Structural Organization and the Transcription Initiation Site of the Human Hepatocyte Growth Factor Gene[†,‡]

Keiji Miyazawa, Akiko Kitamura, and Naomi Kitamura*
Institute for Liver Research, Kansai Medical University, Moriguchi, Osaka 570, Japan
Received April 26, 1991; Revised Manuscript Received July 3, 1991

ABSTRACT: The structural organization of the gene coding for human hepatocyte growth factor (hHGF) has been determined by seven overlapping λ phage genomic clones including three clones that were previously characterized. The gene for hHGF spans about 70 kbp of DNA and consists of 18 exons separated by 17 introns. The coding sequence of hHGF consists of multiple putative domains that are homologous to those observed in plasminogen. These regions were found as separate exons in the gene, and the exon–intron arrangement was similar to that of plasminogen. These results suggest that the genes for hHGF and plasminogen have arisen through gene duplication events from an ancestral gene. The major transcription initiation site of the hHGF gene is located 76 bp upstream of the translational start codon as judged by S1 nuclease mapping and primer extension analyses. A TATA-like element was found 33 nucleotides upstream of the transcription initiation site. Two sequence elements, an interleukin 6 response element (CTGGGA) and a potential binding site for NF-IL6 (TGAGGAAAG), are located near the transcription initiation site. These sequence elements might be involved in the regulation of HGF gene expression.

Human hepatocyte growth factor (hHGF), which was first isolated from the plasma of patients with fulminant hepatic failure (Gohda et al., 1988), is a potent growth stimulator of both human and rat hepatocytes in primary culture (Strain et al., 1991). Human HGF may function as a regulator of liver regeneration because its levels in plasma or sera are higher in patients with liver diseases (Tsubouchi et al., 1989, 1991). Recently, hHGF has been shown to have mitogenic activity on a variety of cells including melanocytes and endothelial cells as well as cells of epithelial origin (Rubin et al., 1991; Kan et al., 1991; Igawa et al., 1991). Human HGF is a heterodimeric glycoprotein consisting of a heavy chain and a light chain with molecular weights of about 65 000 and 35 000, respectively. These chains are linked together by a disulfide bond (Gohda et al., 1988).

The primary structures of human and rat HGF have been established by cDNA cloning and sequencing (Miyazawa et al., 1989; Nakamura et al., 1989; Tashiro et al., 1990; Okajima et al., 1990). Human HGF is synthesized as a precursor form of 728 amino acids. The modified N-terminal amino acid of the heavy chain of hHGF is pyroglutamate derived from glutamine at the 32nd residue from the initiation methionine (Yoshiyama et al., 1991). Thus, the N-terminal 31 amino acid residues of the precursor, which contain many hydrophobic amino acids, probably function as a signal peptide. The heavy chain (463 amino acids) and the light chain (234 amino acids) are derived from the N- and C-terminal parts of the precursor, respectively. The sequence of hHGF is homologous to that of human plasminogen (about 38% identical) (Nakamura et al., 1989). The heavy chain contains four kringle structures, and the light chain is homologous to the catalytic domain of serine proteases. Analysis of the genomic DNA encoding the kringle region of hHGF (Miyazawa et al., 1991) showed that each of the four kringles is encoded by two exons and the exon–intron arrangement is very similar to that of human plasminogen (Petersen et al., 1990).

HGF mRNA is expressed in human placenta (Miyazawa et al., 1989), liver (Nakamura et al., 1989), and leukocytes (Seki et al., 1990) as well as in a variety of rat tissues (Tashiro et al., 1990; Okajima et al., 1990). The mRNA is induced in the rat liver and spleen during liver regeneration following hepatic injury caused by administration of hepatotoxins (Kinoshita et al., 1989; Okajima et al., 1990). In contrast, no obvious mRNA induction has been observed in the rat kidney even though the kidney is one of the major sites of the mRNA production (Okajima et al., 1990). The mechanisms responsible for HGF mRNA induction in these tissues are not understood. Previous studies on many genes have shown that sequences upstream of the 5′ end play an important role in the regulation of gene expression. This report thus concerns the identification of the transcription initiation site of the hHGF gene and characterization of the 5′ flanking region, which may have the regulatory elements. By analyzing the entire hHGF gene, we have also elucidated the relationship between the exon–intron arrangement and protein structural domains and have compared it with that of human plasminogen.

Experimental Procedures

Materials. Reagents were obtained as follows: [α-^{32}P]-dCTP from Amersham; [γ-^{32}P]ATP from New England Nuclear; oligo(dT)–cellulose (type 7) and random primer [pd(N)$_6$] from Pharmacia; the Klenow fragment of *Escherichia coli* DNA polymerase I from Boehringer Mannheim; restriction endonucleases from Takara Shuzo Co. and Toyobo Co.; calf intestine phosphatase and T4 polynucleotide kinase from Toyobo Co.; S1 nuclease from Takara shuzo Co.; and reverse transcriptase from Life Science Inc.

RNA Preparation. PLC/PRF/5 human hepatoma cells and MRC-5 human embryonic lung fibroblasts were cultured in Dulbecco's-modified Eagle's medium supplemented with 10% fetal calf serum. Total RNA was extracted from the cultured cells and a human placenta by the guanidine thiocyanate

[†] This work was supported in part by research grants from the Ministry of Education, Science, and Culture of Japan and the Japan Private School Promotion Foundation.
[‡] The genetic sequence reported in this paper has been submitted to GenBank under Accession Number J05342.
* To whom correspondence should be addressed.

