGAGGGTGAGAGCAGGGAATCCACCTGAGTCAGCAAGGT  -106
TATGGCCTTGGCCAGCATCTTCGTGCCAAAAGGGCAGGGGCAGGGCGGGGCTGGGGGGAGAATGAGGAGTTTTTAAAGGC   -26
TCCCCAAGGACACTATACAGGTCCAAGATCACCACCTGCTTCTCCTGGACACCTGATACCATGTGGCTCCAGATCCTGTC   +55
CCTAGTCCTGTCCGTGGGACGAATTGGTGAGACGGGGGGAAGCAGGCATGGTCAGAGCCCTGACTCTCATGCTGCTTTTG   135
GCCCTGGACATCCCTTTCCCCTCTCCCCTGGCCCCACCAACATTCTAAAGCACCACCTGCTCCTCCCTCTCTTCTCTGCA   215
CCCTGCCAAGTCCCTCACAGTCCTCTCTCCCATGCCAGGCACCTGTTCCCTCCCATTTTCCAGAGCTTTCTCCTTCCTTA   295
GCCACCTCATCATCTCTGGGCTTCTCCAGTCTCTCCAAGGAATTTTTCCAAACCCCGATGAACCTGAGGACCTCACTTGT   375
CTCTCTATTTTGCCGTTTTACCTAGAGCCTAAGATTCCAACCCCTGTGTCAGAAGCTTCCCCAAATCCCACACTAGAGCC   455
CTTGGTGTCCAGCAGCTGTCAGCCTTCAGCTGTGCTCTCTACATCCTGAAAACTAGTCCGCCATCCTCTCAAGGTGCTCA   535
GTTTCTATTGACAGTTTCACTTCGCAAATTGGCTTCTGACATCCGATGAGTCCCCTAACCCTTTCTGGTGTTGGGGATCA   615
GCACATGTCCCTTCTCAATGGCTTCTTGGTTAAGTCTGGTTTTGGGTGGAGGAGGCAGCTCTTACAGGCCAGACAGACAT   695
CAAGGATCAATTGGATAGGGCCTCACTGAATGGGGATTTTCCATTTAGACTAACTTGGCTGTCTGGAAGGTCTTAGGGGA   775
CTCCTACTGCTTCCACCTCTGGTGTTACAGGGGCACACTGCCACACCTGACCTCTTACCTGGGTTCTGGGGAGTTCAGCT   855
CGCGTCCTAATGCTTGCACAGCCAATACTGTAGCAACCAATCCATCTGCCTAGCTCTTGGGAATAAGAGCTGAGCTAGTG   935
TCTCATGGTGTAACTCAGGCTTTTCTCTGGTCTGTGGCAACTCTCCTGCCTCAGCCTTCTGTGTGCTGGGATTACAGAGA  1015
AATAATTGCCAAGCTTCCCTGTTTCCTAACGGAGTGCTTTGGCCCACTCTTCTCTCCCGAGCTGCCCACTCCCAGCCTCA  1095
GGCCCCTTGTTGTTTCTAGCCTCCAGCAGCTCCTGTCCTATCTCGCTGGAACCCCACTGGAACCCTGCACAGCTCTAGTT  1175
CCCATGCTGAGCAGGACATCCCATGACCCCAAAACCTATCCTTAGCACACCACACCAACTGGGTCCTCCTCTCTCCCTCA  1255
CAGTGGTATGACTTTTTAAACTCCTTTTCACCCCTGTGTTGTCAGAGGCTGCCTCCAAGAGAACCCTGAACCTCACCCAT  1335
ACCCTGGCCATTTGCCCCCGCCTACTCCGTGGGAAGAAGGACTGTTCCCTCACATCCCCCAGAGTCTCCTTCTCCCACTG  1415
CTTGTCCAGGCTCAAACTCTGTTCTTAGGGAGCCCCACCCTCATAGGTATCATCTCCACCACAGACCCTAACAGGAGTCC  1495
ACGCCTAAGCCTTAACCCTGTGCCTTTGATCTGGAAACCGCTCATTCTCTCCTATCCCAGCTGGGCACTCCAGGAAACCC  1575
AAAATCATCCCCAAGTTCCCATTTCTGGATTCATTTATTCGCTCAAATCTAGTTATTAGGTGGCCTCACCCTTTCCAGAA  1655
TTCTCAAAACCCAACCAAGGCCCTGATGTCTTCTCAAGCCCTAGAGATTTTAAAGGCCAAAGTGTCTTGGCCTCTGATTT  1735
CACTCTCATTGGACTGAGAATGGACTGGAATCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT  1815
CTCTCTAGCTTGCTGGCTTGCTCTCTCTGCCTGCCATCCTGTAACCTGATTTCGGCTTCAGGAGTTATCTCCGTCCACAC  1895
AGCATTGTTCTCCCTACAGAGGATCAGCCATACTTGAGTCCCTTCCCCACACCCTCTGGCCATGTCCCAGCCCTGCCTAA  1975
CTGTGTCCCCAACCCTACAACCATACCCCCTACCTGCCAACCATAAACCAACCCCCTGCACCTGTTACAGATGCTGCACC  2055
TCCTGGCCAGTCTCGGATTGTTGGAGGATATAAGTGTGAGAAGAATTCCCAACCCTGGCAAGTGGCTGTCATCAATGAAT  2135
ACCTCTGCGGGGGTGTCCTGATAGACCCCAGCTGGGTGATCACGGCTGCCCACTGCTATAGCAAGTGAGTAAGGGAGGGG  2215
ACAGGAATGACTCCAGGACGGGCTGTGGTACTGAATGGCAGGAGGGTGGGATGGGCTCAAAGTCTAAAGAGGTCCAGCAG  2295
GCCTGGACCCATGGTCTGTCCATGACCTCCTGGGTCTCTTTGCTCCTCTCTCCAGGCTTGTGTTCATGTTCTGGTTCTTT  2375
```



