Biochemical and Biophysical Research Communications **261,** 614–621 (1999)
Article ID bbrc.1999.1087, available online at http://www.idealibrary.com on IDEAL®

Human ZHX1: Cloning, Chromosomal Location, and Interaction with Transcription Factor NF-Y

Kazuya Yamada, Richard L. Printz, Haruhiko Osawa, and Daryl K. Granner[1]
Department of Molecular Physiology and Biophysics, 707 Light Hall, Vanderbilt University School of Medicine, Nashville, Tennessee 37232-0615

Received July 8, 1999

NF-YA, B, and C comprise the heterotrimeric transcription factor known as nuclear factor Y (NF-Y) or CCAAT-binding protein (CBF). NF-Y binds many CCAAT and Y box (an inverted CCAAT box, ATTGG) elements. Mutations of these elements that disrupt the binding of NF-Y result in decreased transcription from various tissue-specific and inducible promoters. We employed a yeast two-hybrid system to screen a human liver cDNA library in an effort to isolate proteins that interact with NF-Y and that may play a role in tissue-specific or hormone-inducible promoter activity. Using a fragment of the NF-YA subunit as bait we isolated a cDNA that encodes most of the open reading frame of the human zinc fingers and homeobox 1 (ZHX1) protein. The complete open reading frame was subsequently isolated and found to encode a protein of 873 amino acids that contains two zinc fingers and five homeodomain motifs. Northern blot analysis of poly(A)$^+$ RNA isolated from various tissues revealed two major *ZHX1* transcripts of about 4.5 and 5 kilobases. Both transcripts were expressed ubiquitously, although the 5-kilobase transcript is of greater abundance in most tissues examined. The human *ZHX1* gene is located on chromosome 8q, between markers CHCL.GATA50B06 and CHLC.GATA7G07.

CCAAT boxes are among the many *cis*-acting DNA elements known to be involved in the regulation of transcription in eukaryotic cells (1). Many tissue-specific or hormone-regulated promoters require the binding of the heterotrimeric CCAAT-binding protein (CBF) (1). The Y box element, originally defined as an inverted CCAAT box motif (5′-CTGATTGGYY-3′) in the major histocompatibility complex (MHC) class II genes, binds NF-Y and plays an important role in tissue-specific gene expression (2). The binding protein NF-Y is identical to CBF, and it stimulates gene transcription when bound to either a CCAAT or Y box (2). NF-Y, highly conserved among species, is composed of three subunits, referred to as NF-YA (CBF-B), NF-YB (CBF-A) and NF-YC (CBF-C), all of which are necessary for DNA binding (1, 3). Although the MHC class II genes are expressed in a cell-specific pattern, NF-Y is ubiquitously expressed (1), and it also binds to Y box motifs in non-class II genes, including those that encode thymidine kinase (4), albumin (5), α-collagen (6, 7), and interleukin 4 (8).

We have shown that a Y box, a CCAAT box and a CRE contribute to the basal promoter activity of the rat hexokinase II (HKII) gene promoter (9). NF-Y binds to the homologous Y and CCAAT boxes, whereas CREB/ATF-1 homo- and heterodimers bind to the CRE (9). Of these three basal elements, the CCAAT box and CRE are also involved in the cAMP response of the HKII gene (9). A Y box mediates cAMP responsiveness of the human tryptophan hydroxylase and fatty-acid synthase genes through an interaction with an unknown nuclear factor (10, 11). In addition, a CCAAT box mediates the cAMP response of the G-protein α_{i2} subunit gene, which lacks a typical CRE (12). Based on these observations, we postulate that NF-Y should be added to the list of factors involved in cAMP-mediated transcription.

There is no evidence to suggest that NF-Y is a phosphoprotein, thus the requirement for NF-Y in cAMP-mediated transcription may be a consequence of its ability to interact with other transcription factors. NF-YB interacts with Tax1, a potent activator of human T-cell lymphotropic virus type 1 transcription (13). NF-YB and NF-YC both interact with the TATA-binding protein (14). To further analyze the molecular mechanism of transcription activation by NF-Y, we first determined whether the NF-YA subunit interacts with either a known or a novel transcription factor. Accordingly, a human liver cDNA library was screened in an effort to isolate a protein that interacts with NF-YA in the yeast two-hybrid system. This resulted in the isolation of the cDNA for the human zinc fingers and homeobox 1 (ZHX1) protein. This putative tran-

[1] To whom correspondence should be addressed. Fax: 615-322-7236. E-mail: daryl.granner@mcmail.vanderbilt.edu.

0006-291X/99 $30.00

scription factor specifically interacts with NF-YA both *in vivo* and *in vitro*.

MATERIALS AND METHODS

Plasmid construction. A double-stranded oligonucleotide, consisting of a customized multiple cloning site with *Eco*RI and *Bam*HI cohesive ends (5′-AATTGCATATGGAATTCGCGGCCGCCCCGGGG-ATC-3′), was inserted into the *Eco*RI and *Bam*HI sites of a GAL4 DNA-binding domain (DBD) cloning vector, pGBT9 (Clontech; Palo Alto, CA). The resultant plasmid, pGBT9(N/R/N), has several cloning sites arranged in the following order: *Nde*I, *Eco*RI, *Not*I, *Sma*I, *Bam*HI, *Sal*I, and *Pst*I, respectively. A cDNA fragment that contains the open reading frame (ORF) of NF-YA with *Nde*I and *Eco*RI ends was amplified by the polymerase chain reaction (PCR) and ligated into the *Nde*I and *Eco*RI sites of pGBT9(N/R/N) to generate pGBT9(N/R/N)NF-YA. A deletion construct that encodes the first 269 amino acids of NF-YA was made by digesting pGBT9(N/R/N)NF-YA with *Afl*II and *Eco*RI to remove a 219-bp fragment. The cohesive ends were made blunt-ended using T4 DNA polymerase, and the plasmid was recircularized using blunt-end self ligation by standard techniques (15). This construct, pGBT9(N/R/N)YA1-269, lacks the sequence that encodes the subunit interaction domain and the DBD of NF-YA.

