Primary Structure and Transcriptional Regulation of Rat Pepsinogen C Gene*

(Received for publication, December 22, 1988)

The entire rat pepsinogen C gene has been isolated from a rat genomic library, using the rat pepsinogen C cDNA as a probe. Southern blot analysis showed that there exists at least two rat pepsinogen C genes. The nucleotide sequences of the coding regions and the 5'- and 3'-flanking regions of one of the rat pepsinogen C genes have been determined. This gene is split into 9 exons interrupted by eight intervening sequences. The 5'-flanking region is similar to that of the human pepsinogen C gene, but only the former has the core sequence of the Sp1 binding site.

The amount of transcripts of the rat pepsinogen C genes was found to increase during development, and a similar increase was shown to be induced by injection of hydrocortisone. As a candidate of a factor which regulates the transcription, we found a 25-kDa protein by Southwestern blotting. It binds to a specific site in the 5'-flanking region of the gene only in the presence of Mg2+ ion, and it is present in the nuclear fraction of the gastric mucosa but not of the liver.

Pepsinogen, the inactive precursor of pepsin, is synthesized in the chief cells of gastric glands. It is secreted into the stomach and autocatalytically activated, under acidic conditions, to pepsin, a member of aspartic proteinases. Pepsinogen consists of various isozymogens, of which pepsinogen C is the major component in rat (1). It has been found that the amount of rat pepsinogen C increases during development and that its increase is induced by glucocorticoid hormones during development (2-4). Furthermore, the biosynthesis of pepsinogen is known to be affected by several compounds, such as gastrin (5) and secretin (6), but little has been known about the mechanism of the pepsinogen gene expression.

In 1987, the complete amino acid sequence of the rat pepsinogen C was deduced from the sequence analysis of the cDNA clone (7). Human pepsinogen A gene (8) and pepsinogen C gene (9) were also isolated and characterized in our laboratory. Human pepsinogen genes, however, are not useful for experiments concerning the regulation of gene expression because no cultured cells which synthesize pepsinogen are available. Thus we turned our attention to the rat pepsinogen C gene. As a first step we have isolated the gene and determined its primary structure. Further, we have made it clear that the regulation of the rat pepsinogen gene expression occurs at the transcriptional level during development.

Recently, it has become known that multiple sequence-specific DNA-protein interactions occur in distinct regulatory regions of genes, in most cases in 5'-flanking regions, which largely determine the degree of transcriptional activation of genes (10-12). In the case of pepsinogen genes, however, there have been no reports describing such DNA-binding proteins which specifically bind to the promoter region of the pepsinogen gene. In the present study, we demonstrate, by using the Southwestern blotting procedure, the presence of a DNA-binding protein that specifically binds to the promoter region of the rat pepsinogen C gene.

EXPERIMENTAL PROCEDURES

RESULTS

Isolation of the Rat Pepsinogen C Gene—About 1 × 10^6 plaques from the rat genomic library were screened with the rat pepsinogen C cDNA (pRPC1) (7), and 15 clones of recombinant phage were detected and plaque-purified. These 15 clones were characterized by restriction mapping and Southern blot analysis. These analyses revealed that they were classified into at least two nonoverlapping groups (Fig. 1). One of these groups (group A), consisting of at least seven clones, contained 2.5-, 3.3-, and 4.8-kb EcoRI fragments, each of which hybridized to the cDNA clone. On the other hand, the second group (group B), consisting of at least three clones, contained 2.5-, 3.2-, and 5.3-kb EcoRI fragments, each of which hybridized to the cDNA clone. We found a clone having a different 3'-flanking sequence from that of group A (Fig. 1C). However, since there is an EcoRI site between the 3'-flanking sequence and the pepsinogen gene, it remains to be solved whether this clone is derived from the group B gene or from a third gene or whether it is formed by accidental joining of the pepsinogen gene and an irrelevant DNA fragment during the cloning procedures.

