Human Inter-α-Trypsin Inhibitor Heavy Chain H3 Gene

GENOMIC ORGANIZATION, PROMOTER ANALYSIS, AND GENE LINKAGE*

(Received for publication, May 28, 1998)

Maryam Diarra-Mehrpour‡, Nasrin Sarafan, Jeannette Bourguignon, Florence Bonnet, Frédéric Bost, and Jean-Pierre Martin

From the Laboratoire de Physiopathologie et Génétique Rénale et Pulmonaire, INSERM, Unité 295, Faculté de Médecine, 76183 Rouen Cedex, France and IFR 61 (Physico-Chimie et Biologie des Systèmes Intégrés), Université de Rouen, F-76821 Mont-Saint-Aignan, France

To understand more about the human inter-α-trypsin inhibitor heavy chain H3 (ITIH3) expression and the relationship between this gene and the family of other ITI heavy chain genes, an analysis of the structure of the ITIH3 gene and its promoter region was performed. This gene is a single copy gene, 14 kilobase pair in length and consists of 22 exons. ITIH3 shares highly conserved exon size and intron-exon borders with other ITI heavy chain genes. We determined that the human ITIH1, ITIH3, and ITIH4 genes are closely linked within a 45-kilobase pair. They are arranged in the order of H1-H3-H4, with the ITIH4 gene transcribed in the opposite direction. A model for the evolution of the ITI heavy chain gene family is presented that involves multiple rounds of gene duplication plus inversion events. The minimum promoter region (−135 to +75) is identified in HepG2 cells. The transient transfection study in various cell lines indicates that the activity of the ITIH3 promoter is not liver-specific. DNase I footprinting, mobility shift assays, and cotransfection experiments reveal a functional CCAAT/enhancer-binding protein site (C/EBP, −1344 to −1305) which interacts with C/EBPα and C/EBPβ factors. The latter factors control the transcription of the ITIH3 gene positively.

The inter-α-trypsin inhibitors (α1-PI) are a family of structurally related plasma serine protease inhibitors (1–2) involved in extracellular matrix stabilization (3–5). Remodeling of the extracellular matrix is a critical event in normal processes such as tissue morphogenesis, differentiation, and wound healing; inter-α-trypsin inhibitors are believed to participate in these events. These glycoproteins are also involved in pathological conditions such as tumor invasion (6), metastasis (7), and arthritis (8), which reinforces the interest in understanding their regulation.

This family is composed of multiple proteins made up of a given combination of polypeptide chains after complex post-translational maturation (9–12). Members of the inter-α-trypsin inhibitors family can be divided into two subclasses depending on their composition with or without bikunin (a Kunitz-type protease inhibitor). When bikunin is present, mature heavy chains named HC1, HC2, HC3 are covalently linked to bikunin by an unusual cross-link named protein glycosaminoglycan protein (13–15).

Pre-α-trypsin inhibitor is a member of the bikunin complex subclass. It is a glycoprotein of Mr, 125,000 composed of one heavy chain of Mr, 90,000, called HC3, and one light chain of Mr, 16,000, bikunin carrying the glycosaminoglycan (14). The level of pre-α-trypsin inhibitor, an acute-phase protein, rises in response to trauma and inflammation (16). In the pathological synovial fluid from human arthritis patients, HC3 is covalently bound to hyaluronan (17).

PK120 (18) is a member of the bikunin uncompelled subclass (also known as inter-α-trypsin inhibitors heavy chain-related protein) (19). It is a single chain glycoprotein of 120 kDa, cleaved into 85- and 35-kDa fragments when plasma is incubated at 37 °C. PK120 is highly sensitive to plasma kallikrein, and its mRNA is strongly induced in pig’s liver after a cardiogenic shock (20).

The structures of the different polypeptide forms of the ITI family have been determined by molecular cloning of their respective cDNAs (21–25). The five chains of human ITI are encoded in the liver by five distinct genes located on three chromosomes (26, 27). The structure of the human light chain (ITI L) gene has been established (28). It was mapped by in situ hybridization to 9q2 32–33. Separate exons of the ITI L precursor code for two proteins, a1-microglobulin and bikunin, which have no typical functional feature (21, 28). The exon-intron organization suggests that the regions coding for a1-microglobulin and other members of the lipocalin superfamily evolved from a common ancestral gene that is different from the gene coding for bikunin, whereas bikunin belongs to the superfamily of Kunitz-type protease inhibitors. These data suggest that two distinct ancestral genes fused during evolution (28). A functional analysis of the ITI L gene promoter showed the presence of a potent and liver-specific enhancer that drives a weak promoter of ubiquitous activity (29–30).

