Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 5999–6003, October 1984
Biochemistry

Structure of the human renin gene

(hypertension/aspartyl proteinase/nucleotide sequence/splice junction)

HITOSHI MIYAZAKI*, AKIYOSHI FUKAMIZU*, SHIGEHISA HIROSE*, TAKASHI HAYASHI*, HITOSHI HORI*, HIROAKI OHKUBO†, SHIGETADA NAKANISHI†, AND KAZUO MURAKAMI*‡

*Institute of Applied Biochemistry, University of Tsukuba, Ibaraki 305, Japan; and †Institute for Immunology, Kyoto University Faculty of Medicine, Kyoto 606, Japan

Communicated by Leroy Hood, June 27, 1984

ABSTRACT **The human renin gene was isolated from a Charon 4A human genomic library and characterized. The gene spans about 11.7 kilobases and consists of 10 exons and 9 introns that map at points that could be variable surface loops of the enzyme. The complete coding regions, the 5′- and 3′-flanking regions, and the exon–intron boundaries were sequenced. The active site aspartyl residues Asp-38 and Asp-226 are encoded by the third and eighth exons, respectively. The extra three amino acids (Asp-165, Ser-166, Glu-167) that are not present in mouse renin are encoded by the separate sixth exon, an exon as small as 9 nucleotides. The positions of the introns are in remarkable agreement with those in the human pepsin gene, supporting the view that the genes coding for aspartyl proteinases have arisen as the result of duplication of a common ancestral gene. As in most eukaryotic genes, the putative T-A-T-A and C-A-A-T sequences, which may play a role in the initiation of gene transcription, are found in the vicinity of −29 and −51 nucleotides of the cap site. Further upstream, at nucleotides −456 to −451, is located the hexanucleotide T-G-T-T-C-T, which has recently been suggested as a binding site for the glucocorticoid receptor. In the 3′-flanking region, there is the conserved hexanucleotide sequence A-A-T-A-A-A, thought to be necessary for polyadenylylation. Blot-hybridization analyses of the isolated gene clone and the total cellular DNA after digestion with restriction enzymes revealed that human renin is encoded by a single gene.**

Renin is an aspartyl proteinase synthesized mainly in the juxtaglomerular cells of the kidney, which are located in the afferent arterioles immediately proximal to the glomeruli (1). When the blood flow through the kidney becomes inadequate, renin is secreted into the circulation where it acts on a plasma protein to produce the decapeptide angiotensin I. Under the influence of the converting enzyme, the angiotensin I derived from the larger parent protein angiotensinogen, in turn, becomes angiotensin II, which has potent vasoconstrictor and aldosterone secretion-stimulating activities. Renin is, therefore, the key enzyme of the renin–angiotensin–aldosterone cascade aimed at increasing blood pressure and extracellular fluid volume (2).

Recently we (3) have reported the isolation and sequence analysis of a human renin cDNA clone. From the cDNA sequence, the human renin precursor, preprorenin, was shown to be composed of 406 amino acids. In the present study, we used the human renin cDNA to isolate the corresponding gene sequence from a human genomic library. Analysis of the organization and structure of the renin gene revealed that it exists in a single copy, spans roughly 11.7 kilobases (kb), and consists of 10 exons. We also examined the intron–exon junctions to see if there is a correlation between the intron–exon organization of the gene and the tertiary structure of the protein.

MATERIALS AND METHODS

Materials. All restriction enzymes were obtained from either New England Biolabs or Takara Shuzo (Kyoto, Japan). *Escherichia coli* alkaline phosphatase and T4 DNA ligase were from Takara Shuzo. [γ-^{32}P]ATP (>5000 Ci/mmol; 1 Ci = 37 GBq) and [α-^{32}P]dCTP (≈3000 Ci/mmol) were from Amersham.