Analysis of Rat Total Genomic DNA—Whether or not these groups were derived from the native rat pepsinogen C genes on the chromosome was investigated by total Southern blot analysis. Total rat DNA was digested with EcoRI, BamHI, 1

1 Portions of this paper (including "Experimental Procedures" and Figs. 1-3, 5, 9, and 10) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.

2 The abbreviations used are: kb, kilobase(s); bp, base pair(s); DTT, dithiothreitol; Hpes, N-2-hydroxyethylpiperezine-N'-2-ethanesulfonic acid; PMSF, phenylmethylsulfonyl fluoride; SSC, saline sodium citrate.

* 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 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J04812.
and HindIII, electrophoresed in 0.9% agarose gel, and characterized by subsequent Southern blot analysis (13) with the rat pepsinogen C cDNA in a stringent condition (washed twice with 0.1×SSC for 30 min at 65 °C) (Fig. 2). EcoRI fragments of approximately 4.8, 5.3, and 3.3 kb and two HindIII fragments of about 6.5 kb strongly hybridized to the cDNA. These results could be reproduced when the DNA was prepared from another rat. This analysis revealed that these bands (4.8- and 5.3-kb EcoRI fragments and two 6.5-kb HindIII fragments) should be derived from the fragments of the two groups (group A and group B). EcoRI fragments of about 2.5-kb length which were expected to be hybridized to the cDNA were not detected, because this fragment has only about a 60-base pair length for the first exon which can be hybridized to the cDNA.

The recombinant phage DNA diluted to various concentrations and the rat total DNA digested with restriction endonucleases were electrophoresed and analyzed by Southern blotting on the same filter. When the intensities of hybridizing bands were compared with each other, it was shown that there were only a few rat pepsinogen C genes in a rat haploid genome (data not shown).

We analyzed the rat total DNA by Southern blotting also with human pepsinogen A cDNA as a probe under nonstringent conditions (washed with 2×SSC instead of 0.1×SSC). No hybridizing bands, however, were detected (data not shown).

**Organization and Nucleotide Sequences of the Rat Pepsinogen C Genes**—The two groups of the recombinant clones were further analyzed for detailed characterization. Each EcoRI fragment hybridizing to the cDNA was subcloned to the pBR322 EcoRI site and analyzed by restriction nuclease mapping. This analysis revealed that these groups were very similar to each other (Fig. 3A and B), namely only the KpnI-NcoI fragment of group A (1.6 kb) was different from that of group B (2.1 kb). The clone shown in Fig. 1C had the same restriction map as those in Fig. 1, A and B.

The nucleotide sequences of the coding region and the 5′- and 3′-flanking regions of one of the rat pepsinogen C genes (group A) were determined by the dideoxy method (14) according to the strategy shown in Fig. 3D. Fig. 4 presents the nucleotide sequences of the exons, their flanking regions, and the 5′- and 3′-flanking regions of the rat pepsinogen C gene. The genomic sequence was identical to that of the cDNA except that one base change occurred in the nontranslational region of the first exon, i.e., at the position 11 nucleotides upstream from the A of the initiation codon ATG, where G is present in the genomic DNA and A in the cDNA. All the exon-intron junction sequences were compatible with the GT/AG rule (15), and the positions of introns were conserved in common with the other mammalian (8, 9, 16-18) and chicken (19) aspartic proteinase genes.

The nucleotide sequences of the first exon and its 5′-flanking region of the other rat pepsinogen C gene (group B) were also determined. The sequences of the first exon and its 5′-flanking region of the group B gene were identical to those of group A gene except for one base change which occurred at the same position as in the group A gene and the cDNA. Thus the sequence of the first exon of the group B gene was identical to that of the corresponding region of the cDNA.