The ITI heavy chain (H) gene family consists of at least four highly homologous genes designated ITIH1, ITIH2, ITIH3 (26), and ITIH1L (27), also named ITIH4. The ITIH2 gene has been mapped to 10p14–15, whereas the ITIH1L, ITIH3, and ITIH4 genes map to 3p211-p212 (for review, see Ref. 31). The ITIH1 gene spans 14 kb and is composed of 22 exons (32). The ITIH4.

* This work was supported by Association pour la Recherche sur le Cancer. 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™/EBI Data Bank with accession number(s) X75318 and X99854.

‡ To whom correspondence should be addressed: INSERM, Unité 295, Faculté de Médecine, 22 Boulevard Gambetta, F-76183 Rouen Cedex, France. Tel.: 33 2 35 14 82 80; Fax: 33 2 35 14 82 37; E-mail: maryam.diarra-mehrpour@univ-roen.fr.

† The abbreviations used are: ITI, inter-α-trypsin inhibitor; H, heavy chains; L, light chain; kb, kilobase(s); bp, base pairs; PCR, polymerase chain reaction; CAT, chloramphenicol acetyltransferase; EMSA, electrophoretic mobility shift assays; CHO, Chinese hamster ovary; C/EBP, CAAT enhancer-binding protein; Pipes, 1,4-piperazinediethanesulfonic acid.
Characterization of the Gene for Human ITIH3

The human liver heptatoma HepG2, murine macrophage (p388D1), Chinese hamster ovary (CHO), and African green monkey kidney SV40 transformed (COS 7) cells lines were grown in a modified Eagle's medium fed with 5% fetal calf serum. The HepG2 cells were cultured in RPMI 1640, p388D1 and COS 7 cells were cultured in Dulbecco's modified Eagle's medium. CHO cells were cultured in Isocove's modified Dulbecco's medium. All the media were supplemented with 10% fetal calf serum, 2 mM glutamine, and 1 mM sodium pyruvate.

HepG2, CHO, and COS 7 cells were cultured in 75-cm² flasks. Fresh culture medium was added every 2nd day until cell subconfluence was reached. The cells were then trypsinized with 1 ml of trypsin solution in 10 mM Na₂EDTA and transferred into 6-cm dishes in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum at an initial concentration of 1.5 × 10⁵ cells/dish. The cells were cultured at 50% confluence and were transfected by a calcium phosphate precipitation procedure (41) with 10 μg of plasmids and 1 μg of pSV-β-Gal. The transfected cells were further cultured for 48 h and harvested for luciferase activity according to the Promega instructions. P388D1 cells were plated 24 h before transfection into 6-cm dishes at the subconfluent density. Transient transfection was performed by the LipofectAMINE method as described by the manufacturer (Life Technologies, Inc.) with 10 μg of each plasmid and 1 μg of pSV-β-Gal. The transfection experiments were performed after 6 h of incubation and replaced with fresh medium. The luciferase assay was also performed according to Promega. The protein was quantified with the Bradford assay kit (Bio-Rad) to normalize the luciferase activities. The luciferase activity was corrected for efficiency of transfection by measuring the β-galactosidase activity. Each transfection was repeated at least three times.

Cotransfection experiments were carried out with 3 μg of reporter plasmid and 6 μg of pSV-β-Gal and treated with 10 units of S1 nuclease (Sigma). The S1 nuclease-digested samples were analyzed by polyacrylamide gel electrophoresis.

Characterization of the Gene for Human ITIH3

The human liver heptatoma HepG2, murine macrophage (p388D1), Chinese hamster ovary (CHO), and African green monkey kidney SV40 transformed (COS 7) cells lines were grown in a modified Eagle's medium fed with 5% fetal calf serum. The HepG2 cells were cultured in RPMI 1640, p388D1 and COS 7 cells were cultured in Dulbecco's modified Eagle's medium. CHO cells were cultured in Isocove's modified Dulbecco's medium. All the media were supplemented with 10% fetal calf serum, 2 mM glutamine, and 1 mM sodium pyruvate.

HepG2, CHO, and COS 7 cells were cultured in 75-cm² flasks. Fresh culture medium was added every 2nd day until cell subconfluence was reached. The cells were then trypsinized with 1 ml of trypsin solution in 10 mM Na₂EDTA and transferred into 6-cm dishes in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum at an initial concentration of 1.5 × 10⁵ cells/dish. The cells were cultured at 50% confluence and were transfected by a calcium phosphate precipitation procedure (41) with 10 μg of plasmids and 1 μg of pSV-β-Gal. The transfected cells were further cultured for 48 h and harvested for luciferase activity according to the Promega instructions. P388D1 cells were plated 24 h before transfection into 6-cm dishes at the subconfluent density. Transient transfection was performed by the LipofectAMINE method as described by the manufacturer (Life Technologies, Inc.) with 10 μg of each plasmid and 1 μg of pSV-β-Gal. The transfection experiments were performed after 6 h of incubation and replaced with fresh medium. The luciferase assay was also performed according to Promega. The protein was quantified with the Bradford assay kit (Bio-Rad) to normalize the luciferase activities. The luciferase activity was corrected for efficiency of transfection by measuring the β-galactosidase activity. Each transfection was repeated at least three times.