The *ZHX1* clone selected for study (#111), in the pACT2 vector, was digested with *Eco*RI and *Xho*I. A 0.4-kb *Eco*RI fragment and a 3.5-kb *Eco*RI–*Xho*I fragment were ligated into the corresponding *Eco*RI and *Eco*RI–*Xho* I sites of the pBluescriptII-SK(+) vector (Stratagene). These plasmids were named pBSII-hZHX1E and pBSII-hZHX1E/X. The 3.5 kb *Eco*RI–*Xho*I fragment of *ZHX1* cDNA (described above) was ligated into the *Eco*RI and *Xho*I sites of the pGEX-4T-2 vector (Pharmacia; Piscataway, NJ) to construct an expression vector for the glutathione-*S*-transferase fusion protein designated GST · ZHX1. The resultant plasmid was digested with *Eco*RI and the 0.4-kb *Eco*RI ZHX1 cDNA fragment (described above) was inserted into this vector to produce pGST · ZHX1#111. The insertion sites of this plasmid were determined by DNA sequencing to ensure that the ORF of the GST · ZHX1 fusion protein was maintained. cDNA fragments that contain the ORF of NF-YA and NF-YB with *Nde*I and *Eco*RI ends were amplified by PCR. These were subcloned into the *Nde*I and *Eco*RI sites of the pET-3a to obtain NF-YA · pET-3a and NF-Y · BpET-3b, respectively.

cDNA library screening. The yeast two-hybrid system, Yeastmaker, and the human liver Matchmaker cDNA library were purchased from Clontech. All procedures were performed as described by the manufacturer unless otherwise stated. The yeast strain HF7c, which has HIS3 and lacZ reporters, was transformed with either pGBT9(N/R/N)NF-YA or pGBT9(N/R/N)YA1-269 using a transformation method (16). The HF7c strain, harboring either the pGBT9(N/R/N)NF-YA or pGBT9(N/R/N)YA1-269 plasmid, was used to screen the human liver Matchmaker cDNA library. Approximately 1×10^6 primary transformants were obtained from the human cDNA library. These transformants were selected on SD-His-Trp-Leu plates to identify interactions of NF-YA with other proteins using the HIS3 reporter of the HF7c yeast strain. As the second confirmation of protein · protein interactions involving NF-YA, the lacZ reporter of the HF7c yeast strain was examined in pHis$^+$Trp$^+$Leu$^+$-clones by measuring β-galactosidase (β-gal) activity using a filter assay. Plasmid DNA was prepared from yeast clones that were both His$^+$Trp$^+$Leu$^+$- and β-gal activity positive. Each of these plasmids was transformed by electroporation into the *E. coli* strain, Electromax DH10B (GIBCO/BRL). As the second screen, HF7c harboring the NF-YA encoding plasmid was transformed with each of the cDNA clones to isolate reproducible His$^+$Trp$^+$Leu$^+$- and β-gal activity-positive clones. Clones that were positive for both reporter assays and both screens were further characterized.

Liquid β-galactosidase assays. The yeast strain SFY526 that contains a quantifiable lacZ reporter, and either the pGBT9(N/R/N) or pGBT9(N/R/N)YA1-269 plasmids, was transformed with clone #111 or the parent vector, pACT2. Quantitative β-galactosidase assays, using *o*-nitrophenyl-β-D-galactoside, were carried out on permeabilized cells, as described (15). Units of β-galactosidase activity were defined by the formula $(1000 \times OD_{420})/(t \times V \times OD_{600})$, where OD_{420} is the absorbance of *o*-nitrophenyl-β-D-galactoside at 420 nm; t is the length of incubation (time in min); V is the volume of the yeast culture from which the assayed extract sample was prepared; and OD_{600} is the absorbance at 600 nm of 1 ml of the yeast culture used to prepare the assayed extract.

Nucleic acid sequencing and sequence analysis. The DNA sequence of each clone was determined using a Sequenase kit (Amersham) and a custom-synthesized sequencing primer 5′-TACCACTAC-AATGGATG-3′) that corresponds to the GAL4-activation domain in the pACT2 plasmid. The sequence of the *ZHX1* cDNA was determined using a Sequenase kit (Amersham) or by an automated DNA sequencer, with commercially available sequencing primers or custom-synthesized primers using known *ZHX1* cDNA sequences. Nucleotide sequences from each positive clone were compared to those entered in the GenBank database using the BLAST sequence search and comparison program (17).