Screening. A human genomic library, prepared from partial *Alu* I and *Hae* III digestion and ligated into the *Eco*RI arms of the λ vector Charon 4A, was kindly provided by T. Maniatis (4). A portion of the amplified library plated on *E. coli* LE 392 was screened, by the method of Benton and Davis (5), for sequences that hybridize with cDNA probes [*Pst* I–*Hpa* II and *Hpa* II–*Rsa* I fragments of pHRn 321 (3)] labeled with ^{32}P by nick-translation to a specific activity of 2×10^8 cpm/μg. Hybridization was conducted at 65°C for 12–24 hr and the filters were sequentially washed several times in (*i*) 0.3 M NaCl/0.03 M sodium citrate/0.1% $NaDodSO_4$ at room temperature, (*ii*) 0.3 M NaCl/0.03 M sodium citrate/0.1% $NaDodSO_4$ at 65°C, (*iii*) 15 mM NaCl/1.5 mM sodium citrate/0.1% $NaDodSO_4$ at 65°C, and (*iv*) 0.3 M NaCl/0.03 M sodium citrate at room temperature and exposed to Fuji x-ray film RX-50 with a DuPont Cronex Lightning Plus screen. Positive clones were picked and rescreened at low plaque density to achieve high purity. Preparation of phage DNA was carried out as described (6).

All cloning experiments were performed in accordance with the guidelines for research involving recombinant DNA molecules issued by the Ministry of Education, Science and Culture of Japan.

Subcloning. DNA fragments from the positive phage clones were digested with *Eco*RI and the four fragments spanning the entire length of the gene were isolated by electrophoresis through 0.7% agarose gels. The purified DNA fragments were subcloned into the *Eco*RI site of plasmid pBR322 since larger quantities of the *Eco*RI fragments were required for sequence determination. Transformations of *E. coli* HB101 were performed as described by Morrison (7).

Mapping and Sequencing. DNA samples were digested with various restriction enzymes under conditions specified by the supplier and separated by electrophoresis on 0.7% and 2.5% agarose or 5% polyacrylamide gels. Ethidium bromide-stained gels were photographed under UV light and DNA was transferred to nitrocellulose filters in 0.9 M NaCl/0.09 M sodium citrate as described by Southern (8). Hybridization was carried out at 65°C for 12 hr with nick-translated cDNA probes as described above, and filters were autoradiographed at −70°C with intensifier screens.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: kb, kilobase(s).

‡To whom reprint requests should be addressed.

Appropriate restriction fragments were isolated from polyacrylamide gels and sequenced according to Maxam and Gilbert (9) by using the 5′ end-labeling method. The sequence strategy used is shown in Fig. 1*C*.

RESULTS

Isolation of a Genomic Clone for Human Renin and Subcloning of Its *Eco*RI Fragments. A human genomic library constructed in the λ phage Charon 4A (4) was screened for the renin gene by using a nick-translated cDNA insert as a hybridization probe. After screening 400,000 recombinant phages, 7 positive clones were identified. They were plaque-purified and their DNA was isolated for restriction endonuclease characterization. Preliminary restriction enzyme analysis of these clones indicated that they covered similar regions of about 12 kb of the human genome. Two clones, designated λHRn 42 and λHRn 72, were further analyzed and shown to have three *Eco*RI sites (Fig. 1*A*). The orientation of the gene was determined by using the 5′ and 3′ specific cDNA probes generated by digesting the human renin cDNA at a unique *Hpa* II site. Four *Eco*RI fragments (the 5′-most 7.8-kb, the 3′-most 5.7-kb, and the overlapping 3.4-kb and 0.8-kb fragments shown in Fig. 1*A*), which together comprise 17.7 kb of the human genome, were subcloned for the subsequent sequence analysis.

Soubrier *et al.* (10) have shown that the renin gene exists in a single copy in the human genome. To confirm this and to establish the absence of major sequence rearrangement of the renin gene during cloning and propagation of the λHRn 72 DNA, we analyzed restricted genomic DNA by blot hybridization. When the placental cellular DNA was digested with *Sac* I or *Hin*dIII, the fragments detected by the [^{32}P]cDNA probe were identical with those seen in the cloned λHRn 72 DNA except the fragments containing the left or a portion of the right arm of λ phage DNA. The sizes of these hybridizing fragments correspond exactly to those expected from the partial restriction map shown in Fig. 1*B*. These results suggest that the cloned gene sequence (λHRn 72) represents the original structure of the human renin gene that is present as a single copy in the chromosome.