Cotransfection experiments were carried out with 3 μg of reporter plasmid and 6 μg of pSV-β-Gal and treated with 10 units of S1 nuclease (Sigma). The S1 nuclease-digested samples were analyzed by polyacrylamide gel electrophoresis.
Characterization of the Gene for Human ITIh3

20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 0.5% SDS, 20 mM EDTA, pH 8.0, and 5 μg of tRNA. The labeled DNAs were phenol/chloroform-extracted and separated on a 6% sequencing gel. Maxam Gilbert sequencing reactions were used as size markers.

For EMSA, duplex oligonucleotide (5′-ctgaGTGCTGCGTGTATTTGGCAACTCTCCATCATTGCTCT-3′) used as probe was end-labeled by filling overhanging 5′ ends with [α-32P]dCTP (3000 Ci/mmol) (Amerham Pharmacia Biotech) using the Klenow fragment. 50 ng of radiolabeled probes were incubated with 6 μg of nuclear protein in a buffer containing 12% glycerol, 25 mM Hepes, pH 7.9, 60 mM KCl, 5 mM MgCl2, 0.12 mM EDTA, 0.06 mM dithiothreitol, and 1 μg of poly(dI-dC)-d1 in a final volume of 20 μl. The reaction mixture was kept at room temperature for 30 min, and the nucleoprotein complexes were separated by non-denaturating electrophoresis in a 4 or 6% polyacrylamide gel for 3 h at 20 mA in TBE buffer (89 mM Tris-HCl, 89 mM orthoboric acid, and 2 mM EDTA). The gel was dried and visualized by autoradiography. In competition experiments, excess unlabeled duplex oligonucleotide competitors were incubated with the nuclear extract simultaneously with the labeled probe. For supershift experiments we used C/EBPα and C/EBPβ antisera given by Steven McKnight (Tularik Inc., San Francisco, CA). These antisera were mixed directly with nuclear protein for 90 min at 4 °C before adding binding buffer and a radiolabeled probe.

RESULTS

Cloning and Sequencing of the ITIh3 Gene—During the initial screening of a human genomic library in EMBL-3 for the ITI heavy chain gene, 10 positive phage clones were identified with heavy chain cDNAs. The radiolabeled cDNAs corresponding to the 5′, central, and 3′ regions of heavy chain H3 and appropriate oligonucleotides were used as probes to cross-hybridize these clones. Three clones were selected and further analyzed by restriction mapping and subcloning. The results of this mapping and Southern blot hybridization with the probes described above are summarized in Fig. 1. They showed that the insert of clone G3 contained the 3′ region of the ITIh1 and the 5′ region of ITIh3, but lacked the central and 3′ region of the ITIh3 gene, whereas two clones G21 and G22 were overlapping and contained the 3′ region coding sequence of ITIh3 (Fig. 1), plus additional genomic DNA extending 7 kb downstream from the exonic boundaries of the gene. The lacked central region of the ITIh3 gene was obtained by several polymerase chain reactions (PCR) using human genomic DNA and a series of gene-specific primers.

A complete set of subclones was used to determine the sequence of the human ITIh3 gene. Comparison of the complete genomic sequence with that of the cDNA (24) established the organization of the human ITIh3 gene. This gene spans 14 kb and is composed of 22 exons (Fig. 1). This analysis reveals that the overall coding capacity for this gene is about 20%. Comparison of the exon-intron organization with the protein structure revealed that HC3 begins at exon 2 and ends at the beginning of exon 18 (Fig. 2). The exons and introns range in size from 9 to 248 bp and from 91 to 1000 bp, respectively. The exon/intron boundaries conform to the GT/AG rule for splice site junctions (Fig. 2) as described by Breathnach and Chambon (44). The gene sequence revealed several differences from our previously published cDNA sequence (24) as follows: substitution of C for T at position 4217 (exon 7), G for A at position 5082 (exon 9), C for T at positions 7687 (exon 12) and 12,165 (exon 19), G for C at position 12,166, C for G at position 12,167; presence of a G at position 12,229; presence of two C, one at position 12,217 and another at position 12,236 (exon 19) and presence of an A at position 13,969 (exon 22). To understand the cause of these differences, we utilized a blast search for the exon sequences of the ITIh3 gene and our previously published cDNA sequence with the non-redundant GenBankTM EST data base. This comparison showed that the difference observed in exons 7 and 12 could be due to intraspecies polymorphism or a cloning artifact of cDNA or of genomic DNA, respectively. The discrepancies observed in exons 9, 19, and 22 were due to the mistake in the published cDNA sequence.