GST binding assay for protein · protein interactions. Topp3 cells (Stratagene) that contain the pGEX-4T-2 vector or a GST · ZHX1 fusion protein expression plasmid (pGST · ZHX1#111) were grown in LB media, and fusion protein expression was induced for 4 h at room temperature by the addition of IPTG to a concentration of 0.4 mM. Bacteria collected from a 20-ml culture were resuspended in 1 ml of lysis buffer (50 mM Tris–HCl, pH 8.0, 150 mM NaCl, 100 mM EDTA, 1% Triton X-100, 1 mM PMSF, 10 mM DTT), transferred to a 1.5-ml centrifuge, sonicated three times for 10 s and then centrifuged at 1500*g* for 5 min at 4°C. The supernatant was divided into 100-μl aliquots in 1.5 ml centrifuge tubes and diluted with the addition of 500 μl of lysis buffer. The diluted supernatant was gently mixed for 18 h at 4°C with 40 μl of a 50% slurry of glutathione–agarose (Sigma). The GST-fusion protein–glutathione agarose complex was pelleted by centrifugation at 3000*g* for 30 s at 4°C and washed three times by resuspending the complex in 500 μl of lysis buffer and repeating the centrifugation. The complex was then washed once with interaction buffer (10 mM Tris–HCl, pH 8.0, 50 mM NaCl, 5 mM EDTA, 0.1% Triton X-100, 1 mM PMSF, 1 mM DTT). Each protein to be tested for interaction with the GST-fusion protein was synthesized from a plasmid template and labeled with [^{35}S]methionine using the T7 TNT Quick-coupled Transcription/Translation System (Promega; Madison, WI). NF-YA · pET-3a and NF-YB · pET-3a plasmids were used as templates for the *in vitro* synthesis of labeled NF-YA and NF-YB. Ten microliters of each labeled protein and 200 μl of interaction buffer were used to resuspend the GST-fusion protein–glutathione agarose complex after the final wash. The protein binding/interaction reaction was carried out for 30 min at room temperature with continuous gentle mixing. The newly formed complex was then washed three times with 500 μl of interaction buffer and once with a solution of 10 mM Tris–HCl, pH 8.0, 50 mM NaCl, 5 mM EDTA. After the final wash, the complex was resuspended in 2× SDS–PAGE sample buffer and boiled for 5 min to elute the proteins from the agarose beads. The beads were pelleted by centrifugation and each supernatant was loaded on a 10% SDS–PAGE gel with a prestained molecular weight marker (Sigma). Enhanced detection of ^{35}S-labeled proteins was done as described (18). The gel was dried and exposed to Kodak XAR film at −70°C. The relative purity and amount of each fusion protein were determined by gel-staining with Coomassie brilliant blue.

Isolation of total RNA, reverse transcription and polymerase chain reaction. HepG2 cells, a human hepatoma cell line, were cultured as described previously (19). Total RNA was isolated from HepG2 cells using the acid guanidine-phenol–chloroform method (20). Reverse transcription was performed at 37°C for 1 h in a 50-μl reaction mixture that contained 50 mM Tris–HCl, pH 8.3, 50 mM KCl, 10 mM $MgCl_2$, 0.5 mM spermidine, 10 mM DTT, 35 units of RNase inhibitor

TABLE 1

Confirmation of the Specific Interaction between NF-YA and Clone #111

Plasmids		Growth on	
Bait	Library	SD-TRP-Leu plate	SD-His-Trp-Leu plate
pGBT9(N/R/N)	pACT2	Yes	No
pGBT9(N/R/N)	#111	Yes	No
pGBT9(N/R/N)YA1-269	pACT2	Yes	No
pGBT9(N/R/N)YA1-269	#111	Yes	Yes

(Promega), 1 mM dNTPs, 50 units of AMV reverse transcriptase (Promega), 10 μg of total RNA and 25 pmol of primer P1 (5′-CAAAACTACCATCAAAAGTCAG-3′). The polymerase chain reaction (PCR) was performed in a 100 μl reaction mixture that contained 10 mM Tris–HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin, 2 mM dNTPs, 10 pmol of primer P1 and primer N1 (5′-TGAGCTTTTCCTCATAACATC-3′), 10 μl of the reverse transcription reaction as a template, and 5 units of AmpliTaq DNA polymerase (Perkin–Elmer). The PCR was conducted at 94°C for 1 min, followed by 30 cycles of 94°C for 1 min, 50°C for 1 min and 72°C for 1 min. The PCR product obtained was ligated into the pGEM-T vector (Promega).

Poly(A)$^+$ RNA blot analysis. Human Multiple Tissue Northern and RNA Master Blots (Clontech) were hybridized with a 0.4 kb [α-^{32}P]dATP-labeled *Sal*I and *Eco*RI fragment of the *ZHX1* cDNA that was isolated from the pBSII-hZHX1E plasmid and labeled with the Prime-it II kit (Stratagene). Prehybridization, hybridization and washing procedures were performed according to the protocol provided by the supplier.

Genomic mapping of human ZHX1. The probe used for the RNA blots was hybridized to two BIOS Somatic Cell Hybrid Blots (BIOS-HindIII and BIOS-TaqI restricted), as suggested by the supplier (BIOS). These blots were exposed to Kodak XAR film at −70°C to visualize the chromosome fragments that contain *ZHX1*. A second localization approach involved the use of primer P1 (see above) and primer N2 (5′-CATTTATTGTTTGTTCAAAGATTG-3′) to screen the GeneBridge 4 Panel (Research Genetics). After an initial incubation at 94°C for 10 min, the subsequent PCR incubations were conducted at 94°C for 1 min, 55°C for 1 min, and 72°C for 2 min for a total of 39 cycles. Based on the *ZHX1* cDNA sequence, and the location of the primers in the same exon, the predicted product size is 476 bp. After the PCRs, the samples were run on 2% agarose gels and scored for the presence or absence of the predicted product. These data were submitted to the Whitehead Institute/MIT Center for Genome Research STS mapping server (Cambridge, MA), which provided the chromosomal location of *ZHX1*.

RESULTS

Isolation of human ZHX1 cDNA from a yeast two-hybrid library screen. pGBT9(N/R/N)NF-YA, a plasmid that encodes the entire sequence of NF-YA fused to the GAL4 DNA-binding domain (DBD), was used to screen a human liver cDNA library, by the yeast two-hybrid method. Yeast strain HF7c, which has dual *HIS3* and *lacZ* reporters for detecting protein · protein interactions when grown on SD-His-Trp-Leu plates, was the screening vehicle. Unfortunately, an unacceptably high background was obtained when HF7c cells were transformed with pGBT9 (N/R/N) NF-YA and pACT2 (the library vector plasmid). The C-terminal region of NF-YA contains the DBD and subunit-interaction domains (21, 22), which may allow the GAL4DBD · NFYA protein to interact with HAP3 and HAP5, the yeast homologs of NF-YB and NF-YC, thus creating the high background. To circumvent this possibility, we constructed pGBT9(N/R/N)YA1-269, a plasmid that contains an NF-YA cDNA fragment encoding amino acids 1–269 and lacking the NF-YA DBD and subunit-interaction domains. The HF7c yeast strain containing this plasmid, and pACT2, did not show background activity on an SD-His-Trp-Leu plate. We then proceeded to screen a human liver cDNA library

