Structural Organization of the Human Kininogen Gene and a Model for Its Evolution*

(Received for publication, September 21, 1984)

Naomi Kitamura, Hiroyuki Kitagawa, Daikichi Fukushima, Yoshio Takagaki, Takashi Miyata,†, and Shigetada Nakanishi

From the Institute for Immunology, Kyoto University Faculty of Medicine, Kyoto 606, Japan and the †Department of Biology, Kyushu University Faculty of Science, Fukuoka 812, Japan

The entire human kininogen gene has been isolated as a set of overlapping genomic DNA fragments, and the 11 exons encompassing approximately 27 kilobase paired with a thnic restriction enzyme analysis and nucleotide sequence determination. The nine 5’-terminal exons encode the 5’-untranslated region and the protein-coding region for the signal peptide and the heavy chain, which are common for high molecular weight (HMW) and low molecular weight (LMW) prekininogen mRNAs. Exon 10 consists of the common sequence for bradykinin and the immediately following unique sequence for HMW prekininogen mRNA. Exon 11 is then located following a 90-nucleotide sequence downstream from exon 10 and precisely specifies the sequence unique to LMW prekininogen mRNA. This, together with the hybridization analysis of total human cellular DNA, leads us to conclude that human HMW and LMW prekininogen mRNAs are produced from a single gene as a consequence of alternative RNA processing events. The structural analysis of the kininogen gene also shows that each of the nine 5’-terminal exons discretely specifies the nine protein domains observed in the amino-terminal portion of the kininogens. Furthermore, these nine genetic domains can be characterized by a thrice repeated pattern of three genetic segments, and two sets of these three domains, encompassing exons 3–5 and exons 6–8, are most closely related to each other. Therefore, we have proposed two successive duplication mechanisms as a model for the generation of the structure of the kininogen gene.

Mammalian kininogens, kinin precursors, are a useful model for studying the expression of the eukaryotic gene as well as for elucidating early events in its evolution. There are two types of kininogens, designated HMW1 and LMW kininogens, in mammalian organisms, and both of them are single-chain glycoproteins consisting of an amino-terminal H chain, a bradykinin moiety, and a carboxyl-terminal L chain (1–3). In the preceding paper (4), we described the cloning of the cDNAs for human HMW and LMW kininogen precursors (prekininogens). Analyses by sequence determination, RNA blot-hybridization, and S1 nuclease mapping indicated that the two mRNAs share an identical sequence throughout the 5’-untranslated region and the protein-coding region up to the sequence encoding the first 12 amino acid residues of the L chain, but they completely diverge from each other in the following 3’ sequences. This characteristic structural relationship between the two mRNAs suggested that the two mRNAs are produced from the same gene. In our previous study (3), an examination of a genomic clone containing the 3’ portion of the bovine kininogen gene indeed supported the view that the two prekininogen mRNAs are encoded by a single gene.

We also reported that the amino-terminal sequence preceding the bradykinin moiety possesses a characteristic structure composed of 9 structural domains, one for the signal peptide region and the other eight probably each forming a loop structure linked by two adjacent cysteine residues. Furthermore, we found that two sets of the three internal domains, corresponding to the 3rd–5th and 6th–8th domains, exhibit a remarkable sequence homology including three disulfide linkages present in the homologous region (4). This finding raises an interesting question concerning the relationship between the protein domains of kininogen and the exon-intron arrangements of its coding gene.

The study presented in this paper thus concerns an examination of the structural organization of the human kininogen gene to disclose the structural basis for the generation of the two prekininogen mRNAs and to investigate the evolutionary process which formed the primordial kininogen gene. We report here the isolation and the characterization of genomic clones containing the whole kininogen gene. Restriction mapping of the genomic DNA and nucleotide sequence determination of the exon sequences and their surrounding regions demonstrate that the two prekininogen mRNAs are encoded by a single gene. Furthermore, a good coincidence of the protein domains of the kininogens with the genetic domains of the kininogen gene and comparisons of the nucleotide sequences of the homologous domains which are separated from each other by genetic segments allow us to propose the evolutionary mechanisms which generated the primordial kininogen gene.