0006-2960/91/0430-9170$02.50/0

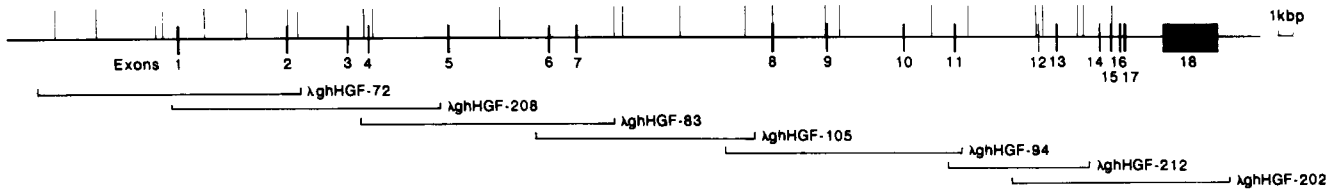

FIGURE 1: *Eco*RI restriction map and location of the exons in the hHGF gene. The seven overlapping λ phage clones were aligned according to restriction mapping and sequence determination. The 18 exons are shown by boxes. The *Eco*RI restriction sites are shown by vertical bars. λ ghHGF-83, -105, and -94 have been previously characterized and published by Miyazawa et al. (1991).

method (Chirgwin et al., 1979). Poly(A) RNA was isolated by subjecting the total RNA to oligo(dT)–cellulose chromatography (Aviv & Leder, 1972).

Isolation and Characterization of Genomic Clones. The human genomic DNA library obtained from Clontech is a collection of recombinant phage carrying human placenta DNA fragments generated by partial digestion with *Sau*3AI and joined to EMBL3 arms. The phage were screened by hybridization in situ (Benton & Davis, 1977) with appropriate restriction fragments derived from the hHGF cDNA clone. Hybridization-positive phage clones were isolated by repeated plaque purification. The inserts of the clones were analyzed by restriction mapping and Southern blotting (Southern, 1975). Appropriate DNA fragments were subcloned into pUC or M13 vectors and sequenced by the chain termination method (Sanger et al., 1977) with use of a kit from United States Biochemicals.

Isolation and Characterization of the cDNA Clone. A cDNA library was constructed by the method of Okayama and Berg (1983), with 12.6 μg of poly(A) RNA and 4.2 μg of the vector–primer DNA. *E. coli* HB101 was used for transformation, and ampicillin-resistant transformants were screened by hybridization with the 834-bp *Eco*RI fragment derived from the hHGF cDNA as a probe. A positive colony was replated and rescreened until all colonies hybridized to the probe. The insert of the plasmid was characterized by restriction mapping and sequencing.

Southern Blot Hybridization Analysis of Total Cellular DNA. Human leukocyte cellular DNA was digested with *Eco*RI, *Bam*HI, or *Kpn*I, electrophoresed on a 0.6% agarose gel, and transferred to a nylon membrane (Biodyne). The filter was hybridized for 16 h at 60 °C with the ^{32}P-labeled cDNA fragments as probes in a solution containing 10 × Denhardt's solution (0.2% bovine serum albumin, 0.2% poly(vinylpyrrolidone), and 0.2% Ficoll), 6 × SSC, 50 mM Tris-HCl (pH 7.5), 1% SDS, and 50 μg/mL sonicated salmon testis DNA. The filter was washed with 2 × SSC containing 0.1% SDS at 60 °C. The probe was labeled with ^{32}P by the random labeling method (Feinberg & Vogelstein, 1983).

Northern Blot Hybridization Analysis. Poly(A) RNA was denatured with 2.2 M formaldehyde/50% (v/v) formamide, electrophoresed on a 1% agarose gel containing 2.2 M formaldehyde (Lehrach et al., 1977), and transferred to a nylon membrane (Biodyne). The filter was hybridized for 16 h at 42 °C with the ^{32}P-labeled probe in a solution containing 50% formamide, 5 × Denhardt's solution (0.1% bovine serum albumin, 0.1% poly(vinylpyrrolidone), and 0.1% Ficoll), 5 × SSC, 50 mM sodium phosphate (pH 6.5), 0.1% SDS, and 100 μg/mL sonicated salmon testis DNA (Alwine et al., 1977). The filter was washed with 1 × SSC containing 0.1% SDS at 50 °C.

S1 Nuclease Mapping and Primer Extension Analyses. The DNA probe used for the S1 nuclease mapping analysis was prepared as follows. A 124-bp *Bgl*II–*Bam*HI fragment was excised from the genomic clone and inserted into pUC18 vector at the *Bam*HI site. The plasmid was digested with *Bam*HI, dephosphorylated with calf intestine phosphatase, and then labeled with ^{32}P at the 5′ end. The labeled fragment was digested with *Eco*RI, and the 145-bp fragment containing the 124-bp genomic and the 21-bp vector sequences was isolated by electrophoresis on a 7 M urea/6% polyacrylamide gel. The 145-bp fragnment (1×10^5 cpm) was hybridized to 20 μg of poly(A) RNA in a solution containing 80% (v/v) formamide, 0.4 M NaCl, 1 mM EDTA, and 40 mM Pipes (pH 6.4) at 42 °C for 12 h, and the DNA–RNA hybrid was digested with S1 nuclease (8 units/μg of RNA) at 13 °C for 90 min (Berk & Sharp, 1977; Weaver & Weissman, 1979). For the primer–extension experiment, an oligonucleotide was chemically synthesized with the use of an Applied Biosystems 380B DNA synthesizer and labeled with ^{32}P at the 5′ end. The labeled oligonucleotide (1×10^5 cpm) was hybridized to 20 μg of poly(A) RNA in the solution described for S1 nuclease analysis at 32 °C for 12 h. The DNA–RNA hybrid was then reverse transcribed (Sambrook et al., 1989). The S1 digestion products and the primer-extended cDNAs were electrophoresed in parallel with the sequence ladder on 7 M urea/6% polyacrylamide gel.