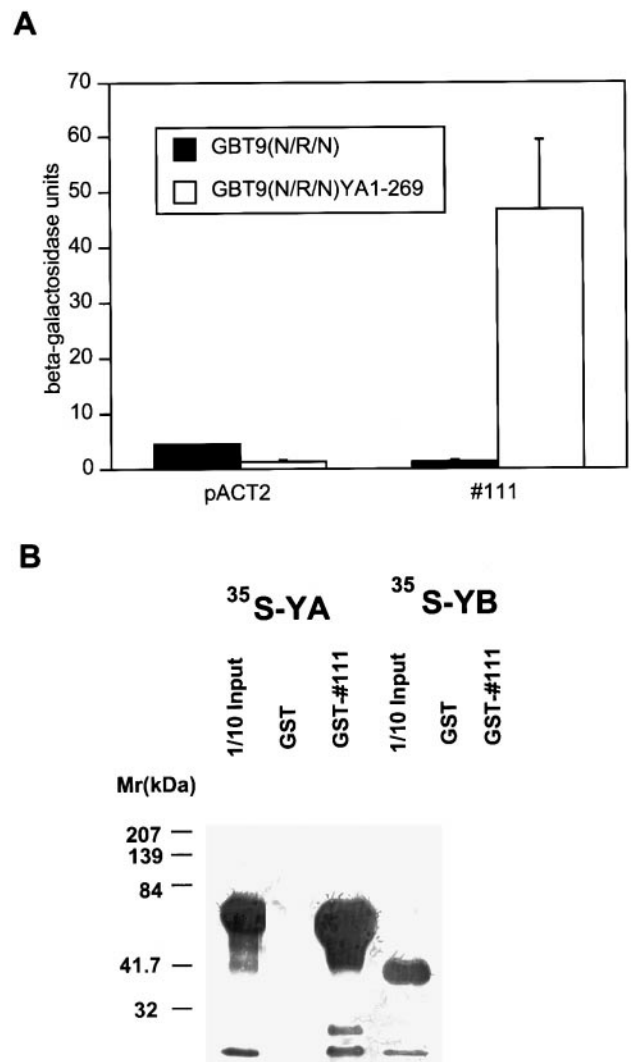
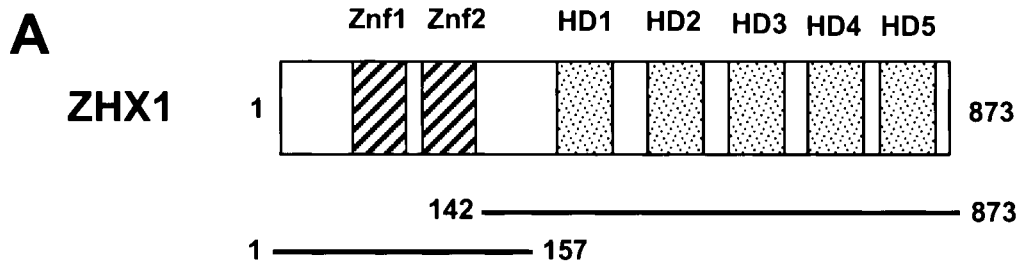


FIG. 1. Determination of a specific interaction between NF-YA and ZHX1 #111 protein using the yeast two-hybrid system or a GST binding assay. (A) β-galactosidase activity was determined in yeast containing the combination of bait plasmids and library plasmids. (B) *In vitro* translated, ^{35}S-labeled, NF-YA or NF-YB were incubated with agarose beads containing bound GST or the GST · ZHX1 #111 (GST-#111) fusion protein. The beads were washed and the bound protein was eluted and analyzed by SDS–PAGE and autoradiography. The signal in the lane marked "Input" represents 10% of the protein added to the reactions shown in the other lanes.



