Hepatocyte Nuclear Factor-4 Is Responsible for the Liver-specific Expression of the Gene Coding for Hepatocyte Growth Factor-like Protein*

Susan E. Waltz, Fiona K. Gould, Ellen L. Air, Susan A. McDowell, and Sandra J. Friezner Degen‡
From the Division of Developmental Biology, Children's Hospital Research Foundation, Cincinnati, Ohio 45229-3039

In an attempt to understand the molecular mechanism regulating the expression of the gene coding for human hepatocyte growth factor-like protein/macrophage stimulating protein (HGFL), our laboratory has isolated and characterized approximately 4200 bp of the 5'-flanking region of the HGFL gene. To determine the location of sites which may be critical for the function of the HGFL gene promoter, we constructed a series of hybrid genes containing serial deletions of this region attached to the coding sequences for chloramphenicol acetyltransferase. Expression of these chimeric plasmids was examined by transient transfection of HepG2 and 293 cells. Our results suggest that the transcriptional activity of the HGFL promoter is modulated in HepG2 cells by one positive element at position −135 to −105 (−135/−105). In contrast, only background levels of chloramphenicol acetyltransferase expression have been detected in 293 cells. The −135/−105 region appears to bind a liver-specific transcription factor essential for expression of this gene. Gel mobility shift experiments with antibodies against hepatocyte nuclear factor-4 (HNF-4) and transactivation of the HGFL promoter by a HNF-4 cDNA expression vector suggest that HNF-4 binds to the −135/−105 region and is responsible for the liver-specific expression of HGFL.

We previously isolated a human gene which is located at the D3F15S2 locus on human chromosome 3 (3p21), a region believed to code for one or more tumor suppressor genes because this area is deleted in DNA from small cell lung carcinomas, renal carcinomas, and other forms of cancer (1, 2). The corresponding cDNA for this gene codes for a protein with a domain structure similar to that of a known growth factor, hepatocyte growth factor (HGF). Both proteins contain four kringle domains followed by a serine protease-like domain and are approximately 50% identical. Based on these similarities, the protein encoded by this cDNA was designated as “hepatocyte growth factor-like protein” (HGFL).

HGFL is a multifunctional protein that elicits different biological responses in a cell-type and tissue-specific manner (3). HGFL can function as either a growth factor or tumor suppressor for a broad spectrum of tissues and cell types including various epithelial cells. HGFL has been shown to be identical to scatter factor (SF), a mesenchymal cell-derived cytokine that dissociates cohesive colonies of epithelial cells into individual units (4). The cellular responses of HGFL/SF are elicited by c-met, a tyrosine kinase receptor involved in signal transduction (5). Recently, HGFL/SF was shown to be required in mice for liver and placental development, since mice lacking this gene die in utero with defects primarily affecting these organs (6, 7).

The human HGFL gene is 4960 base pairs (bp) in length (from the codon for the putative initiator methionine to the polyadenylation site), containing 18 exons and 17 intervening sequences (2). The expression pattern of HGFL has been shown by in situ hybridization analysis of mouse tissues to be restricted to hepatocytes in the liver (8). The translated amino acid sequence of the gene and cDNA coding for human HGFL predict a protein of 80,325 dalton molecular mass containing 711 amino acids (2). Western analysis using polyclonal antibodies to HGFL has shown that this protein is secreted and is present in human, mouse, and rat plasma with a molecular mass of approximately 90,000 daltons (9).

Although the precise biological function(s) of HGFL remains to be elucidated, two functions for HGFL have been reported. The HGFL protein has been shown to function as an inflammatory mediator based on the identity of its translated amino acid sequence with the amino acid sequence of macrophage-stimulating protein (MSP) (9, 10). This function is further enforced by the observation that conditioned medium from COS-7 cells expressing human HGFL mRNA specifically contained a factor which activates macrophages (10). Based on the structural homology to a known growth factor (HGF), HGFL may also exert its effects as a regulator of cell growth. Gaudino et al. (11) have shown that recombinant HGFL induces phosphorylation of RON, a tyrosine kinase receptor homologue to the HGF receptor gene (c-met), in the epithelial cell line T47D. This phosphorylation is followed by a stimulation in DNA synthesis.

Apart from the progress on identifying a putative function for the HGFL protein, no information regarding the regulation of the expression of the HGFL gene is yet available. A wide range of liver-specific gene products have been found to be controlled at the level of transcription (12). This control may be governed by the interaction of cis-acting DNA sequences in the 5'-flanking region of many liver-specific genes and their cognate trans-acting factors. In this report, we have used this
strategy to investigate the regulation of the expression of the HGFL gene. We have cloned and characterized the 5'-flanking region of the HGFL gene. A positive regulatory element has been identified using transient transfection analyses into various cell lines with HGFL promoter-CAT chimeric plasmids. Sequence and transfection analyses suggested that HNF-4 may play a critical role in HGFL regulation. HNF-4 is a liver-enriched transcription factor in the steroid/thyroid/retinoic acid superfamily which has been shown to regulate liver-specific expression of a variety of genes. Antibody reactivity and competition gel mobility shift assays demonstrate that HNF-4 binds to the HGFL promoter and transactivation experiments show that HNF-4 is sufficient for HGFL reporter gene expression.