```
GGGATATAACTGCAGCTGAGAGTATTGGAGGACTTTCAGAGGAGAGTCCCTGCCAAGACTCAGGCAGTAAGGAGTAGGCC  4830
AA**C***CACA***G********C*******************G****.......**************A**C*G****

CTAGAGTGGGGAGTGAGGGTTGCCACCTGTGTCCCCAGGGTGCTACCTGCTCTCAGGGGAAGTATGGGACCTCAAGGGAG  4910
T**A***********G**C*************************************************************G*****

GTCTGACCATAACTGAGCCACAGGGACCGAGGTAGAGGCATCAGGCTGCGGGGACTCATCAGTGTCCTGAGAAGTTGTGG  4990
******TGGA**************************** ************************************C*T

TGTCCGTCTGTTGGGGAGGC.CCTGATAGGTCACAAAGCTGCTTGCAGTAGACAGCTTCTGGGCGGGCTCC.AAGCTTCT  5068
C****C**************T**A*G***T*********************************TC******A******

GTCCGATACGGACGGCTTCTCCAAGTGTAATGTGTGGTGTTGCAAGGCCCAACAGAAGTAGCTGAGCCTCTGCTCCATGA  5148

AATGCGTTCTGTAGCCCAGTGGGAACTCACAGCTGCCCAGTGCATCAGCAGGTAGGATGGGCTCAGGTGTGGCTGTGGCC  5228

GGAGCACTTGCCTGACAAGCTAGGGTCCCTGGGATCCACCTCCCAAACACATTAGTAGGAGGACAATTGGGGTGAGGGGA  5308

GCTCCTCTTCTTCAAGGAAGACTATGCCACCCTCTTTCTTGGGTACCTCATGGGAGGTAGATGTCTGGGGAATGAAGGTT  5388

AGTACAGGGTCAGAAGGGCAGAGTGGTGGGTGGGCCATGAGCACCTCAGTGTCACCTTATTTCTCCGTTTTCCTGGCCCG  5468

CAGTCTTGTGTGATCCTACTCTTTGTGTGACTATGAGAAACTGTGTTTTTGTCTCTCCTCTTTCCCTGAATCCATGCCTC  5548

TGCATCTGTCCTATCTCTGTCCTCCTCTGGGTGCCTCTATGACCATCTGTCACCATTGTTCCCCACTTCCAGATCTCTAC  5628

TCAACTCTCTCTGGGTCTGGGTTCCACTCTCCATCCCCATCACTACTGACCACAGTGAAGAACAGGATTCACTGCACCCT  5708

AAGAGGCAGGCATTGGGGGACCATATGCCAACGTGAGGGAAGTCCTCACATTCTCCAGACCAACCTGGCACTGTAGGATC  5788

CTGTCCAGGCTGGGGGACACCAAGGGAGAGGTTGGACAGAGCTGTGTCAATACAGCTGGATGAGCACAGAGACAGGGCAG  5868

GGTACTGGGAGGAAGTGGGTTCAAAGCTTGGGCTGTGACCATATCACAGGTACCAGGCAAGCTCAGGCTCCCAGTGTACC  5948

CAAGTGGCTCTCAGGCCTGCCCCTTCCCTGTTCCATTTCCCCTTCTCTCTTTACTTCCTGCCTCATTGCCCACTTCCTCT  6028

GTTTGTCTCTGCCTCTTTTCCGTCCCCAGTCCCTTCTTCTCTCTCATGTACTTACTGCTCCCATGTGTCTCCACCTGCTT  6108

TCCCATGGTGCTTCACACTGCCCTTCCCCCTGCTACCTTTCTAGTTCTCTTGATTTAATCTTGTGTGTGTGGTGCATGTG  6188

TGATACTGTTGATGAAGTGTGTGTGAGCGTGTGTGTGTGTGTATGGAGGCCAGGAGCAGACATCATCTGTCTTCCTCTCT  6268

TGCTCTCCAGCTTGTGTTTTCTAAGTTGATTTATCCCCTTTATTTGTTTTGTATGTGTGAGTGTGTGAGTCTGTGTATGT  6348

GCAAGTATGTATGACACTGCATGAGCGTGGAGATCACGAGACAGCCTGCAGGAGTCAGTTCCCTCAATCCACCATGTAGG  6428

TCTTGGGATTGAAATCCAGTCACGATGCTTGGCTGCAAGAGCCCCTACCTACTGAGCCGCTTCACTGCCTGCATCATTTT  6508

TCTTTAATCACCTTAATAAAAATCCTGTATATGAGTGTATTGCTTCTCTGTGTTTCTATGCACTTTGAACTAGAGTTACA  6588

GACACTTGTGAGCCAGCATGTTTGTGCTGGGAATGTGTTGTTGTAAAAATATAAAAATAAAAATGTATTGTTGGTCTTTT  6668

ACCCCGCTGGGTCTAACCTCGGTGTCCAGGATATCTGCTAGATATCTTGGTGGAAACACATCCCAGCCGCACACTTTTTT  6748

ACCCTCAAACCCTCACATAAAAGAACACACAACACAATAGTCTCAGATCCAATTGGTAAGATATAATTGTCCACTTAAAC  6828

ATACAATGTCCGGTACTATCTATCCCTTGAGAACATTAATAACAACCTGTAAATACACAGAGCAGAATCTTAACATCACC  6908

TGCCATCTTGTCCTGTCACGGCTTCTCTCCCTCTCTCCCTCCTGTCTTTTCCTCTCTCCCTTAGTCTCCTCTTCTTCTTT  6988

CAAACTTCTCTCCCACCCATCCTTCCTTCTCCTCCAATGACAGACCTCCTTCTATCCTGTACCTGCCCCTCACCAGTACT  7068

TTACAAATTCAATGGAGAGGGGGTTCTGGTGAAGTCACCTGAGTTCTGAGTCCTTGACTAGGCAACTGTCTTGGGGCAGT  7148

GGAATTAGCATCAAAATACAGATAACTTCAGGGCAAACCACAACAGGAATGACCCAGGTCCTCTGAAAAAGCAGTTAGTG  7228
```

FIGURE 2: Primary structure of the tonin (RSKG-5) gene maximally aligned with a partial sequence of the RSKG-50 gene. The upper sequence is the tonin gene, and the lower sequence is RSKG-50. The cap site of the tonin gene is assigned as +1. An asterisk represents a nucleotide sequence identity between the two genes. Dots represent deleted nucleotides in either of the genes. The underlined sequences indicate the exon regions of the genes. The GTs and AGs of the splicing donor and acceptor are boxed. The variant TATA and CAAT boxes are double-underlined. The AATAAA polyadenylation signal is underlined in bold. The T base with a closed box over the tonin gene sequence is the nucleotide that differs from the S2 cDNA.

mp18 or mp19, and their nucleotide sequences were determined. Figure 1 shows the restriction maps and sequencing strategies used for analysis of RSKG-5 and RSKG-50. Approximately 15 kilobases (kb) of the clone, RSKG-5, were sequenced. Analysis of the nucleotide sequence revealed the complete tonin gene as 4.2 kb of the clone. RSKG-50 contains an incomplete kallikrein-like gene which is truncated at an *Eco*RI site in the second intron. An additional 1000 base pairs (bp) 3′ to the truncated gene were also sequenced.