**5′-Flanking Region of the Rat Pepsinogen C Gene**—To determine the starting point of the transcription of the rat pepsinogen C gene, S1 nuclease mapping of the rat pepsinogen C mRNA was carried out. As the probe, the NheI-FokI fragment of the group A gene (Fig. 1), which carried the 5′-flanking region and the first exon, was used. This analysis revealed that the transcriptional initiation site was assigned to the residue "C" at 63 nucleotides upstream from a putative initiation codon (Fig. 5). Twenty-eight nucleotides upstream of the putative cap site, there is the 5′-TATAAA-3′ sequence which may correspond to the Hogness box (20). Further upstream, there was the Sp1 binding site core sequence (GC box = GGGCGG) (10) which could affect the transcriptional regulation of the rat pepsinogen C gene.

**Transcriptional Regulation of the Rat Pepsinogen C Gene during Development**—In the gastric mucosa of rats, an increase in the potential peptic activity of pepsinogen was known to begin to occur around 15 days after birth (2). Whether or not this change in expression of the rat pepsinogen C is regulated at the transcriptional level was investigated by RNA dot hybridization. RNAs at different developmental stages were prepared and analyzed by RNA dot hybridization using the rat pepsinogen C cDNA as a probe. This analysis (Fig. 6) revealed that the mRNA of the rat pepsinogen C began to appear 18 days after birth. Furthermore, when rats were fed hydrocortisone acetate, a kind of glucocorticoid hormone precursor, from day 8 to day 10 after birth, mRNA of the rat pepsinogen C began to increase 2 days after the start of the injection (Fig. 7). These results were in agreement with previous reports (2-4) on the effect of glucocorticoid hormone on the appearance of pepsinogen activity during development of the rat.

**Identification of DNA-binding Proteins by Protein Blotting**—By Southern blotting (21, 22), we looked for sequence-specific DNA-binding proteins which may specifically bind to the promoter region of the rat pepsinogen C gene. Nuclear extracts of tissues were electrophoresed and transferred to filters, and the filters were incubated with a 32P-labeled DNA probe. The probe used in this experiment was the NheI-Stul fragment of the 5′-flanking region which does not contain the putative Hogness box. Fig. 8 is a representative result of analysis of nuclear proteins from the rat gastric mucosa and liver. Lanes 1 and 2 of Fig. 8 show the pattern of protein staining with Coomassie Blue, and lanes 3 and 4 of Fig. 8, the pattern of autoradiographic visualization of the probe binding with nuclear proteins. Despite many proteins present in the gel (as seen in Fig. 8, lanes 1 and 2), only a limited number of them bound to the promoter region of the rat pepsinogen C gene. Furthermore, one of these proteins, which migrated at an apparent molecular weight of about 25,000 in nonreducing condition, existed in the nuclear extract of gastric mucosa but not in that of liver. Thus, this tissue-specific nuclear protein could regulate the tissue-specific transcription of the rat pepsinogen C gene. This protein was detected only when the filter was incubated in the presence of 5 mM Mg2+ ion (data not shown). Whether or not this tissue-specific protein binds specifically with the probe was investigated. The filter was incubated in the binding buffer containing the 32P-labeled probe with or without an excess (about 10 times) of an unlabeled probe or an unlabeled probespecific DNA (Fig. 9). Densitometric analysis of the bands on film showed that the relative intensities of the 25-kDa band to the about 31-kDa band were about 80% in lanes 1 and 2 and about 40% in lane 3. Thus, the intensity of the band of the gastric mucosa-specific DNA-binding protein was markedly reduced only when the filter was incubated with the unlabeled probe. On the other hand, when the filter was incubated with another 32P-labeled probe (HindIII-Stul fragment), which contained sequences further upstream of the promoter region of the rat pepsinogen C gene, the binding of
the tissue-specific protein to the DNA probe was not detected. This gastric mucosa-specific nuclear protein was therefore concluded to be the probe-specific DNA-binding protein.