MATERIALS AND METHODS

Phage Screening—A human genomic DNA library was kindly provided by Dr. T. Maniatis (Harvard University) and was a collection of recombinant phage that carried human fetal liver DNA fragments generated by partial digestion with HaeIII and AluI and joined to the λ Charon 4A arms (5). Human genomic DNA fragments digested with BamHI were isolated with the use of the λ cloning system. Human placental DNA (0.2 mg) was digested completely with 1000 units of BamHI for 4 h and electrophoresed on a gel of 0.5% agarose. A DNA fraction that was expected to contain a desired fragment on the basis of blot-hybridization analysis was extracted.

* This work was supported in part by research grants from the Ministry of Education, Science and Culture of Japan and Japan Heart Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1754 solely to indicate this fact.

† The abbreviations used are: HMW, high molecular weight; LMW, low molecular weight; H chain, heavy chain; L chain, light chain; kb, kilobase pair(s); bp, base pair(s).
and ligated to the purified BamHI arms of λ Charon 28 (6). The recombinant DNA was packaged in vitro into viable phage (6). The phage were screened by hybridization in situ (7) with an appropriate probe labeled by nick translation (8). The following hybridization probes were used for screening phage carrying the human kininogen gene (see also "Results"). For the initial screening of the human genomic library cloned DNA of pk2200, which contained the cDNA sequence for bovine HMW prekininogen (3), was digested with HindIII and Hpal, and two large HindIII and Hpal-Hpal fragments were isolated and used as a probe. These two fragments covered the 5' cDNA sequence up to the middle portion of the 3'- untranslated region of bovine HMW prekininogen mRNA. For the second screening, the PstI-PumII fragment, which contained the cDNA sequence corresponding to the 3' portion of human HMW prekininogen mRNA together with the 3'G-GC and 3'-AAG tails flanking the cDNA sequence, was isolated from clone pKHG4 (4) and digested with BstNI. The three fragments, i.e. the PstI-BstNI fragment, the BstNI fragment, and the BstNI-PumII fragment, were isolated by polyacrylamide gel electrophoresis to eliminate the other BstNI fragment containing the repetitive Alu family sequence (4). A mixture of these three DNA fragments was labeled and used as a probe. For the screening of phage containing the BamHI-digested genomic DNA fragments, the BstNI fragment encoding the bradykinin sequence corresponding to the 3'-flanking region was isolated from a genomic subclone derived from XhKG-D and used as a probe. Hybridization-positive phage clones were isolated by repeated plaque purification (9). Subcloning of genomic DNA fragments in plasmid pBR322 was performed as described in Maniatis et al. (9). All of the cloning procedures were conducted in accordance with the guidelines for recombinant DNA molecules issued by the Ministry of Education, Science and Culture of Japan.

Restriction Mapping, Blot-Hybridization, and Nucleotide Sequence Analysis—Restriction endonucleases were purchased from Takara Shuzo Co., Bethesda Research Laboratories, and New England Biolabs. Reactions were carried out under the conditions recommended by the suppliers. Restriction fragments were isolated by electrophoresis on agarose gel (10) or on polyacrylamide gel (11). Blot-hybridization analysis of cloned genomic DNA and human placental cellular DNA was conducted as described by Southern (12). 5' End labeling of restriction fragments and DNA sequencing were carried out by the methods of Maxam and Gilbert (11).