The tonin (RSKG-5) gene is 4146 bp in length, including five exons and four introns (Figure 2). The 5 exons are 81, 154, 287, 137, and 199 nucleotides in length, which encode 15, 52, 95, 48, and 51 amino acid residues, respectively. The untranslated regions in the first and the fifth exons are 35 and 46 bases, respectively. The 4 intervening sequences of the tonin gene are 1964, 871, 96, and 357 nucleotides long, respectively. The consensus sequence for splicing proposed by Mount (1982) and the GT-AG rule of Breathnach et al. (1978) is obeyed at the 5′ and 3′ termini of each intervening sequence.

An adenosine base is tentatively assigned as the mRNA transcription start site (nucleotide 1 in Figure 2) because it is flanked appropriately by pyrimidine residues as that found in other genes (Corden et al., 1980). The consensus sequence, TTTAAA, found 28–34 bp upstream from the transcription

start site of the tonin gene transcript, is most likely a variant promoter necessary for correct initiation of RNA synthesis by RNA polymerase II. The same consensus sequence is also seen in the rat renal kallikrein gene (RSKG-7) (Chen et al., 1988), the mouse glandular kallikrein-related gene (mGK-1) (Mason et al., 1983), the subunits of mouse nerve growth factor genes (Evans & Richards, 1985), and the epidermal growth factor binding protein genes (Drinkwater et al., 1987). A region matching the transcriptional factor Sp1 target element, GGGCGG (Dynan & Tijan, 1985), is present at −56 to −61. The sequence CCAAA, centered at −77, is probably a variant of the CAAT box (Benoist et al., 1980). The sequence AATAAA, required for polyadenylation and/or termination of mRNA transcription by RNA polymerase II (Proudfoot & Brownlee, 1976), is located in the 3′-noncoding region of the fifth exon, 22 bases 3′ to the termination codon of the gene.

The four bases in the coding strand of the tonin gene are present in unequimolar amounts in both the gene and its upstream sequence. In the gene, the four base compositions are as follows: A, 20.4%; C, 30.0%; G, 22.5%; T, 27.1%. For the 3.7-kb upstream sequence of the gene, the four bases are as follows: A, 26.3%; C, 22.5%; G, 23.6%; T, 27.6%. However, the base composition in the sequence spanning 7.5 kb of the 3′ downstream of the gene is nearly equimolar. Recent studies have found that the conformation of a double-helix DNA is determined by its nucleotide composition. For example, sequences with tandem poly[d(TG)·d(AC)] can adopt the "Z"-form DNA structure (Hamada et al., 1984a), and sequences with poly[d(GA)·d(TC)] will adopt an "H"-form structure (Mirkin et al., 1987). Also, Sen and Gilbert (1988) discovered that DNA containing short guanine-rich motifs can self-associate at physiological salt concentrations to make four-stranded structures in which the strands run in parallel fashion. Thus, the changes in A and C content in the body of the tonin gene and the upstream region suggest that different structures may be adopted between these two regions. Whether they are related to the gene expression and regulation is not known at present. The richness of C and T bases in the tonin gene is also a very interesting phenomenon. It perhaps is a general characteristic of the rat tissue kallikrein-like genes, but its function remains to be explored.

A second kallikrein-like gene was found in clone RSKG-50. The results of the restriction mapping and the Southern blot hybridization analysis revealed that the gene in RSKG-50 is truncated at an *Eco*RI site in the second intron excluding the first and the second exons. However, this truncated gene was sequenced along with approximately 900 bp 3′ to the gene. The sequencing data show the third, fourth, and fifth exons of the gene are 275, 137, and 199 base pairs, respectively. The third and fourth introns are 96 and 364 base pairs, respectively. Sequence comparison shows that the nucleotide sequences in the RSKG-50 gene and its 3′-flanking region are very similar to that in the corresponding regions of the tonin gene, except a low-homology region of 200 bp in the second intron (Figure 2). The overall sequence identity between these two genes is about 80%. Not only the sequence but also the sizes and the locations of the introns and exons are highly conserved between these two genes, implying that these two genes might have recently evolved from a common ancestral gene by duplication.

The coding regions of the tonin and RSKG-50 genes were compared with those of the rat tissue kallikrein-like cDNA clones (PS, S_1, S_2, S_3) (Ashley & MacDonald, 1985a) and those of two rat renal kallikrein genomic clones (RSKG-3 and RSKG-7) (Chen et al., 1988). The results of sequence comparisons reveal that the coding sequence of the tonin gene is identical with that of the tonin cDNA (S2) reported by Ashley and MacDonald (1985a), indicating that the RSKG-5 gene is the rat tonin gene. The only difference found in the nucleotide sequences of the exons between these two clones is that a C in S2 is replaced by a T in the 3′-untranslated region of the tonin gene (Figure 2).