EXPERIMENTAL PROCEDURES

Cloning and Sequencing of the Promoter Region of the Human HGFL Gene—The plasmid pL5Bam6 was previously isolated as described by Han et al. (2). The DNA sequence of the 5'-flanking region of the human gene coding for HGFL was determined by both the chemical modification procedure (13) and the quasi-end-labeling modification of the 5'-terminus procedure (13) and the quasi-end-labeling modification of the 5'-terminus procedure (13).

RNA Preparation and Primer Extension Analysis—Liver RNA was isolated using Trizol reagent (Life Technologies, Inc.). A 30-base oligonucleotide transcript from pL5Bam6 containing nucleotides 1273 to 1616 was cloned into pBluescript downstream of a CAT promoter upstream of the CAT reporter gene. pL5CAT5(105/135), pL5CAT5(135/160), and pL5CAT5(160/190) were all PCR-amplified with the n and o primers followed by digestion with HindIII and XhoI and ligation into pBluescript. All of the DNA sequences of the promoter region of the human HGFL gene reported in this paper were determined by these methods.

Oligonucleotides used in cloning HGFL promoter-CAT constructs

| Primer | Nucleotides | HGFL Sequence | Strand |
|--------|-------------|---------------|--------|
| a      | -309 to 320 | 5' - ATACCGGTATCCCTTGCTGAGACGAGGATGATACCTG - 3' | Coding |
| b      | -167 to 168 | 3' - AGTGCCTGTTGGCGGCTGCCCATGGA - 5' | Coding |
| c      | -155 to 156 | 5' - TCTGTGGTTGGGCGGCTGCCCATGGA - 3' | Coding |
| d      | -1272 to 1273 | 5' - ACTACTCCGCGCTGCTGAGACGAGGATGATACCTG - 3' | Coding |
| e      | -1075 to 1076 | 5' - GCGCGGCTGCCCATGGA - 3' | Coding |
| f      | -10 to 11 | 5' - ATACCGGTATCCCTTGCTGAGACGAGGATGATACCTG - 3' | Coding |
| g      | -43 to 44 | 5' - GATTCGAGTCAAGGCTGCGCTGAGACGAGGATGATACCTG - 3' | Coding |
| h      | -43 to 44 | 5' - GATTCGAGTCAAGGCTGCGCTGAGACGAGGATGATACCTG - 3' | Coding |
| i      | -1554 to 1555 | 5' - ATACCGGTATCCCTTGCTGAGACGAGGATGATACCTG - 3' | Coding |
| j      | -1554 to 1555 | 5' - ATACCGGTATCCCTTGCTGAGACGAGGATGATACCTG - 3' | Coding |
| k      | -1554 to 1555 | 5' - ATACCGGTATCCCTTGCTGAGACGAGGATGATACCTG - 3' | Coding |

1 The numbers correspond to the HGFL sequence shown in Fig. 1.
2 Underlined regions refer to mutated nucleotides. Boldface nucleotides correspond to restriction endonuclease sites present in the primer (refer to text for more detail).

Liver-specific Expression of HGFL

pL5CAT(−272/+1) was created by digestion of the −793/+1 PCR product with PstI and XhoI and contains nucleotides −272 to +1 of the HGFL promoter. Constructs pL5CAT(−135/+1), (−104/+1), (−69/+1) and (−56/+1) were made by PCR amplification using the primers a, i, j, k, respectively with primer t and digestion with XbaI and HindIII and ligation into pBluescript. The vector containing nucleotides −104 to −56 was cloned into pBLCAT6 downstream of a BamHI/XbaI 2.5 kb fragment from pL5Bam6 containing nucleotides −4154 to −1555. The entire 4.1 kb insert was excised by digestion with HindIII and cloned into pBLCAT6 to create pL5CAT(−4154/+43).

Plasmids containing PCR products were confirmed by restriction enzyme and DNA sequence analyses.

The HGFL promoter regions for plasmids pL5CAT(−1554/−43) and (−43/−1554) were amplified using the n and o primers followed by digestion with SstI and XbaI and ligation into pBluescript. The insert was excised by digesting with HindIII and cloned into pBLCAT6 in both orientations. This same fragment, containing nucleotides −1554 to −43 was cloned into pBluescript downstream of a BamHI/XbaI 2.5 kb fragment from pL5Bam6 containing nucleotides −4154 to −1555. The entire 4.1 kb insert was excised by digestion with HindIII and cloned into pBLCAT6 to create pL5CAT(−4154/+43).

Plasmid pL5CAT(−135/−105) was created by annealing oligonucleotides p and q, followed by digestion with HindIII and XbaI and ligation into pBLCAT5 (16). pBLCAT5 contains the herpes simplex virus tk promoter upstream of the CAT reporter gene. pL5CAT(−135/+105) was synthesized in an analogous manner with primers r and s. The plasmid pMT2.HNF4(17) was described elsewhere and was a generous gift from Dr. Francis M. Sladek (University of California, Riverside).

