# A Human $\alpha$ -Fetoprotein Enhancer-Binding Protein, ATBF1, Contains Four Homeodomains and Seventeen Zinc Fingers

TOMONORI MORINAGA,<sup>1</sup>\* HISATAKA YASUDA,<sup>1</sup> TOMOKO HASHIMOTO,<sup>2</sup> KANJI HIGASHIO,<sup>1</sup> and TAIKI TAMAOKI<sup>3</sup>

Research Institute of Life Science, Snow Brand Milk Products Co., Ltd., 519 Shimoishibashi, Ishibashi-machi, Tochigi 329-05,<sup>1</sup> and Department of Genetics, Hyogo College of Medicine, Nishinomiya, Hyogo 663,<sup>2</sup> Japan, and Department of Medical Biochemistry, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada T2N 4N1<sup>3</sup>

Received 28 May 1991/Accepted 6 September 1991

We have isolated a full-length cDNA encoding a protein (ATBF1) that binds to an AT-rich motif in the human  $\alpha$ -fetoprotein gene enhancer. The amino acid sequence deduced from the cDNA revealed that this is the largest DNA-binding protein (306 kDa) known to date, containing four homeodomains, 17 zinc finger motifs, and a number of segments potentially involved in transcriptional regulation. Although the exact function of this protein has not been determined, these structural features suggest that ATBF1 plays a transcriptional regulatory role.

Tissue-specific expression of the human  $\alpha$ -fetoprotein (AFP) gene is strongly stimulated by an enhancer present 3.3 to 4.9 kb upstream of the transcription initiation site (32, 35, 43). We have identified one of the enhancer elements containing an AT-rich core sequence (AT motif) (35). This element is active in HuH7 cells and other human hepatoma cells producing AFP but inactive in HeLa and other cell lines not producing AFP (35, 43). Similar AT-rich sequences are also present in the promoters of a number of liver-specific genes, including AFP, albumin, and  $\alpha$ 1-antitrypsin. This promoter element has been shown to be regulated by a liver-enriched factor termed HNF1, LF-B1, or APF (7-10, 12, 17, 23, 26). Competition binding assays have shown that the same factor binds to the human enhancer AT motif (35), 10-fold less efficiently than it binds to the promoter AT-rich element (12). In addition, the AT motif in the human AFP enhancer is not conserved in the mouse AFP enhancer (16). Thus, the possibilities that HNF1 (LF-B1) is a transcription factor restricted to the promoter region (12) and that a protein other than HNF1 (LF-B1) also interacts with the AT motif in the human AFP enhancer cannot be excluded.

In an attempt to analyze a nuclear factor in HuH7 cells which interacts with the human AFP enhancer AT motif, we screened an HuH7 cDNA expression library with an AFP enhancer fragment which bears the AT motif. In this paper, we report the isolation of a cDNA that can code for an AT motif-binding factor, termed ATBF1. This is the largest DNA-binding protein known to date and the first protein shown to contain multiple homeodomains and multiple zinc finger motifs.

## MATERIALS AND METHODS

Cells. HuH7 (31) and HepG2 (25) are human hepatoma cell lines producing AFP. HuH7 cells were grown in RPMI 1640 medium (GIBCO) containing 0.3% lactalbumin hydrolysate. HepG2 and HeLa cells were grown in Dulbecco's modified Eagle medium containing 5 to 10% fetal calf serum. T24 bladder carcinoma cells (27) were grown in RPMI 1640 medium containing 10% fetal calf serum.

\* Corresponding author.

Preparation of cDNA libraries. Total RNA was prepared from HuH7 cells by the guanidinium isothiocyanate-cesium chloride procedure (15), and poly(A) RNA was separated by oligo(dT) cellulose chromatography. Randomly primed cDNA was synthesized by using the cDNA synthesis kit (Pharmacia) by adding 2.5 µg of poly(A) RNA and 0.4 µg of random primer [p(dN)<sub>6</sub>; Takara] to the 32-µl first-strand reaction mixture. The second-strand synthesis and adaptor ligation were done according to the supplier's protocol. Oligo(dT)-primed cDNA was synthesized by using the cDNA synthesis system (Amersham) and was rendered double stranded according to the supplier's protocol. Double-stranded cDNA was ligated to Agt11 arms and packaged with a commercial packaging extract (Stratagene). Escherichia coli Y 1090(r<sup>-</sup>) was transfected with the recombinant phage and plated in accordance with the standard procedures (20).

Screening of expression libraries. The cDNA expression libraries were screened as described by Singh et al. (37), except that poly(dI-dC) poly(dI-dC) was replaced by sheared and denatured calf thymus DNA (5 µg/ml) (39). To prepare a screening probe, an HgiAI-BstNI AFP enhancer fragment containing the AT motif was excised from a cloned genomic DNA (43), treated with mung bean nuclease, and blunt-end ligated to pUC18 at the BamHI site, which had been filled in with the Klenow fragment of DNA polymerase I. The resultant plasmid, pAFE(H/B)1, was double digested with XbaI and PstI and was treated with exonuclease III and then with mung bean nuclease to shorten the insert DNA from the BstNI end. XhoI linkers were attached to the deleted ends, and the plasmid was recircularized. The 31-bp NlaIV-XhoI fragment, which contains the protein binding site, was treated with the Klenow fragment of DNA polymerase I, self-ligated to form multimers, and cloned into pUC18 at the *Hin*cII site. A clone containing six copies of the fragment (hexamer) was isolated and termed pAFE(N/ X)6. The hexamer was isolated from this plasmid by double digestion with BamHI and HindIII, labeled at the 3' end by using 5'- $[\alpha$ -<sup>32</sup>P]dCTP (Amersham) and the Klenow fragment of DNA polymerase I, and used as a screening probe.

For the isolation of cDNA clones covering the entire coding region, primary recombinants were screened with

<sup>32</sup>P-labeled  $\lambda^2$  cDNA by using the standard procedure of plaque hybridization (4). Clones  $\lambda$ 488,  $\lambda$ 475, and  $\lambda$ 476 were isolated from a library generated from the oligo(dT)-primed cDNA.  $\lambda$ 130 was isolated from a library generated from the randomly primed cDNA.  $\lambda$ 537 and  $\lambda$ 659 were isolated from oligo(dT)-primed and randomly primed cDNA libraries, respectively, by using portions of  $\lambda$ 488 as probes.

Nucleotide sequence determination. The nucleotide sequence was determined by the dideoxy chain termination procedure (34) with standard sequencing primers and custom-made primers. All the isolated clones were entirely sequenced on both strands.

SDS-polyacrylamide gel electrophoresis and Southwestern blot analysis. Y 1089(r<sup>-</sup>) lysogens harboring recombinant or control phage were generated as described by Huynh et al. (20). Cells from 1.25-ml aliquots of induced lysogen cultures were rapidly pelleted and resuspended in 100  $\mu$ l of sodium dodecyl sulfate (SDS)-polyacrylamide gel loading buffer. A 20-µl portion of the sample was heated at 100°C for 5 min and electrophoresed in an SDS-7.5% polyacrylamide gel, and proteins were detected by staining with Coomassie brilliant blue R-250. For Southwestern (DNA-protein) blotting, the separation gel was soaked in buffer containing 25 mM Tris base and 192 mM glycine (pH 8.3) for 30 min, and proteins were electrophoretically transferred onto a Durapore membrane (Millipore). The proteins bound to the membrane were denatured according to the procedure of Celenza and Carlson (6). The membrane was then screened with the hexamer probe in the manner described above for recombinant phage.