The coding sequence of the RSKG-50 gene shares 86% identity with that of the rat pancreatic kallikrein cDNA clone, PS, and even a higher degree of similarity, up to 98%, with that of the cDNA clone, S1. Compared to the pancreatic kallikrein cDNA clone, 2 deletions containing a total of 12 nucleotides are found in the third exon of the RSKG-50 gene. The deletions span nine and three bases, respectively. The three-base deletion is also found in the S1 cDNA at the corresponding site.

Analysis of the Nucleotide Sequences in the Flanking Regions of the Tonin Gene. The nucleotide sequences in the 5′-upstream 3.7 kb and the 3′-downstream 7.5 kb of the rat tonin gene were also determined. Sequence analysis revealed that in both of the flanking regions, there are several open-reading frames which could encode polypeptides more than 60 amino acid residues in length. However, none shares a significant homology with the kallikreins. Thus, the results of sequencing 7.5 kb of the flanking region of the tonin gene suggest that the intergenic distances between the members of the rat kallikrein gene family exceed those of mouse kallikrein genes recently reported by Evans et al. (1987) as from 3.3 to 7 kb with only two exceptions.

TG Dinucleotide Repeat Rich Regions of RSKG-5 and RSKG-50. Sequence comparison among the four rat kallikrein-like genes, RSKG-5, RSKG-50, RSKG-3, and RSKG-7 (Chen et al., 1988), shows a high degree of homology in all of the exons and most of the introns, except a small region in the second intron showing a low homology. Figure 3 shows sequence analysis in this low-homology region containing TG dinucleotide repeats of different length. The nucleotide sequences from a conserved *Sst*I site in the second intron to the splicing acceptor site of the third exon of each gene were aligned and compared. As shown in Figure 3, the sequences at both 5′ and 3′ ends of this region are highly conserved among all four kallikrein genes. However, gaps of different lengths must be introduced to obtain maximal identities, thus indicating that various portions in this region of the genes have been deleted. In addition, a TC repeat spanning 54 bp was identified in the first intron of the tonin gene (Figure 2).

Tonin and the RSKG-50 Gene Encode Kallikrein-like Proteins. The preproenzyme of the tonin gene product consists of 259 amino acid residues, including a signal peptide of 17 residues and a zymogen peptide of 7 residues. Alignment of the nucleotide sequences of the genomic and the cDNA clones shows that the signal peptide of tonin is split between the first and the second exons and the seven residues of the zymogen peptide are encoded by the second exon. Each of the active-site residues, His-41, Asp-96, and Ser-189, characteristic of serine proteases is encoded separately in the second, the third, and the fourth exons, respectively, of the tonin gene.

The nucleotide sequences of the three exons present in RSKG-50 encode a polypeptide of 188 amino acid residues corresponding to the C-terminal portion of tissue kallikrein (Figure 4). The deduced amino acid sequence is 80% homologous to that of rat pancreatic kallikrein and 75% to that of tonin in the corresponding regions. The Asp-96 and Ser-189 residues, which are part of the catalytic triad of the serine protease, are found in the RSKG-50 gene product at the

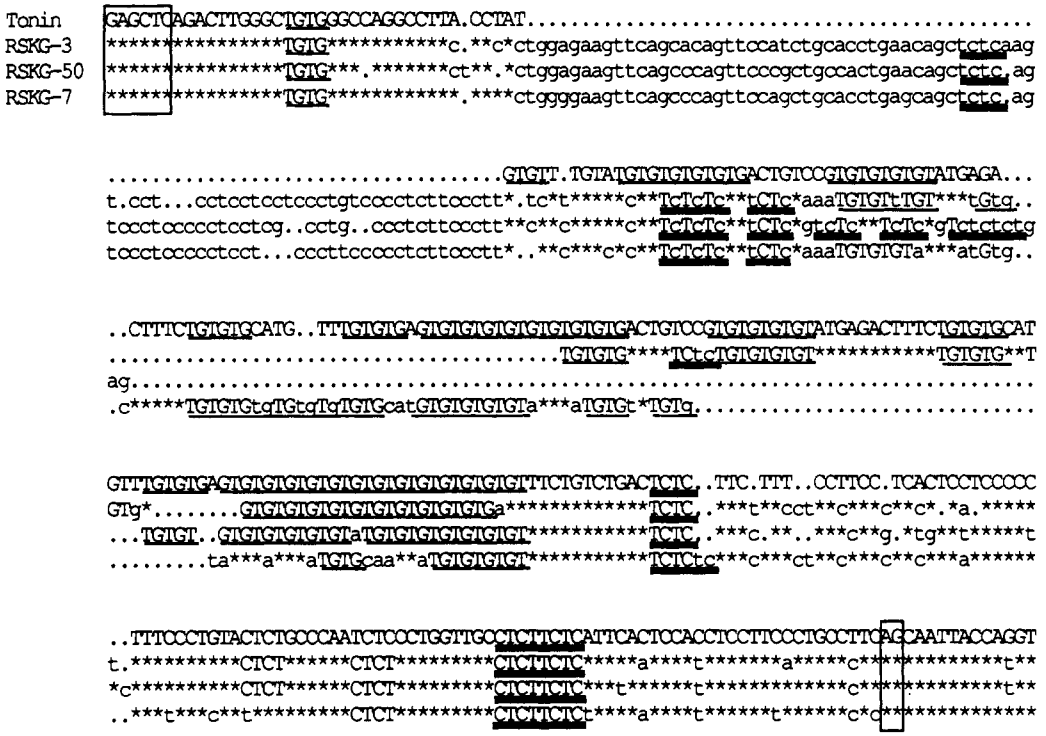


FIGURE 3: Sequence comparison in TG dinucleotide repeat rich regions of the rat kallikrein-like genes RSKG-5, RSKG-3, RSKG-50, and RSKG-7. The nucleotide sequences from the conserved *SstI* sites in the second introns to the splicing acceptor sites of the third exons of these four genes are aligned. Gaps have been introduced for maximal sequence identity and are represented by a dot. Sequences identical with the tonin gene are denoted with an asterisk, and the mismatched nucleotides are in lower case. The boxed GAGCTCs and the AGs are the conserved *SstI* sites and the splicing acceptor sites of the third exon, respectively. The TG dinucleotide repeats are underlined, and the TC dinucleotide repeats are underlined in bold.