**Primer Extension and S1 Nucleic Mapping Analysis—**Poly(A) RNA was isolated from a human liver as described in the preceding paper (4). Fractions containing either HMW or LMW prekininogen mRNA were obtained by sucrose density gradient centrifugation as described (4). For the primer extension experiment, the 106-bp HindII fragment was obtained from clone position 243 to 349 (4) and digested with SmaI. The 44-bp HindII-SmaI fragment corresponding to the 5'-terminal portion of the 5'-untranslated region of human prekininogen mRNA was isolated and labeled with 32P at the 5' ends. For the S1 nuclease mapping analysis, the ~1.3-kb Hpal-EcoRI fragment was isolated from the genomic subclone containing the ~5.2-kb EcoRI fragment, which was derived from λhKG-B and was digested with BstNI. The 229-bp Hpal-Sau3A1 fragment corresponding to the 5'-untranslated region of human prekininogen mRNA and its 5'-flanking genomic sequence was isolated and labeled with 32P at the 5' ends. The 243-nucleotide antisense strand was then isolated by electrophoresis on a 5% polyacrylamide gel as described (11). The Hpal-Sau3A1 fragment (1.0 × 10^6 cpm) was hybridized to several RNA preparations in a solution containing 80% (v/v) formamide and 0.4 M NaCl at 39°C for 9 hr, and the DNA-RNA hybrids were digested with S1 nuclease (50 units/μg RNA) at 40°C for 30 min (13, 14). The Hpal-Sau3A1 fragment (4 × 10^6 cpm) was denatured at 75°C for 5 min, hybridized to human liver poly(A) RNA or yeast tRNA at 39°C for 9 hr, and the solution described for S1 nuclease analysis, and then subjected to reverse transcription reaction (15). The S1 digestion products and the primer-extended cDNAs were electrophoresed parallely with chemically degraded products of the 229-bp Hpal-Sau3A1 fragment on 7 M urea/50% polyacrylamide gel.

**Statistical Analysis—**For the calculation of sequence homology, the nucleotide sequences were aligned by using a matrix alignment program. The validity of a given homology between two nucleotide sequences was evaluated as follows. For each of the two sequences to be evaluated, we randomly generated 100 pairs of sequences which were adjusted to correspond to the evaluated sequences with respect to length, base composition, codon preference, constraint on amino acid composition, and absence of termination codon. Then we calculated the mean (h) and the standard deviation (σ) of the per cent homologies of these 100 sequences. The probability that the observed per cent homology (hobs) of the two real sequences might arise by chance was then evaluated by introducing the index (hobs - h)/σ, whose value is expected to be normally distributed with a mean of zero and a standard deviation unity (16).

**RESULTS**

**Cloning of the Human Kininogen Gene—**An initial attempt to isolate phage carrying the kininogen gene was performed by hybridization in situ of a human genomic DNA library with restriction fragments derived from the cDNA clone insert for bovine HMW prekininogen as a probe. Three hybridization-positive clones, λhKG-A, λhKG-B, and λhKG-C, were isolated from approximately 6 × 10^9 plaque. Restriction endonuclease digestion and DNA blot-hybridization analyses revealed that the three clones covered the DNA sequence corresponding to the 5'-mRNA sequence and its 5'-flanking genomic sequence but did not encode the bradykinin sequence nor the following 3' sequence (Fig. 1a). Therefore, a further attempt was made to isolate clones containing the remaining 3' portion of the kininogen gene. The cDNA probe used in the second screening was restriction fragments derived from the cloned cDNA for human HMW prekininogen, which was made available to us at this time. Because human HMW prekininogen mRNA contains a part of the repetitive Alu family sequence (4), the sequence containing the Alu family sequence was eliminated from the cDNA probe used. In the second screening, one clone, λhKG-D, was isolated from approximately 5 × 10^9 plaque, and this clone was found to encode the 3' portions of both HMW and LMW prekininogen mRNAs. The total 4 clones thus obtained covered the whole mRNA sequences for both HMW and LMW prekininogens.

The genomic DNA inserts of λhKG-B and λhKG-C and those of λhKG-A and λhKG-D were cloned and sequenced. The genomic DNA inserts of 2 clones of each of the 2 clones were aligned into an approximately 46-kb length of continuous genomic DNA, as shown in Fig. 1a.

**Blot-hybridization analysis of human placental cellular DNA exhibited hybridization-positive fragments as anticipated from the restriction map, and no other hybridization-positive bands were detected (data not shown). These results indicate that the cloned DNA segments (data not showing) contain the kininogen gene sequence organization found in the cellular DNA. They further suggest that there are no other closely related genes or pseudogenes in the human genome. Appropriate restriction fragments which covered the whole.
kininogen gene was subcloned in pBR322, and DNA fragments containing exons and their surrounding regions were subjected to nucleotide sequence analysis according to the strategy indicated in Fig. 1, b–j.