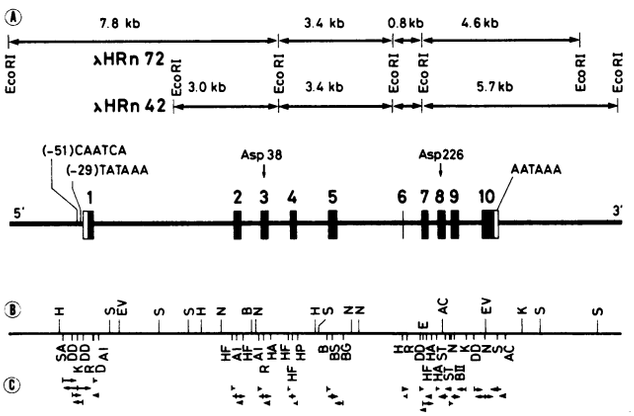

FIG. 1. Restriction map and strategy for sequencing the human renin gene. (*A*) Location of *Eco*RI sites in recombinant phage clones. (*B*) Detailed restriction map of the renin gene and its flanking regions. The restriction sites utilized for sequencing are indicated by downward lines. (*C*) Regions of the DNA that were sequenced; the horizontal arrows indicate the direction and length of each sequence analysis. The structure of the renin gene is illustrated schematically with solid (coding regions) and open bars (untranslated regions) and solid lines (introns and flanking regions). All sites of the following restriction enzymes are shown: B, *Bam*HI; EV, *Eco*RV; H, *Hin*dIII; K, *Kpn* I; S, *Sac* I. Restriction sites used for sequencing are AC, *Acc* I; AI, *Ava* I; BII, *Ban* II; BG, *Bgl* I; BS, *Bst*EII; D, *Dra* I; DD, *Dde* I; E, *Eco*RI; HA, *Hae* III; HF, *Hin*fI; HP, *Hpa* II; N, *Nco* I; R, *Rsa* I; SA, *Sau*3AI; ST, *Stu* I (these do not cover all of the sites present in the gene).

Organization of the Cloned Gene. From the results of single and double digestion experiments using the subcloned fragments, a partial restriction map of the human renin gene was constructed (Fig. 1*B*). The restriction fragments containing an exonic sequence and its surrounding regions, identified by blot hybridization with a nick-translated [^{32}P]cDNA probe, were subjected to nucleotide sequence analysis to determine the exon–intron junctions. The subcloned 0.8-kb fragment that did not hybridize with the cDNA probe was also sequenced to cover the entire coding sequence. As shown in Figs. 1 and 2, human renin gene consists of 10 exons and 9 introns. Central parts of the introns were not sequenced. The exon sequences are identical with those of the corresponding region of the cDNA (3). In accordance with the general G·T-A·G rule, all of the introns begin with the dinucleotide G·T and end with A·G. More extensive consensus sequences have been proposed for the RNA splice sites by analyzing a great number of eukaryotic genes that are transcribed by RNA polymerase II (12, 13). The donor ($\frac{C}{A}$-A-G/G-T-$\frac{A}{G}$-A-T-G-T) and the acceptor [$(\frac{T}{C})_{11}$-N-$\frac{C}{T}$-A-G/G] consensus sequences are reasonably well matched by the intron–exon junctions of the renin gene (Fig. 2).

When compared to mouse renin, human renin contains three extra amino acids (Asp-165, Ser-166, Glu-167; the numbering system begins from Leu-1 of mature renin) at the middle of its polypeptide chain. Surprisingly, these three additional amino acids are nearly fully encoded by a separate exon (as defined in Fig. 2) of only 9 nucleotides.

Comparison between the gene structure and a proposed tertiary structure of human renin (14) revealed that all of the intron–exon boundaries follow the rule, proposed by Craik *et al.* (15), that splice junctions tend to map at protein surfaces.

Flanking Regions. The 3′- and 5′-flanking regions, with a special emphasis on the 5′ portion, were also sequenced since these regions may be important in the regulation of gene expression. The transcription initiation site or the cap site was tentatively assigned to the residue A labeled +1 based on the fact that an "A", which is preceded by a "C" in most cases, is the preferred cap site and assuming that our cDNA covers nearly the full length of its template—namely, human renin mRNA. This assignment is supported by the relative position of the putative regulatory sequences such as the Goldberg–Hogness box ("TATA") found upstream of the 5′ terminus. Inspection of the region upstream from the start site revealed the presence of the promoter sequence T-A-T-A-A-A between positions −29 and −24. At positions −51to −46 there is the sequence C-A-A-T-C-A, which is homologous to the consensus "CAAT" box. In the region immediately upstream from the TATA box, several palindromic sequences are observed (Fig. 2). Further upstream is located a hexanucleotide sequence, T-G-T-T-C-T, which has been suggested to be a preferential binding site for the glucocorticoid receptor by using cloned DNA fragments of mouse mammary tumor virus (11). That the human renin gene is regulated by the steroid hormone is an exciting, but unproven, possibility.