Cell Culture, DNA Transfections, and CAT Assays—HepG2 (human hepatocellular carcinoma) cells and 293 (transformed human embryonal kidney) cells were grown in Dulbecco’s modified Eagle’s medium supplemented with 10% fetal calf serum, 2 mEq l-glutamine, and 50 μg/ml gentamicin. For transient transfection analysis, the cells were seeded at approximately 40% confluency 24 h prior to transfection. The cells were transfected with the various human HGF promoter-CAT plasmids using the DNA calcium phosphate method (18).

In each experiment, the cells were co-transfected with equal molar amounts (4.75 × 10−12 M) of each HGFL-CAT promoter construct and 5 μg of the β-galactosidase reference plasmid pRSV/βgal as an internal standard for transfection efficiency. The presence of the DNA-calcium phosphate co-precipitate for 24 h. After removal of the precipitate, the cells were incubated for an additional 48 h. HepG2 and 293 cells were harvested and resuspended in 0.25 M tris-HCl. Whole cell extracts were prepared by disrupting the cells with three freeze-thaw cycles. The supernatant was assayed for the amount of CAT protein produced using CAT enzyme-linked immunosorbent assay kits (5 Prime − 3 Prime, Boulder, CO) according to the manufacturer's instructions. All experiments were repeated at least four times (in duplicate), and CAT protein production was normalized for β-galactosidase activity (19).

RNA Preparation and Primer Extension Analysis—Total RNA from cells transiently transfected with HGFL promoter-CAT constructs was isolated using Trizol reagent (Life Technologies, Inc.). A 30-base oligo-
nucleotide complementary to nucleotides 59–88 of the coding strand of pBLCA6 (16) was 5'-end-labeled with T4 polynucleotide kinase and γ-32P-ATP. The end-labeled primer (500,000 cpm) was hybridized in 2× hybridization salts (20 mM Tris-HCl pH 7.5, 0.6 M NaCl, 4 mM EDTA) to 20 µg of total HepG2 RNA transfected with either pSCAT−1348/−1 or pSCAT−1554/−1 at 44°C overnight after denaturation for 5 min at 90°C. Nucleic acids were purified and redissolved in 20 µl of reverse transcriptase mix (20). Complementary DNA was synthesized using avian myeloblastosis virus reverse transcriptase (Life Science Inc., St. Petersburg, FL). The DNA products were separated on a 6% denaturing polyacrylamide gel along with a sequencing ladder for accurate size determination. The results were visualized by autoradiography after a 16-h exposure at −80°C with intensifying screens.

Preparation of Nuclear Extracts—For nuclear extract preparation, tissue culture cells were washed twice with ice-cold phosphate buffered saline. The cells were collected and resuspended in 5 volumes of a buffer containing 10 mM HEPES, pH 7.9, 1.5 mM MgCl2, 10 mM KCl, 0.5 mM dithiothreitol, and 25 µM phenylmethylsulfonyl fluoride. Dithiothreitol and phenylmethylsulfonyl fluoride were added just prior to use. The cells were incubated on ice for 10 min, pelleted, and resuspended in 3 volumes of the same buffer. Nonidet P-40 was added to 0.05%, and the suspension was incubated on ice for 10 min. The supernatant was pelleted, and resuspended in 3 volumes of the same buffer. Nonidet P-40 was added to 0.05%, and the supernatant was quick frozen in aliquots and stored at −80°C. The protein concentration (as determined by Bradford assays) ranged from 2 to 6 mg/ml.

Gel Retardation Assay—In order to obtain the −135/−105 fragment, complementary oligonucleotides encompassing this region of the HGFL promoter were synthesized, annealed, and end-labeled with T4 polynucleotide kinase and γ-32P-ATP. Nuclear extracts (2 to 5 µl) were incubated in a binding buffer containing 15% glycerol, 20 mM HEPES pH 7.9, 5 mM MgCl2, 50 mM KCl, 0.1 mM EDTA, and poly(dI-dC) (0.1 mg/ml final concentration) for 10 min. Subsequently, 2000–5000 cpm of probe were added (together with the unlabeled competitors in the case of competition experiments), and the binding reaction was continued for another 30 min. For supershift experiments, antiseraum was added 10 min after the addition of probe, and the reactions were continued from this point. The reactions were loaded on precooled 0.25× TBE acrylamide gels containing 0.5× TBE as running buffer. The gels were run in the cold at 150 V, dried, and exposed to film. For competition experiments, the sequence of the HNF-4 oligonucleotide is 5′-GGCCAAGGT-TCATATTGTTAG-3′ (21), and for the human fibrinogen β-chain gene (23). Mutant oligonucleotides encompassing the −135/−105 region are shown in Table II.