Results

Isolation of Genomic Clones for hHGF Gene. We have previously described the isolation and characterization of the genomic clones (λ ghHGF-83, -105, and -94) coding for the four kringle structure regions of hHGF (Miyazawa et al., 1991). In a previous screening of the *Alu*I/*Hae*III genomic library of Lawn et al. (1978), we isolated another clone (λ ghHGF-72) that covered the N-terminal region of hHGF but we could not isolate the other portions of the gene. We thus screened another genomic library by appropriate restriction fragments derived from the hHGF cDNA and isolated three more clones (λ ghHGF-208, -212, and -202). As shown in Figure 1, a total of seven clones containing the entire hHGF gene were arranged into a continuous genomic DNA of about 70 kbp.

Southern blot analysis of human leukocyte cellular DNA exhibited hybridization-positive fragments as expected from the restriction map, and no other hybridization-positive bands were detected (data not shown). These results indicate that the cloned DNA segments containing the hHGF gene retain the 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 were subcloned into the pUC vector, and DNA fragments containing exons and their surrounding regions were sequenced.

Isolation of cDNA for the 3′ Noncoding Region of hHGF mRNA. Placenta hHGF is mainly encoded by a 6.3-kb mRNA (Miyazawa et al., 1991). The cDNA clone for hHGF that we previously isolated lacked almost all of the 3′ noncoding region of the mRNA (Miyazawa et al., 1989). Thus, we isolated the cDNA covering the 3′ noncoding region to map

Exon	Exon Size (bp)	3'Exon Junction	Intron	5'Exon Junction	Exon	Junction type
1	165	*Tyr Ala G*[30] TAT GCA G	GTTAGTTCCC---(A,~7.1kbp)---TATTTAATAG	[30]*lu Gly Gln* AG GGA CAA	2	I
2	166	*Thr Cys Ly*[85] ACT TGC AA	GTAAGTTACT---(B,~3.9kbp)---TTTTTTTCAG	[85]*s Ala Phe* G GCT TTT	3	II
3	113	*Asn Lys A*[123] AAC AAA G	GTAACTGACT---(C,~1.4kbp)---TCTTCAATAG	[123]*sp Tyr Ile* AC TAC ATT	4 (K1a)	I
4 (K1a)	115	*Glu His Se*[161] GAA CAC AG	GTAAGAACAG---(D,~5.1kbp)---TTAAAACTAG	[161]*r Phe Leu* C TTT TTG	5 (K1b)	II
5 (K1b)	143	*Ser Glu V*[209] TCA GAA G	GTAAATAAAC---(E,~6.6kbp)---ATTTTATTAG	[209]*al Glu Cys* TT GAA TGC	6 (K2a)	I
6 (K2a)	121	*Pro Glu Ar*[249] CCT GAA AG	GTAAAATATT---(F,~1.6kbp)---TGGTTTGCAG	[249]*g Tyr Pro* A TAT CCC	7 (K2b)	II
7 (K2b)	119	*Thr Cys A*[289] ACA TGC G	GTAAGTGAAG---(G,~13.0kbp)---TTGCCAATAG	[289]*la Asp Asn* CT GAC AAT	8 (K3a)	I
8 (K3a)	175	*Lys Cys Ly*[347] AAG TGC AA	GTGAGTAAAG---(H,~3.4kbp)---TTCTCAATAG	[347]*s Asp Leu* G GAC CTA	9 (K3b)	II
9 (K3b)	128	*Gly Gln A*[390] GGA CAA G	GTAATAGCTG---(I,~5.0kbp)---TTATGTCTAG	[390]*sp Cys Tyr* AT TGT TAT	10 (K4a)	I
10 (K4a)	103	*Leu His Ar*[424] TTA CAT CG	GTGTGTAAAT---(J,~3.4kbp)---GTTTCCGCAG	[424]*g His Ile* T CAT ATC	11 (K4b)	II
11 (K4b)	134	*Ser Arg C*[469] TCT CGT T	GTAAGTACAG---(K,~5.4kbp)---ATTTTTACAG	[469]*ys Glu Gly* GT GAA GGT	12	I
12	39	*Leu Asp H*[482] TTA GAC C	GTAAGTAATA---(L,~1.2kbp)---CCTGTTTCAG	[482]*is Pro Val* AT CCC GTA	13	I
13	97	*Arg Tyr Ar*[514] AGA TAC AG	GTAATTATTA---(M,~2.8kbp)---TTTTTCACAG	[514]*g Asn Lys* A AAT AAA	14	II
14	75	*Pro Ser Ar*[539] CCT TCT CG	GTAAAGTGTT---(N,~0.7kbp)---TTCAAAACAG	[539]*g Asp Leu* A GAC TTG	15	II
15	141	*Leu Ala Ar*[586] CTT GCC AG	GTTAGTTACT---(O,~0.6kbp)---GTGTTTTCAG	[586]*g Pro Ala* G CCT GCT	16	II
16	107	*Thr Gly L*[622] ACT GGA T	GTAAGCTAGT---(P,~0.2kbp)---ATCAATCTAG	[622]*eu Ile Asn* TG ATC AAC	17	I
17	146	*Pro Cys Glu*[670] CCA TGT GAG	GTAAAAAGGA---(Q,~2.7kbp)---TTACTCCTAG	[671]*Gly Asp Tyr* GGG GAT TAT	18	0
18	~3600					