The 3′ end of the human renin gene was determined by comparing the nucleotide sequence of the 3′-untranslated region of the gene with the sequence of cDNA that contains the complete 3′-noncoding region (3). The potential polyadenylylation signal (A-A-T-A-A-A) found in the majority of eukaryotic genes is located 20 nucleotides upstream from the poly(A) addition site or 180 nucleotides downstream from the termination codon. Several short direct repeats were found in the 3′-flanking region by a local homology search using the method of Korn and Queen (16). For example, at 3–9 nucleotides downstream from the termination codon is the se-

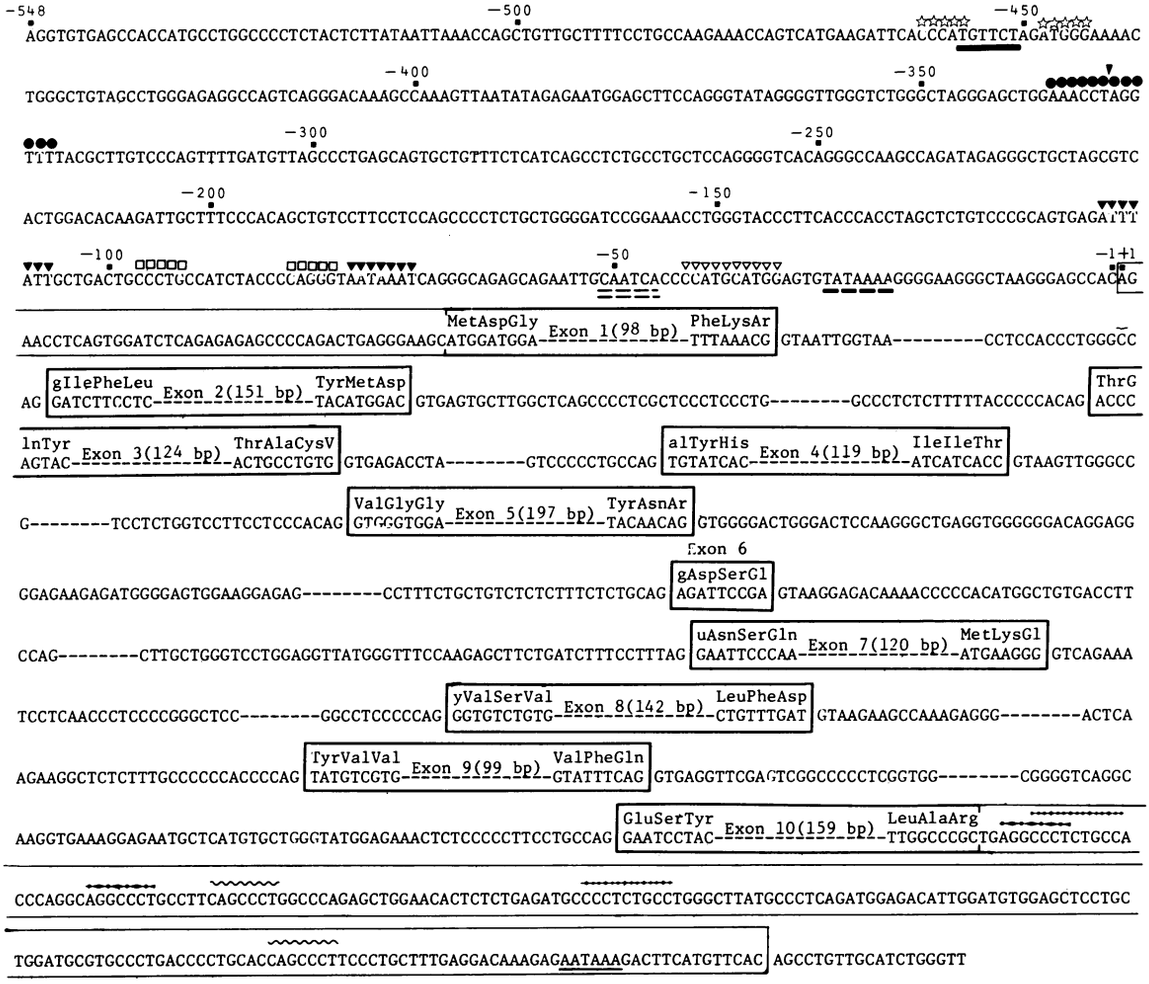

FIG. 2. DNA sequence of the human renin gene. Position +1 corresponds to the probable transcription initiation site. Exons are boxed. The C-A-A-T, T-A-T-A, and A-A-T-A-A-A sequences are underlined. The hexanucleotide T-G-T-T-C-T, which is considered to be the glucocorticoid–receptor-binding sequence (11), is indicated by a bold line. Palindromic sequences in the 5′-flanking region are marked by dots, stars, closed triangles, and open squares. Repeated sequences in the 3′-noncoding region are indicated by various lines. bp, Base pairs.