**Table II**

| Mutant HGFL sequences | WT | Mutant 1 | Mutant 2 | Mutant 3 | Mutant 4 | Mutant 5 | Mutant 6 | Mutant 7 | Mutant 8 | Mutant 9 | Mutant 10 |
|-----------------------|----|----------|----------|----------|----------|----------|----------|----------|----------|----------|---------|
|                       |    | LAC      | LAT      | LEA      | ISA      | ICAT     | ICA      | ITA      | ITA      | ITA      | ICAT    |
|                        |    |          |          |          |          |          |          |          |          |          |         |
|                        |    | wild type |          |          |          |          |          |          |          |          |         |

1 The wild type −135/−105 HGFL promoter sequence is shown in boxes on the first line. Homologous sequences are shown by --- and variations are indicated. The oligonucleotides were designed based on the sequence reported in Han et al. (2). The G residue reported at −106 was found to be deleted in this study (Refer to Experimental Procedures for more detail). Oligonucleotides with or without this residue gave similar binding and competition results (data not shown).

2 For the competition assays, +++ indicates strong competition, ++ intermediate competition, + slight competition, and − indicates a lack of competition. Refer to text for more details.

**RESULTS**

Comparison of the 5′-Flanking Regions of the Mouse and Human HGFL Genes—Sequence for 4154 bp of 5′-flanking sequence, immediately upstream from the initiator methionine, has been determined for the human gene coding for HGFL protein (Fig. 1; including 272 bp previously reported by Han et al. (2)). Three regions of repetitive DNA were identified by searching GenBankTM; including one Alu repetitive sequence (Fig. 1). Comparison of the 5′-flanking region of the human gene coding for HGFL protein with 1285 bp of 5′-flanking sequence of the mouse gene coding for the same protein (1) identified five regions of homology (Fig. 1).

Analysis of the Human HGFL 5′-Flanking Region—Examination of the 5′-flanking region of the HGFL gene reveals a number of potential regulatory elements (those found within the first 1 kb of the iMet are shown in Fig. 1). Most prominent, within approximately the first 1 kb upstream of the HGFL gene, are six potential liver factor A1 (LF-A1/HNF-4) nucleotide binding sites. LF-A1/HNF-4 has been shown to be important for the expression of a number of liver specific promoters including the α1-antitrypsin (α1-AT) gene (22). Furthermore, no apparent TATA box was found in this region. The precise role of the multiple potential regulatory sites in the flanking DNA of the HGFL gene remains to be determined. However, the numerous potential regulatory elements in the promoter region of the human HGFL gene suggests that the expression of this gene may be regulated by a variety of cytokines and steroid hormones, analogous to HGF (24).
+1) which contains an additional 43 bp suggesting that these extra bases were not important for CAT protein production from these constructs.

As shown in Fig. 2B, the region containing nucleotides from −135 to +1 of the HGFL promoter appeared to contain the minimal amount of flanking DNA required to give strong liver-specific promoter activity. When the sequence between −135 to −105 was deleted to create pl5CAT(−104+1), CAT protein production decreased to approximately background levels. Another less prominent positive regulatory element may also be present in the region between nucleotides −21075 and −2914. CAT protein production was reduced by approximately 25% when this region is deleted (Fig. 2B). Further upstream, the nucleotides between −24154 to −23050 appear to contain a negative regulatory element since removal of this region relieves transcriptional suppression.

Primer Extension Analysis—In order to determine the transcriptional start site of transfected HGFL-CAT plasmids, primer extension analysis was performed. These chimeric constructs were utilized since the transcriptional start point of the endogenous locus could not precisely be mapped. Fig. 3 shows the primer extension results obtained from HepG2 cells transfected with pL5CAT (−1348+1) using an oligonucleotide primer complementary to part of the coding region of the CAT gene. Two bands of approximately equal intensity were obtained of 117 and 118bp in length (lane 2). These results place the transcription start site at −275 and −276bp upstream of the iMet (Fig. 1). Identical results were obtained with RNA prepared from HepG2 cells transfected with other HGFL-CAT chimeric plasmids. No primer extension products were obtained.
tained from RNA isolated from untransfected HepG2 cells (data not shown) or when tRNA was used as a negative control with the same primer (lane 1).

Localization of Nuclear Factor Binding Sites by Gel Retardation Analysis—To investigate the possible interaction of trans-acting factors with the -135/-105 region of the HGFL gene, oligonucleotides were synthesized spanning both strands of the -135/-105 region and gel electrophoretic mobility shift assays were performed with crude nuclear extracts from HepG2 and 293 cells (Fig. 4). A specific protein/DNA complex is formed with this region in HepG2 cell extracts (lanes 2 and 8) but is not present in 293 cells (lane 9). This complex can be specifically competed by a nonradiolabeled self-competitor (lanes 3–5) and not with a nonrelated fragment homologous to an HNF-1 binding site (lanes 6 and 7). These results suggest that the region encompassing nucleotides -135/-105 is involved in the liver-specific expression of HGFL.

Since the region between nucleotides -135/-105 appeared to be crucial for the expression of the HGFL promoter-CAT constructs in HepG2 cells, the nucleotide sequence in this region was examined for known protein binding motifs. The most striking putative transcription factor binding site in this area was for LF-A1 (Fig. 1) (22). LF-A1 is thought to be identical to HNF-4 based on DNA binding data and on antiserum reactivity (26). To test indirectly if the protein binding to the -135/-105 was HNF-4, oligonucleotides were synthesized complementary to a reported HNF-4 consensus sequence (21) and to a region of the α₃-AT promoter containing nucleotides -130 to -101.
which has been shown to bind HNF-4 (22). These oligonucleotides were used in competition gel mobility shift experiments shown in Fig. 5A. The α1-AT competitor almost completely competed for the wild type −135/−105 region at 50- and 100-fold molar excess (lanes 4 and 5). The HNF-4 consensus competitor competed to a lesser extent (lanes 6–8).