The ITIh3 Gene Is Physically Linked to ITIh4—To determine if the ITIh3 gene is physically associated with the ITIh4 gene located on chromosome 3p21, we hybridized the genomic clones isolated in this study with three different probes corresponding to the 5′, central, and 3′ regions of each ITIh3 and ITIh4 cDNA. Two of them (A21 and A22) simultaneously hybridized with the 5′ portion of ITIh4 and the 3′ region of ITIh3. These clones were analyzed by Southern blot hybridization and subcloning. DNAs from these clones were digested with EcoRI or NcoI or a combination of EcoRI and NcoI and hybridized with two 32P-end-labeled 22-mer synthetic oligonucleotides chosen on the mRNA of each ITI gene. These probes were localized in exon 22 of ITIh3 and the putative last exon of ITIh4 (positions 2850–2877 on the cDNA (25)). None of the fragments hybridized simultaneously with both probes, whereas a fragment EcoRI-EcoRI of 1.8 kb hybridized with ITIh3 and a fragment EcoRI-EcoRI of 0.7 kb hybridized with the ITIh4 probes (Fig. 1). The results indicate that the ITIh4 gene is at the 3′ end of the ITIh3 gene. Identical results were obtained with the human genomic DNA, indicating that these clones contain the 3′ regions of the ITIh3 and ITIh4 genes. To determine the distance from the end of the ITIh3 gene to the end ITIh4, a PCR was performed using the same 22-residue primer described and the human genomic DNA mentioned above. After electrophoretic analysis of the PCR products, ethidium bromide staining showed a fragment of 4.5 kb (Fig. 3, lane 2). The last ITIh4 exon is located just downstream from the last ITIh3 exon.

To gain insight into the evolution of the ITI heavy gene family, we compared the genes of three members of this family: ITIh1, ITIh3, and ITIh4 (Fig. 4). A comparison of amino acid
sequence (Fig. 4A), nucleotide identity, and intron/exon organization (Fig. 4B) reveals that the ITIH1 and ITIH3 genes are very close in structure and are similar to the ITIH4 gene. Whereas ITIH1 and ITIH3 have 22 exons interrupted by 21 introns in identical locations and phases, the ITIH4 gene is composed of 24 exons. The 12 exons from 2 to 13 of the ITIH4 gene were very similar to those (from 3 to 14) of each of the other ITI genes as regards nucleotide length and intron phasing. Intron 2 of the ITIH1 and ITIH3 genes has no counterpart in the ITIH4 gene (Fig. 4A). This intron has either been lost in the ITIH4 genes or gained in the ITIH1 and ITIH3 genes since their divergence.

Determination of the Transcription Start Site of ITIH3—To determine the transcription start site, an S1 nuclease experiment was performed. We used a 169-nucleotide uniformly labeled single-stranded probe complementary to the gene sequence; its 3' and 5' ends were at positions -268 and +101, respectively (Fig. 5C). This fragment was annealed to human liver poly(A)-rich RNAs. After the S1 nuclease digestion, one fragment of 100 nucleotides was protected by the RNA (Fig. 5A). To ensure that a splice site did not occur in the 5'-non-coding region, a primer extension was performed using the same 25-mer primer and poly(A)-rich RNA preparation used in the S1 nuclease analysis. One fragment of 100 nucleotides was also observed (Fig. 5B).

Promoter Analysis—We previously reported that the ITIH1 gene is located 2721 bases upstream from the ITIH3 gene. We also determined the complete sequence of the fragment containing the last exon of ITIH1 and the first exon of ITIH3 (34). To determine whether the putative promoter region of ITIH3 is functional, we now constructed a series of 5'-deletion mutants extending from -22928 toward the transcription start site (Fig. 6A). Fragments were linked to the luciferase reporter gene in the pGL-2 vector. These vectors were introduced into the human hepatoma HepG2 cells, and the transient expression of Luc activity was measured. In each experiment cells were transfected with the promoterless pGL2-Basic as a negative control and with pGL2 promoter as a positive control, which was highly active in the HepG2 hepatoma cells. As shown in Fig. 6B, compared with the promoterless pGL2-Basic and antisense constructs, whose activities were almost undetectable, the construct P-135/175 Luc showed a luciferase activity, whereas the constructs deleted further (P-78/175 Luc and P-20/175) lacked any significant luciferase activity, indicating that the 5'-flanking region (-2135 bp) contains all the necessary elements to support basal transcriptional activity and serve as a "minimal" promoter. A 30% decrease of luciferase activity was observed when the sequence between -22958 and -21578 was deleted (Fig. 6B), whereas the deletion of region from -21578 to -21045 resulted in a 70% decrease of activity. So, there could be a positively regulated region located between -21578 and -21045.