FIGURE 2: Nucleotide sequences at the splice junctions. The exon-intron boundaries are positioned according to the GT–AG rule for splice junctions. K1a–K4a: N-terminal half of the kringle. K1b–K4b: C-terminal half of the kringle. Junction type refers to the placement of intron sequence in the codon triplet. A type 0 occurs between codons, a type I occurs between the first and second base, and a type II occurs between the second and third base of the codon (Sharp, 1981). The amino acid residues are numbered in Miyazawa et al. (1989). The sequences at the splice junctions of exons 4–11 have been previously published by Miyazawa et al. (1991).

the mRNA sequence in the genomic sequence. A human placenta cDNA library was constructed by the method of Okayama and Berg and screened with the 834-bp *Eco*RI fragment corresponding to the 3′ end of the previously isolated hHGF cDNA. A positive clone (phHGF-26) was isolated from about 2×10^5 colonies. The isolated clone carried the 4.3-kbp insert. The overlapping sequence of the clone and the previously isolated clone was about 1.2 kbp long. The entire cDNA sequence that corresponds to the 6.3-kb mRNA was thus obtained by combining these clones.

Structural Organization of the hHGF Gene. Comparison of the genomic DNA sequence with the cDNA sequence indicated that the gene consisted of 18 exons (exon 1–18) interrupted by 17 introns (A–Q). The sequence of the exon–intron boundaries and the sizes of the exons and introns are summarized in Figure 2. The sequence of all the intron–exon splice junctions agreed with the GT–AG rule (Mount, 1982). All introns interrupt the protein-coding regions of the gene. The first exon contained the 5′ noncoding region and coded for a putative signal peptide. Intron A was located between the nucleotide sequence coding for the signal peptide and the N-terminal region of the heavy chain. The N-terminal region of the heavy chain that precedes the kringle structures was encoded by exons 2 and 3, which were interrupted by intron B. As previously described, each of the four kringles was coded by two separate exons (exon 4–11) (Miyazawa et al., 1991). The site for the proteolytic processing to form the heavy and light chains was located in exon 13. The serine protease like sequence of the light chain consisted of five exons (exon 14–18). Exon 18 was the largest and encoded 58 amino acid residues of the light chain and the 3′ noncoding region of the mRNA.

Human HGF mRNA Production in MRC-5 Human Embryonic Fibroblasts. A cell culture system is simple and useful to study the regulation of gene expression. We thus examined hHGF mRNA production in several human cell lines by Nothern blot analysis. As shown in Figure 3, a significant amount of hHGF mRNA was produced in MRC-5 cells (lane 2) but not in PLC/PRF/5 human hepatoma cells (lane 3) or in U937 human histiocytic lymphoma cells (data not shown). Four mRNAs of 6.3, 3.1, 2.3, and 1.5 kb observed in human placenta (lane 1) were also detected in MRC-5 cells. We have previously shown that the 1.5-kb mRNA was generated in human placenta by alternative RNA processing events from

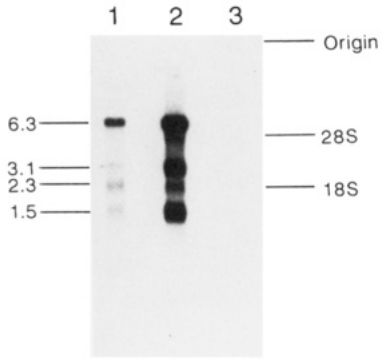

FIGURE 3: Northern blot analyses of poly(A) RNA from MRC-5 and PLC/PRF/5 cells. Five micrograms of poly(A) RNA from MRC-5 (lane 2) and PLC/PRF/5 (lane 3) cells was analyzed. The positive control was 5 μg of human placenta poly(A) RNA (lane 1). The hybridization probe was the 870-bp *Bam*HI–*Sca*I fragment from the hHGF cDNA. Numbers on the left side indicate derived estimated sizes (in kilobases) of the mRNA. The size markers used were human ribosomal RNA.

the hHGF gene (Miyazawa et al., 1991). The 1.5-kb mRNA production in MRC-5 cells indicates that this alternative RNA processing occurred also in the cell culture system.