Structural Organization of the Human Kininogen Gene—Comparison of the genomic DNA sequence with the cDNA sequence reported in the preceding paper (4) enabled us to construct a structural organization of the human kininogen gene. The sequences of the exon-intron boundaries and the sizes of the exons and introns are summarized in Fig. 2, and the detailed sequences of exons 10 and 11 and their flanking regions are shown in Fig. 3. All introns interrupt the protein-coding regions of the gene. Introns A–I split the common H chain-coding sequence between the triplets encoding amino acid residues 65–66, 102–103, 131, 188–189, 224–225, 253, 310–311, 346–347, and 375–376, respectively. Exons 2–9 thus consist of 111, 85, 173, 108, 85, 173, 108, and 87 bp, respectively. Exon 10 can be divided into two portions. The 5' portion of exon 10, termed exon 10BK, is composed of 78 bp and encodes the common bradykinin sequence for HMW and LMW prekininogens, while the 3' portion of exon 10, termed exon 10HMW, consists of 2015 bp and encodes the HMW prekininogen mRNA-specific sequence. Exon 11, which is located following a 90-bp sequence downstream from exon 10HMW, specifies the 248-bp sequence precisely corresponding to the divergent 3'-terminal region of LMW prekininogen mRNA. Thus, the complete unraveling of the genomic structure for kininogens indicates that the sequences unique to HMW prekininogen mRNA and to LMW prekininogen mRNA are continuous, with the interval of the 90 bp, and that the exon 10HMW and its flanking 90-bp sequence serve as an intervening sequence (intron J) in the process of the generation of the LMW prekininogen mRNA.

In the preceding paper (4), we described that the amino-terminal sequence preceding the bradykinin moiety can be divided into nine structural domains which consist of the
FIG. 2. Exon-intron arrangements of the human kininogen gene and sequences at the splice junctions. The sequences at the splice junctions are determined by comparing the genomic DNA sequence with the cDNA sequence. The exon-intron boundaries are positioned according to the GT/AG rule for the splice junctions (19). The amino acid residues are numbered as in Ref. 4.

FIG. 3. The nucleotide sequences of exons 10 and 11 and their surrounding regions. Boxes indicate the exons. Wavy lines show the boundary sequence of intron J, which corresponds to the consensus sequence for the donor and acceptor sites of RNA splicing (19).

The structural organization of the human kininogen gene is depicted. The exons are numbered sequentially from the 5' end to the 3' end. The introns are indicated by lines between the exons. The nucleotide sequences of the exons and introns are shown. The exon-intron boundaries are determined by the GT/AG rule. The amino acid sequences are numbered as in Ref. 4.

The three genetic domains encompassing exons 3-5 and exons 6-8 are extremely related, with a 49% nucleotide match between exons 4 and 7, a 52% match between exons 5 and 8, and a 44% match between exons 3 and 6. Furthermore, not only the exon sizes but also the positions of the introns interrupting the amino acid sequences are identical for all three paired exons. The sequences of exon 2 and exon 9 are also highly homologous to those of exons 5 and 8 and to those of exons 3 and 6, respectively. In the former case, exon 2 differs slightly in size from exons 5 and 8, and one amino acid gap is necessary to align the amino acid sequence encoded by exon 2 with those encoded by exons 5 and 8, yet 41 and 48% nucleotide matches are observed between exons 2 and 5 and between exons 2 and 8, respectively. In the latter case, although the pattern of intron interruption at the 3' end of exon 9 differs from those observed at the 3' ends of exons 3 and 6, the two cysteine residues are located at exactly equivalent positions among the three sequences, and 35 and 37% nucleotide matches are observed between exons 9 and 3 and between exons 9 and 6, respectively.