Gel retardation assay and DNase I footprint analysis. Lysogen extracts were prepared as described by Singh et al. (37). pAFE(H/B)1 was double digested with SmaI and HincII. The Smal-HincII fragment containing the AT motif was isolated, labeled with <sup>32</sup>P, and incubated with lysogen extracts (2 µg of protein) in 25 µl of gel retardation reaction mixture (38). The DNA-protein complexes formed were resolved in a 5% polyacrylamide gel (acrylamide-bisacrylamide weight ratio of 30:1) containing 50 mM Tris base, 50 mM boric acid, and 1 mM EDTA. The gel was then dried and autoradiographed. DNase I footprint analysis was done basically as described by Singh et al. (38). Gel retardation mixture prepared as described above was incubated with DNase I (0.5  $\mu$ g/ml) for 1 min at 25°C in the presence of 2.5 mM MgCl<sub>2</sub>, and DNA-protein complexes were resolved in a 5% polyacrylamide gel. The DNA in the retarded bands was eluted from the gel with 0.3 M ammonium acetate-10 mM magnesium acetate-0.1% SDS-1 mM EDTA, precipitated with ethanol, and analyzed on an 8% polyacrylamide-7 M urea sequencing gel.

Southern blot analysis. Genomic DNA was digested with restriction enzymes, fractionated on a 0.7% agarose gel, and blotted to Hybond-N membrane (Amersham) in 0.5 N NaOH-0.5 M NaCl. The blot was then hybridized with a <sup>32</sup>P-labeled  $\lambda$ 659 cDNA probe at 65°C and autoradiographed as described previously (29).

**PCR.** For reverse transcriptase-linked polymerase chain reaction (PCR), total cellular RNA was isolated by the guanidinium isothiocyanate procedure (15), treated with RNase-free DNase, and annealed with  $oligo(dT)_{12-18}$  to synthesize cDNA with reverse transcriptase (Life Sciences, Inc.). The cDNA formed was amplified by 40 cycles of PCR by using *Taq* polymerase (Cetus) and several sets of primers listed below. The PCR product was analyzed on a 1% agarose-2 to 3% NuSieve (FMC Corp.) gel. Amplification of genomic DNA was conducted in a similar manner. The primers used were as follows: F29, 5'-ACTGACACTGAC



FIG. 1. Analysis of the polypeptide encoded by  $\lambda 2$  cDNA and its DNA binding specificity. (A) Induced  $\lambda gt11$  and  $\lambda 2$  lysogens were suspended in SDS loading buffer, denatured at 100°C for 5 min, and analyzed on an SDS-7.5% polyacrylamide gel, along with molecular size markers (M). (B) Southwestern blot analysis of  $\lambda 2$  lysogen extract. Proteins shown in panel A were transferred to a nylon membrane and incubated with  $^{32}$ P-labeled hexamer, as described in Materials and Methods.

TGGGACAG-3' (2850 to 2869); F41, 5'-TGGAGGGAGCA TGGAAGGCT-3' (3189 to 3170); F54, 5'-ATGAACCTGG GCGAGAGCTT-3' (169 to 188); F60, 5'-CTTCAATGGC TTCTTCTGGG-3' (850 to 831); F62, 5'-TCGTGCCTGGA GTTGACCGT-3' (581 to 562); B10, 5'-TTGTACCAGGCT TTTCTCCT-3' (7319 to 7338); and B21, 5'-GGAACAATG AAGGGCGTA-3' (7739 to 7720).

Nucleotide sequence accession number. The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL, and GenBank nucleotide sequence data bases under accession no. D90395.

## RESULTS

ATBF1 is the largest DNA-binding protein so far reported. We screened a  $\lambda$ gt11 cDNA expression library derived from poly(A)-containing RNAs of HuH7 cells with an oligonucleotide consisting of six tandem copies of a 31-bp enhancer fragment which bears the AT motif (35). By screening 3  $\times$  $10^5$  phage from a library representing  $3 \times 10^6$  recombinants, a positive clone,  $\lambda 2$ , was isolated. Extracts prepared from  $\lambda 2$ lysogen gave a band of 160 kDa on an SDS-polyacrylamide gel, compared with a 120-kDa band for the control  $\lambda$ gt11 lysogen, which corresponded to  $\beta$ -galactosidase (Fig. 1A). This indicates that  $\lambda 2$  cDNA encodes a polypeptide of 40 kDa. Southwestern blot analysis confirmed the binding of the fusion protein to the AT-motif-containing DNA fragment (Fig. 1B). Gel retardation analysis showed that the  $\lambda 2$ lysogen extract formed two retarded bands, with an enhancer fragment carrying the AT motif (Fig. 2). DNase I footprint analysis demonstrated that the AT motif is involved in the formation of both bands (Fig. 3). This suggests that the fast-moving and slowly moving complexes correspond to the monomer and dimer forms of the fusion protein bound to the AT motif, respectively. In DNase I footprint analysis, the region protected by the fusion protein was smaller in the lower strand than in the upper strand. This



FIG. 2. Gel retardation analysis of  $\lambda 2$  lysogen extracts. Extracts prepared from induced  $\lambda 2$  and  $\lambda gt11$  lysogens were incubated with a <sup>32</sup>P-labeled 96-bp enhancer fragment containing the ATBF1-binding site. DNA-protein complexes formed were resolved in a 5% poly-acrylamide gel. Probe DNA without incubation with lysogen extract was run in lane 1. A and B, slowly moving and fast-moving bands, respectively.

may be due to a spatial hindrance caused by the fused  $\beta$ -galactosidase portion. Recent studies of the sequence recognition by the POU-homeodomain have shown that the homeodomain itself protects only several nucleotides on each strand (42).

The  $\lambda 2$  cDNA is a partial clone, about 1 kb in length, lacking both the 5' and the 3' ends (Fig. 4). Using  $\lambda 2$  as a probe, we isolated two cDNAs containing the 3' end ( $\lambda$ 475 and  $\lambda$ 476) and one clone having a 5' extension ( $\lambda$ 488) (Fig. 4). By using  $\lambda$ 488 as a probe, a clone containing the 5' end  $(\lambda 537)$  was isolated (Fig. 4). The complete nucleotide sequence of the mRNA determined from these overlapping clones and the amino acid sequence deduced from it are shown in Fig. 5. Several nucleotide sequence heterogeneities were found in overlapping clones. Most of these differences involve single nucleotides in the third positions of codons, with no changes in the amino acids encoded. They include cytosine in  $\lambda$ 537 versus thymine in  $\lambda$ 488 at nucleotide 2316. thymine in  $\lambda 659$  versus cytosine in  $\lambda 488$  at nucleotide 4338, thymine in  $\lambda$ 130 versus cytosine in  $\lambda$ 488 at nucleotide 5565, cytosine in  $\lambda 475$  versus thymine in  $\lambda 476$  at nucleotide 7508, and guanine in  $\lambda 475$  versus thymine in  $\lambda 476$  at nucleotide 8403. In  $\lambda$ 488, an adenine is deleted at position 4378. This is probably an artifact of cloning, since the nucleotide sequences of genomic DNAs prepared from HuH7, HeLa, and IMR90 (human lung fibroblast) cells, when determined by direct sequencing of polymerase-amplified DNA, agreed with that of  $\lambda$ 659, which contains adenine at this position (data not shown). Several major differences between two 3'-terminal clones,  $\lambda$ 475 and  $\lambda$ 476, were found. (i) At position 7508, cytosine is present in  $\lambda$ 475, coding for alanine, whereas thymine is present in  $\lambda$ 476, encoding value. (ii) In comparison with  $\lambda 475$ ,  $\lambda 476$  has deletions of 3 consecutive nucleotides at two places (nucleotides 7903 to 7905 and 7921 to 7923) and an insertion of 24 nucleotides (GG CAA CTA CAG CAG CAG CAG CAG C) between nucleotides 7528



FIG. 3. DNase I footprint analysis demonstrating recognition of the AT motif by the fusion protein. In both upper- and lower-strand panels: lane 1, products of chemical cleavage at adenines and guanines of the probe DNA (G/A); lane 2, DNA without incubation with  $\lambda 2$  lysogen extract (free); lane 3, DNA extracted from band A in Fig. 2 (band A); lane 4, DNA extracted from band B in Fig. 2 (band B). Summary of the analysis is shown below the footprints.

and 7529. (iii) In  $\lambda$ 476, a poly(A) tail is attached to cytosine at nucleotide 8521, whereas in  $\lambda$ 475, it is attached to cytosine at nucleotide 8588. These alterations result in an increase of six amino acids in the polypeptide encoded by  $\lambda$ 476. The



FIG. 4. Restriction endonuclease map of ATBF1 cDNA. Open bar, full-length cDNA. ATG and TAA, translation initiation and termination sites, respectively. Sizes and positions of partial cDNA clones are shown above the map.