corresponding sites. In addition, specific residues involved in the determination of substrate specificity of rat tissue kallikrein, such as Tyr-93, Asp-183, Trp-205–Gly-206, and Gly-218, are also present in the RSKG-50 gene product. This suggests that the enzyme encoded by RSKG-50 is a kallikrein-like serine protease with specific cleavage sites at basic amino acid residues.

Figure 4 shows the alignment of the amino acid sequences encoded by the tonin and the RSKG-50 genes with previously reported rat and mouse glandular kallikreins including the rat pancreatic kallikrein (PS), S1, and S3-encoded products (Ashley & MacDonald, 1985a), rat renal kallikreins RSKG-3 and RSKG-7 (Chen et al., 1988), the mouse mGK-1 gene product (Mason et al., 1983), epidermal growth factor binding protein (Lundgren et al., 1984), and the γ-subunit of nerve growth factor (γ-NGF) (Thomas et al., 1981). The amino acid sequences of the rat genes, RSKG-7, RSKG-50, and the cDNA S1, are closely related to that of the pancreatic kallikrein gene (PS), and those of the RSKG-3, tonin, and the cDNA S3 genes are highly homologous to that of the tonin (S2) cDNA. This suggests a close evolutionary relationship among the members of the tissue kallikrein family. Variations in the key residues responsible for lining the substrate binding pocket may play a crucial role in determining substrate specificities among these tissue kallikreins.

Sex-Dimorphic Expression of Tonin. The level of the tonin gene expression in rat submandibular gland is sex-linked, with expression in male approximately 2–3-fold higher than in female (Figure 5). The sexual dimorphism of the tonin gene expression in rat submandibular glands has been demonstrated previously by radioimmunoassay using a tonin-specific radioimmunoassay (Shih et al., 1986).

Discussion

We report here the molecular cloning and nucleotide sequence analysis of two tissue kallikrein-like genes (RSKG-5 and RSKG-50). Sequence analysis of the RSKG-5 gene reveals that it is the rat tonin gene. Further, a specific oligonucleotide probe shows sexually dimorphic expression of tonin in rat submandibular gland with levels higher in adult male rats. Tonin has been isolated and purified from the rat submandibular gland (Boucher et al., 1972; Demassieux et al., 1976), and its amino acid and cDNA sequences have been completely determined (Lazure et al., 1987; Ashley & MacDonald, 1985a). This is the first report of a complete nucleotide sequence for the rat tonin gene (RSKG-5). In addition, we have also sequenced and characterized a genomic clone encoding a kallikrein-like gene (RSKG-50).

In the clone RSKG-50, only part of the gene was found. However, the presence of residues forming the catalytic triad and those responsible for the kallikrein-like substrate binding pocket in the RSKG-50 gene product indicate that the predicted polypeptide is a kallikrein-like enzyme. The deletion of four amino acid residues (positions 59–61 and position 91) does not include those residues forming the active site nor the residues determining the substrate binding specificity of the enzyme. However, the shortening of four residues may cause a conformational change in the domain of the polypeptide chain and thus affect the activity of the enzyme. Three possible glycosylation sites are located in the deduced polypeptide chain encoded by the RSKG-50 gene. Two sites, Asn-Leu-Ser (positions 55–58) and Asn-His-Thr (positions 84–86), are the potential glycosylation sites; the Asn residue in the sequence Asn-Pro-Ser (positions 136–138) cannot be glycosylated due to the presence of a vicinal Pro residue (Bause & Legler, 1981).

A high degree of sequence homology is found at either the nucleotide or the amino acid level when the tonin gene and the RSKG-50 gene are compared with other reported rat tissue kallikrein-like genes. This suggests a close evolutionary re-

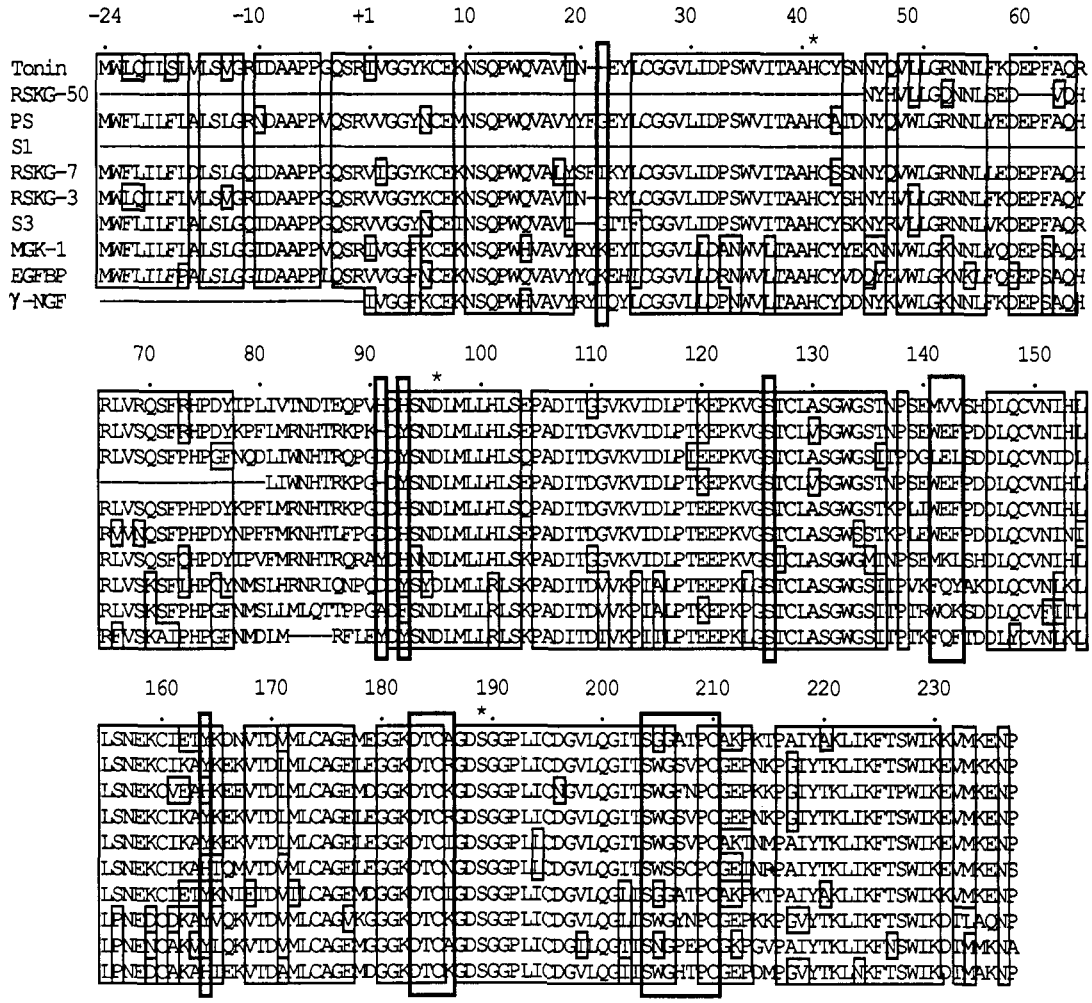


FIGURE 4: Comparison of translated amino acid sequences of rat and mouse kallikrein genes. The numbering scheme starts at the amino terminus of the predicted active enzyme form encoded by the tonin gene. Residues that are identical with the tonin gene product are boxed. Residues boxed in bold are involved in the formation of substrate binding pockets. The asterisks denote the positions of the histidine, aspartate, and serine of the catalytic triad. Dashes indicate where gaps have been introduced in order to maximize alignment. The amino acid sequences deduced from rat cDNAs (PS, S1, and S3) (Ashley & MacDonald, 1985a), rat renal kallikrein (RSKG-3, RSKG-7) (Chen et al., 1988), mouse mGK-1 gene product (Mason et al., 1983), the γ-subunit of mouse nerve growth factor (γ-NGF) (Thomas et al., 1981), and mouse epidermal growth factor binding protein (EGF-BP) (Lundgren et al., 1984) are included.