B

```
human ZHX1 MASRRKSTTP CMVLASEQDP DLELISDLDE GPPVLTPVEN TRAESISSDE EVHESVDSDN  60
mouse Zhx1 ---------- ---------- ---------- ---I------ AK---V---- ---G------

                   * ********** ********** ***          ********* **********
human ZHX1 QQNKKVEGGY ECKYCTFQTP DLNMFTFHVD SEHPNVVLNS SYVCVECNFL TKRYDALSEH 120
mouse Zhx1 ---------- ---------- ---------- ---------- ---------- ----------

           *****
human ZHX1 NLKYHPGEEN FKLTMVKRNN QTIFEQTIND LTFDGSFVKE ENAEQAESTE VSSSGISISK 180
mouse Zhx1 ---------- ---------- ---------- ---------- --T--G--ID ----------

human ZHX1 TPIMKMMKNK VENKRIAVHH NSVEDVPEEK ENEIKPDREE IVENPSSSAS ESNTSTSIVN 240
mouse Zhx1 ---------- ------T--- --A-GTS--- --GV-ASQ-- NA-SV----L -------TI-

human ZHX1 RIHPSTASTV VTPAAVLPGL AQVITAVSAQ QNSNLIPKVL IPVNSIPTYN AALDNNPLLL 300
mouse Zhx1 -V---P---- ---T------ ---------- -----L---- ---------- ----------

human ZHX1 NTYNKFPYPT MSEITVLSAQ AKYTEEQIKI WFSAQRLKHG VSWTPEEVEE ARRKQFNGTV 360
mouse Zhx1 ---------- ---------- ---------- ---------- ---------- ----------

human ZHX1 HTVPQTITVI PTHISTGSNG LPSILQTCQI VGQPGLVLTQ VAGTNTLPVT APIALTVAGV 420
mouse Zhx1 ---------- ----T----- ---------- ---------- ---------- ----------

human ZHX1 PSQNNIQKSQ VPAAQPTAET KPATAAVPTS QSVKHETALV NPDSFGIRAK KTKEQLAELK 480
mouse Zhx1 -N-T-V---- ------ATD- --------S- P--RP-A--- ---------- ----------

human ZHX1 VSYLKNQFPH DSEIIRLMKI TGLTKGEIKK WFSDTRYNQR NSKSNQCLHL NNDSSTTIII 540
mouse Zhx1 ---------- ---------- ---------- ---------- ---------- -----A----

human ZHX1 DSSDETTESP TVGTAQPKQS WNPFPDFTPQ KFKEKTAEQL RVLQASFLNS SVLTDEELNR 600
mouse Zhx1 ------P-P- AARAS-Q--- -------A-- ---------- -A-------- -------V--

human ZHX1 LRAQTKLTRR EIDAWFTEKK KSKALKEEKM EIDESNAGSS KEEAGETSPA DESGAPKSGS 660
mouse Zhx1 ---------- ---------N -T----D--I -V----V--- ---P--S--G --TV-----G

human ZHX1 TGKICKKTPE QLHMLKSAFV RTQWPSPEEY DKLAKESGLA RTDIVSWFGD TRYAWKNGNL 720
mouse Zhx1 ---------- ---------- ------A--- ----E----- ---------- ----------

                              .......... .......
human ZHX1 KWYYYYQSAN SSSMNGLSSL RKRGRGRPKG RGRGRPRGRP RGSKRINNWD RGPSLIKFKT 780
mouse Zhx1 --------S- ---L------ -R-------- ---------- --G--M-T-- -V--------

                                                    +  ++++++++++
human ZHX1 GTAILKDYYL KHKFLNEQDL DELVNKSHMG YEQVREWFAE RQRRSELGIE LFEENEEEDE 840
mouse Zhx1 ---------- ---------- -----R---- ---------- ---------- ----------

           ++++++++++ +          .......
human ZHX1 VIDDQEEDEE ETDDSDTWEP PRHVKRKLSK SDD                              873
mouse Zhx1 -V-------- ---------- ------NV-- ---
```

FIG. 2. Schematic diagram of human *ZHX1* cDNA and a comparison of the ZHX1 amino acid sequences of human and mouse. (A) A schematic representation of human ZHX1 is presented. The two zinc finger domains (Znf1 and Znf2) and the five homeodomains (HD1–HD5) are illustrated. Diagram 142–873 represents the amino acids included in clone #111. Amino acids 1–157 were obtained by PCR, as described under Materials and Methods. (B) The ZHX1 amino acid sequences of human and mouse are compared. The dashes indicate amino acid identity. The two zinc fingers (Znf1 and Znf2) and five homeodomains are indicated by the asterisks and underlines, respectively. The acidic region and putative nuclear localization signals are marked with plus signs and closed circles (●), respectively.

using pGTB9(N/R/N)YA1-269 as the bait. We obtained 123 putative clones using the *HIS3* reporter in the initial screen, but only 12 clones showed β-galactosidase activity from the *lacZ* reporter. These 12 clones were subjected to a second screen. Five clones exhibited reproducible $His^+Trp^+Leu^+$ and β-galactosidase activity.

We determined the nucleotide sequence of these five cDNA clones starting from the region just 3′ of the GAL4 activation domain (AD) coding sequence. A translation stop codon is located near the adapter sequence used for cDNA synthesis in clones, #9 and #109. The other three clones, #111, #112, and #120, have a long open reading frame inframe with the GAL4AD coding sequence. The nucleotide sequences of clones #112 and #120 are identical. These clones are currently undergoing further analysis and are not the topic of this manuscript. The DNA sequence of clone #111 was compared with the GenBank database using the BLAST search program. The deduced amino acid sequence is nearly identical with that of the mouse zinc fingers and homeobox 1 (Zhx1) protein (23) and the hamster very divergent homeobox (vdx) protein (24), which are themselves nearly identical. Therefore, we designated clone #111 as a human homolog of *Zhx1*.[2]

ZHX1 interacts with NF-YA. The specificity of the interaction between NF-YA and the ZHX1 #111 protein was tested using two different yeast two-hybrid assays and a GST binding assay. The first yeast two hybrid test employs a HIS3 reporter. pGBT9(N/R/N), which encodes the GAL4 DNA-binding domain and the TRP1 gene, and pACT2, which encodes the GAL4-activation domain and the LEU2 gene, are the parent vectors for this two-hybrid system. When the yeast HF7c strain containing pGBT9(N/R/N) was transformed with pACT2, or clone #111, transformants grew on SD-Trp-Leu plates but not on SD-His-Trp-Leu plates (Table 1). However, when HF7c, which contains the pGBT9(N/R/N)YA1-269 plasmid, was transformed with pACT2 or clone #111, transformants grew on SD-Trp-Leu plates, but only yeast transformed with clone #111 grew on a SD-His-Trp-Leu plate (Table 1).

The specificity of the interaction between NF-YA and the ZHX1 #111 protein was next quantitated with a liquid β-galactosidase assay. A yeast strain (SFY526) with a quantitative lacZ reporter gene was used. As shown in Fig. 1A, only the yeast SFY526 strain harboring both the pGBT9(N/R/N)YA1-269 and clone #111 exhibited high levels of β-galactosidase activity, ~10-fold greater than that noted from a vector-only combination.