1	CGCGGCCCGAGCGCCTCTTTTCGGGATTAAAAGCGCCGCCAGCTCCCGCCGCCGCCGCCGCCGCCGCGCGCG	
121	CCCCTAGACATGCGGCTCGGGGGCGGGGGGGGGGGGGGG	( 37
241	AAGTTCACGACGACAACCTGGACATGCTGGGCCTGCACATGAACGTGGAGGCGCACCTGTCGGAGGACGAGTGGAGGGGGTGATGGGGGACTCATACCAGTGCAAGCTCTGCCGCTAC K F T T D N L D M L G L H M N V E R S L S E D E W K A V M <u>G D S Y Q C K L C R Y</u>	(77
361	AACACCCAGCTCAAGGCCAACTTCCAGCTGCACTGCAAGACAGAC	( 117
481	GTGGCCATCGGCAACCCCGTGCACCTCAAGTGCAACGCCTGGGCTACCACCACCCGAGAAGCTGCGGCTGCGCGCTGCACCCGGGCCAGCCTGAGGCCAGCCTGAGGTGCAGCCTGAGGCCAGCCTGAGGCCAGCCTGAGGCCAGCCTGAGGCCAGCCTGAGGCCAGCCTGAGGCCAGCCTGAGGCCAGCCTGAGGCCAGCCTGAGGCCAGCCTGCGGCTGCAGCCGCGCGCG	(157
601	AAGCACCTGCAGCAGCAGCATGAGAGTGGTGTAGAAGGTGAGAGGCTGCTACTACCACTGCGTTCTGTGCAACTACTCCACCAAGGCCAAGCTCAACCTCATCCAGCATGTGCGCCTCATGAGA K H L Q Q H E S G V E G E <u>S C Y Y H C V L C N Y S T K A K L N L I Q H V R S M K</u>	( 197
721	CACCAGCGAAGCGAGAGCCTGCGAAAGCTGCAGCGGCTGCAGAAGGGCCTTCCAGGGGGAGCACGAGGACCTGGGGCAGATCTTCACCATCCGCAGGTGCCCCTCCACGGACCCAGAAGAA H Q R S E S L R K L Q R L Q K G L P E E D E D L G Q I F T I R R C P S T D P E E	(237
841	GCCATTGAAGATGTTGAAGGACCCAGTGAAACAGCTGCTGATCCAGAGGAGCTTGCTAAGGACCCAAGAGGGGGGAGCATCGTCCAGCCAAGCAGGAGAAGGAGCTGACAGATCTCCCTGCA A I E D V E G P S E T A A D P E E L A K D Q E G G A S S S Q A E K E L T D S P A	(277
961	ACCTCCAAACGCATCTCCCTTCCCAGGTAGCTCAGAGTCTCCCCTCTCTTCGAAGCGACCAAAAACAGCTGAGGAGATCAAACCGGAGCAGATGTACCAGTGTCCCTACTGCAAGTACAGT T S K R I S F P G S S E S P L S S K R P K T A E E I K P <u>E Q M Y Q C P Y C K Y S</u>	( 317
1081	AATGECEGATGTCAACCEGGETGCATGECCATGACGCAGCAGCACTCEGTGCAACCCATGCTCGCCGCCGCGCCCCGTGCCCCGGGCATGCTCAACAACAAGATCCACCTCCAGCTGCACCCA N_A_D_V_N_R_L_R_V_H_A_M_T_Q_H_S_V_Q_P_M_L_R_C_P_L_C_Q_D_M_L_N_N_K_I_H_L_Q_L_H_L_	( 357
1201	ACCCACCTCCACAGCGTGGCACCTGACTGCGGGGAGAAGCTCATTATGACGGTGACCACCCCTGAGATGGTGATGCCAAGCAGCATGTTCCTCCCAGCAGCGTGTTCCAGATGGA <u>I I I I S V A P D</u> C V E K L I M T V T T P E M V M P S S M F L P A A V P D R D G	( 397
1321	AATTCCAATTTGGAAGAGGCAGGGAAAGCAGCCTGAAACCTCAGAGGATCTGGGAAAGAACATCTTGCCATCCGCAAGCAA	( 437
1441	GGCTCTGTGAGAGAAGACTCAGGCTTCATCTGCTGGAAGAAGGGGTGCAACCAGGTTTTCAAAACTTCTGCTGCCCTTCAGAGGCGCATTTTAATGAAGTGCATGCCAAGAGGCCTCAGCTG G S V R E <u>D S G F I C W K K G C N Q V F K T S A A L Q T H F N E V H A K R P Q</u> L	( 477
1561	CCGGTGTCAGATCGCCATGTGTACAAGTACCGCTGTAATCAGTGTAGCCTGGCCTTCAAGACCATTGAAAAGTTGCAGCTCCATTCTCAGTACCATGGATCAGAGCTGGCACCATGTGC P V S D R H <u>V Y K Y R C N Q C S L A F K T I E K L Q L H S Q Y H V L R A A T M C</u>	( 517
1681	TGTCTTTGTCAGCGCAGTTTCCGAGCTTTCCAGGCTCTGAGAAGAAGCACCCTGGAGCTGAGTGAG	(557
1801	GACCTCCTGGCAATGGGAGACCCCACTCTGGCTGAGGACCATACCATAATTGTTGAGGAAGAAGAAGAGGAGGAGTGACTTGGAAGATAAACAGAGCCCAACGGGCAGTGACTCTGGG D L L A M G D P T L A E D H T I I V E E D K E E E S D L E D K Q S P T G S D S G	(597
1921	TCAGTACAAGAAGACTCGGGCTCAGAGCCAAAGAGAGGCTCTGCCTTTCGAGAAAAGGTCCCAATTTTACTATGGAAAAGTTCCTAGACCCTTCTGCCCCTACAGGTGTACCGTCTGCAAG S V Q E D S G S E P K R A L P F R K G P N F T M E K F L D P <u>S R P Y K C T V C K</u>	(637
2041	GAATCTTICACTCAAAAGAATATCCTGCTAGTACACTACAATTCTGTCTCCCCCCCC	(677