To examine the DNA sequence requirements for protein binding to the −135/−105 region, mutant oligonucleotides were synthesized and tested for their ability to compete for trans-acting factor binding in gel mobility shift assays. Table II shows the sequence of the mutant oligonucleotides used in the assay, and Fig. 5B shows the results of the competition experiments. At 100-fold molar excess, mutants 4, 5, and 6 (lanes 6–8) were unable to compete for protein binding to the −135/−105 region of the HGF promoter. Mutant 2 almost completely competed for specific protein binding (lane 4) and mutants 1, 3, 7, 8, 9, and 10 competed to a lesser extent (lanes 3, 5, 9, 10, 13, and 14, respectively). Therefore it appears that the most important sequence involved in protein binding to the −135/−105 region is TCAGGCAG at nucleotides −126 to −118 (Table II; Fig. 1). However, the remaining nucleotides in this region also appear to be involved in protein binding but to a lesser extent.

HNF-4 Binds to the −135/−105 Region of the HGF Promoter—To directly demonstrate that HNF-4 binds to the −135/−105 region of the HGF promoter, gel mobility shift experiments were performed with the −135/−105 region and antiserum directed against a synthetic peptide derived from the carboxyl terminus of the HNF-4 protein (26). The antiserum retarded the mobility of the protein-DNA complex formed with the −135/−105 probe and HepG2 cell extracts (Fig. 6A, lane 4). Antiserum against HGF was used as a negative control and did not react with the protein-DNA complex (lane 3). Preimmune serum also had no effect (data not shown).

As further evidence for HNF-4 binding to the −135/−105 region of the HGF promoter, 293 cells were transfected with an expression vector for HNF-4, pMT2.HNF-4 (17). Nuclear extracts were prepared from the transfected 293 cells and tested for binding to the −135/−105 region (Fig. 6B). Extracts prepared from untransfected 293 cells were unable to retard the mobility of the −135/−105 region (lane 2), whereas extracts from 293 cells transfected with pMT2.HNF-4 (lane 3) produced a mobility shift identical to that of extracts from HepG2 cells (lane 1).

Transactivation by HNF-4—In order to test the −135/−105 region for heterologous promoter activity and transactivation by HNF-4, oligonucleotides spanning this region were synthesized and cloned into the multiple cloning site of pBLCAT5 (16) creating pL5CAT5(−135/−105). pBLCAT5 contains the herpes simplex virus tk promoter 5′ to the CAT gene. As depicted in Fig. 7, this region was able to stimulate transcription from the heterologous promoter approximately 16-fold over the parental plasmid in HepG2 cells. Additionally, a heterologous promoter construct, pL5CAT5(−135/−105mut), containing three base pair changes compared to the wild type −135/−105 HGF sequence (GGT to TTG at nucleotides −123 to −121; mutant 5 in Table II) was tested. The mutant construct stimulated transcription from the tk promoter approximately 6-fold compared to pBLCAT5. In contrast, 293 cells transfected with the −135/−105 heterologous promoter construct gave levels of CAT production similar to pBLCAT5. The mutant −135/−105 region, however, gave approximately a 4-fold stimulation compared to both the wild type sequence and pBLCAT5 in this cell type.

Gel mobility shift experiments performed with the radiolabeled
Liver-specific Expression of HGFL

Fig. 6. HNF-4 binding to the -135/-105 region of the HGFL promoter. A, gel retardation assays with the -135/-105 HGFL promoter sequence and nuclear extracts from HepG2 cells. Lane 1, probe alone. Lanes 2–4, probe with HepG2 extracts. Lane 3, with the addition of an antibody against HNF-4. Lane 4, with an antiserum against HNF-4. A indicates no antibody present. The letters F, C, and S represent free probe, protein-DNA complex, and antibody supershift, respectively. B, the -135/-105 region was tested for protein binding with 293 extracts made from cells transfected with the expression vector, pMT2.HNF-4 which contains a cDNA for HNF-4. Gel mobility shift experiments were performed with extracts from HepG2 cells (lane 1), from 293 cells (lane 2), and from 293 cells transfected with the expression vector (lane 3). All other notations are as in A.

The HGFL gene was identified based on its similarity to a probe coding for the kringle domains of bovine prothrombin (2) and named for its structural similarity to HGF. Based on the translated cDNA sequence for this gene, it has become apparent that HGFL protein is identical to MSP (9, 27). HGFL/MSP has been shown to be a chemoattractant for mouse resident peritoneal macrophages (28), to induce morphological changes in macrophages (10), and to inhibit induction of nitric oxide production by lipopolysaccharide- or cytokine-induced macrophages (29). Furthermore, the expression of HGFL has been shown to be up-regulated during liver regeneration and inflammation (30). This investigation has focused on understanding the factors involved in the regulation of the HGFL gene.