lationship among the members of the tissue kallikrein family. On the basis of sequence identity in the coding regions and sequence homology in the key residues determining the substrate binding pockets, the kallikrein gene family can be classified into two subgroups. The kallikrein subgroup includes PS (pancreatic kallikrein), S1, RSKG-7, and RSKG-50, while the tonin subgroup includes S2 (tonin cDNA), S3, RSKG-3, and the tonin gene (RSKG-5). In the kallikrein subgroup, the nucleotide sequence in the partial coding region of RSKG-50 shares 98% and 85% identity with S1 and PS cDNA, respectively. In the tonin subgroup, the coding sequence of the tonin gene is almost identical with that of the S2 cDNA with the exception of one base mismatch at the 3′-untranslated region. It is interesting to note that the sequence identity among S2 cDNA, S3 cDNA, and RSKG-3 gene is as high as 85%. Furthermore, a six-nucleotide deletion found in S2 cDNA, when it is compared against the sequence of PS cDNA, is also seen in all of the tonin subgroup genes at exactly the same site. This close sequence relationship among the tonin subgroup indicates that they may have recently evolved from a common ancestral gene.

Previous studies suggested that the number of tissue kallikrein-like genes varies widely among different species. Evans et al. (1987) reported recently that 24 tissue kallikrein-like genes are present in the mouse genome. Studies of the rat kallikrein-like genes suggest that rat may have 10 (Ashley & MacDonald, 1985a) −17 genes (Gerald et al., 1986a). The results of genomic Southern blot hybridization analysis with human DNA suggest that there are probably only three to four tissue kallikrein-like genes in the human genome (Schedlich et al., 1987). A rat tissue kallikrein genomic library containing 34 clones was isolated previously (Gerald et al., 1986a). The 34 clones were classified into 6 groups according to restriction mapping and Southern blot analysis using a partial kallikrein cDNA clone, RSK1105, as a probe. At present, at least eight different rat tissue kallikrein-like gene family members have been identified and studied in detail (Gerald et al., 1986a; Ashley & MacDonald, 1985a; Chen et al., 1988; this paper). The similar restriction maps of the 34 genomic clones and the high degree of sequence identity among the members in each group indicate their close evolutionary relationships.

Two particular nucleotide sequence elements in the intron regions are noteworthy. A region composed of TC dinucleotide repeats is present in the first intron of the RSKG-5 and RSKG-7 genes. The tertiary structure of the DNA in this region may be very unusual due to the fact that one strand is totally composed of pyrimidines and the other one totally of purines (Leslie et al., 1980). Theoretically, the poly[d-(TC)]·poly[d(GA)] may form an unusual double-helical structure for several turns. However, the nature of this

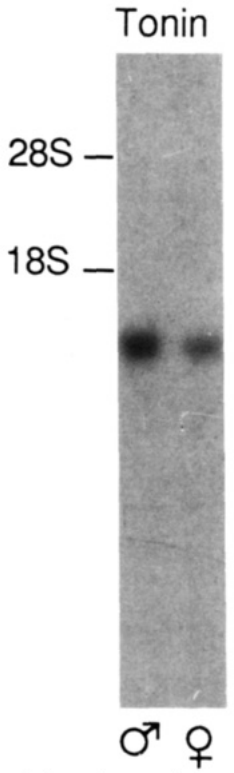


FIGURE 5: Expression of the tonin gene in rat submandibular gland. Twenty five micrograms of total submandibular RNA from either male or female adult rats was electrophoretically separated in a 1.5% agarose–formaldehyde gel, transferred onto nitrocellulose, and followed by hybridization with a tonin gene-specific oligonucleotide probe.

structure and its role, if any, in the function of these genes are not known. A region containing TG dinucleotide repeats is found in the second introns of the tonin and RSKG-50 genes, and also in two renal kallikrein genes, RSKG-7 and RSKG-3. The TG dinucleotide repeats are of 4–30 bases in size and distributed discontinuously in a region spanning 150–200 bp in the second intron of these genes. DNA composed of long stretches of TG dinucleotide repeats has been reported to adopt the "Z"-form DNA structure (Singleton et al., 1982). "Z"-form DNA was reported to mediate recombination events (Stringer, 1985) and function as a weak enhancer (Hamada et al., 1984). Weinreb et al. (1988) have shown that both the TC and TG dinucleotide repeats are involved in sister chromatid exchanges for the IgG_{2a} gene in mouse. The presence of both the TC and TG dinucleotide repeats in rat tissue kallikrein genes raises the possibility that recombinatorial events may have played a role in the evolution of this rat gene family. Why the nucleotide sequences in this TG-rich region are deleted so often and irregularly is not clear at the present time. We suggest that these deletions may result from recombinations among the members of the tissue kallikrein gene family during evolution. Furthermore, the recombinations may be mediated by these TG dinucleotide repeat elements.