2161	GACAACAAACCTTITAAGTGTAACACTTGTAATGTGGCCTACAGCCAGAGTTCCACTCTGGAGATCCATATGAGGTCTGTGTTACATCAAACCAAGCCCGGGCAGCCAAGCTGGAGGCT D N K P F K C N T C N V A Y S Q S S T L E I H M R S V L H Q T K A R A A K L E A	(717
2281	GCAAGTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTATTTCCTTGAGCCCCCCAGGCGCAGCCGCGCAGTAGCACCTTTACCACCCCCCAATCCAAGCAGTGCT A S G S S N G T G N S S S I S L S S S T P S P V S T S G S N T F T T S N P S S A	(757
2401	GGCATTGCTCCAAGCTCTAACTTACTAAGCCAAGTGCCCACTGAGAGTGTAGGGATGCCACCCTGGGGGAATCCTATTGGTGCCAACATTGCTTCCCCTTCAGAGCCCAAAGAGGCCAAT G I A P S S N L L S Q V P T E S V G M P P L G N P I G A N I A S P S E P K E A N	(797
2521	CGGAAGAAACTGGCAGATATGATTGCATCCAGGCAGCAACAACAACAACAACAACAACAACAACAACA	(837
2641	GTTCAAGCTCACCTGCAGCAGGAGGCTGCAGCAACAGGCTGCCCGATCCAGCTGTTTAACCCCACCCTCCTCCTCACTTCCCCATGACAACTGAGACCCTGCTGCAACTACAG V Q A H L Q Q E L Q Q Q A A L I Q S Q L F N P T L L P H F P M T T E T L L Q L Q	(877
2761	CAGCAGCAGCACCTCCTCTTCCTTCTACATCCCCAGTGCTGAGTCTCAGCTTAACCCCGAGGGAGCTGGCAGTGAGCAGTGGGGCACTGACACTGACACTGGGACAGGCCCAGGCCTG Q Q H L L F P F Y I P S A E F Q L N P E V S L P V T S G A L T L T G T G P G L	(917
2881	CTGGAAGATCTGAAGGCTCAGGTTCAGGTCCCACAGCAGAGCCATCAGCAGAATCTTGCCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCCAGC	( 957
3001	CAGCACCCCGAAAAGAAGAACAAATTGGTCATCAAAAGAAAAGGAAAAAGAAAAGACAGGAGAGAGA	( 997
3121	GCCTTGAAGGCCAAAGAGAAGAAGAGTGGCACCAGGGGGGGG	(1037
3241	TTTGGCTTTGAGTTGGTCATCCAGTATAATGAGAACAAGCAGAAGAAGGTGCAGAAAAGAATGGGAAGACTGGCAGGAGAGAACCTGGAAAAGCTCGAGTGGACTCCTGCGGCAAGTTG F G F E L V I Q Y N E N K Q K V Q K K N G K T D Q G E N <u>L E K L E C D S C G K L</u>	(1077
3361	TTTTCCAACATCTTGATTTTAAAGAGTCATCAAGAGCACGTTCATCAGAATTACTTTCCTTTCCAAACAGCTCGAGAGGGTTTGCCAAACAGTACAGAGAGCACTACGATAAACTGTACCCA F_S_N_L_L_L_K_S_H_Q_E_H_Y_H_Q_N_Y_F_P_F_K_Q_L_E_R_F_A_K_Q_Y_R_D_H_Y_D_K_L_Y_P	(1117
3481	CTGAGGCCCCAGAGCCCACCACCACCACCTCCCCCCCCCC	(1157
3601	ACCTCACCTACACTGCACCGGCCCAGCCATCAGTGCCGCTCCACCCAGCTCTCCGGAGGAGCTGCCCATCTTCTCGCCGCTGATGATGCAGACGATGCCGCTGCAGACCTTGCCG T S P T I A P A Q P S V P L T Q L S M P M E L P I F S P L M M Q T M P L Q T L P	(1197
3721	GCTCAGCTAGCTGCGGAGCTGGGGACCTGGGGGCCTGGCCGGACCTGGCCCAACTCTACCGAGCAGCAACCCAGCTGCTGCTGCAGCAGCAGGAGGAGGGCCTCGCACCAGG A Q L P P Q L G P V E P L P A D L A Q L Y Q H Q L N P T L L Q Q Q <u>N K R P R T R</u>	(1237
3841	ATCACAGATGATCAGCTCCGAGTCTTGCGGCAATATTTTGGCATTAACAACTCCCCCGGGTGAAGAGGAGAGGAGGAGAGGGGGGGG	(1277)
3961	TGGTTCAGGAACACTCTCTTCAAAGAGAGGCAGCGTAACAAGGACTCCCCTTACAACTTCAGTAATCCTCCTATCAACCAGCCTGGAGGAGCTCAAGATTGACTCCGCGGCCCCCTTCGCG W F R N T L F K E R Q R N K D S P Y N F S N P P I T S L E E L K I D S R P P S P	(1317)
4081	GAACCTCCAAAGCAGGAGTACTGGGGAAGCAAGAGGAGTCTCTAAGAACAAGGTTTACGGACTACCAGCTGAGGGTCTTACAGGACTTCTTCGATGCCAATGCTTACCCAAAGGATGATGAA E P R Q E Y W G S <u>K R S S R T R F T D Y Q L R V L Q D F F D A N A Y P K D D E</u>	(1357)
4201	TTTGAGCAACTCTCTAATTTACTGAACCTTCCAACCCGAGTGATGATGAGTGGGTGTGGTTTCAGAATGCCCGACAGAAGAATTATGAGAATCAGGGAGAGGGCAAAGATGGAGAG F E Q L S N L L N L P T R V I V V V F Q N A R Q K A R K N Y E N Q G E G K D G E	(1397)