Transient transfection analyses with sequential deletions of the HGFL flanking DNA fused 5′ to the CAT gene were performed to identify potential regulatory sequences governing the expression of this gene. Since the endogenous transcriptional start point (tsp) of this gene could not be precisely determined due to the duplicated copies of this locus (32), the 3′ end point of the majority of deletion constructs was set at the A nucleotide of the potential iMet (+1 in Fig. 1) with a few exceptions. The plasmids pL5CAT(-4154/-43) and pL5CAT(-1554/-43) have their 3′ end point at the A nucleotide (nucleotide -43 in Fig. 1) of an inframe ATG codon upstream of the ATG at +1. An upstream ATG codon is also present in the mouse; however, this codon is not in frame. Transient transfection analyses of pL5CAT(-1554/-1) and pL5CAT(-1554/-43) gave indistinguishable results (Fig. 2B), suggesting that these first 43 bp were not necessary for the transcription of CAT. From these results, it may be inferred that the tsp is not within the first 43 bp, however, an alternative start site can not be excluded.

In order to determine the tsp of the HGFL gene, primer extension analyses were performed using RNA isolated from HepG2 cells transfected with the chimeric HGFL promoter-CAT construct pL5CAT(-1348/+1). This allowed us to assay for RNA initiated specifically from the HGFL promoter and not from RNA initiated from highly homologous regions of the amplified chromosome 1 loci (32). The sequence of one copy of a homologous gene has been determined (9) and was found to be 97% identical to the gene coding for HGFL, including 2200 bp of the 5′-flanking sequence. All but the 3′ end of the HGFL gene is homologous to sequences on chromosome 1 (2). Primer extension experiments performed on total RNA from HepG2 cells using endogenous HGFL sequence primers resulted in a multitude of putative start sites most likely originating from the combination of HGFL and related chromosome 1 sequence start points. The two bands obtained in Fig. 3 correspond to RNA transcripts potentially initiated from -75 and -76 bp upstream of the iMet of the HGFL promoter (Fig. 1). This same combination of bands was observed using several HGFL promoter-CAT constructs. The tsp identified in the human HGFL flanking DNA is within 20 bp of the unique tsp identified in the (nonamplified) mouse gene (1) when the two sequences are aligned at the iMet residue.

Transient transfection analyses implicated several regions which may be critical for expression of HGFL. One region, between nucleotides -135 to +1, contained the minimal sequence of the human HGFL flanking DNA that could drive cell type-specific expression (Fig. 2B). Given the results of the primer extension experiments, this suggests that the minimal promoter required may only span 60 bp from nucleotides -135 to -75. Furthermore, there is no apparent TATA box or Sp1 site near the tsp (Fig. 1). TATA-less promoters are found in many housekeeping genes (e.g. genes encoding the enzymes of intermediary metabolism). Genes with TATA boxes in the promoter region generally initiate transcription at well defined sites, whereas transcription of TATA-less promoters generally occurs over an extended region. The primer extension results presented in Fig. 3 suggest a well defined tsp contrary to the ambiguous tsp of many promoters devoid of TATA boxes.

When comparing the mouse and human HGFL sequences, several areas of homology were observed; one of which overlaps with the -135 to +1 region. Nucleotides -164 to +1 have a high degree of similarity, approximately 76%, between the two species. Furthermore, the -135/-105 region is 71% identical between the mouse and human. Similarities in DNA sequence

---

3 S. J. F. Degen, S. A. McDowell, S. E. Waltz, F. Gould, L. A. Stuart, and B. Carritt, submitted for publication.
Liver-specific Expression of HGFL

The −135/−105 region appeared to be the most critical region implicated in our transient transfection studies since it is both required for promoter activity and is able to confer this activity in a cell type-specific manner. Among the multiple putative binding motifs in the −135/−105 region, sequence analyses identified a putative recognition site for LF-A1/HNF-4 in the −135/−105 region. HNF-4 is a potent transcriptional activator that controls the expression of a variety of genes among different species are thought to represent important regulatory regions.

The −135/−105 region as shown in Fig. 6B. Furthermore, antibodies directed specifically against HNF-4 were able to completely supershift the protein-DNA complex formed in endogenous HepG2 cell extracts (Fig. 6A), suggesting that all of the protein complexes contain at least one HNF-4 monomer. These results indicate that HNF-4 does in fact bind to this region. HNF-4 thus far has been reported to bind exclusively as a homodimer (26, 34). The results of our supershift experiments, in which all of the protein-DNA complex is recognized, are consistent with this observation, although heterodimer formation with another protein(s) cannot be excluded. Interestingly, HNF-4 is not exclusively found in the liver, but its expression has also been reported in intestine and kidney tissue (21, 35).

Preliminary northern analyses of HepG2 and 293 RNA indicate that HNF-4 is not present in the kidney derived 293 cells used in our studies.