Finally, we performed a GST pull-down assay to confirm the specific interaction between the nontruncated form of NF-YA and the ZHX1 #111 protein. The ZHX1 #111 protein was expressed in *E. coli* as a GST fusion protein. GST and the GST · ZHX1 #111 fusion protein were immobilized on glutathione agarose beads. *In vitro*-translated S^{35}-labeled NF-YA or NF-YB were incubated with this immobilized fusion protein–bead complex. Bound S^{35}-labeled NF-YA or NF-YB was eluted and analyzed by SDS/PAGE and autoradiography. NF-YA interacted with GST · ZHX1 #111 but not GST alone, and NF-YB did not interact with either GST or GST · ZHX1 #111 (Fig. 1B). The results from these three experiments suggest that NF-YA and ZHX1 #111 specifically interact *in vitro* and *in vivo.*

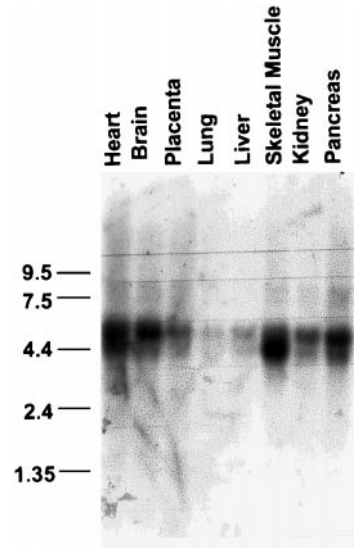


FIG. 3. Tissue distribution of human *ZHX1* mRNA. A human multiple tissue northern blot (Clontech) was hybridized with labeled *ZHX1* cDNA, as described above. Each lane contains 2 μg of poly A^+-RNA isolated from the indicated tissues. Size markers are shown on the left in kb.

Sequence analysis of the human ZHX1 cDNA and its deduced amino acid sequence. The DNA sequence of clone #111 contains a poly(A) stretch of 116 nucleotides at the 3′-terminus of the cDNA that most likely corresponds to a portion of the poly$(A)^+$ tail of the *ZHX1* mRNA. However, the 5′-terminus of cDNA #111 has a sequence that corresponds to the codon for the threonine at position 142 of mouse *Zhx1.* The human and mouse cDNA sequences are found in several human expressed sequence tags (EST) database entries. Of particular interest are the entries T86901, W60281, and W60373, which appear to be homologous to the mouse *Zhx1* cDNA and which contain 5′-noncoding sequence and the first 250 bp of coding sequence. Using the combination of a primer (N1) derived from common nucleotide sequences in T86901, W60281, and W60373, and a primer (P1) derived from the clone #111 sequence, a cDNA fragment was obtained by RT-PCR using RNA isolated from HepG2 human hepatoma cells as a template. This allowed for the isolation of the 5′-noncoding sequence and the portion of the *ZHX1* cDNA ORF that is missing in the #111 clone. The complete amino acid sequence was deduced from the combination of the sequences derived from this PCR fragment and clone #111.[3] The full-length human

[2] This symbol has been approved by the HUGO nomenclature committee with the name zinc fingers and homeobox 1.

[3] The nucleotide sequence of the *ZHX1* cDNA was submitted to GenBank Accession No. AF106862.