4321	* R R E L T N D R Y I R T S N L N Y Q C K K C S L V F Q R I F D L I K H Q K K L C (1437)	)
444]	TACAAGGATGAGGATGAGGAGGGGGGGGGGGGGGGGGGG	)
4561	. AGGGCCCCAGCACCATCAGCCATAATACAGCTICCTCCGCCTTICTIGCAGCTIACAGCGGAGGCGGGGGGAGCGGGCCACCTTCAATTCAA	)
4681	. CTGGGGGAGGCTCCCAGGCAAACCAAACCAAAGCCAAAAGCAAAGCCAGAGCGGCAGCA	)
4801	. CCCCAGCTGGTGTCCCTGCTTGCTACAGCGCCTCCTCAGAGGGCCCCTACCCGAGCCCGAGCCCGAGCCTGCCGAGCTGCCCAGCTGCCCCGCCCTGAGGCC PQLVSLPSLPQPPPQAPPQCPPQCSPSPSQLSHLPLKP(1597)	)
4921	. CTCCACACATCAACTCCTCCAACAGCTGGCAAACCTACCCCTACCAGTGTGACCAGTGTAAGCTGGCATTTCCGTCATTTGAGCACTGGCAGGAGCATCAGCAGCTC L H T S T P Q Q L A N L P P Q <u>L I P Y Q C D Q C K L A F P S F E H W Q E H Q Q L</u> (1637)	)
5041	CACTTCCTGAGGGGGAGCAGTTCATCCACCCCAGTTTTTGGACAGGTCCCTGGATATGCCTTTCATGCTCTTTGATCCCAGTAACCCACTCCTGGCCAGCTGCTCTCTGGG H_F_L_S_A_Q_N_Q_F_I_H_P_Q_F_L_D_R_S_L_D_M_P_F_M_L_F_D_P_S_N_P_L_L_A_S_Q_L_L_S_G_(1677)	)
5161	GCCATACCTCAGATTCCAGCAGCACGCCACTCTCCACTTCAACTCCACACTCCACATGAACACTCTCAAGAGGAAGCTGGAGGAAGGCCAGTGCCAGCCCTGGCGAAAACGACAGT A I P Q I P A S S A T S P S T P T S T M N T L K R K L E E K A S A S P G E N D S (1717)	)
5281	GGGACAGGAGGAGAGAGGACCTCAGAGAGAAGAGCATCGGAGAACAACCATCACACCGGAACAACTAGAAATTCTCTACCAGAAGTATCTACTGGATTCCAATCGACTCGAAAGAATGTTG G T G G E E P Q R <u>D K R L R T T I T P E Q L E I L Y Q K Y L L D S N P T R K N L</u> (1757)	)
5401	GATCACATTGCACACGAGGTGGGGCTTGAAGAAACGTGTGGTACAAGTCTGGTTTCAGAACACCCGAGCTCGGGAAAGGAAAGGACAGTTCCGGGCTGTAGGCCCAGGGCCCACAGG D H I A H E V G L K K R V V Q V M F Q N T R A R E R K G Q F R A V G P A Q A H R (1797)	)
5521	AGATGCCCTTTTTGCAGAGCGCTCTTCAMAGCCAAGACTGCTCTTGAGGCTCATATCCGGTCCCGTCACTGGCATGAGCCAGAGAGCTGGCTACACCTAACTCTGTCTG	)
5641	TTAGACTGTGATGGGGGGCTCCAGATGAAAGGAGATATTTTGACGGAACTAGCTTTCCCACCCA	)
5761	ATGGAATTGTCACCCAGAACTCTTCTAAGCCCCTCCCCCATTAAGGTGGAAGGGATTGAAGACTTTGAAAGCCCCTCCATGTCCTCAGTTAATCTAAACTTTGACCAAACTAAGCTGGAC N E L S P R T L L S P S S I K V E G I E D F E S P S M S S V N L N F D Q T K L D (1917)	)
5881	AACGATGACTGTTCCTCTGTCAACACAGCAATCACAGATACCACAACTGGAGACGAGGGCAACGCAGGTAACGACAGTGCAACGGGAATAGCAACTGAAACCAAATCCTCTTCTGCACCC N D D C S S V N T A I T D T T T G D E G N A D N D S A T G I A T E T K S S S A P (1957)	)
6001	AACGAAGGGTTGACCAAAGCGGCCATGATGGCAATGTCTGAGTATGAAGATCGGTTGTCATCTGGTCTGGTCAGCCCGGCCCCGAGCTTTTATAGCAAGGAATATGACAATGAAGGTACA N E G L T K A A M M A M S E Y E D R L S S G L V S P A P S F Y S K E Y D N E G T (1997)	)
6121	GTGGACTACAGTGAAACCTCAAGCCTTGCAGATCCCTGCTCCCCCGAGTCCTGGTGGGAGGGGAGGGGATCTGCAGGGGAAAACGTTTTCGCACT V D Y S E T S S L A D P C S P S P G A S G S A G K S G D S G D R P G <u>Q K R F R T</u> (2037)	)
6241	CANATGACCAATCTGCAGCTGAAGGTCCTCAAGTCATGCTGCTAGGTCTCAGGACACCCACTATGCTAGGATGTGAGGTCCTGGGCAATGACATTGGACTGCCAAAGAGAGTCGTTCAG Q M T N L Q L K V L K S C F N D Y R T P T M L E C E V L G N D I G L P K R V V Q (2077)	)
6361	GTCTGGTTCCAGATGCCCGGGCAAAAGAAGAAGAAGAAGTCCAAGTTAGGCATGGCCAAGCATTTGGTATAAACCAAACGAGTTATGAGGGACCCAAAACAGAGTGCACTTTGTGTGGGCATC <u>V W F Q N A R A K E K K S K</u> L S M A K H F G I N Q T S Y E <u>G P K T E C T L C G I</u> (2117)	)
6481	AAGTACAGCGCTCGGCTGTCTGTACGTGACCATATCTTTTCCCCAACAGCATATCTCCAAAGGTAAAGACACCATTGGAAGCAGCGGGGACAAGGAGAAGAATACTTTGACCCAGCCACC K Y S A R L S V R D H I F S Q Q H I S K V K D T I G S Q L D K E K E Y F D P A T (2157)	)
6601	GTACGTCAGTTGATGGCTCAACAAGAGTTGGACCGGATTAAAAAGGCCAACGAGGCCTTGGACTGGCAGCTCAGCAGCGAGGGGATGTTTGACAACACCCCTCTTCAGGCCCTTAACCTT V R Q L M A Q Q E L D R I K K A N E V L G L A A Q Q G M F D N T P L Q A L N L (2197)	)
6721	CCTACAGCATATCCAGCGCTCCAGGGCATTCCTCCTGTGTTGCTCCCCGGGCCTCAACAGCCCTCCTTGCCAGGCTTTACTCCATCCA	)
6841	ATGGGTCTGGCCCAGCACACTGTTCCTTCCCCTGGCCTCCCCACTTCTGGATTACCAATAAACCGTCCTCAGCGTCGCTGAGCTCCCCAACCCCAGCACAAGCCACGATGGATG	)
6961	CCTCAGCAACCCCCCCCAGCAGCAGCAGCAGCAGCAGCAGCAG	)
7081	CGCAAGGACAAGACAGTGAGAAAGTAAAGGACAAGGACAAAGGGAAAGGGGAAACGCGGAACCCCTGCCTG	)
7201	ATCTCAGCCCCGCTGCCCACCATGGAGTATGCGGTAGACCCTGCACGCCGCGGCGGCGGTGACTTCGGACCCCACAGCATTGCTCACAAGCCAGTTCCTTACTTT I S A P L P T M E Y A V D P A Q L Q A L Q A A L T S D P T A L L T S Q F L P Y F (2397)	)
7321	GTACCAGGCTTTTCCCTTATTATGCCCCCAGATCCCCGGCGCCCCGCAGAGCGCGGTACCCGCAGCCCTATGTATG	)
7441	CTGATGGGGGCTGTCCCCAGGCTCCCTACTGCAGCAGTACCAGCAATACCAGCAGGAGGCAGGTCGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	)
7561	CAMACCCCAGTCCCCCCGGGGCTCCTTCCCCCAGACAMAGACCCCCCCAAAGAACAGAAGAACAGAAAAAACACCCCCC	)
7681	GAMGAGCCAGAAGCAGAAAGCAAAAGCGAGGACTCCCCCTACGACCCCCTTCATTGTTCCAAAGGTGCAGTACAAGTTGGTCTGCCGCCAAGTGCCAGGCGGGCTTCAGCGACGAGGAGGCA E E P E A E S K S A D S L Y D P F I V P K V <u>Q Y K L V C R K C Q A G F S D E E A</u> (2557)	
7801	GCGAGGAGCCACCTGAAGTCCCTCTGCTTCTTCGGCCAGTCTGTGGTGAACCTGCAAGAGATGGTGCTTCACGTCCCCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC	
7921	GCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGCCGCTGCTG	
8041	ATCACGAGAGCAGCAGAGAAAACGCCAAAGAGCACCCTAGTTTATTACCTCACTCGCCTGCCT	
8161	CCTCCCCCGTCGGCCGCCCCCCTCCTCCGCCTCCCCACGCCACGCAAGTGGTCTCCCGGCGCTCCGGCGCGCGC	
8281	TCTTCAACGGTTACCTCAAGTTCATGCAGCACCTCAGGGGTTCAGCCCTCGATGCCAACAGACGACTATTCGGAGGAGTCTGACACGGATCTCAGCCAAAAGTCCGACGGGAGCCGGAC S S T V T S S S C S T S G V Q P S M P T D D Y S E E S D T D L S Q K S D G P A S (2757)	
8401	CCGGTGGAGGGTCCCAMAGACCCCACCGCCCAAGGACAGTGGTCGACCAGTGTAGGACCCTTCAGATTGTAAGCTTTGAAGATGAACAATACAAACAA	
8521	CAMMATTANTANCAMICCANTTCAMMATAGACTAACTGCANTTCCAMAGCTTCTAACCAMMAACAMMAMAAAA	