Further experiments were undertaken to demonstrate that not only could HNF-4 bind to the −135/−105 region, but that binding of this protein resulted in transcriptional activation. Cotransfection experiments with the vector pBLCAT5 and pMT2.HNF-4 gave results identical to those of pBLCAT5 alone in both cell types. Transactivation experiments shown in Fig. 7 demonstrated that the −135/−105 region was sufficient to promote activation of a heterologous promoter construct. More importantly, HNF-4 expression was solely able to activate transcription of CAT through the −135/−105 promoter region in 293 cells. These results demonstrate that HNF-4 is necessary and sufficient to both bind to (Fig. 6) and activate transcription from the −135/−105 region of the HGFL promoter.

A mutant −135/−105 sequence which changes the important HNF-4 AGGTCA motif to ATTTGCA, decreased transcription of CAT in HepG2 cells compared to the wild type sequence. This mutant sequence was shown not to compete, even at 100-fold molar excess, with the wild type sequence in gel mobility shift assays (mutant 7, Table II and Fig. 5B). In contrast to the wild type −135/−105 heterologous promoter construct which did not stimulate transcription in 293 cells, the mutated sequence was able to confer some transcriptional activity in 293 cells. The activity in both cell types, however, appeared not to be stimulated to a significant extent by HNF-4. Gel mobility shift experiments performed with this mutant sequence showed that protein was capable of binding to the mutated sequence in both cell extracts (data not shown) and may account for the activation seen. The protein-DNA complex formed with the wild type

**Fig. 7. Heterologous promoter activity of the −135/−105 region.** The −135/−105 region of the HGFL promoter was inserted upstream of the herpes simplex virus tk promoter in pBLCAT5 to create pLSCAT5(−135/−105). A mutant −135/−105 region which changes 3 bp (GGT to TTT) at nucleotides −123 to −121; Table II, mutant 7) was also doned 5’ to the tk promoter in pBLCAT5 to create pLSCAT5(−135/−105mut). The amount of CAT protein produced was determined after transfection into HepG2 (striped bars) and 293 (solid bars) cells and normalized for β-galactosidase activity. In each cell type, the activity of pLSCAT5 was set at 1. The fold stimulation of CAT protein production of the −135/−105 and −135/−105mut clones over pBLCAT5 after transfection into is shown. The experiments were performed with and without cotransfections of an HNF-4 cDNA expression vector, pMT2.HNF-4.
sequence had a different electrophoretic mobility (larger) than the mutant complex. Furthermore, the activity was approximately equivalent (5.7-fold versus 4.2-fold) in both HepG2 and 293 cells. This was not entirely unexpected since weak binding to direct repeats (DR +1 elements) or near matches to these repeats (as is apparent in this case) has been reported for family members of the type 2 nuclear receptor family, specifically 9-cis-retinoic acid response element homo- and heterodimers (31, 34, 36). Initial attempts investigating the protein/DNA complexes formed on the −135/+1 promoter region using DNAasel footprinting analyses with extracts from 293 cells with and without overexpression of HNF-4 have been unsuccessful thus far. However, it appears as though the majority of the −135/+1 region of the promoter may be occupied in various cell types. Efforts are underway to purify HNF-4 and to more precisely map the important contact residues for HNF-4 binding.

After examination of the DNA sequence of the 5′-flanking region of the human HGFL gene, various putative regulatory elements found in inducible genes were identified (Fig. 1). Of these elements, there are several potential liver-specific C/EBP transcription factors and multiple potential HNF-4 binding sites. Both of these proteins have been found to regulate a number of liver-specific promoters. There are a number of potential regulators in the −135 to −105 region. There are various hormone responsive elements including: a putative estrogen response element half-site, a site for regulation by various hormone responsive elements including: a putative interferon response element, a site for the thyroid hormone and/or derivatives of retinoic acid through the 9-cis-retinoic acid response element, a site for regulation by estrogen response element half-site, a site for regulation by potential regulators in the −105 region. There are several potential liver-specific C/EBP elements found in inducible genes were identified (Fig. 1). Of these putative binding sites occur within or overlap the −135/−105 region. There are more than the competition or availability of these factors with their recognition sequences.

Based on our results, it can be concluded that HNF-4 is necessary and sufficient for the liver-specific expression of HGFL in HepG2 cells. Antibody reactivity and transactivation experiments confirm that HNF-4 binds to the −135/−105 region and is the sole factor required for stimulating transcription from the wild type −135/−105 region in 293 cells. Furthermore, mutations of this sequence result in the loss of HNF-4 binding and tissue specificity. These studies represent an initial effort to unravel the mechanisms governing expression of HGFL and provide a basis for further study of transcriptional regulation of this gene. Current studies are under way to determine the inducibility of this gene, to more precisely map protein/DNA contact sites and to further map upstream regulators.

Acknowledgments—We thank Dr. Francis M. Sladek for generously providing the HNF-4 antisera and cDNA expression vector, Dr. Michelle Barton and Dr. Thomas Bugge for helpful suggestions and critical reading of the manuscript, Rebecca Muraoka for Northern analyses on 293 and HepG2 cells, and Lori Stuart for technical assistance.