DISCUSSION

We have isolated at least two of the rat pepsinogen C genes by screening a rat genomic library with the rat pepsinogen C cDNA as a probe. The genomic Southern blotting revealed that the difference between the two genes was derived from the chromosomal sequences and not due to artifacts of cloning. The nucleotide sequences of one of these two genes and part of the other gene were determined. So far as investigated, the two sequences were identical (even in the nontranslational region) to each other except for one base change in the first exon, and the nucleotide sequence of the first exon of the group B gene was identical to that of the cDNA. As all intron-exon junction sequences of the group A gene were compatible with the GT/AG rule (17) and as the translational region of this gene was identical to the cDNA, the group A gene should not be a pseudogene. Although we have not sequenced all the exons of the group B gene, we think it is probably not a pseudogene either. The two genes are very similar to each other and seem to have been duplicated so recently in the course of evolution that the probability of having mutations that inactivate the group B gene is very low. The predicted amino acid sequence is shown below the corresponding nucleotide sequence of the coding regions and the 5′- and 3′-flanking regions of the rat pepsinogen C gene. The predicted amino acid sequence is shown below the corresponding nucleotide sequence of the coding regions and the 5′- and 3′-flanking regions of the rat pepsinogen C gene. The predicted amino acid sequence is shown below the corresponding nucleotide sequence of the coding regions and the 5′- and 3′-flanking regions of the rat pepsinogen C gene.

**Fig. 4.** Nucleotide sequences of the coding regions and the 5′- and 3′-flanking regions of the rat pepsinogen C gene. The predicted amino acid sequence is shown below the corresponding nucleotide sequence of the coding regions and the 5′- and 3′-flanking regions of the rat pepsinogen C gene.
pepsinogen C gene (9) but different from that of human pepsinogen A gene (8) and the human prorenin gene (15, 16). As shown in Fig. 10, the promoter regions of the rat pepsinogen C gene and the human pepsinogen C gene have several blocks of identical sequences. The major species of pepsinogen isozymogens in rat gastric mucosa is pepsinogen C (1), while in human this zymogen is only a minor species (23). The differences between the nucleotide sequences of the promoter regions of the human and rat genes could reflect the expression pattern of the isozymogens. In this context, it is interesting to note that the Sp1 binding site core sequence (GC box) was present in the rat but not in the human gene.

Experiments on the expression of rat pepsinogen C in this work revealed that the pepsinogen C mRNA is developmentally regulated, and the increase of the pepsinogen C mRNA is induced by injection of hydrocortisone acetate. These results correlate well with the changes during development in the amount of pepsinogen estimated by assaying the potential pepsin activity after injection of hydrocortisone reported by Furihata et al. (2). Thus our results indicate that the regulation of the rat pepsinogen C gene expression occurs at least mainly at the transcriptional level. Also it is known that the differentiation of the gastric mucosa is enhanced in the rat that has been treated with hydrocortisone acetate during development (3, 4). Therefore, the induction of pepsinogen with glucocorticoid hormone may be caused not only by the direct effect of the glucocorticoid hormone through its receptor but also by the maturation of the chief cells which synthesize pepsinogen. So far as we investigated, we could not find the sequence of GRE (glucocorticoid-responsive element) (24, 25) in the 5'-flanking region of the rat pepsinogen C gene, but it may exist further upstream. In this experiment, however, the accumulation of the pepsinogen mRNA was detected, so the possibility that this accumulation is due to

---

the slow degradation of the pepsinogen mRNA cannot be ruled out.

Recently, the gel shift assay has been used in most cases to identify specific DNA-binding proteins (26). We tried the same technique in the present study, but no sequence-specific retarded band was observed in the presence of an excess EDTA. In addition, the probe was degraded when Mg\textsuperscript{2+} ion existed. For this reason we used the Southwestern blotting procedure (21, 22) to identify DNA-binding proteins in the nuclear extract which bind to the promoter region of the rat pepsinogen C gene. Indeed the probe-specific DNA-binding proteins existed in the nuclear extract of the gastric mucosa but not in that of liver. This DNA-binding protein is a candidate for a factor that increases the rate of transcription of the rat pepsinogen C gene. The accurate nucleotide sequences to which this protein binds are not clear, so we will try to elucidate those sequences by Southwestern blotting under conditions with the oligonucleotides as competitors. We think it important to make sure that the DNA-binding protein binds specifically to the probe under physiological conditions. In the Southwestern blotting procedure, it is not clear if the DNA-binding protein was renatured during transfer. Hence it is preferable to try the gel shift assay after overcoming the problem of degradation of the probe in the presence of the Mg\textsuperscript{2+} ion, probably by partial purification of the DNA-binding protein by column chromatography. It is also important to identify the cis-acting elements that affect the transcriptional regulation and show that the DNA-binding protein not only binds to the 5′-flanking DNA but also actually enhances the rate of transcription. Since it is difficult to perform the CAT assay (27) due to lack of the appropriate pepsinogen-producing cell culture system (28), the in vitro transcription system (29) of the gastric mucosa may be the best choice for this purpose. The pepsinogen gene is one of the most active genes in the gastric mucosa and is expected to become a good model system to study the regulation of gene expression.