quence A-G-G-C-C-C-T, which is repeated at nucleotides 23–29. The sequence C-C-C-T-C-T-G-C-C beginning at nucleotide 6 is reiterated around the 72nd nucleotide from the termination codon. The sequence C-A-G-C-C-C-T at nucleotide 35 is also found at nucleotide 151.

DISCUSSION

Blake (17) has recently pointed out that there is a good correlation between the number of exons and the length of polypeptide chain, with 2 or 3 exons per 100 amino acid residues. The human renin gene, having 10 exons per 406 residues, is a representative example of this rule. However, the presence of the unusually small exon, which consists of 9 nucleotides and codes essentially completely for the extra 3 amino acids not found in mouse renin (Fig. 3), is a unique feature of the human renin gene. Cloning and sequence analysis of the renin gene provide (as discussed below) (*i*) a clue to the evolution of the gene family to which it belongs and (*ii*) tools for clarifying the control mechanism of expression of the gene.

In addition to the nucleotide sequences of all 10 exons and their exon–intron junctions, we have sequenced about 500 nucleotides flanking the 5′ end of the human renin gene in the hope of learning about some of the mechanisms involved in the regulation of its expression. In addition to the putative TATA and CAAT boxes, this region is expected to contain unique recognition sequences (to be identified by future studies). The fact that hexanucleotide T-G-T-T-C-T, found ≈450 base pairs upstream from the start of transcription, is identical to the suggested binding site of DNA for the glucocorticoid–receptor complex offers intriguing hints for future exploration. An interesting alternative possibility is that the nucleotide sequence responsible for the binding of the aldosterone–receptor complex might reside in this controlling region or upstream for the following reason. Since high aldosterone levels in the circulation are achieved by the expense of stored renin, it seems possible to speculate that aldosterone can, in turn, trigger the biosynthesis of renin to compensate the decrement of renin stored in the juxtaglomerular cells. The sequence presented here and the cloned 5′-most fragment of the gene containing the promoter and the region immediately upstream would be valuable for exploring the control mechanism of the transcription of the renin gene.

Structural Comparison of Human Renin and Pepsin Genes Suggests That They Are Descendants of a Common Ancestral Gene. The primary and tertiary structures of aspartyl proteinases have been studied extensively (3, 14, 18–29) and there is clear evidence that they have a closely resembling bilobal structure that accommodates the two catalytically important and strictly conserved aspartyl residues in the right position. The hypothesis that the members of the aspartyl proteinase family have arisen from a single ancestral gene