Nucleotide sequences over 15 kb in the tonin gene have been determined, but only one gene was found. This information suggests that the intergenic distances between rat kallikrein-like genes are longer than those in mouse (Mason et al., 1983; Evans et al., 1987). A recent report on the study of human tissue kallikrein genes discussed the cloning and sequencing of a human kallikrein gene, hGK-1 (Schedlich et al., 1987). The complete hGK-1 gene sequence was accomplished by the isolation of two overlapping clones. The inserts of these two overlapping clones covered a region of 30 kb, but only one gene was found. This finding indicates that like rat genes, the human tissue kallikrein genes are not as closely linked as are the mouse genes. Thus, we believe that the distance between two closely linked rat genes would be longer than 10 kb. In order to uncover the organization of the rat tissue kallikrein genes, chromosomal walking with the use of DNA fragments at the extremities of each clone as probes is necessary.

Registry No. Serine protease, 37259-58-8; tonin, 53414-68-9; tonin (rat clone RSKG-5 reduced), 97071-43-7; protonin, 120788-22-9; protonin (rat clone RSKG-5 reduced), 120788-23-0; preprotonin, 85876-29-5; preprotonin (rat clone RSKG-5 reduced), 97071-42-6; DNA (rat clone RSKG-5 tonin gene), 120788-17-2; RNA (rat clone RSKG-5 tonin-specifying messenger), 120788-20-7.

REFERENCES

Ashley, P. L., & MacDonald, R. J. (1985a) *Biochemistry 24*, 4512–4520.

Ashley, P. L., & MacDonald, R. J. (1985b) *Biochemistry 24*, 4520–4527.

Bause, E., & Legler, G. (1981) *Biochem. J. 195*, 639–644.

Benoist, C., O'Hare, K., Breathnach, R., & Chambon, P. (1980) *Nucleic Acids Res. 8*, 127–143.

Bothwell, M. A., Wyndham, H., & Shooter, E. M. (1979) *J. Biol. Chem. 254*, 7287–7294.

Boucher, R., Saidi, M., & Genest, J. (1972) in *Hypertension* (Genest, J., & Koiw, E., Eds.) pp 512–523, Springer-Verlag, New York.

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

Chao, J. (1983) *J. Biol. Chem. 258*, 4434–4439.

Chao, J., Shimamoto, K., & Margolius, H. S. (1984) *Hoppe-Seyler's Z. Physiol. Chem. 365*, 1137–1141.

Chen, Y., Chao, J., & Chao, L. (1988) *Biochemistry 27*, 7189–7196.

Chretien, M., Lee, C. M., Sandberg, B. E. B., Iversen, L. L., Boucher, R., Seidah, N. G., & Genest, J. (1980) *FEBS Lett. 113*, 173–176.

Corden, J., Wasyluk, B., Buchwalder, A., Sasone-Corsi, P., Kedinger, C., & Chambon, P. (1980) *Science 209*, 1405–1414.

Davis, L. G., Dibner, M. D., & Battey, J. F. (1986) in *Basic Methods in Molecular Biology*, pp 130–146, Elsevier, New York.

Demassieux, S., Boucher, R., Grise, C., & Genest, J. (1976) *Can. J. Biochem. 54*, 788–795.

Drinkwater, C. C., Evans, B. A., & Richards, R. I. (1987) *Biochemistry 26*, 6750–6756.

Dynan, W., & Tjian, R. (1985) *Nature (London) 316*, 774–778.

Evans, B. A., & Richards, R. I. (1985) *EMBO J. 4*, 133–138.

Evans, B. A., Drinkwater, C. C., & Richards, R. I. (1987) *J. Biol. Chem. 262*, 8027–8034.

Gerald, W. L., Chao, J., & Chao, L. (1986a) *Biochim. Biophys. Acta 886*, 1–14.

Gerald, W. L., Chao, J., & Chao, L. (1986b) *Biochim. Biophys. Acta 867*, 10–13.

Gutkowska, J., Corvol, P., Thibault, G., & Genest, J. (1982) *Can. J. Biochem. 60*, 843–846.

Hamada, H., Petrino, M. G., Kakunage, T., Seidman, M., & Stollar, B. D. (1984a) *Mol. Cell. Biol. 4*, 2610–2612.

Hamada, H., Petrino, M. G., & Kakunaga, T. (1984b) *Mol. Cell. Biol. 4*, 2622–2630.

Ikeda, M., & Arakawa, K. (1984) *Hypertension (Dallas) 6*, 222–228.

Lazure, C., Leduc, R., Seidah, N. G., Thibault, G., Genest, J., & Chretien, M. (1987) *Biochem. Cell Biol. 65*, 321–337.

Leslie, A. G. W., Arnoff, S., Chandrasekaran, R., & Ratliff, R. L. (1980) *J. Mol. Biol. 143*, 49–72.
Lundgren, S., Ronne, H., Rask, L., & Peterson, P. A. (1984) *J. Biol. Chem. 259*, 7780–7784.
Maniatis, T., Fritsch, E. F., & Sambrook, J., Eds. (1982) in *Molecular Cloning: A Laboratory Manual*, pp 383–389, 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.
Messing, J. (1983) *Methods Enzymol. 101*, 20–78.
Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dorbrynin, Y. N., Filippov, S. A., & Frank-Kamenetskii, M. D. (1987) *Nature (London) 330*, 495–497.
Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459–472.
Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.
