Nucleic Acids Research, Vol. 20, No. 4 797–803

Tissue restricted expression and chromosomal localization of the *YB-1* gene encoding a 42kD nuclear CCAAT binding protein

Dimitri D.Spitkovsky+, Brigitte Royer-Pokora[1], Hajo Delius, Fjodor Kisseljov[2], Nancy A.Jenkins[3], Debra J.Gilbert[3], Neal G.Copeland[3] and Hans-Dieter Royer*
Angewandte Tumorvirologie, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 506, D 6900 Heidelberg, [1]Institut für Humangenetik und Anthropologie der Universität Heidelberg, FRG, [2]All Union Cancer Research Center, 115478 Moscow, Russia and [3]Mammalian Genetics Laboratory, ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, MD 21702, USA

Received November 11, 1991; Revised and Accepted January 24, 1992

ABSTRACT

YB-1 cDNA clones were isolated by binding site screening of a Hela expression library using a human papillomavirus type 18 enhancer oligonucleotide. YB-1 belongs to a family of transcription factors which bind to recognition sequences containing a core CCAAT element. YB-1 bound to its single stranded recognition sequence on the sense strand but not to the anti-sense strand. A synthetic peptide antiserum derived from the predicted YB-1 amino acid sequence identified a 42kD nuclear protein in immunoblots. A protein with the same size was detected by binding site blotting experiments using the HPV18 enhancer oligonucleotide which bound YB-1. *YB-1* gene expression was restricted in tissues from a human 24 week old fetus. High levels of YB-1 mRNA were present in heart, muscle, liver, lung, adrenal gland and the brain, in contrast, low amounts of YB-1 mRNA were found in thymus, kidney, bone marrow and spleen. In pancreas, bladder, stomach and testis YB-1 mRNA could not be detected by Northern hybridization. Finally, we have identified four *YB-1* related loci in the mouse genome and have mapped these loci to four different mouse chromosomes by interspecific backcross analysis.

INTRODUCTION

The CCAAT box is a sequence element which is present in many gene control regions. Multiple CCAAT box binding proteins exist: CP1/NFY (1, 2), CP2 (3), CTF/NF1 (4), C/EBP (5), CBF (6), YB-1/dbpB, dbpA (7, 8) and H1TF2 (9), and sequence comparison of cDNA clones encoding CP1/NFY (10), CTF/NF1 (11), C/EBP (12), CBF (6), dbpA and YB-1/dbpB (7, 8) shows that these factors are not derived from one large gene family but are distinct. A common feature between these factors is the recognition of CCAAT boxes. CP1/NFY, CP2, and CTF/NF1 bind to consensus sequences (1); CP1/NFY: A/G A/G CCAAT, CP2: A/G RCCAAT, CTF/NF1: NGCCAAN (half site). CBF and YB-1/dbp1 are also sequence specific DNA binding proteins recognizing the CCAAT box because mutations in the CCAAT sequence abolishes DNA binding (6, 7). C/EBP binds to the CCAAT motif and to a GCAAT motif with high affinity (5).

CTF/NF1 consists of a series of polypeptides ranging in relative molecular mass (Mr) from 52000 to 66000 (13). The different CTF/NF1 polypeptides are encoded by multiple messenger RNAs containing alternative coding regions, apparently as a result of differential splicing (11). CTF/NF1 are sequence specific transcription initiation proteins (4, 13) and are also involved in initiating Adenovirus DNA replication (14). C/EBP is a 42kD protein (15) which displays several interesting functions: It activates transcription of genes in the liver and in adipocytes where it is involved in regulating energy metabolism (16), it is only expressed in terminally differentiated cells, which are growth arrested (17) and it is directly involved in inducing terminal differentiation and growth arrest (17). NFY/CP1 (1, 2) binds as heterodimer to its recognition sequence which was originally identified in murine major histocompatibility (MHC) class-II gene promoters (2). The NFY subunits NFY-A and NFY-B are 42kD and 35kD proteins (10) which are functionally and structurally related to the *Saccharomyces cerevisiae* HAP2 and HAP3 gene products (3, 10). NFY proteins act as transcriptional activators for various genes (10). CCAAT binding proteins which have not been characterized in detail are responsible for cell cycle regulated expression of the human thymidine kinase gene (18), and are responsible for serum regulated expression of the human heat