Identification of Nuclear Factors That Interact with the Enhanced Responsive Element of the 5'-Flanking Region of the ITIH3 Gene—To determine the molecular basis of the activation was performed. We used a 169-nucleotide uniformly labeled single-stranded probe complementary to the gene sequence; its 3' and 5' ends were at positions -68 and +101, respectively (Fig. 5C). This fragment was annealed to human liver poly(A)-rich RNAs. After the S1 nuclease digestion, one fragment of 100 nucleotides was protected by the RNA (Fig. 5A). To ensure that a splice site did not occur in the 5'-non-coding region, a primer extension was performed using the same 25-mer primer and poly(A)-rich RNA preparation used in the S1 nuclease analysis. One fragment of 100 nucleotides was also observed (Fig. 5B).

Promoter Analysis—We previously reported that the ITIH1 gene is located 2721 bases upstream from the ITIH3 gene. We also determined the complete sequence of the fragment containing the last exon of ITIH1 and the first exon of ITIH3 (34). To determine whether the putative promoter region of ITIH3 is functional, we now constructed a series of 5'-deletion mutants extending from -2928 toward the transcription start site (Fig. 6A). Fragments were linked to the luciferase reporter gene in the pGL-2 vector. These vectors were introduced into the human hepatoma HepG2 cells, and the transient expression of Luc activity was measured. In each experiment cells were transfected with the promoterless pGL2-Basic as a negative control and with pGL2 promoter as a positive control, which was highly active in the HepG2 hepatoma cells. As shown in Fig. 6B, compared with the promoterless pGL2-Basic and antisense constructs, whose activities were almost undetectable, the construct P-135/175 Luc showed a luciferase activity, whereas the constructs deleted further (P-78/175 Luc and P-20/175) lacked any significant luciferase activity, indicating that the 5'-flanking region (-135 bp) contains all the necessary elements to support basal transcriptional activity and serve as a "minimal" promoter. A 30% decrease of luciferase activity was observed when the sequence between -2958 and -1578 was deleted (Fig. 6B), whereas the deletion of region from -1578 to -1045 resulted in a 70% decrease of activity. So, there could be a positively regulated region located between -1578 and -1045.
FIG. 4. Comparison of the three members of the human ITI heavy chain family. A, deduced amino acid sequence of H1 (upper line) (23), H3 (middle line) (24), and H4 (lower line) (25). Horizontal dashes correspond to gaps introduced in order to maximize similarity scores. Numbering begins with the Met initiation and is in the right margin. The reverse print corresponds to identical residues and shading to Cys residues. The arrows indicate the location of the introns in the genes; those of the ITIH1 and ITIH4 genes are based on previously published reports (32, 33), and those of the ITIH3 gene are based on the sequence described in this paper (Fig. 2). The introns of ITIH1 and H3 are numbered 1–21 and are
Characterization of the Gene for Human ITIH3

Fig. 5. Analysis of the 5’ end of the human ITIH3 gene. A, the 169-nucleotide probe synthesized with a primer corresponding to nucleotides +77 to +101 was used for S1 nuclease mapping of the 5′ region of the ITIH3 gene. Lane 1, S1 nuclease digestion of the probe in the absence of poly(A) RNAs. Lane 2, untreated radioactive probe plus human liver poly(A) RNAs. Lane 3, untreated radioactive probe in the absence of RNA. Lane 4, S1 digestion of the probe after hybridization to human liver poly(A) RNAs. The arrow on the right identifies the size of the protected fragment. B, primer extension was performed with the same primer and mRNA preparation used for the S1 nuclease analysis. Lane 1, primer extension reaction in the presence of human liver poly(A) RNAs. Lane 2, primer extension reaction in the absence of RNA. The arrow on the left identifies the size of the extended fragment. As a size marker, the appropriate dideoxy sequencing reactions were electrophoresed in parallel (lanes T, C, G, A). C, schematic summary of the result with numbers referring to the start of exon 1.

A

T C G A 1 2 3 4

B

1 2 A G C T

S1 Nuclease

169 nt probe

C

Pat I

- 68

+ 1

PE

5’

3’

The functional interaction of C/EBPα and C/EBPβ factors with H3 RE was further demonstrated by cotransfection experiments (Fig. 8C). Transfection of HepG2 cells with a pBLCAT2 plasmid containing one copy of the H3 RE site (sense or antisense) in front of the thymidine kinase (tk) promoter together with both vectors expressing C/EBPα and C/EBPβ resulted in a stimulation of CAT expression (8-fold, Fig. 8C). This effect was specific to the H3 RE sequence, since it was not observed with the pBLCAT2 and pcDNA3/CAT plasmids.