FIG. 5—Continued.

polypeptide based on  $\lambda 476$  consists of 2,789 amino acid residues with a molecular mass of 307 kDa, whereas the polypeptide based on  $\lambda 475$  consists of 2,783 amino acid residues with a molecular mass of 306 kDa. The nucleotide sequence shown in Fig. 5 is based on  $\lambda 475$ .

ATBF1 contains four homeodomains and 17 zinc finger

**motifs.** Analysis of the amino acid sequence of ATBF1 shows several novel structural features (Fig. 6). First, four homeodomains (HD1 to HD4) are present in this protein. They show 27 to 43% sequence identities with homeodomains in several transcription factors (Fig. 7A). The homologies are somewhat higher with LIM group proteins



FIG. 6. Potential functional domains of ATBF1. Ovals, HD1 to HD4. Closed circles, zinc finger motifs. Rectangles, segments rich in acidic amino acids (Ac), serine and threonine (S/T), glutamine (Q), proline (P), and proline and glutamine (P/Q).

(30 to 43%) than with POU group proteins (27 to 35%). In all cases, the highest sequence conservation is found in the third helix (recognition helix). Several key residues in this helix, such as Trp (48th residue), Phe (49th residue), Asn (51st residue), and Arg (53rd residue), are conserved in HD2, HD3, and HD4 (Fig. 7A). HD1 contains all these residues except Arg (53rd residue). HD1 is also characterized by the presence, 131 amino acids upstream, of a sequence homologous to the A box of the POU-specific domain (Fig. 7B).

Second, ATBF1 contains 17 zinc finger motifs (ZF1 to ZF17). Ten of the 17 zinc fingers are clustered in the amino-terminal region, and the rest are distributed throughout the molecule (Fig. 6). ZF12 and ZF16 belong to the CC-HC-type fingers, whereas others are CC-HH-type fingers (5) (Fig. 7C). All these fingers contain a constant spacing of 12 amino acids between cysteine 9 and histidine 22. Phenylalanine 13 and leucine 19 are conserved in 9 and 13 zinc fingers, respectively.

Third, ATBF1 has a number of regions that may poten-

tially play a role in transcriptional regulation. They include (i) six acidic domains (three within the cluster of zinc fingers in the amino-terminal portion, two in the middle portion, and one near the carboxyl terminus), (ii) two regions rich in both serine and threonine (51%, amino acids 719 to 773; 38%, amino acids 2650 to 2737), (iii) a region rich in glutamine (34%), including a stretch of 19 glutamine residues (amino acids 809 to 958), and (iv) a region rich in proline (39%) with a stretch of 12 proline residues (amino acids 1117 to 1211). Clusters of glutamine and proline residues from amino acids 1516 to 1617, 2182 to 2317, and 2339 to 2537, and a stretch of 22 glycine residues interrupted by a single serine residue (amino acids 2585 to 2607) may also have transcriptional regulatory functions.

The ATBF1 gene is a single-copy gene, and the transcripts are detected in non-AFP-producing as well as AFP-producing cells. Southern blot analysis with  $\lambda 659$  cDNA as a probe gave a single hybridizable band with HuH7 DNA digested with *BalI*, *AvaIII*, *BglII*, *EcoRI*, *XbaI*, *Bam*HI, and *Hin*dIII (Fig. 8). This suggests that the ATBF1 gene is a single-copy gene. Similar experiments with genomic DNAs from HeLa, Chang (human liver cell line), and IMR90 cells gave identical hybridization patterns, indicating that there is no rearrangement involving the ATBF1 gene (data not shown).

In order to examine whether ATBF1 mRNA is expressed in cells other than HuH7 cells, we analyzed RNAs from HepG2, HeLa, and T24 cells using reverse transcriptaselinked PCR. Four combinations of primers were used, two corresponding to regions near the 5' end, one in a middle

Α		Helix 1		Helix 2	Helix 3		Seq	uence	identi	ity (%)
	10	20	30	40	50	60	HD1	HD2	HD3	HD4
HD1	NKRPRTRITD	DQLRVERQYE	DINNSPSEEQ	IKEMADKSGĒ	POKVIKHWFR	NTLFKERORN	-	38	35	33
HD2	KRSSRTRFTD	YQLRVLQDFF	DANAYPKDDE	FEQLSNLLNL	PTRVIVVWFO	NAROKARKNY	38	-	30	43
HD3	DKRLRTTITP	EQLEILYQKY	LLDSNPTRKM	LDHIAHEVGL	KKRVVQVWPQ	NTRARERKGO	35	30	-	43
HD4	QKRFRTQMTN	LQLKVEKSCF	NDYRTPTMLE	CEVLGNDIGL	PKRVVQVWFQ	NARAKEKKSK	33	43	43	_
Pit1	KRKRRTTISI	AAKDALERHF	GEHSKPSSQE	IMRMAEELNL	EKEVVRVWFC	NRROREKRVK	27	30	33	30
Oct1	RRKKRTSIET	NIRVALEKSF	LENQKPTSEE	ITMIADQLNM	EKEVIRVWRC	NRRQKEKRIN	32	32	28	28
Unc86	KKRK <b>RT</b> SIAA	PEKREBEQFF	KQQPRPSGER	IASIADRLDL	KKNVVRVWFC	NOROKOKRDF	35	30	33	30
lin-11	RRGPRTTIKA	KQLETEKNAF	AATPKPTRHI	REQLAAETGE	NMRVIQVWFQ	NRRSKERRMK	33	40	43	42
Isl-1	TTRVRTVLNE	KQLHTERTCY	AANPREDALM	KEQLVEMTGL	SPRVIRVWFQ	NKRCKDKKRS	30	38	33	40
mec-3	RRGPRTTIKQ	NQLDVLNEMF	SNTPKPSKHA	RAKLALETGL	SMRVIQVWFQ	NRRSKERRLK	37	37	38	38
	200000	· 96 40	.3%	SNL		ini 196 198				
LF-B1	GRRNRFKWGP	ASQQIEFQAY	ERQKNËSKEE	RETLAQGLOV	TEVRVYNWFA	NRRKEEAFRH	23	20	28	23
			VEECH	RAECIQRGVS	PSQ					
R				C						
U	1100		1125		6	9 13	19 2	2	28	
ATBF1	<b>QLERFAKQYRI</b>	OHYDKLYPLRP(	TPEP	ZF1	GDSYQČKL-	-ČRYNTQLKAN	IFQLÌ	İCKTI	окнос	QKYQ
Pit1	ELEQUANEFK	/RRIKLGYTQTI	IVGBA	ZF2	PVHLKCNA-	-CDYYTNSLER	LRLE	HTVN:	SRHE	ASLK
Oct1	ELEQUARTFK	QRRIKLGFTQGI	VGLA	ZF3	SCYYHCVL-	-CNYSTKAKLN	LIQI	IVRS	4KHQI	RSES
Unc86	QUET PACHFK(	QRRIKLGVTQAL	VGKA	ZF4	EQMYQCPY-	-CKYSNADVNF	LRŶI	TMAE	2-HS	VQPM
Cfla	DUBAFAKQFK	QRRIKEGFTQAI	DVGLA	ZF5	QPMLRCPL-	-CQDMLNNKIH	LQLI	ILTH	L-HSV	VAPD
HNF1	TKKG	GRRNRFKWGPAS	SQQIL	ZF6	DSGFICWKK	GCNQVFKTSAA	LQTH	IFNE	7-HAI	KRPQ
LF-B1)	<b></b>			ZF7	VYKYRCNQ-	-CSLAFKTIER	LQLI	ISQY-	HV	IRAA
	POU-specifi	ic domain A-box		ZF8	RAATMCCL-	-CQRSFRTFQ#	LKKI	ILETS	S-HLI	SLSE
				ZF9	SRPYKCTV-	-CKESFTQKN1	LLVI	IYNS	/SHL	IKLK
				ZF10	NKPFKCNT-	-CNVAYSQSS1	LEIF	IMRS	LHQ?	<b>FKAR</b>
				ZF11	LEKLECDS-	-CGKLFSNIL1	LKSI	IQEHV	7–HQI	NYFP
				ZF12	NLNYQCKK-	-CSLVFQRIFE	LIKI	IQKKI	L-CYI	CDED
				ZF13	LIPYQCDQ-	-CKLAFPSFEH	WQEI	IQQL-	HFI	SAQ
				ZF14	QAHRRCPF-	-CRALFKAKTA	LEAF	IIRSF	R-HWI	IEAK
				ZF15	GPKTECTL-	-CGIKYSARLS	VRDI	IIFSC	QHIS	SKVK
				ZF16	QYKLVCRK-	-CQAGFSDEEA	ARSI	ILKSI	-CFI	GQS
				ZF17	GGSYHCLA-	-CESALCGEEA	LSQE	ILES	LHKI	IRTI