besides the above sequence homologies among eight of the exons, we noted that the 3'-terminal portion up to about 80 nucleotides upstream from the 3' end of exon 1 is considerably homologous to the 3'-terminal regions of exon 4 with a nucleotide match of 42%, whereas no such homology is observed between the 5'-terminal sequence encoding the signal peptide region in exon 1 and the 5'-terminal region of exon 4. Similarly, the 3'-terminal sequences of exons 1 and 7 (36% homology) are more related than the 5'-terminal sequences of exons 1 and 7 (30% homology), even though this difference is not statistically significant. In addition, three protein structures of this homologous family possess a canonical glycosylation site Asn-X-Thr/Ser (17) at the almost corresponding location. For example, the sequences of exons 4 and 7 differ slightly in size from exons 5 and 8, yet 41 and 48% nucleotide matches are observed between exons 2 and 5 and between exons 2 and 8, respectively. In the latter case, although the pattern of intron interruption at the 3' end of exon 9 differs from those observed at the 3' ends of exons 3 and 6, the two cysteine residues are located at exactly equivalent positions among the three sequences, and 35 and 37% nucleotide matches are observed between exons 9 and 3 and between exons 9 and 6, respectively.

Multiple 5' Ends of Prekinogen mRNAs—The 5' termini of the prekinogen mRNAs were identified by S1 nuclease

<table>
<thead>
<tr>
<th>Exon Number</th>
<th>Exon Size (bp)</th>
<th>3' Exon Junction</th>
<th>Intron</th>
<th>5' Exon Junction</th>
<th>Exon Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>327, 328</td>
<td>Thr Lys Thr</td>
<td>GTCAGTAAAT---(Intron A, -2.2 kbp)--TCGTTCCTAG</td>
<td>Ala Gly Ser</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>111</td>
<td>ACT AGC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>85</td>
<td>GTGGCTGGCTT-----(Intron C, -2.7 kbp)-----CTGTTLTNG</td>
<td>Ala Thr Gly</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>173</td>
<td>Gln Ala CAA</td>
<td>GTTGTCTCT-----(Intron D, +1.9 kbp)-----ATTGTTCCTAG</td>
<td>Ala Thr Gln</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>108</td>
<td>Trp Asn Gly</td>
<td>GTAAGACC-----(Intron E, -4.1 kbp)-----AAAACCGG</td>
<td>Val Ala Gln</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>253</td>
<td>Tyr Pro G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>173</td>
<td>Arg Val Gln</td>
<td>GTGGAAGCTA-----(Intron F, +0.7 kbp)-----ATCCCTTCTAG</td>
<td>Val Gly Ser</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>108</td>
<td>Lys Leu Gly</td>
<td>GTAATGCTGCT-----(Intron H, 121 bp)-----TACAATATAG</td>
<td>Val Gly Ser</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>87</td>
<td>Leu Gly Met</td>
<td>GTTGGCTGGCAT-----(Intron I, 2.4 kbp)----TCTGTGTAG</td>
<td>Val Gly Ser</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>78</td>
<td>Glu Thr Thr</td>
<td>GTAAGGCCC-----(Intron J, 2,105 bp)-----ATACCTTACG</td>
<td>Val Gly Ser</td>
<td>9</td>
</tr>
<tr>
<td>10*</td>
<td></td>
<td>GAS ACA ACT</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The nucleotide sequences of exons 10 and 11 are shown. The boundary sequence of intron J, which corresponds to the consensus sequence for the donor and acceptor sites of RNA splicing (19), is also shown.

The nucleotide sequences of exons 10 and 11 and their surrounding regions are shown. The boundary sequence of intron J, which corresponds to the consensus sequence for the donor and acceptor sites of RNA splicing (19), is also shown.
Fig. 4. Alignment of sequence homologies within related exons. The nucleotide sequences of related exons are aligned with spaces between in-frame codons. Within each comparison, homologies between the upper sequence and the middle sequence and between the middle sequence and the lower sequence are indicated by vertical lines, and homologies between the upper sequence and the lower sequence are shown by asterisks above the upper sequence. The positions corresponding to cysteine residues and to possible glycosylation sites conforming to the Asn-X-Thr/Ser sequence (17) are enclosed in boxes and indicated by wavy lines, respectively.