ACKNOWLEDGEMENTS—We are very grateful to Dr. M. Nobuhara (Mochida Pharmaceutical Co., Tokyo) for kindly providing us with the human pepsinogen A cDNA. We wish to thank Dr. M. Ichinose (Faculty of Medicine, the University of Tokyo) for his helpful discussions.

REFERENCES

Continued on next page.
Experimental procedures

Materials - Reagents used were purchased as follows: [α-32P]dCTP (5000Ci/mmol) and [γ-32P]ATP (5000 Ci/mmol) from the Radiochemical Centre Amersham, England; restriction endonucleases, T4 polynucleotide kinase, the Klenow fragment of DNA polymerase I, T4 DNA ligase, and T4 alkali-sensitive phosphatase from Takara Shuzo (Kyoto, Japan); Nippon gene (Tokyo, Japan), and Novagen (Madison, USA); and [γ-32P]ATP (500 Ci/mmol) from Seibutsu (Tokyo, Japan). Recombinant DNAs - All manipulations were conducted in accordance with the Guidelines for Research Involving Recombinant DNA Molecules issued by the Ministry of Education, Science, and Culture of Japan.

Preparation and screening of the Rat Genetic Library - A rat genomic library of Charas 44 bacteriophage was prepared from random partial Eco R1 digests of Wistar-inuclgen rat total DNA. The library was screened with a rat pepcinogen C cDNA (pRCl) (7) as a probe. The hybridization was performed as described before (9).

Restriction Endonuclease Analysis and Southern Blot Analysis - Restriction endonuclease analysis and Southern blot analysis were performed as described before (4), except that the probe was labeled by random priming procedure instead of nick-translation.

DNA Sequencing Methods - DNA sequencing was performed by the dideoxy chain termination method of Sanger et al. (14).

Preparation of RNA - Rat gastric mucosa total RNA was isolated from the tissue of Wistar-Inuclgen rats by the guanidine thiocyanate/citric acid method as described by Chirgwin et al. (31). 

Nuclease mapping - The transcriptional initiation site of the rat pepcinogen C gene was determined by nuclease mapping by the method of Yen and Sharp (32) as modified by Hayano et al. (9).

RNA Hybridization of Rat Pepsinogen C cDNA Development - The Wistar-inuclgen rats were used. Hydroxycytosine (250 μg/kg of body weight) was injected subcutaneously into the back every 12 h from day 8 to 10 after birth. Control animals received injections of saline. Animals were anesthetized with diethyl ether and the stomach was quickly cut out for preparation of RNA. Total RNA of the stomach was isolated as described above, and applied to nitrocellulose filters. The filters were hybridized to the [32P]-labeled rat pepcinogen C cDNA.