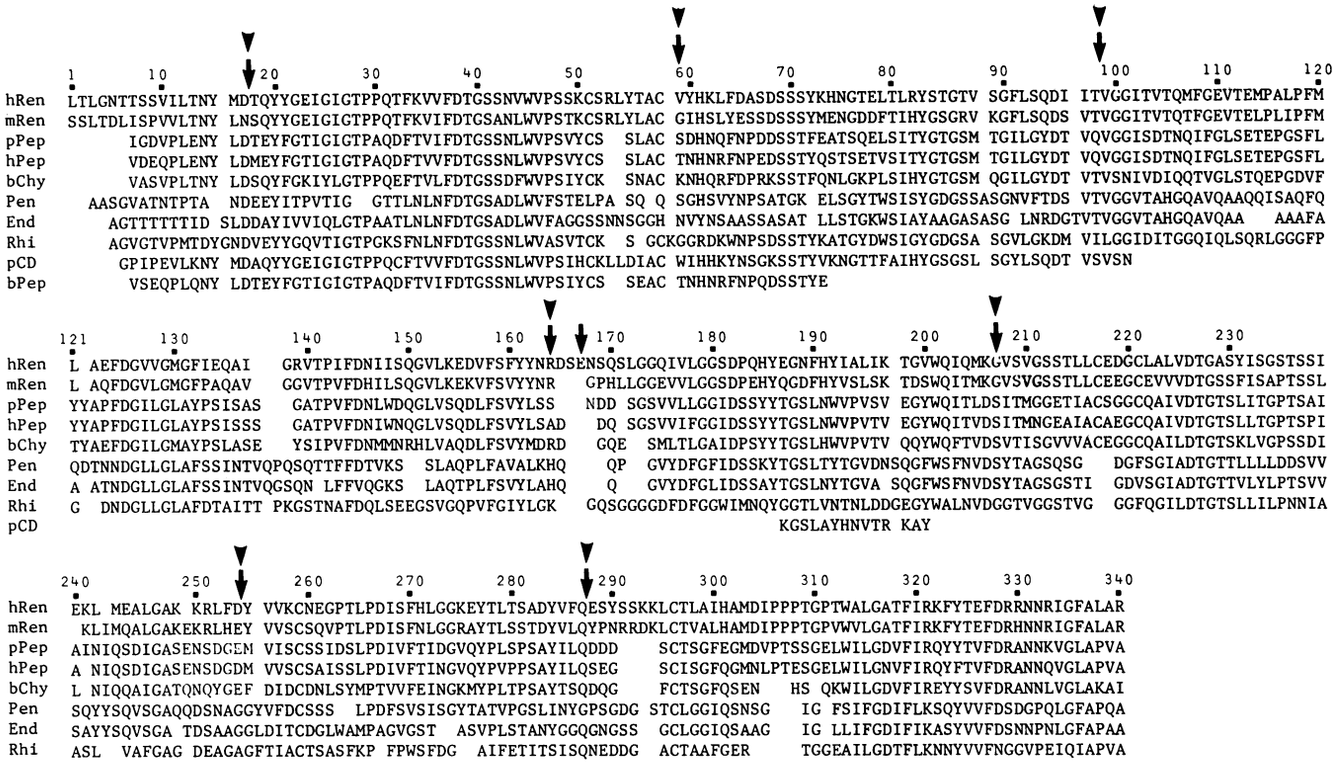

FIG. 3. Alignment of amino acid sequences of aspartyl proteinases. Locations of splice junctions of human renin and pepsin genes are indicated by arrows and arrowheads, respectively. The alignment is based on the primary structural homologies among human renin (hRen) (3), mouse renin (mRen) (23, 24), pig pepsin (pPep) (18, 22), human pepsin (hPep) (29), bovine chymosin (bChy) (18, 22), penicillopepsin (Pen) (25), endothiapepsin (End) (18), rhizopuspepsin (Rhi) (18), pig cathepsin D (pCD) (26, 27), and bovine pepsin (bPep) (18).

by duplication and subsequent divergence has been formulated based on these earlier observations. When the gene structure of human renin determined in the present study is compared with that of human pepsinogen gene reported by Sogawa *et al.* (29), similarities are striking; both genes have 9 exons of similar sizes, with human renin containing an additional exon with an unusually small size. Furthermore, the intron positions of the two genes coincide very well; the locations of all introns (except the first one and those flanking the 9-nucleotide-long exon of human renin gene) show a nearly perfect match (Fig. 3). For example, the introns in the human renin gene interrupt the coding region[§] at codons 33, 83/84, 125, 164/165, 230, 233, 273, 320/321, and 353/354 and those in the pepsinogen gene occur at codons 19, 73/74, 113, 152/153, 219, 258, 306/307, and 339/340 (note that the active site aspartyl residues of the renin and pepsinogen correspond to codons 104 and 292 and codons 94 and 276, respectively, according to this numbering system). These genetic data lend strong support to the above hypothesis.

Association of the Splice Junctions with the Surface Loops That Show Length Variation. Craik *et al.* (30) have recently examined the intron positions of the genes for serine proteinases and dihydrofolate reductases and found that intron–exon junctions frequently occur at surface regions where the length of the polypeptide chains is significantly altered by addition or deletion of a short segment of amino acids. This finding led them to propose a sliding junction model or an exon shuffling model that reasonably explains the observed coincidence of splice junctions with variability in surface loops on related proteins; according to this model, the length variation is caused by junctional sliding and the presence of intronic/exonic structure provides an enhanced possibility for variation of the gene products, thereby speeding the evolution. To see if this attractive hypothesis can be applicable to the members of the aspartyl proteinase family, we compared the splice junctions of the human renin gene with the amino acid sequences, aligned properly to maximize homologies, of microbial and mammalian aspartyl proteinases, including human renin and pepsin (Fig. 3). Among 8 introns interrupting the coding sequence for mature human renin, 7 map at positions where length polymorphism occurs; only 1 intron at Gly-207 exists within a region of strict length conservation (Fig. 3). Thus, the comparison of gene structure and protein structure of the aspartyl proteinase family also supports the proposed mechanism of gene diversification.