Table I
Sequence homologies among exons

<table>
<thead>
<tr>
<th>Paired exons</th>
<th>Per cent homology</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 versus 4*</td>
<td>21.1</td>
<td>0.82 (NS)*</td>
</tr>
<tr>
<td>1 versus 4'</td>
<td>42.3</td>
<td>3.4 x 10^-2</td>
</tr>
<tr>
<td>1 versus 7*</td>
<td>29.5</td>
<td>0.20 (NS)</td>
</tr>
<tr>
<td>1 versus 7'</td>
<td>35.9</td>
<td>0.11 (NS)</td>
</tr>
<tr>
<td>4 versus 7</td>
<td>49.1</td>
<td>&lt;10^-9</td>
</tr>
<tr>
<td>2 versus 5</td>
<td>41.0</td>
<td>6.9 x 10^-4</td>
</tr>
<tr>
<td>2 versus 8</td>
<td>47.6</td>
<td>6.1 x 10^-6</td>
</tr>
<tr>
<td>5 versus 8</td>
<td>51.9</td>
<td>&lt;10^-9</td>
</tr>
<tr>
<td>3 versus 6</td>
<td>43.5</td>
<td>3.2 x 10^-4</td>
</tr>
<tr>
<td>3 versus 9</td>
<td>35.3</td>
<td>9.8 x 10^-2</td>
</tr>
<tr>
<td>6 versus 9</td>
<td>36.5</td>
<td>3.1 x 10^-2</td>
</tr>
</tbody>
</table>

* Comparisons between the indicated exons at the 5'-terminal sequences of the 95 nucleotides.
° NS, not statistically significant.
* Comparisons between the indicated exons at the 3'-terminal sequences of the 78 nucleotides.

mapping and primer extension analyses (Fig. 5). For primer extension analysis, a CDNA fragment corresponding to the 3'-terminus of the 5'-untranslated region of prekininogen mRNA (the 44-bp HinfI-Sau3A1 fragment) was isolated and labeled at the 5' ends with [γ-32P]ATP. The labeled fragment was denatured, hybridized to total liver poly(A) RNA, and then subjected to reverse transcription reaction. For S1 nuclease mapping analysis, a genomic DNA fragment containing the 5'-untranslated region of the corresponding mRNA and the 5'-flanking region of the kininogen gene (the 239-bp HpaI-Sau3A1 fragment) was labeled at the 5' ends, and the anti-message strand was isolated by polyacrylamide gel electrophoresis after denaturation. The isolated single-stranded DNA was hybridized to total liver poly(A) RNA, and the RNA-DNA hybrids were then analyzed by S1 nuclease mapping. To relate the start sites of the mRNAs to the DNA sequence, the nucleotide sequence of the exon sequence and its 5'-flanking region was determined according to the strategy indicated in Fig. 1b, and the result of the sequence determination is presented in Fig. 5c.

Fig. 5a shows an autoradiograph of the primer extension and S1 nuclease mapping analyses of total human poly(A) RNA. Both analyses revealed a length heterogeneity at the 5' end of the prekininogen transcript. The relative intensities and the locations of the observed bands generally agreed between the two methods, but some bands differed in their sizes by a few nucleotides. Because a cap structure at the 5' terminus of the mRNA as well as A-T base pairs of the 5' end of the DNA-RNA hybrid may affect the protection of the DNA probe by S1 nuclease digestion (14), the 5' ends of the prekininogen mRNAs were assigned on the basis of results obtained by the primer extension analysis. Accordingly, three major and one minor 5' termini of prekininogen mRNAs were mapped at 132, 133, and 184 bp, and at 155 bp upstream from the 5' end of the initiation codon AUG, respectively. Because those assignments were supported by the S1 nuclease analysis regardless of several nucleotide differences in the assignment, we conclude that the 5'-terminal gene sequence is uninterrupted, and thus exon 1 encodes the whole 5'-untranslated region of the mRNA as well as the protein-coding region corresponding to the signal peptide and the amino-terminal portion of the H chain. This assignment also leads to the conclusion that the human kininogen is approximately 27 kb long.

In the 5'-flanking region of the kininogen gene, two AT-rich sequences, 5'-TTAATAATGT'TT-3' and 5'-ATTTAA-3', were found at 19 nucleotides upstream from the major 5' end of the prekininogen mRNA assigned at the most 5'-extreme region and at 24 nucleotides upstream from the minor 5' end of the mRNA, respectively. The two AT-rich sequences differ from the typical "TATA" box which is commonly found at the corresponding positions of eukaryotic genes (18, 19). Moreover, no AT-rich sequence was found within 45 nucleotides upstream from the start sites of the two other major mRNA species.