Transcription Initiation Site of the hHGF Gene. The site of transcription initiation was determined by S1 nuclease mapping and primer-extension analyses. Probes for both experiments were labeled at the same site to facilitate direct comparison of the products obtained by the two techniques. For S1 nuclease mapping analysis, a probe containing the 124-bp *Bgl*II–*Bam*HI genomic DNA fragment was labeled at the 5′ end with ^{32}P, and the antimessage strand was isolated by polyacrylamide gel electrophoresis after denaturation. The isolated DNA was hybridized to poly(A) RNA, and the RNA–DNA hybrids were then analyzed by S1 nuclease mapping. The primer used in the primer-extension experiment was a 27mer oligonucleotide whose sequence is indicated in Figure 5A. The oligonucleotide was labeled at the 5′ end with ^{32}P. The labeled primer was hybridized to poly(A) RNA and reverse transcribed. The products of both reactions were run on a DNA sequence gel in parallel with DNA sequence ladders generated from the same site. Figure 4 shows an autoradiograph of the S1 nuclease mapping and primer-extension analyses. To relate the start site of the mRNA to the DNA sequence, the nucleotide sequence of the exon sequence and its 5′-flanking region was determined, and the result of the sequence determination is presented in Figure 5A.

The S1 nuclease analysis of human placenta RNA generated two major protected DNA fragments of 54 and 55 nucleotides and a minor fragment of 128 nucleotides (lane 2). The primer extension analysis of human placenta RNA produced a DNA fragment that corresponded to the 55-nucleotide S1 nuclease protected fragment (lane 5). The S1 nuclease and primer-extension analyses of RNA derived from MRC-5 cells generated essentially the same DNA fragments observed with human placenta RNA although some additional fragments were detected (lanes 3 and 6). No products were detected with either technique in control reactions containing RNA derived from PLC/PRF/5 cells (lanes 4 and 7). Because a cap structure at the 5′ end 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 (Weaver & Weissman, 1979), the transcription initiation site was assigned by the primer-extension analyses. Accordingly, the major transcription initiation site of the hHGF gene was mapped at 76 nucleotides upstream of the translational initiation codon in both human placenta and MRC-5 cells.

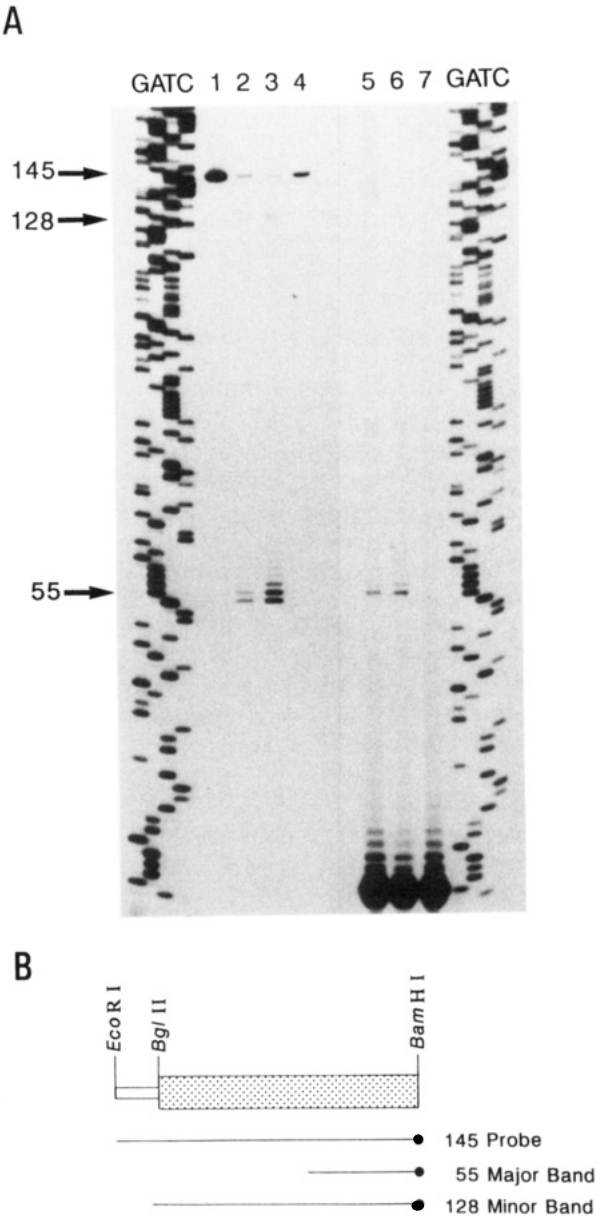

FIGURE 4: Identification of the transcription initiation site of the hHGF gene by S1 nuclease mapping and primer extension analyses. In the autoradiograph (A), the following DNAs were electrophoresed: the probe for S1 nuclease mapping (lane 1), the S1 digestion products using poly(A) RNA from human placenta (lane 2), MRC-5 cells (lane 3), PLC/PRF/5 cells (lane 4), and primer-extended cDNAs using poly(A) RNA from human placenta (lane 5), MRC-5 cells (lane 6), and PLC/PRF/5 cells (lane 7), and the sequence ladders (the four right and four left lanes) as a chain length marker. In (B), the diagram of the DNA probe for S1 nuclease mapping is indicated. The dotted and white boxes are derived from the sequences of the genomic DNA and the vector DNA, respectively. Under the diagram, the S1 nuclease protected fragments are indicated.