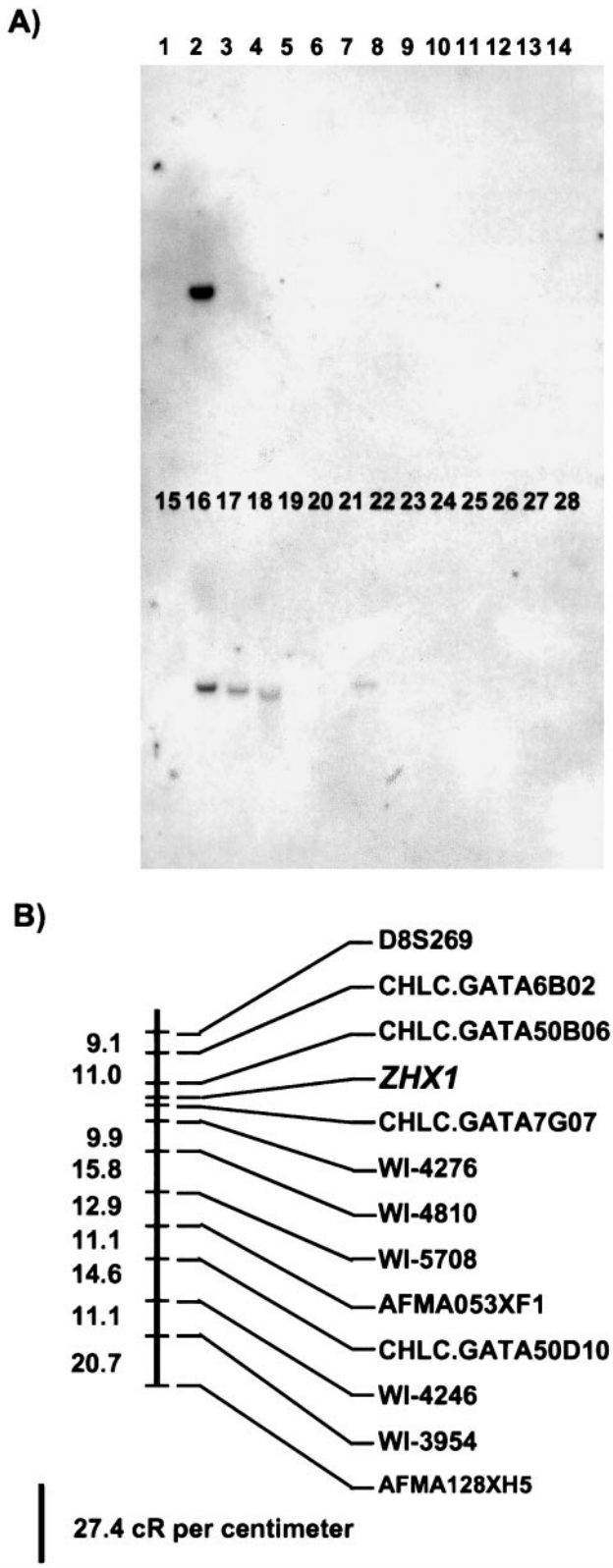


FIG. 4. Chromosomal location of *ZHX1*. (A) A Southern blot analysis of a HindIII-digested murine-human somatic cell hybrid panel was performed using a ^{32}P-labeled *ZHX1* cDNA probe. The probe hybridized to human control DNA (lanes, 2 and 16) and hybrid cell lines 803 (lane 17), 811 (lane 18), and 909 (lane 21). These all contain fragments of chromosome 8. (B) A PCR analysis with *ZHX1* specific primers of the GeneBridge 4 panel (see Materials and Methods) yielded the scoring code data, 00100000010010000000000011010000001101000110001101100001012101000101111110000200110111000000, which was submitted to the Whitehead Institute/MIT Center for Genome Research STS mapping server. The chromosome location identified is illustrated as a portion of chromosome 8q, with *ZHX1* placed on the radiation hybrid framework map. The numbers along the left of this chromosome map indicate the distance between markers shown on the right. 1 cR = 1% frequency of breakage between the markers. The data were obtained from Whitehead Institute Center for Genome Research and are available at http://www-genome.wi.mit.edu.

ZHX1 cDNA encodes a protein of 873 amino acids, which is shown in Fig. 2 in comparison with mouse Zhx1. The human ZHX1 protein has a predicted molecular mass of 98 kilodaltons (kDa) and an isoelectric point of 5.7. The human ZHX1 sequence contains motifs for two C-x-x-C-(x)12-H-x-x-x-x-H zinc fingers and five homeodomains, two nuclear localization signals starting at amino acid positions 740 and 860 (see Fig. 2), and a C-terminal cluster of acidic amino acids (830 to 851) that may act as a transcription activation domain. These characteristics suggest that ZHX1 may function as a transcription factor, however, we have no evidence to suggest that ZHX1 binds to DNA.

Tissue distribution of ZHX1 mRNA. The tissue distribution of human *ZHX1* mRNA was determined by northern blot analysis (Fig. 3). Two major *ZHX1* transcripts of about 4.5 and 5 kb are found in all tissues examined, however, the ratio and intensity of these transcripts varies among these tissues. The 5-kb transcript is highly expressed in heart, brain, pancreas, kidney, placenta and skeletal muscle, and in lower amounts in lung and liver. The 4.5 kb transcript is highly expressed in skeletal muscle, heart and pancreas. Although the mRNA species were not distinguished on the RNA master dot blot (see Materials and Methods), *ZHX1* mRNA is expressed in all tissues examined (data not shown).

Chromosomal localization of human ZHX1 gene. The location of human *ZHX1* was determined by Southern blot analysis of somatic cell hybrids and by PCR screening of a radiation hybrid panel. As shown in Fig. 4A, hybridization with the *ZHX1* cDNA probe was observed in lanes 2 and 16 from human genomic DNA, but no hybridization occurred with hamster DNA (lanes 1 and 15). Hybridization signals also were noted in lanes 17, 18, and 21, which contain genomic DNA from hamster somatic cell hybrids that contain some portion of human chromosome 8. No signals were observed in lanes lacking DNA from chromosome 8.

As a second independent approach, the radiation hybrid panel GeneBridge 4 was screened by PCR to localize *ZHX1* to a particular chromosome (see Materials and Methods). The presence or absence of the

expected 476 bp *ZHX1* PCR fragment was scored for each DNA sample. Analysis of this data indicated that *ZHX1* is located between CHCL.GATA50B06 and CHLC.GATA7G07 on chromosome 8q (Fig. 4B).

DISCUSSION

The mouse *Zhx1* cDNA was originally isolated by expression screening with a monoclonal B92 antibody (23). This antibody recognizes a 100-kDa nuclear protein in most tissues, and a 49-kDa nuclear protein in cells of hematopoietic lineage, and it is particularly abundant in stem cells (25). It is unclear whether the 49-kDa protein represents a closely related protein or is a processed form of the 100-kDa protein. The predicted molecular mass of human ZHX1 is 98 kDa, which is consistent with that of the protein recognized by the monoclonal B92 antibody. The *ZHX1* coding sequences of human and mouse are 89% identical at the nucleotide level and 91% identical at the amino acid level, and the two zinc finger and five homeodomain sequences are conserved. The human *ZHX1* cDNA and the *vdx* cDNA of the hamster pancreas have very similar nucleotide sequences, however, the latter has not been completely analyzed, so it is not possible to determine whether *vdx* cDNA is the hamster homolog of *ZHX1* (24).

The *ZHX1* mRNA is ubiquitously expressed in adult and fetal tissues, but there are differences in tissue expression, as noted above. *ZHX1* is expressed as two major transcripts of 5 and 4.5 kb. The location of *ZHX1* to a single location on human chromosome 8 suggests that these transcripts arise from a single gene. It remains to be determined whether these transcripts result from different transcription initiation sites, from alternative processing of the primary transcript, or from the utilization of different polyadenylation sequences in the 3′ noncoding sequence. This may also account for the observed difference in the sizes of the mouse *Zhx1* transcripts, compared to the human forms (23).

Transcription factors often have zinc fingers or homeodomains, and some have both of these features. For example, Isl-1 has two zinc fingers and a homeodomain (26) and α-fetoprotein enhancer-binding protein, ATBF1, has seventeen zinc fingers and four homeodomains (27). Although homeodomain-containing proteins are known to bind to AT-rich DNA sequences, we have no evidence that ZHX1 recognizes AT-rich sequences, or even that it binds to DNA. We have shown that ZHX1 specifically interacts with NF-YA, a subunit of the DNA binding protein NF-Y. This protein · protein interaction does not involve the zinc finger regions of ZHX1, since these domains are not present in the protein (from the #111 clone) used in the original yeast two hybrid isolation system. ZHX1 may therefore not be a DNA binding protein, especially if it acts as a transcription co-regulator with NF-Y.