The Activity of the Promoter Region Is Not Tissue-specific—To determine whether the liver-specific pattern of ITIH3 mRNA expression detected by Northern blot analysis in normal human tissues was mediated by the liver-specific activity promoter of the ITIH3 gene, the P-135/101 and P-2928/101 constructs (data

shown in boldface, whereas the introns of ITIH4 are numbered 1–23 and are shown in italics. B, genomic organization of the ITIH1, ITIH3, and ITIH4 genes. Exons and introns are schematized as described in Fig. 1. The exons that represent a higher than 65% identity are linked by lines. The dashed lines link the exons of ITIH1 with ITIH4. The solid lines link the exons of ITIH3 with ITIH1 and ITIH4. ● corresponds to the L1 primate repetitive sequence, and —— corresponds to Alu repeats. A fragment encompassing the sequence between bp −1423 and −1105 (Fig. 2) was end-labeled and incubated with nuclear extracts prepared from HepG2 cells. One major protected region was detected in this fragment, and Fig. 7 shows the protected sequence that contains two C/EBP response elements characterized by the consensus sequences (T(T/G)NNG-NAA(T/G)) with one mismatch and ATTGC (45).

In order to identify the trans-acting factors that interact with target sequences located by the footprinting analysis, EMSA were conducted with a synthetic double-strand oligonucleotide probe named H3RE. In EMSA, this probe produces a complex pattern of bands (Fig. 8A, six complexes I, II, III, IV, V, and VI; lanes 2), suggesting that the site is bound by more than one protein.

To investigate the binding specificity of the multiple complex, we studied the competition between the H3 RE probe and the self-competitor or heterologous competitors H1 RE (32) and C/EBP Bakker (46). The formation of the six complexes was abolished by the presence of a 100-fold molar excess of unlabeled self-competitor (Fig. 8A, lane 5). The I, II, and III complex formation was abolished by the presence of a 100-fold molar excess of the oligonucleotide with H1 RE and C/EBP Bakker (Fig. 8A, lane 8 and 11). The oligonucleotide with the H1 RE sequence was unable to compete and to abolish the IV, V, and VI complex formations (Fig. 8A, lane 8). The C/EBP Bakker sequence seems able to compete and to decrease the IV, V, and VI complex formations (Fig. 8A, lanes 9–11). These results suggest that the C/EBP family may be part of the protein-DNA complex formed with H3 RE.

Although all these data strongly indicated that C/EBP-related proteins recognized H3 RE, they were indirect, precluding the precise identification of which members of the C/EBP family are involved. To address this issue, we used specific antisera raised against C/EBPα and C/EBPβ. Incubation of a HepG2 extract with the C/EBPα and C/EBPβ antisera and binding to H3 RE resulted in the appearance of a supershift band (Fig. 8B, lanes 3 and 4). Moreover, the intensity of the signal corresponding to complex III and V decreased (Fig. 8B, lane III and V, lanes 3 and 4). When we used normal rabbit serum, no super shift band was observed, yet the intensity of the signal corresponding to complex III decreased (Fig. 8B, lane 5). These results indicate that the various factors that interact with H3 RE site include C/EBPα and C/EBPβ factors.

We first performed DNase I footprinting experiments. A fragment encompassing the sequence between bp −1423 and −1105 (Fig. 2) was end-labeled and incubated with nuclear extracts prepared from HepG2 cells. One major protected region was detected in this fragment, and Fig. 7 shows the protected sequence that contains two C/EBP response elements characterized by the consensus sequences (T(T/G)NNG-NAA(T/G)) with one mismatch and ATTGC (45).
DISCUSSION

In this report we show that the gene for human ITIH3 is closely linked to ITIH1 and ITIH4. This gene is made of 22 exons spread along 14 kb of the genome. We observed that the average size of the introns is much smaller than that encountered in vertebrates (47); this makes the genomic ITIH3 a compact gene, with a high coding/non-coding ratio.

The genomic structure of ITIH3 is very similar to that of ITIH1. The intron/exon boundaries occur at exactly the same positions in these two genes. Primer extension and nuclease S1 revealed only a single transcription start site, instead of two for the ITIH1 gene.