The 5′ end of the hHGF cDNA (λ hHGF-21) obtained from a human placenta cDNA library corresponded to 98 nucleotides upstream of the translational initiation codon (Miyazawa et al., 1989). It is located upstream of the major transcription initiation site. S1 nuclease analysis of human placenta RNA generated the minor protected fragment of 128 nucleotides that corresponded to the *Bgl*II site. The result indicated that the synthesis of the minor fraction of hHGF mRNA was initiated at the site upstream of the *Bgl*II site. The cDNA clone obtained was probably transcribed from the minor mRNA. The estimation of the S1 protected fragments by scanning densitometry of the autoradiogram indicated that

FIGURE 5: Nucleotide sequence around the 5′- and 3′-terminal regions of the hHGF gene. (A) The 5′-terminal region. The nucleotide residues are numbered, beginning with the major transcription initiation site, and the nucleotides on the 5′ side of residue 1 are indicated by negative numbers. The restriction sites for the S1 mapping probe (*Bgl*II–*Bam*HI fragment) are indicated. The sequence for which the complementary sequence was chemically synthesized for the primer extension analysis is shown by a thick bar. The consensus sequence for IL6 response element (IL6-RE) (Hattori et al., 1990) and NF-IL6 binding element (NF-IL6) (Akira et al., 1990), and the TATA-like box are underlined. Three palindromes (P1–P3) are indicated by arrows. (B) The 3′-terminal region. The polyadenylation site of hHGF mRNA is indicated by an arrow. The potential polyadenylation signal sequence is underlined.

the minor fraction of hHGF mRNA was 5–10% of the total in both human placenta and MRC-5 cells.

Nucleotide Sequences of the 5′- and 3′-Flanking Regions of the hHGF Gene. In the 5′-flanking region of the hHGF gene (Figure 5A), no consensus TATA, CAAT, or GC boxes, which are sequence elements common to a variety of regulated eukaryotic gene promoters, were found proximal to the transcription initiation site. However, an AATAAAA sequence was found 33 nucleotides upstream from the major transcription initiation site. The sequence differs from the typical TATA box but may function as a TATA-like element. A sequence element of CTGGGA, previously identified as a sequence conserved among several acute-phase genes (Fey et al., 1989), was found at positions −304 to −299 from the major transcription initiation site. This hexanucleotide has been defined as the interleukin 6 (IL6) response element of the rat α_2-macroglobulin gene (Hattori et al., 1990). Another sequence element, TGAGGAAAG, which matched the NF-IL6 (a nuclear factor for IL6 expression) binding consensus sequence T(T/G)NNGNAA(T/G) (Akira et al., 1990), was located at positions −151 to −143. Other characteristic features of the sequences were three palindromes (P1–P3) with six or seven continuously matched nucleotides that were located near the major transcription initiation site.

In the 3′ end of the hHGF gene (Figure 5B), a potential polyadenylation signal (AATAAA) (Proudfoot & Brownlee, 1976) was found 11 nucleotides upstream from the polyadenylation site. Consensus sequences of CAYTG (Berget, 1984) and YGTGTTYY (McLauchlan et al., 1985), which are required for efficient formation of the 3′ terminus of mRNA, were not present within 85 nucleotides downstream from the polyadenylation site.

Discussion

This report describes the structural organization and the transcription initiation site of the hHGF gene. The gene is approximately 70 kbp in length and consists of 17 introns and 18 exons. Two different cDNAs for hHGF have been independently isolated from human placenta (Miyazawa et al., 1989) and human liver cDNA libraries (Nakamura et al., 1989). The deduced amino acid sequences showed 14 single amino acid substitutions scattered throughout the molecule. Because no more than one hHGF gene was identified in the genomic DNA, these differences may be the results of genetic polymorphisms in the hHGF gene. However, the protein-coding sequence of the exons of the genomic clones completely matched that of the human placenta cDNA. Further analyses of genomic DNAs from many individuals will be required to demonstrate genetic polymorphisms in the hHGF gene.

The major transcription initiation site of the hHGF gene is located 76 nucleotides upstream of the translational initiation codon in both placenta and MRC-5 cells. Since the HGF gene is induced during liver regeneration after hepatic injury (Kinoshita et al., 1989; Okajima et al., 1990), regulatory