FIG. 7. Homeodomains, POU-specific domain A box, and zinc finger motifs in ATBF1. (A) Sequence comparison of homeodomains in ATBF1 and other transcription factors. Amino acid sequences are as described by Scott et al. (36) for Pit1, Oct1, Unc86, and mec-3; by Freyd et al. (14) for lin-11; by Karlsson et al. (24) for Isl-1; and by Frain et al. (13) for LF-B1 in the arrangement suggested by Nicosia et al. (33). Conserved amino acid residues are shaded. The extent of sequence identity of ATBF1 homeodomains among themselves and with other homeodomains is shown on the right. (B) Amino acid sequence alignment of ATBF1 with the A box of POU-specific domains of several transcription factors. The conserved amino acid residues are shaded. Amino acid sequences are as described by Herr et al. (19) for Pit1, Oct1, and Unc86; by Johnson and Hirsh (22) for Cf1a; and by Baumhueter et al. (3) for HNF1 (LF-B1). (C) Amino acid sequences of the 17 zinc finger motifs in ATBF1. Cysteines (C) and histidines (H) which are thought to be involved in the coordination of zinc are indicated by dots above the sequences.



FIG. 8. Southern blot analysis of HuH7 DNA. Genomic DNA prepared from HuH7 cells was digested with restriction enzymes, and the fragments were separated on a 0.7% agarose gel, blotted onto a nylon membrane, and hybridized with a <sup>32</sup>P-labeled  $\lambda$ 659 cDNA probe. The restriction enzyme used is given above each lane.

portion, and one near the 3' end. All RNA preparations yielded PCR products whose sizes match exactly those estimated from ATBF1 cDNA (Fig. 9). Several other nonhepatic cell lines, J82, HTB9, and MCF7, also produced similar PCR fragments (data not shown). These results indicate that the expression of ATBF1 transcripts is not restricted to hepatomas. These results as well as the sequence analysis of the overlapping clones confirm that ATBF1 mRNA is in fact large, encoding all of the domains mentioned above.

When genomic DNAs of HuH7 and HeLa cells were used as templates, three of the four sets of primers used produced bands of the same sizes as the bands produced by cDNA (Fig. 9A, C, and D). On the other hand, the combination of primers F54 and F60 yielded no detectable bands (Fig. 9B, lanes 6 and 7). This may suggest the presence of long noncoding sequences between these two primers. This interpretation was consistent with the result of Southern blot analysis, which suggested the presence of at least two introns between sites represented by these primers (data not shown).

In the case of RNA and genomic DNA of HuH7 cells, the combination of primers B10 and B21 generated three PCR fragments (Fig. 9D, lanes 2 and 6). The fastest-moving band matches the band amplified with cDNA clone  $\lambda$ 475, and the middle faint band matches the band amplified with  $\lambda$ 476 (data not shown). The band amplified with  $\lambda$ 476 also matches the bands amplified by using RNAs from the other three cell lines. These results suggest that HuH7 cells contain at least two kinds of alleles of ATBF1, one normal and the other truncated.

#### DISCUSSION

We report here the isolation of a cDNA clone encoding a human AFP enhancer-binding protein, ATBF1, which shows several unique features. First, this is the largest DNAbinding protein so far reported, having a molecular mass of 306 kDa. The largest previously known DNA-binding protein is PRDII-BF1 (298 kDa), which binds to the human beta interferon gene promoter (11).

Second, ATBF1 is the first protein known to contain four



. . . . . . .

FIG. 9. Agarose-NuSieve gel electrophoresis of PCR products synthesized with total RNAs and genomic DNA templates. Both cDNA-RNA hybrids synthesized with total RNAs from HuH7, HepG2, HeLa, and T24 cells and genomic DNA isolated from HuH7 and HeLa cells were subjected to PCR amplification, and the products were analyzed on agarose-NuSieve gels.  $\phi$ X174 DNA digested with *HaeIII* was used as a size marker (M). The combinations of primers are F54 (nucleotides 169 to 188 of the cDNA) and F62 (nucleotides 581 to 562) (A), F54 (nucleotides 2850 to 2869) and F41 (nucleotides 3189 to 3170) (C), and B10 (nucleotides 7319 to 7338) and B21 (nucleotides 7739 to 7720) (D).

homeodomains in addition to multiple zinc finger motifs. We have shown that the  $\lambda$ 2-encoded polypeptide carrying HD4 binds to the AT motif in the AFP enhancer. Sequence comparison shows that 10 of 11 amino acids in helix 3 (amino acids 42 to 52 in Fig. 7A) are conserved in HD3 and HD4 and that 8 are conserved in HD2 and HD4. In particular, these three homeodomains share glutamine at the ninth position of helix 3 (amino acid 50 in Fig. 7A), a residue known to play a crucial role in the recognition of a target sequence (18, 41). It is therefore possible that HD2 and HD3 also interact with the AT motif. HD1, on the other hand, shows the least sequence similarity with HD4 in helix 3 (4 of 11 residues), with a different amino acid (arginine) occupying the ninth position. HD1 is also unique in that it is preceded by a POU-specific domain A box.

Third, ATBF1 has 17 zinc finger motifs. It is likely that they are involved in DNA binding, either by themselves or in conjunction with neighboring homeodomains. Conceivably, they serve to build up the specific DNA-binding function of ATBF1, as has been postulated for POU-specific and LIM domains (14, 24, 40).

Fourth, ATBF1 has many potential transcriptional regulatory domains which include acidic domains and regions rich in glutamine, proline, and hydroxyl amino acids (serine and threonine) (28). These domains may allow ATBF1 to interact with other factors to regulate transcription.

The ATBF1 protein described here is a second protein found to interact with the AT motif in the human AFP enhancer. This is not surprising, since it is well established that a given DNA element can be recognized by multiple factors (21, 28). The biological function of ATBF1 is not clear at present, although its structural features described above suggest a transcriptional regulatory role for this protein. Judging from the presence of multiple DNA-binding domains, we speculate that ATBF1 regulates gene transcription in a complex manner. Functional specificity for ATBF1 also remains unclear at this time, since ATBF1 transcripts were detected in nonhepatic cells, such as HeLa cells, in addition to hepatoma cells. However, HeLa cells apparently lack functional ATBF1 molecules, since no protection of the AT motif is observed with HeLa nuclear extracts in footprint analysis (35). It is possible, therefore, that the formation of functional ATBF1 molecules is regulated posttranscriptionally, as has been shown with several other transcription factors (1, 2, 30).

### ACKNOWLEDGMENTS

We thank H. Nakabayashi for technical advice.

This work was supported by grants from the Medical Research Council of Canada and the National Cancer Institute of Canada. T.T. is a recipient of a Terry Fox Cancer Research Scientist award of the National Cancer Institute of Canada.