Sanger, F., Nicklen, S., & Coulson, A. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Schachter, M. (1963) *Ann. N.Y. Acad. Sci. 104*, 108–115.
Schreir, P. H., & Cortese, R. (1979) *J. Mol. Biol. 129*, 169–172.
Schedlich, L. J., Bennetts, B. H., & Morris, B. J. (1987) *DNA 6*, 429–437.
Seidah, N. G., Chen, J. S. D., Mardini, G., Benjannet, S., Chretin, M., Boucher, R., & Genest, J. (1979a) *Biochem. Biophys. Res. Commun. 86*, 1002–1013.
Seidah, N. G., Chan, J. S. D., Routhier, R., Gossard, F., Chretien, M., Boucher, R., & Genest, J. (1979b) in *Peptides* (Gross, E., & Meienhofer, J., Eds.) pp 921–924, Pierce Chemical Co., Rockford, IL.
Sen, D. G., & Gilbert, W. (1988) *Nature (London) 334*, 364–366.
Shih, H. C., Chao, L., & Chao, J. (1986) *Biochem. J. 238*, 145–149.
Singleton, C. K., Klysik, J., Stirdivant, S. M., & Wells, R. D. (1982) *Nature (London) 299*, 312–316.
Stringer, J. R. (1985) *Mol. Cell. Biol. 5*, 1247–1259.
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.
Thomas, K. A., Baglan, N. C., & Bradshaw, R. A. (1981) *J. Biol. Chem. 256*, 9156–9166.
Weinreb, A., Katzenberg, D. R., Gilmore, G. L., & Birshtein, B. K. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 529–533.

Kinetic and Structural Effects of Activation of Bovine Kidney Aldose Reductase†

C. E. Grimshaw,* M. Shahbaz, G. Jahangiri, C. G. Putney, S. R. McKercher, and E. J. Mathur‡

Division of Biochemistry, Division of Preclinical Neuroscience and Endocrinology, Department of Molecular and Experimental Medicine, Research Institute of Scripps Clinic, Scripps Clinic and Research Foundation, La Jolla, California 92037

Received October 20, 1988; Revised Manuscript Received March 16, 1989

ABSTRACT: Aldose reductase, purified to homogeneity from bovine kidney, is converted in a temperature-dependent process from a low-K_m/low-V_{max} form to a high-K_m/high-V_{max} form of the enzyme. Activation, which results in significant changes in the protein secondary structure, as detected by fluorescence spectroscopy, circular dichroism, and thiol modification with 5,5′-dithiobis(2-nitrobenzoic acid), has no effect on the apparent M_r, pI, or homogeneity of the enzyme, as judged by sodium dodecyl sulfate–polyacrylamide gel electrophoresis and agarose isoelectric focusing. V_{max}, which varied less than 3-fold for a series of aldehyde substrates with either activation state of the enzyme, increased an average of (17 ± 4)-fold upon activation of the enzyme. $V/K_{aldehyde}$ increased or decreased up to 4-fold, depending on the substrate. Activation desensitized the enzyme to inhibition by aldose reductase inhibitors, with the apparent K_i value increasing from 2-fold for Epalrestat [ONO-2235, (*E*)-3-(carboxymethyl)-(*E*)-5-[2-methyl-3-phenylpropenylidene]-rhodanine] to 200-fold for AL-1576 (spiro[2,7-difluorofluorene-9,4′-imidazolidine]-2′,5′-dione). Biphasic double-reciprocal plots for the aldehyde substrates and biphasic Dixon plots for inhibition by AL-1576 and Statil [ICI-128 436; 3-[(4-bromo-2-fluorobenzyl)-4-oxo-3*H*-phthalazin-1-ylacetic acid], observed during the course of activation, are quantitatively accounted for by the individual contributions of the two enzyme forms. On the basis of an analysis of the kinetic data, a mechanism is proposed in which isomerization of the free enzyme limits the rate of the forward reaction for the unactivated enzyme and is the primary step affected by activation.

Aldose reductase (ALR2;[1] alditol:NADP$^+$ 1-oxidoreductase; EC 1.1.1.21; also referred to as "low-K_m" aldehyde reductase) catalyzes the NADPH-dependent reduction of a wide variety of aldehydes to the corresponding alcohols:

$$NADPH + RCHO + H^+ \leftrightarrow NADP^+ + RCH_2OH$$

The kinetic properties of this reaction have been a subject of controversy, with conflicting reports of either Michaelis–Menten type kinetics (Boghosian & McGuinness, 1979; Wermuth et al., 1982; Branlant, 1982; Hara et al., 1983; Cromlish & Flynn, 1983a,b; Morjana & Flynn, 1989) or nonlinear double-reciprocal plots displaying apparent negative

† This work was supported by a grant to C.E.G. from the NIH (DK 32218) and the Olive H. Whittier Fund. Support for the VAX 11/750 in the General Clinical Research Center, Scripps Clinic and Research Foundation, was provided by a Division of Research Resources Grant (RR 00833) from the NIH. This is Publication No. 5211 BCR from the Research Institute of Scripps Clinic, Scripps Clinic and Research Foundation.

* Address correspondence to this author.
‡ Present address: Stratagene Cloning Systems, La Jolla, CA 92037.

0006-2960/89/0428-5343$01.50/0