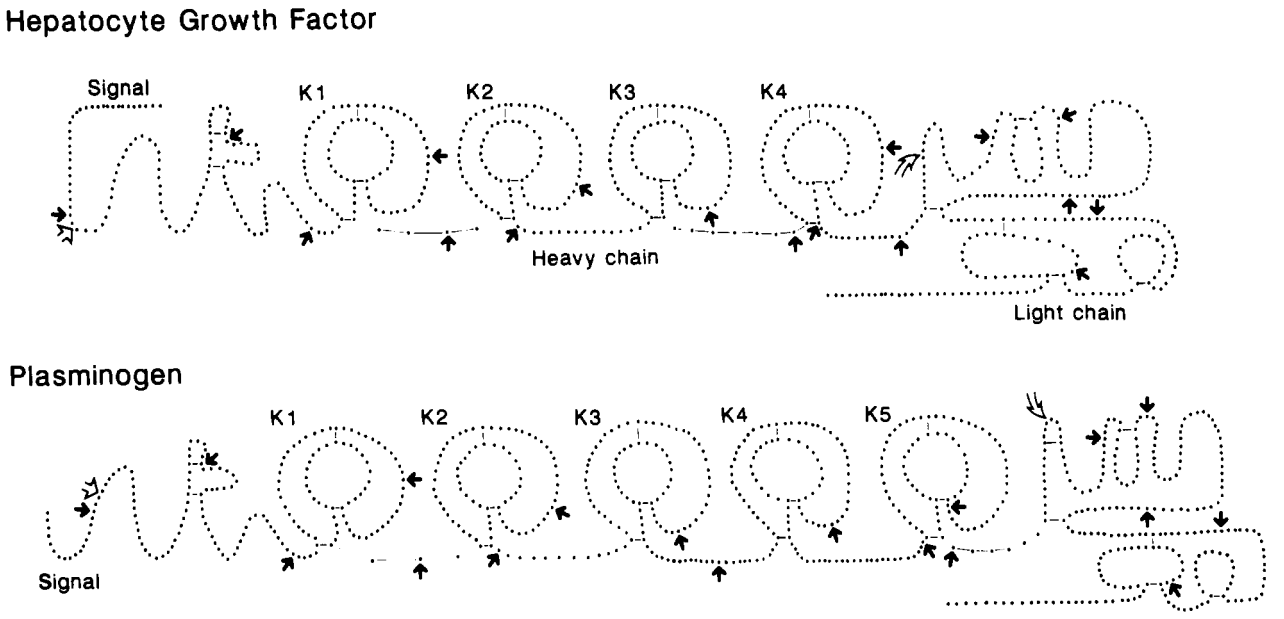

FIGURE 6: Location of the introns in the genes for hHGF and human plasminogen. The positions of introns are indicated by solid arrows. The cleavage sites between the putative signal peptide and the heavy chain are shown by open straight arrows. The cleavage sites between the heavy and light chains are shown by open curved arrows. K1–K5 refer to kringles 1–5. Data for human plasminogen are taken from Petersen et al. (1990).

elements may exist in the promoter region of the gene. In this connection, two interesting sequence elements were found near the transcription initiation site. One was the hexanucleotide CTGGGA, which was found in the promoter regions of several acute-phase genes including rat α_2-macroglobulin (Tsuchiya et al., 1987), rat α_1-acid glycoprotein (Prowse & Baumann, 1988), and rat fibrinogen genes (Fowlkes et al., 1984). These genes are induced by IL6. Recently, Hattori et al. have defined the sequence as an interleukin 6 response element (IL6-RE) and have observed an IL6-induced characteristic protein–DNA complex between the element and nuclear proteins (Hattori et al., 1990). Thus, IL6-RE is likely important in mediating IL6 activation of those acute-phase genes. The other sequence element was the nonanucleotide TGAGGAAAG, which matched the NF-IL6 binding consensus sequence. NF-IL6 was first identified as a nuclear factor that bound to an IL1 response element in the IL6 gene (Isshiki et al., 1990). Akira et al. (1990) showed that NF-IL6 bound to the consensus sequence in the promoter regions of several acute-phase genes and cytokine genes including tumor necrosis factor (TNF), IL8, and granulocyte-colony stimulating factor (G-CSF) genes. The NF-IL6 binding sequence may be involved in the regulation of these genes. Although the HGF gene is not a prototype acute-phase gene, it is induced by liver injury. The induction is likely mediated by cytokine-like molecules, which are released from the injured liver. The IL6 related consensus sequences might be the regulatory elements involved in the induction by such molecules of the HGF gene. Functional studies will be necessary to determine whether those sequences actually serve regulatory roles.

Human HGF and human plasminogen bear structural homologies but share only small regions of amino acid sequence identity (about 38% identity). However, all cysteine residues but one (Cys-561) in hHGF are located at positions essentially equivalent to those of human plasminogen. Therefore, both protein molecules probably exhibit similar three-dimensional structure. We previously analyzed the genomic DNA encoding kringle regions of hHGF (Miyazawa et al., 1991). From those results, we suggested that the hHGF gene is closely related to the human plasminogen gene. In this study we further elucidated the exon–intron arrangements of the hHGF gene encoding the regions other than kringles. Figure 6 shows the domain structure and placements of intron sequences with respect to the amino acid sequences of hHGF and human plasminogen. The results showed that the exon–intron arrangements for N-terminal portion and the light chain region of hHGF are similar to those for human plasminogen. The N-terminal regions of the heavy chains of both hHGF and plasminogen are encoded by two exons. Those are interrupted by an intron at the amino acid residue next to the equivalent cysteine residue, and the codon is interrupted between the second and third bases. The light chains of hHGF and plasminogen consist of five exons that are interrupted by introns inserted into the same positions in triplet codons. The light chain of plasminogen contains the catalytic triad residues histidine, aspartic acid, and serine. The histidine and serine residues are replaced by glutamine and tyrosine, respectively, in the light chain of HGF. Thus, HGF seems to have no serine protease activity. The exon–intron arrangements of the light chain regions are highly conserved between hHGF and human plasminogen even though the functions of the domains are different. The similar exon–intron arrangements suggest that the genes of hHGF and human plasminogen have probably arisen through gene duplication events from an ancestral gene.

REFERENCES

Akira, S., Isshiki, H., Sugita, T., Tanabe, O., Kinoshita, S., Nishio, Y., Nakajima, T., Hirano, T., & Kishimoto, T. (1990) *EMBO J. 9*, 1897–1906.

Alwine, J. C., Kemp, D. J., & Stark, G. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5350–5354.

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

Benton, W. D., & Davis, R. W. (1977) *Science 196*, 180–182.

Berget, S. M. (1984) *Nature (London) 309*, 179–182.

Berk, A. J., & Sharp, P. A. (1977) *Cell 12*, 721–731.

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

Feinberg, A. P., & Vogelstein, B. (1983) *Anal. Biochem. 132*, 6–13.