It has been shown previously that ITIH1, ITIH3, and ITIH4 reside at chromosome 3, p21–22 (26–27). This region is deleted in most human cancers (48–50). It is clear from this study that these genes occur as single copies closely linked in a 45-kb in the order of 5′-ITIH1-ITIH3-ITIH4-3′. The tight clustering of the three genes is reminiscent of that seen in a number of eukaryotic gene families that have arisen by duplication and divergence from an ancestral gene (34). Based on the amino acid and nucleotide sequence homologies among the three ITI genes as well as on similarities of the exon/intron domains, a possible mechanism for the generation of the ITIH1, ITIH3, and ITIH4 genes is proposed in Fig. 10. Thus, the ancestral gene was first duplicated, and one of the duplicated genes was translocated on chromosome 10 to yield the ITIH2 gene. This mechanism is in agreement with the suggestion that the separation of the H2 and H1/H3 genes dates back 300 million years, whereas the ITIH1 and ITIH3 genes diverged later, about 230 million years ago (51). Another duplicated gene, the primordial H1/H3/H4, then produced a primordial H1/H3 and a primordial H4 by duplication. Taking into account (i) that the ancestral genomic structure has been maintained in the ITIH3 gene, (ii) the loss of exon 2 of ITIH1/H3 in ITIH4, (iii) the presence of a transposable element in intron 2 (L1 primate repetitive sequence) and intron 13 (Alu repeat) of ITIH3 (Fig. 4B), we propose that the appearance of primordial H4 was accompanied by the transposition and inversion of exons 3–13 of the primordial H1/H3/H4. The primordial ITIH4 then acquired exon 1 (homologous to the transcriptional activator hepatocyte nuclear factor 6 (52)) and the 3′ end sequences (homologous to the highly conserved repetitive mouse gene designated lrep3) to produce the ITIH4 gene. The lrep3 gene has a propensity to form pseudogenes in mammals (53) and is closely related to the yeast omnipotent informational suppressor sup44 which encodes the yeast ribosomal protein S4 (54) and the ribosomal protein S2 gene in human tumors (55). The presence of the 3′ end of ITIH4 in this region of chromo-
FIG. 8. Electrophoretic mobility of DNA-protein complexes. A, binding of nuclear proteins to the H3 RE sequence. The assay was carried out with 6 μg of nuclear protein extract from HepG2 cells. Lane 1, probe alone; lanes 2–11, probe incubated with the protein extract from HepG2 cells. The competitors used were as follows: cold self-competitor (lanes 3–5), H1 RE (lanes 6–8), and C/EBP Bakker (lanes 9–11). Increasing concentrations of unlabeled competitors were used. Lanes 3, 6, and 9 contain 20-fold, lanes 4, 7, and 10 contain 50-fold, and lanes 5, 8, and 11 contain 100-fold unlabeled competitors. B, immunological identification of proteins binding to H3 RE. The assay was carried out with 6 μg of
Characterization of the Gene for Human ITIH3

A blast search with the 5′-flanking sequences of the ITIH3 reveals several features of the region. An Alu sequence was found from −1327 to −1501 (34). A significant sequence identity (62%) was observed between the promoter region of the ITIH3 and c-MYC extending from −59 to −112. This sequence identity suggests that this region may contain functionally identical regulatory elements. No significant homology was observed in the 5′-flanking sequences of the three human heavy chain ITI genes. However, several potential common cis-acting elements were noted in the promoter regions for the human ITIH1, H3, and H4 genes. The three genes contain the hepatocyte nuclear factor (LF-A1) and CCAAT/enhancer-binding protein (C/EBP) elements. But the location, number, and orientation of these sites differ somewhat.

HepG2 cells are known to express ITIH3 mRNA (35). Therefore, the cell line was used for a transient transfection study of the ITIH3 promoter. In this cell line, the promoter region of the ITIH3 gene (−2928 to +75), which spans at least up to the last exon of the ITIH1 gene, showed a high activity of the reporter gene. By using the series of the deletion, we identified that the 5′-flanking region extending 135 bp upstream from the cap site contains all the necessary elements to support the basal transcriptional activity and serve as a minimal promoter. The promoter region contains the information needed for the efficient expression of the reporter gene in murine macrophage cells (P388D1 cell line), Chinese hamster ovary cells (CHO), indicating that the ITIH3 promoter is not liver-specific. However, this promoter has no activity in African monkey kidney SV40-transformed cells (COS 7). The SV40 T-antigen is expressed in this cell line. The fact that the SV40 T-antigen, associated physically to p53, inhibits a p53-mediated transcriptional activation (56) has been described. The putative p53-binding element is located between −28 and −37 of the cap site of the ITIH3 gene (Fig. 2). This cis element might be involved in basal transcriptional activity of the ITIH3 minimal promoter. The absence of expression of the ITIH3 gene in the COS line may result from the inhibition of the transcriptional transactivation of this gene by p53.