REFERENCES

1. Degen, S. J. F., Stuart, L. A., Han, S., and Degen, S. J. F. (1991) Biochemistry 30, 9781-9791
2. Han, S., Stuart, L. A., and Degen, S. J. F. (1991) Biochemistry 30, 9768-9780
3. Michalopoulos, G. K., and Zarnegar, R. (1992) Hepatology 15, 149-155
4. Weidner, K. M., Arakaki, N., Vandekerekjove, J., Weingart, S., Hartmann, G., Rieder, H., Fonatycz, C., Tsoubouchi, H., Hisida, T., Daikuwara, Y., and Birdicnner, W. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 7001-7005
5. Bottaro, D. P., Rubin, J. S., Faletto, D. L., Chan, A. M.-L., Knieck, T. E., Vande Woude, G. F., and Aaronson, S. A. (1991) Science 251, 802-804
6. Schmidt, C., Blatt, F., Goedeke, S., Brinkmann, V., Zschiesche, W., Sharpe, M., Gherardi, E., and Birdicnner, C. (1995) Nature 373, 699-702
7. Uehara, Y., Minowa, O., Mori, C., Shiota, K., Kunj., Nada, T., and Kitamura, H. (1995) Nature 373, 702-705
8. Bezerra, J. A., Witte, D. P., Aronow, B. J., and Degen, S. J. F. (1993) Hepatology 18, 394-399
9. Bottaro, D. P., Han, S., Danton, M. S., and Degen, S. J. F. (1993) Protein Sci. 2, 666-668
10. Yoshimura, T., Yuki, N., Wang, M. H., Skeel, A., and Leonard, E. J. (1993) J. Biol. Chem. 268, 15461-15468
11. Gaulino, G., Folenzeni, A., Naldini, L., Collesi, C., Santoro, M., Gallo, K. A., Godowski, P. J., and Comoglio, P. M. (1994) EMBO J. 13, 3524-3532
12. Derman, E., Krüther, W., Walling, L., Weinberger, C., Ray, M., and Darnell, J. E., Jr. (1981) Cell 23, 731-739
13. Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 499-560
14. Duncan, C. H. (1985) NEN Product News 4, 6-7
15. Queen, C., and Korn, L. J. (1984) Nucleic Acids Res. 12, 581-599
16. Bosshart, M., Kluppm, M., Schmidt, A., and Luckow, B. (1992) Gene (Amst.) 110, 129-130
17. Mietus-Snyder, M., Sladek, F. M., Ginsburg, G. S., Kuo, C. F., Ladis, J. A. A., Darnell, J. E., Jr., and Karathanasis, S. K. (1992) Mol. Cell. Biol. 12, 1708-1718
18. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., pp. 16.32-16.36, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
19. Nielsen, D. A., Chou, J., Mackrell, A. J., Casadaban, M. J., and Steiner, S. F. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 5190-5202
20. Degen, S. J. F., and Davie, E. W. (1987) Biochemistry 26, 6165-6177
21. Xanthopoulos, K. G., Prezioso, V. R., Chen, W. S., Sladek, F. M., Cortese, R., and Darnell, J. E., Jr. (1993) Proc. Natl. Acad. Sci. U. S. A. 88, 3807-3811
22. Hardon, E. M., Frain, M., Paone, G., and Cortese, R. (1988) EMBO J. 7, 1711-1719
23. Courtois, G., Baumann, S., and Crabtree, G. R. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 937-941
24. Matsumoto, K., Tajima, H., Okazaki, H., and Nakamura, T. (1992) J. Biol. Chem. 267, 24917-24920
25. Birrer, K. L., and Sharp, P. A. (1985) Nucleic Acids Res. 13, 841-857
26. Sladek, F. M., Zhong, W., Lai, E., and Darnell, J. E., Jr. (1990) Genes & Dev. 4, 2353-2365
27. Shimamoto, A., Kimura, T., Matsumoto, K., and Nakamura, T. (1993) FEBS Lett. 333, 61-66
28. Leonard, E. J., and Skeel, A. H. (1978) Exp. Cell Res. 114, 117-126
29. Wang, M. H., Cox, G. W., Yoshimura, T., Shefler, L. A., Skeel, A., and Leonard, E. J. (1994) J. Biol. Chem. 269, 14027-14031
30. Bezerra, J. A., Lane, D. W., Jr., and Degen, S. J. F. (1994) Biochem. Biophys. Res. Commun. 203, 666-673
31. Levine, E., Kantner, R., and Chambon, P. (1992) Trends Biochem. Sci. 17, 427-433
32. Welch, H. M., Darby, J. K., Pilz, A. J., Ko, C. M., and Carritt, B. (1989) Genomics 5, 423-430
33. Costa, R. H., Grayson, D. R., and Darnell, J. E., Jr. (1989) Mol. Cell. Biol. 9, 1415-1425
34. Jiang, G., Nenowamu, L., Hopkins, K., and Sladek, F. M. (1995) Mol. Cell. Biol. 15, 5131-5143
35. Sladek, F. M. (1993) Receptor 3, 223-232
36. Leblanc, B. P., and Stunnenberg, H. G. (1995) Genes & Dev. 9, 1811-1816