* To whom correspondence should be addressed

+ On leave from the All Union Cancer Research Center, Moscow, Russia

shock (Hsp 70) gene (19). The histone H1 gene regulatory region is characterized by a histone H1 subtype specific consensus element which contains a CCAAT motif. This motif binds a 47kD protein (H1TF2) which activates H1 histone gene transcription in vitro (9).

YB-1/dbpB has been identified by binding site screening of lambda gt11 expression libraries with a fragment from the epidermal growth factor receptor gene (*EGF*) enhancer (8) and the major histocompatibility complex class-II Y box containing an inverted CCAAT box (7). Little is known about the function, tissue type expression and the nature of cellular YB-1 protein. In the present communication we have addressed some of these questions. We have identified a YB-1 recognition site in the human papillomavirus type 18 (HPV18) enhancer. We have isolated YB-1 cDNA clones with a HPV18 enhancer oligonucleotide containing this site. We have identified YB-1 protein in Hela cells with a peptide antiserum derived from the predicted primary aminoacid sequence from YB-1 cDNA. We have determined the expression of the *Yb-1* gene in different tissues of a human 24 week old fetus. YB-1 cDNA was used in mouse chromosomal mapping studies that identified four independent loci on different mouse chromosomes.

MATERIALS AND METHODS

Cell culture and cell lines

Human fibroblasts, Hela and Cgl3 cells (20) were maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal calf serum.

Extraction of Nuclear and Cytoplasmic Proteins

Nuclear extracts of human fibroblasts, Hela and Cgl3 cells were prepared according to our published procedure (21). After detergent lysis with 0.65% NP40 nuclei were prepared by low speed centrifugation at 4°C and proteins were eluted with 530mM NaCl with slight agitation. For storage the eluates were dialyzed against a buffer containing 50% glycerol,50mM NaCl, 10mM Hepes (pH7.9), 0.5mM PMSF and 0.5mM DTT. Protein concentrations were determined by a colorimetric assay (Biorad) using serum albumin as a standard. Cytoplasmic proteins were prepared after cell lysis and removal of nuclei by low speed centrifugation. Cytoplasmic proteins were dialyzed as described above for nuclear proteins.

Synthetic oligonucleotides

Single stranded oligonucleotides and the complementary strands were synthesized on Applied Biosystems DNA synthesizers and purified by preparative denaturing acrylamide gel electrophoresis. Full length bands were identified by UV shadowing, excised, eluted by diffusion in 500mM ammoniumacetate and ethanol precipitated. To generate double stranded oligonucleotides complementary strands were annealed at temperatures minus 3°C below the specific melting points (Tm). Radiolabeled double stranded oligonucleotides were prepared with polynucleotide kinase and ^{32}P-gamma ATP. The sequence of HPV18 enhancer oligonucleotide RP3 has been described (21). The mutant long Y-box oligonucleotide: 5'-ATTTTTCTGCTGGGCCAAA-G-3'was selected from a published sequence (7), the position of the mutated inverted CCAAT motif is underlined.

Binding Site Blotting

Nuclear proteins were size fractionated on 8% SDS polyacrylamide gels and transferred to nitrocellulose at 150 mA overnight at room temperature in a horizontal blotting chamber (IBI) with a buffer containing 192 mM glycine and 25 mM Tris (pH 8.3). After blotting bound proteins were denatured in situ by 6 M guanidinium hydrochloride and renatured by sequential dilution of guanidinium as described (21, 22). Nitrocellulose membranes were blocked with 5% nonfat dry milk (Carnation) for 30 min at room temperature and subsequently incubated with end labeled double stranded oligonucleotides at 5×10^5 cpm/ml in the presence of 5 μg/ml poly(dIdC)(dIdC) as nonspecific competitor. The buffers for DNA binding and conditions for washing binding site blots were as described (21).