By using DNase I footprinting, mobility shift assays, and cotransfection experiments, a functional C/EBP site, located at −1344 to −1305, was identified; it interacts with C/EBPα and C/EBPβ factors. The latter factors control the transcription of the ITIH3 gene positively. There are four major proteins in the C/EBP family, currently named C/EBPα, C/EBPβ, C/EBPδ, and C/EBP-related protein 1 (CRP1) (for a review, see Ref. 57), which interact with T(T/G)NNGNAA(T/G) site. These C/EBP family proteins, constitutively expressed and induced, can form homodimers or heterodimers through the basic leucine zipper domain and exhibit similar DNA binding specificity. Moreover, the heterodimerization of C/EBP proteins leads to transcription synergism, further enhancing their effects. The C/EBP family proteins are the transcription factors known to mediate the regulation of several acute-phase protein genes and cytokine genes. This study provide the first evidence for the functional importance of the C/EBP isoforms in ITIH3 gene expression and is in agreement with the transcriptionally up-regulated ITIH3 gene under the influence of interleukin-6, previously described.

some 3 is leading up to the transposition and inversion of exons 2–12 of ITIH4. The primordial H1/H3 gene then duplicated 230 million years ago to yield individual H1 and H3 genes.

The computer scan of the 5′-flanking region with the transcription factor data base, with 0 mismatch, from GenBank™ disclosed several putative binding sites for transcription factors. These include, in particular, the CAAT enhancer-binding protein (C/EBP) site, Sp-1 sites, the p53-binding element, the Ets-like element family (Ets-1, PuF, PEA3) AP1 site, AP2 sites, NF-κB, serum response element, and PPAR (peroxisome proliferator activated receptor, steroid hormone receptor binding element). An equivalent TATA box (TATAAG) is located at position −26 to −21 from the transcription start site. This element is only present on the promoter region of the ITIH3 gene among the ITI heavy genes sequenced.

nuclear protein extract from HepG2 cell in the absence (lane 2) or in the presence of antisera against C/EBPα (lane 3), C/EBPβ (lane 4), and normal rabbit serum (NRS) (lane 5). Lane 1 is the probe alone. The position of a slower moving immune complex is indicated (Supershift). C, functional interaction of C/EBPα and C/EBPβ with H3 RE. Left, reporter constructs. The arrows in front of the thymidine kinase (tk) promoter in the pBLCAT-H3RE sense and pBLCAT-H3RE antisense plasmids indicate the orientation of site H3 RE inserts. Right, the constructs were cotransfected without or with the cDNA coding for C/EBPα and C/EBPβ. The fold induction of CAT activity in cDNA cotransfected cells over cDNA untransfected control cells is shown. Values are means ± S.D. of three independent experiments.
Acknowledgments—We thank Hitoshi Nishimura for providing the ITIH4 cDNA. We thank S. McKnight for the generous gift of antisera raised against C/EBPα and -β and for the expression vectors MSV/EBPα and -β. We are indebted to M. Raymondjean for valuable comments and helpful suggestions. We thank N. Porchet and S. Paris for their excellent technical assistance and J. C. Barre for careful reading of the manuscript.

REFERENCES

1. Hochstrasser, K., Breitgel, G., Feuth, H., Hilla, W., and Lempart, K. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 153–162
2. Potempa, J., Kwon, K., Chawla, R., and Travis, J. (1989) J. Biol. Chem. 264, 15109–15114
3. Chen, L., Mao, S. J. T., and Larsen, W. J. (1992) J. Biol. Chem. 267, 292–296
4. Odum, L. (1992) Int. J. Biochem. 10, 12380–12386
5. Thøgersen, J. B., and Enghild, J. J. (1995) J. Biol. Chem. 270, 3823–3827
6. Tobe, T., Saguchi, K., Hashimoto, K., Miura, N. H., Tomita, M., Li, F., Wang, Y., Minoshima, S., and Shimizu, N. (1995) Cytogeten. Cell Genet. 71, 239–244
7. Diarra-Mehrpour, M., Bourguignon, J., Sesboué, R., Salier, J. P., Léveilllard, T., and Martin, J. P. (1990) Eur. J. Biochem. 191, 131–139
8. Rouet, P., Ragenet, G., Tronche, F., Yaniv, M., N’Guyen, C., and Salier, J. P. (1992) J. Biol. Chem. 267, 20763–20773
9. Tobe, T., Saguchi, K., Hashimoto, K., Miura, N. H., Tomita, M., Li, F., Wang, Y., Minoshima, S., and Shimizu, N. (1995) Cytogeten. Cell Genet. 71, 239–244
10. Diarra-Mehrpour, M., Bourguignon, J., Sesboué, R., Salier, J. P., Léveilllard, T., and Martin, J. P. (1990) Eur. J. Biochem. 191, 131–139
11. He´ron, A., Bourguignon, J., Calle ´, A., Borghi, H., Sesbou ¨e´, R., Diarra-