#### REFERENCES

- 1. Baeuerle, P. A., and D. Baltimore. 1988. Activation of DNAbinding activity in an apparently cytoplasmic precursor of the NF-κB transcription factor. Cell 53:211-217.
- 2. Baeuerle, P. A., and D. Baltimore. 1988. IκB: a specific inhibitor of the NF-κB transcription factor. Science 242:540–546.
- Baumhueter, S., D. B. Mendel, P. B. Conley, C. J. Kuo, C. Turk, M. K. Graves, C. A. Edwards, G. Courtois, and G. R. Crabtree. 1990. HNF-1 shares three sequence motifs with the POU domain proteins and is identical to LF-B1 and APF. Genes Dev. 4:372-379.
- Benton, W. D., and R. W. Davis. 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196:180–182.
- Berg, J. M. 1986. Potential metal-binding domains in nucleic acid binding proteins. Science 232:485–487.
- 6. Celenza, J. L., and M. Carlson. 1986. A yeast gene that is essential for release from glucose repression encodes a protein kinase. Science 233:1175–1180.
- Cereghini, S., M. Blumenfeld, and M. Yaniv. 1988. A liverspecific factor essential for albumin transcription differs between differentiated and dedifferentiated rat hepatoma cells. Genes Dev. 2:957-974.
- Courtois, G., S. Baumhueter, and G. R. Crabtree. 1988. Purified hepatocyte nuclear factor 1 interacts with a family of hepatocyte-specific promoters. Proc. Natl. Acad. Sci. USA 85:7937– 7941.
- 9. Courtois, G., J. G. Morgan, L. A. Campbell, G. Fourel, and G. R. Crabtree. 1987. Interaction of a liver-specific nuclear factor with the fibrinogen and  $\alpha$ 1-antitrypsin promoters. Science 238:688–691.
- 10. De Simone, V., G. Ciliberto, E. Hardon, G. Paonessa, F. Palla, L. Lundberg, and R. Cortese. 1987. Cis- and trans-acting elements responsible for the cell-specific expression of the human  $\alpha$ 1-antitrypsin gene. EMBO J. 6:2759–2766.
- Fan, C.-M., and T. Maniatis. 1990. A DNA-binding protein containing two widely separated zinc finger motifs that recognize the same DNA sequence. Genes Dev. 4:29–42.
- 12. Feuerman, M. H., R. Godbout, R. S. Ingram, and S. M. Tilghman. 1989. Tissue-specific transcription of the mouse  $\alpha$ -fetoprotein gene promoter is dependent on HNF-1. Mol. Cell. Biol. 9:4204-4212.
- 13. Frain, M., G. Swart, P. Monaci, A. Nicosia, S. Stampfli, R.

Frank, and R. Cortese. 1989. The liver-specific transcription factor LF-B1 contains a highly diverged homeobox DNA binding domain. Cell 59:145–157.

- Freyd, G., S. K. Kim, and H. R. Horvitz. 1990. Novel cysteinerich motif and homeodomain in the product of the Caenorhabditis elegans cell lineage gene lin-11. Nature (London) 344:876– 879.
- Glisin, V., R. Crkvenjakov, and C. Byns. 1974. Ribonucleic acid isolated by cesium chloride centrifugation. Biochemistry 13: 2633–2637.
- Godbout, R., R. S. Ingram, and S. M. Tilghman. 1988. Finestructure mapping of the three mouse α-fetoprotein gene enhancers. Mol. Cell. Biol. 8:1169–1178.
- 17. Guertin, M., H. LaRue, D. Bernier, O. Wrange, M. Chevrette, M.-C. Gingras, and L. Bélanger. 1988. Enhancer and promoter elements directing activation and glucocorticoid repression of the  $\alpha_1$ -fetoprotein gene in hepatocytes. Mol. Cell. Biol. 8:1398– 1407.
- Hanes, S. D., and R. Brent. 1989. DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. Cell 57:1275–1283.
- Herr, W., R. A. Strum, R. G. Clerc, L. M. Corcoran, D. Baltimore, P. A. Sharp, H. A. Ingraham, M. G. Rosenfeld, M. Finney, G. Ruvkun, and H. R. Horovitz. 1988. The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2 and Caenorhabditis elegans unc-86 gene products. Genes Dev. 2:1513–1516.
- Huynh, T. V., R. A. Young, and R. W. Davis. 1985. Constructing and screening cDNA libraries in λgt10 and λgt11, p. 49–78. *In* D. M. Glover (ed.), DNA cloning—a practical approach, vol. 1. IRL Press, Oxford.
- Johnson, P. F., and S. L. McKnight. 1989. Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem. 58:799–839.
- Johnson, W. A., and J. Hirsh. 1990. Binding of a Drosophila POU-domain protein to a sequence element regulating gene expression in specific dopaminergic neurons. Nature (London) 343:467-470.
- Jose-Estanyol, M., and J.-L. Danan. 1988. A liver-specific factor and nuclear factor 1 bind to the rat α-fetoprotein promoter. J. Biol. Chem. 263:10865-10871.
- 24. Karlsson, O., S. Thor, T. Norberg, H. Ohlsson, and T. Edlund. 1990. Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. Nature (London) **344**:879–882.
- Knowles, B. B., C. C. Howe, and D. P. Aden. 1980. Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. Science 209:497-499.
- Lichtsteiner, S., and U. Schibler. 1989. A glycosylated liverspecific transcription factor stimulates transcription of the albumin gene. Cell 57:1179–1187.
- Marshall, C. J., L. M. Franks, and A. W. Carbonell. 1977. Markers of neoplastic transformation in epithelial cell lines derived from human carcinomas. J. Natl. Cancer Inst. 58:1743– 1747.
- Mitchell, P. J., and R. Tjian. 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding protein. Science 245:371–378.
- 29. Morinaga, T., A. Fotedar, B. Singh, T. G. Wegmann, and T. Tamaoki. 1985. Isolation of cDNA clones encoding a T-cell receptor β-chain from a beef insulin-specific hybridoma. Proc. Natl. Acad. Sci. USA 82:8163–8167.
- Mueller, C. R., P. Maire, and U. Schibler. 1989. DBP, a liver-enriched transcriptional activator, is expressed late in ontogeny and its tissue specificity is determined posttranscriptionally. Cell 61:279–291.
- Nakabayashi, H., K. Tanaka, K. Miyano, T. Yamane, and J. Sato. 1982. Growth of human hepatoma cell lines with differentiated functions in chemically defined medium. Cancer Res. 42:3858–3863.
- 32. Nakabayashi, H., K. Watanabe, A. Saito, K. Sawadaishi, and T. Tamaoki. 1989. Transcriptional regulation of  $\alpha$ -fetoprotein expression by dexamethazone in human hepatoma cells. J. Biol. Chem. 264:266–271.

- 33. Nicosia, A., P. Monaci, L. Tomei, R. De Francesco, M. Nuzzo, H. Stunnenberg, and R. Cortese. 1990. A myosin-like dimerization helix and an extra-large homeodomain are essential elements of the tripartite DNA binding structure of LF-B1. Cell 61:1225-1236.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- 35. Sawadaishi, K., T. Morinaga, and T. Tamaoki. 1988. Interaction of a hepatoma-specific nuclear factor with transcription-regulatory sequences of the human  $\alpha$ -fetoprotein and albumin genes. Mol. Cell. Biol. 8:5179–5187.
- Scott, M. P., J. W. Tamkun, and G. W. Hartzell III. 1989. The structure and function of the homeodomain. Biochim. Biophys. Acta 989:25-48.
- 37. Singh, H., J. H. LeBowitz, A. S. Baldwin, and P. A. Sharp. 1988. Molecular cloning of an enhancer binding protein: isolation by screening of an expression library with a recognition site DNA. Cell 52:415-423.
- 38. Singh, H., R. Sen, D. Baltimore, and P. A. Sharp. 1986. A

nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature (London) **319:**154–158.

- Staudt, L. M., R. G. Clerc, H. Singh, J. H. LeBowitz, P. A. Sharp, and D. Baltimore. 1988. Cloning of a lymphoid-specific cDNA encoding a protein binding the regulatory octamer DNA motif. Science 241:577-580.
- 40. Strum, R. A., and W. Herr. 1988. The POU domain is a bipartite DNA-binding structure. Nature (London) **336**:601–604.
- Treisman, J., P. Gönczy, M. Vashishtha, E. Harris, and C. Desplan. 1989. A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell 59:553-562.
- Verrijzer, C. P., A. J. Kal, and P. C. van der Vliet. 1990. The oct-1 homeo domain contacts only part of the octamer sequence and full oct-1 DNA binding activity requires the POU-specific domain. Genes Dev. 4:1964–1974.
- Watanabe, K., A. Saito, and T. Tamaoki. 1987. Cell-specific enhancer activity in a far upstream region of the human α-fetoprotein gene. J. Biol. Chem. 262:4812–4818.