NF-Y plays a role in sterol-dependent regulation of the farnesyl diphosphate synthase and 3-hydroxy-3-methylglutaryl-coenzyme A synthase genes (28), in cell cycle-dependent transcription of the human cdc25C gene (29) and in basal and hormone-regulated expression of the hexokinase II gene (9). NF-Y and Sp1-binding sites may be functionally synergistic in the fatty acid synthase and cdc25C genes (30). In this regard, it is of interest to note that Sp1 and NF-YA also form protein · protein complexes (30). NF-Y, Sp1 and ZHX1 are all ubiquitously expressed, so could be involved in the formation of a highly organized nucleoprotein complexes that participate in the regulation of many genes.

ACKNOWLEDGMENTS

This study was supported by NIH Grant DK46867, the Vanderbilt Diabetes Research and Training Center (DK20593), and the Vanderbilt Cancer Center (CA68485).

REFERENCES

1. Maity, S. N., and de Crombrugghe, B. (1998) *Trends Biochem. Sci.* **23,** 174–178.
2. Abdulkadir, S. A., and Ono, S. J. (1995) *FASEB J.* **9,** 1429–1435.
3. Sinha, S., Maity, S. N., Lu, J., and de Crombrugghe, B. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 1624–1628.
4. Arcot, S. S., Flemington, E. K., and Deininger, P. L. (1989) *J. Biol. Chem.* **264,** 2343–2349.
5. Maire, P., Wuarin, J., and Schibler, U. (1989) *Science* **244,** 343–346.
6. Hasegawa, T., Zhou, X., Garrett, L. A., Ruteshouser, E. C., Maity, S. N., and de Crombrugghe, B. (1996) *Nucleic Acids Res.* **24,** 3253–3260.
7. Maity, S. N., Golumbek, P. T., Karsenty, G., and de Crombrugghe, B. (1988) *Science* **241,** 582–585.
8. Li-Weber, M., Davydov, I. V., Krafft, H., and Krammer, P. H. (1994) *J. Immunol.* **153,** 4122–4133.
9. Osawa, H., Robey, R. B., Printz, R. L., and Granner, D. K. (1996) *J. Biol. Chem.* **271,** 17296–17303.
10. Boularand, S., Darmon, M. C., Ravassard, P., and Mallet, J. (1995) *J. Biol. Chem.* **270,** 3757–3764.
11. Rangan, V. S., Oskouian, B., and Smith, S. (1996) *J. Biol. Chem.* **271,** 2307–2312.
12. Kinane, T. B., Shang, C., Finder, J. D., and Ercolani, L. (1993) *J. Biol. Chem.* **268,** 24669–24676.
13. Pise-Massion, C. A., Dittmer, J., Clemens, K. E., and Brady, J. N. (1997) *Mol. Cell. Biol.* **17,** 1236–1243.
14. Bellorini, M., Lee, D. K., Dantonel, J. C., Zemzoumi, K., Roeder, R. G., Tora, L., and Mantovani, R. (1997) *Nucleic Acids Res.* **25,** 2171–2181.
15. Ausubel, F. M., Brent, C. R., Kingston, R., Moore, D., Seidman, J. J., Smith, J., and Struhl, K. (1994) Current Protocols in Molecular Biology, Wiley, New York.
16. Yamada, K., Wang, J. C., Osawa, H., Scott, D. K., and Granner, D. K. (1998) *Biotechniques* **24,** 596–598, 600.
17. Madden, T. L., Tatusov, R. L., and Zhang, J. (1996) *Methods Enzymol.* **266,** 131–141.

18. Osawa, H., Printz, R. L., Whitesell, R. R., and Granner, D. K. (1995) *Diabetes* **44,** 1426–1432.
19. O'Brien, R. M., Lucas, P. C., Yamasaki, T., Noisin, E. L., and Granner, D. K. (1994) *J. Biol. Chem*. **269,** 30419–30428.
20. Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem*. **162,** 156–159.
21. Sinha, S., Kim, I.-S., Shon, K.-Y., de Crombrugghe, B., and Maity, S. N. (1996) *Mol. Cell. Biol.* **16,** 328–337.
22. Kim, I.-S., Sinha, S., de Crombrugghe, B., and Maity, S. N. (1996) *Mol. Cell. Biol.* **16,** 4003–4013.
23. Barthelemy, I., Carramolino, L., Gutierrz, J., Barbero, J. L., Marquez, G., and Zaballos, A. (1996) *Biochem. Biophys. Res. Commun*. **224,** 870–876.
24. Rudnick, A., Ling, T.-H., Odagiri, H., Rutter, W. J., and German, M. S. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 12203–12207.
25. Lee, B.-C., Shac-Tal, Y., Peled, A., Gothelf, Y., Jiang, W., Toledo, J., Polemancher, R. R., Haran-Ghera, N., and Zipori, D. (1996) *Blood* **87,** 2283–2291.
26. Karlsson, O., Thor, S., Norberg, T., Ohlsson, H., and Edlund, T. (1990) *Nature* **344,** 879–882.
27. Morinaga, T., Yasuda, H., Hashimoto, T., Higashino, K., and Tamaoki, T. (1991) *Mol. Cell. Biol.* **11,** 6041–6049.
28. Jackson, S. M., Ericsson, J., Osborne, T. F., and Edwards, P. A. (1995) *J. Biol. Chem*. **270,** 21445–21448.
29. Zwicker, J., Gross, C., Lucibello, F. C., Truss, M., Ehlert, F., Engeland, K., and Muller, R. (1995) *Nucleic Acids Res.* **23,** 3822–3830.
30. Roder, K. H., Wold, S. S., and Schweizer, M. (1997) *Biochem. Soc. Trans.* **25,** 72S.