It is possible that each of the observed multiple 5' ends of the mRNAs may uniquely specify either HMW or LMW prekininogen mRNA. In order to investigate this possibility, total human liver poly(A) RNA was fractionated by sucrose density gradient centrifugation, and fractions containing either HMW or LMW prekininogen mRNA were separately subjected to S1 nuclease analysis. The autoradiograph of such an analysis is presented in Fig. 5b. HMW and LMW prekininogen mRNAs yielded identical multiple S1 nuclease-resistant bands, indicating that the same multiple initiation sites are used for the generation of the two prekininogen mRNAs.

Discussion
This report describes the structural organization of the human kininogen gene. The 11 exons encompassing about 27
Fig. 5. Identification of the 5′ termini of human prekininogen mRNAs by S1 nuclease mapping and primer extension analyses (a and b) and the nucleotide sequence around exon 1 of the kininogen gene (c). In an autoradiograph of a, the following DNAs were electrophoresed: the S1 digestion products using yeast tRNA (lane 1) and total liver poly(A) RNA (lane 2), the G-specific degradation products of the HpaI-Sau3AI fragment used as a chain length marker (lane G), and the primer-extended cDNAs using total liver poly(A) RNA (lane 3) and yeast tRNA (lane 4). In an autoradiograph of b, the following DNAs were electrophoresed: the S1 digestion products using fractions containing LMW prekininogen mRNA (lane 1) and containing HMW prekininogen mRNA (lane 2), total poly(A) RNA (lane 3) and yeast tRNA (lane 4), and the chemically degraded products of the HpaI-Sau3AI fragment used as a chain length marker (the 4 right lanes). In the nucleotide sequence around exon 1 shown in c, the 5′ ends of prekininogen mRNAs are indicated by triangles above the nucleotide sequence, and AT-rich regions are indicated by asterisks under the nucleotide sequence. The restriction sites used for the isolation of the 44-bp HinfI-Sau3A1 fragment and the 243-nucleotide single-stranded HpaI-Sau3AI fragment are indicated under the nucleotide sequence. Note that both fragments were labeled at the same Sau3A1 site present next to the initiation codon AUG. Experimental details for the S1 nuclease mapping and primer extension analyses are described under “Materials and Methods.”

kb were mapped by restriction enzyme analysis and nucleotide sequence determination. Based upon the exon-intron arrangements of the kininogen gene together with the primary structures of two prekininogen mRNAs (4), Fig. 6 schematically illustrates alternative arrangements for the generation of HMW and LMW kininogens. The 9 exons located at the 5′-terminal region of the kininogen gene encode the common sequence for HMW and LMW prekininogen mRNAs, and these exons cover the 5′-untranslated region and the protein-coding region for the signal peptide and the H chain. Exon 10 consists of the common sequence encoding bradykinin (exon 10km) and the immediately following unique sequence for HMW prekininogen mRNA (exon 10a,mw). Exon 11 is then located following a 90-nucleotide sequence downstream from exon 10 and precisely specifies the sequence unique to LMW prekininogen mRNA. Thus, taking into account the fact that no more than one human kininogen gene was identified by blot-hybridization analysis of human cellular DNA, we conclude that HMW and LMW prekininogen mRNAs are produced from a single gene.

The sequence 5′-GTAAGT-3′ located at the 5′ end of exon 10km and the sequence 5′-CTTACAG-3′ flanking the 5′ end of exon 11 correspond to the consensus sequences for the donor and acceptor sites of RNA splicing (19). Thus, the mature LMW prekininogen mRNA can be produced by splicing out the exon 10km sequence and its flanking 90-nucleotide sequence, while the HMW prekininogen mRNA can be generated by including the continuous exon 10km sequence but not the separated exon 11 sequence. Similar alternative RNA processing events for the generation of multiple protein
products from a single gene have been reported for some viral and cellular genes (20–25). In the case of immunoglobulin genes, the splicing of alternative exons results in the expression of either the secreted or membrane type of immunoglobulin \(\mu\) chain and \(\delta\) chain (22–24). The calcitonin gene is also reported to have alternative splicing events, which result in an additional peptide, referred to as calcitonin gene-related peptide (25). A characteristic feature of the kininogen gene is that alternative splicing events still maintain the biologically active peptide bradykinin in the resultant two protein products and nevertheless provide one of the products for an additional function involved in the contact activation reactions. Thus, alternative expression of the kininogen gene seems to serve in the production of dual functions involved in the inflammation reaction.