Preparation of Nuclear Extracts - The nuclear extracts of the rat liver and gastric mucosa were prepared following the method by Parker and Topol (33) with modification. All steps were performed at 4°C. The tissue was minced and suspended in 5-10 times its volume of buffer A: 15 mM KCl, 10 mM Hepes-NaOH (pH 7.6), 2 mM MgCl2, and 0.1 mM EDTA, and centrifuged at 3,000 x g for 10 min. The supernatant was transferred to polycarbonate centrifuge tubes for Beckman type 50 rotor. The homogenate was centrifuged at 4,000 x g for 10 min. The supernatant was centrifuged at 100,000 x g for 1 h. The resulting supernatant was collected and proteins were precipitated by the addition of 0.55 g ammonium sulfate per ml supernatant. The precipitated proteins were collected by centrifugation at 100,000 x g for 30 min at 4°C. The pellet was resuspended in 5 times its volume of buffer A containing with 1 mM PMSF with a teflon homogenizer. The homogenate was adjusted to isotonicity by the addition of 1/10 volume of buffer B: 1 M KCl, 50 mM Hepes-NaOH (pH 7.6), 2 mM MgCl2, 0.1 mM EDTA, and 1 mM DTT. The homogenate was centrifuged at 4,000 x g for 10 min. The resulting supernatant was centrifuged at 100,000 x g for 1 h and centrifuged for 1.5 h at 300,000 x g. The clear supernatant was collected and proteins were precipitated by the addition of 0.55 g ammonium sulfate per ml supernatant. The precipitated proteins were collected by centrifugation at 100,000 x g for 30 min at 4°C. The pellet was resuspended in 5 times its volume of original tissue volume of buffer C: 10% (vol./vol.) glycerol, 25 mM Hepes-NaOH (pH 7.6), 40 mM KCl, 0.1 mM EDTA, and 1 mM DTT. The protein concentration was determined by the method of Bradford et al. (34).

Detection of DNA-Binding Proteins by Protein Blotting - Nuclear extracts were mixed with an equal volume of the sample buffer (5% SDS, 3 M Tris-HCl (pH 6.8), 30% (vol./vol.) glycerol, and 0.05% bromphenol blue). Samples were not boiled before electrophoresis. SDS-polyacrylamide gel electrophoresis was carried out as described by Laemmli (35). The proteins were transferred from the gel to Durapore filter (Millipore Co., Ltd) electrophoretically in 25 mM Tris/190 mM glycine. The blots were stored at 4°C in the blotting buffer. Binding of labeled DNA to the blots was performed as described by Misamatsu et al. (31) with a slight modification. The filter strips were covered with 5% (wt./vol.) non-fat dry milk in 10 mM Hepes-NaOH (pH 7.4), and incubated for 1 h at room temperature. The filters were incubated for 3 h at room temperature in the binding buffer (10 mM Hepes-NaOH (pH 7.6), 10 mM MgCl2, 1 mM EDTA, 250 μg/ml salmon sperm DNA, and 0.2% non-fat dry milk) containing 5 x 10^-6 cpml/m of 5'-end labeled DNA. The filters were then washed in two changes of the binding buffer for 30 min and exposed to Kodak X-OMAT film with an intensifying screen at -70°C. The relative intensity of the bands on film was analyzed by a densitometer (Fuji Laszyme).
Rat Pepsinogen C Gene: Structure and Its Regulation

**Fig. 5.** SI nuclease mapping analysis of the rat pepsinogen C mRNA.

To determine the transcriptional initiation site of the rat pepsinogen C gene, the ShAl-fgl fragment which contains part of the 5'-flanking region and the first exon was annealed to the rat gastric mucosa total RNA and treated with SI nuclease. The products were electrophoresed with the degradation products of the same fragment by the Maxam-Gilbert sequencing method.

Lanes 1-3: These lanes show the fragments produced by SI nuclease treatment, using 50, 100, and 200 units of SI nuclease, respectively.

Lanes 4-7: These lanes show the degradation fragments obtained by the Maxam-Gilbert sequencing method.

**Fig. 9.** Probe specificity of the DNA-binding proteins.

To determine whether the gastric mucosa-specific DNA-binding proteins are probe-specific or not, the South-Western blotting analysis was performed with or without the probe-specific and nonspecific DNA as competitors.

Lane 1: The result without competitors.

Lanes 2 and 3: The result with addition of unlabeled probe-specific DNA (as about 250 bp HpaIII - ShAl fragment just upstream of the probe) (lane 2) or the unlabeled probe (lane 3) as competitors.