Fey, G. H., Hattori, M., Northemann, W., Abraham, L. J., Baumann, M., Braciak, T. A., Fletcher, R. G., Gauldie, J., Lee, F., & Reymond, M. F. (1989) *Ann. N.Y. Acad. Sci. 557*, 317–331.

Fowlkes, D. M., Mullis, N. T., Comeau, C. M., & Crabtree, G. R. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2313–2316.

Gohda, E., Tsubouchi, H., Nakayama, H., Hirono, S., Sakiyama, O., Takahashi, K., Miyazaki, H., Hashimoto, S., & Daikuhara, Y. (1988) *J. Clin. Invest. 81*, 414–419.

Hattori, M., Abraham, L. J., Northemann, W., & Fey, G. H. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 2364–2368.

Igawa, T., Kanda, S., Kanetake, H., Saitoh, Y., Ichihara, A., Tomita, Y., & Nakamura, T. (1991) *Biochem. Biophys. Res. Commun. 174*, 831–838.

Isshiki, H., Akira, S., Tanabe, O., Nakajima, T., Shimamoto, T., Hirano, T., & Kishimoto, T. (1990) *Mol. Cell. Biol. 10*, 2757–2764.

Kan, M., Zhang, G., Zarnegar, R., Michalopoulos, G., Myoken, Y., McKeehan, W. L., & Stevens, J. I. (1991) *Biochem. Biophys. Res. Commun. 174*, 331–337.

Kinoshita, T., Tashiro, K., & Nakamura, T. (1989) *Biochem. Biophys. Res. Commun. 165*, 1229–1234.

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

Lehrach, H., Diamond, D., Wozney, J. M., & Boedtker, H. (1977) *Biochemistry 16*, 4743–4751.

McLauchlan, J., Gaffney, D., Whitton, J. L., & Clements, J. B. (1985) *Nucleic Acids Res. 13*, 1347–1368.

Miyazawa, K., Tsubouchi, H., Naka, D., Takahashi, K., Okigaki, M., Arakaki, N., Nakayama, H., Hirono, S., Sakiyama, O., Takahashi, K., Gohda, E., Daikuhara, Y., & Kitamura, N. (1989) *Biochem. Biophys. Res. Commun. 163*, 967–973.

Miyazawa, K., Kitamura, A., Naka, D., & Kitamura, N. (1991) *Eur. J. Biochem. 197*, 15–22.

Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459–472.

Nakamura, T., Nishizawa, T., Hagiya, M., Seki, T., Shimonishi, M., Sugimura, A., Tashiro, K., & Shimizu, S. (1989) *Nature (London) 342*, 440–443.

Okajima, A., Miyazawa, K., & Kitamura, N. (1990) *Eur. J. Biochem. 193*, 375–381.

Okayama, H., & Berg, P. (1983) *Mol. Cell. Biol. 3*, 280–289.

Petersen, T. E., Martzen, M. R., Ichinose, A., & Davie, E. W. (1990) *J. Biol. Chem. 265*, 6104–6111.

Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.

Prowse, K. R., & Baumann, H. (1988) *Mol. Cell. Biol. 8*, 42–51.

Rubin, J. S., Chan, A. M.-L., Bottaro, D. P., Burgess, W. H., Taylor, W. G., Cech, A. C., Hirschfield, D. W., Wong, J., Miki, T., Finch, P. W., & Aaronson, S. A. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 415–419.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Seki, T., Ihara, I., Sugimura, A., Shimonishi, M., Nishizawa, T., Asami, O., Hagiya, M., Nakamura, T., & Shimizu, S. (1990) *Biochem. Biophys. Res. Commun. 172*, 321–327.

Sharp, P. A. (1981) *Cell 23*, 643–646.

Strain, A. J., Ismail, T., Tsubouchi, H., Arakaki, N., Hishida, T., Kitamura, N., Daikuhara, Y., & McMaster, P. (1991) *J. Clin. Invest. 87*, 1853–1857.

Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.

Tashiro, K., Hagiya, M., Nishizawa, T., Seki, T., Shimonishi, M., Shimizu, S., & Nakamura, T. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 3200–3204.

Tsubouchi, H., Hirono, S., Gohda, E., Nakayama, H., Takahashi, K., Sakiyama, O., Miyazaki, H., Sugihara, J., Tomita, E., Muto, Y., Daikuhara, Y., & Hashimoto, S. (1989) *Hepatology 9*, 875–881.

Tsubouchi, H., Niitani, Y., Hirono, S., Nakayama, H., Gohda, E., Arakaki, N., Sakiyama, O., Takahashi, K., Kimoto, M., Kawakami, S., Setoguchi, M., Tachikawa, T., Shin, S., Arima, T., & Daikuhara, Y. (1991) *Hepatology 13*, 1–5.

Tsuchiya, Y., Hattori, M., Hayashida, K., Ishibashi, H., Okubo, H., & Sakaki, Y. (1987) *Gene 57*, 73–80.

Yoshiyama, Y., Arakaki, N., Naka, D., Takahashi, K., Hirono, S., Kondo, J., Nakayama, H., Gohda, E., Kitamura, N., Tsubouchi, H., Ishii, T., Hishida, T., & Daikuhara, Y. (1991) *Biochem. Biophys. Res. Commun. 175*, 660–667.

Weaver, R. F., & Weissman, C. (1979) *Nucleic Acids Res. 7*, 1175–1193.