Examples of multiple transcription initiations have been reported for several viral systems (26–28) as well as some cellular genes such as \(\epsilon\)-globin (29) and lysozyme gene (30). Although the 5'—flanking region of the kininogen gene contains the AT-rich sequences, these sequences differ from the canonical "TATA" box which is believed to be involved in specific transcription initiation (18, 19). Therefore, the observed heterogeneity of both HMW and LMW prekininogen mRNAs may reflect an imprecision in the polymerase nucleotide selection mechanism.

The present investigation has revealed a remarkable coincidence of the protein domains of the kininogens with the genetic domains of the kininogen gene. Furthermore, the sequence comparison of the 9 exons in the H chain-coding region has clearly indicated that this region is characterized by a thrice repeated pattern of the three genetic domains. It is also evident that the two sets of the three domains encompassing exons 3–5 and exons 6–8 are most closely related, not only in the nucleotide sequence homology but also in the sizes of paired exons and the positions of introns interrupting the amino acid sequences. Therefore, based upon the relatedness of sets of genetic domains, we propose two successive duplication mechanisms as a model for the generation of the structure of the kininogen gene, and our model is schematically illustrated in Fig. 7. We assume that the primordial H chain polypeptide consisted of three segments. An unequal crossover between the 5'-flanking and the 3'-flanking regions of a set of the three domains resulted in the formation of duplicated sets of the three domains. A subsequent homologous recombination event then gave rise to a thrice repeated pattern of the three domains as a result of a duplication of the internal three domains.

The sequence comparison of exon 1 and its related exons 4 and 7 has indicated that the 3'-terminal sequence of exon 1 of about 80 nucleotides is closely related to the corresponding regions of the latter two exons, whereas no such sequence relatedness is observed between the 5'-terminal sequence encoding the signal peptide region in exon 1 and the 5'-terminal sequences of exons 4 and 7. It is, therefore, possible that the signal peptide exon had once been physically separated from the rest of the gene, and after the formation of either the duplicated or triplicated structure of the H chain-coding region, the signal peptide exon was fused to the 5'-terminal exon of the duplicated or triplicated structure by removal of parts of the exon sequences as well as the intervening sequence between the two exons. Alternatively, the signal peptide sequence could have been generated by the sequence divergence of the first exon after the duplication or triplication of the three genomic segments, although it seems difficult to envision how this particular exon underwent such a rapid divergence to acquire a specific function for the signal peptide.

The last two exons, which lie outside the repeated genetic domains, bear the functionally distinct polypeptides from those of the first 9 exons, in that they encode the bradykinin moiety and the L chains of the two kininogens. Although
these exons are included in the first primordial gene structure shown in Fig. 7, we have no clue to deduce the order of the occurrence of these two exons relative to that of the two successive duplications of the three primordial genetic segments. With respect to the alternative splicing events for the generation of the two prekininogen mRNAs, such events could have begun with a mutation in the intervening sequence preceding the LMW prekininogen-specific exon. This mutation caused some RNA transcripts of the kininogen gene to be polyadenylated prior to the sequence of exon 11, and mRNA produced from these transcripts could have contained a new 3' terminus derived from part of the former intron. Subsequent evolution presumably provided the carboxyl-terminal portion of the resultant protein product with a new function involved in the contact activation reactions in inflammation. Alternatively, the LMW prekininogen-specific exon could have been included in the kininogen gene by an exon-shuffling mechanism (31), followed by the generation of a new RNA splice junction within the coding region for HMW prekininogen. However, this possibility seems less likely because such a mechanism is composed of a two-step evolutionary event.

REFERENCES