We thank Mrs. Yutaka Enokizono, Yoshihisa Sasagawa, Takeshi Chyo, and Keisuke Tanigawa for their technical assistance. This work was supported (in part) by a grant for cancer research from the Ministry of Education, Science and Culture of Japan.

1. Oparil, S. & Haber, E. (1974) *N. Engl. J. Med.* **291,** 389–401.
2. Reid, I. A., Morris, B. J. & Ganong, W. F. (1978) *Annu. Rev. Physiol.* **40,** 377–410.
3. Imai, T., Miyazaki, H., Hirose, S., Hori, H., Hayashi, T., Kageyama, R., Ohkubo, H., Nakanishi, S. & Murakami, K. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 7405–7409.
4. Lawn, R. M., Fritsch, E. F., Parker, R. C., Brake, G. & Maniatis, T. (1978) *Cell* **15,** 1157–1174.
5. Benton, W. D. & Davis, R. W. (1977) *Science* **196,** 180–182.
6. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15,** 678–701.

[§]Note that the numbering system used here starts from the first amino acid of the precursor proteins and, therefore, it is different from that of Fig. 3.

7. Morrison, D. A. (1979) *Methods Enzymol.* **68,** 326–331.
8. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
9. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
10. Soubrier, F., Panthier, J.-J., Corvol, P. & Rougeon, F. (1983) *Nucleic Acids Res.* **11,** 7181–7190.
11. Scheidereit, C., Geisse, S., Westphal, H. M. & Beato, M. (1983) *Nature (London)* **304,** 749–752.
12. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
13. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472.
14. Umeyama, H., Akahane, K., Nakagawa, S., Moriguchi, I., Hirose, S. & Murakami, K. (1984) *Hypertension*, in press.
15. Craik, C. S., Sprang, S., Fletterick, R. & Rutter, W. J. (1982) *Nature (London)* **299,** 180–182.
16. Korn, L. J. & Queen, C. L. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4401–4405.
17. Blake, C. (1983) *Nature (London)* **306,** 535–537.
18. Foltmann, B. & Pederson, V. B. (1977) in *Acid Proteases Structure, Function and Biology*, ed. Tang, J. (Plenum, New York), pp. 3–22.
19. Subramanian, E., Swan, I. D. A., Liu, M., Davies, D. R., Jenkins, J. A., Tickle, I. J. & Blundell, T. L. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 556–559.
20. Tang, J., James, M. N. G., Hsu, I. N., Jenkins, J. A. & Blundell, T. L. (1978) *Nature (London)* **271,** 618–621.
21. Andreeva, N. S. & Gustchina, A. E. (1979) *Biochem. Biophys. Res. Commun.* **87,** 32–42.
22. Tang, J. (1979) *Mol. Cell. Biochem.* **26,** 93–109.
23. Misono, K. S., Chang, J.-J. & Inagami, T. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4858–4862.
24. Panthier, J.-J., Foote, S., Chambraud, B., Strosberg, A. D., Corvol, P. & Rougeon, F. (1982) *Nature (London)* **298,** 90–92.
25. James, M. N. G. & Sielecki, A. R. (1983) *J. Mol. Biol.* **163,** 299–361.
26. Takahashi, T., Schmidt, P. G. & Tang, J. (1983) *J. Biol. Chem.* **258,** 2819–2830.
27. Takahashi, T. & Tang, J. (1983) *J. Biol. Chem.* **258,** 6435–6443.
28. Blundell, T., Sibanda, B. L. & Pearl, L. (1983) *Nature (London)* **304,** 273–275.
29. Sogawa, K., Fujii-Kuriyama, Y., Mizukami, Y., Ichihara, Y. & Takahashi, K. (1983) *J. Biol. Chem.* **258,** 5306–5311.
30. Craik, C. S., Rutter, W. J. & Fletterick, R. (1983) *Science* **220,** 1125–1129.