Electrophoretic mobility shift analysis (EMSA)

EMSA was carried out as described (21). The incubations of radiolabeled oligonucleotides with YB-1 fusion protein contained 3 μg IPTG induced E.coli extract and 2.5 μg poly(dIdC)(dIdC) as nonspecific competitor. For EMSA oligonucleotides were incubated with the protein in binding buffer (21), for 25 min at room temperature and loaded on a low salt polyacrylamide gel (4% total monomer, 30:1 acrylamide/N,N'-methylenebisacrylamide ratio). Electrophoresis was at 11 V/cm for 90 min at room temperature. Gels were dried and exposed to X ray film overnight.

Isolation of YB-1 cDNA from a Hela expression library

The Hela expression library in vector lambda gt11 was obtained from P.Angel. The library was plated on square plates (Nunc, bioassay dish) at a density of 2×10^5 pfu per plate on E.coli Y1090 until small plaques appeared, usually 6–8 hours after plating. Fusion proteins were induced by placing an IPTG (isopropylthiogalactoside) soaked nitrocellulose membrane on top of the soft agar layer for four hours. Nitrocellulose filters were denatured and renatured with guanidinium hydrochloride as described (22). For binding site screening membranes were blocked with 5% dry milk for 30 min at room temperature and subsequently incubated with 5×10^5 cpm/ml radiolabeled RP3 for two hours. Membranes were washed and exposed to X ray film. All positive signals were plaque purified twice. Lambda DNAs were prepared from liquid lysates and analyzed after EcoRI digestion and agarose gel electrophoresis

Induction of β-galactosidase-YB-1 fusion protein

Escherichia coli strain Y1090 was infected with clones lambda gt11-N4 or lambda gt11-N19 and grown at 37°C until partial lysis was achieved. Isopropyl thiogalactoside (IPTG) was added to a final concentration of 10 mM, and the incubation was continued for one hour. Phenylmethylsulfonyl fluoride (PMSF) was added to a final concentration of 200 μg/ml, and the incubation was continued for another 30 min at 37°C. The bacteria were centrifuged at $4550\times g$ for 5 min at 4°C. The supernatant was used.

Isolation of RNA from fetal and adult tissues

RNA from fetal and adult tissues was isolated in one step by the guanidinium thiocyanate extraction method (23). For RNA extraction frozen tissues were pulverized in liquid nitrogen using a mortar and pestle. Frozen tissue powder was lysed with

guanidiniumisothiocyanate, filtered through cheesecloth, and centrifuged over a 5.7M CsCl cushion. The RNA pellet was solubilized in water and stored frozen.

Northern hybridization

For Northern analysis 5 μg of total RNA per sample were size separated on 1% agarose gels in the presence of 2.2 M formaldehyde and transferred with 10×SSC Gene Screen membranes (Du Pont) and UV crosslinked with a Stratalinker (Stratagene). Hybridization conditions were 42°C, 50% formamide, 5×SSC, and 2.5×10^6 cpm/ml of radiolabeled YB-1 cDNA. For radiolabeling a 1 kB long EcoRI fragment of clone N4 containing most of the YB-1 open reading frame was isolated by preparative gel electrophoresis and labeled with radioactive triphosphates using a random primer labeling kit (Amersham). The actin probe was radiolabeled as described above using a human β-actin cDNA clone.

DNA sequencing and computer analysis

cDNA was excised from lambda gt11 phage by EcoRI digestion and subcloned into pBluescript II KS+ (Stratagene). Sequencing reactions were performed by double-stranded dideoxynucleotide sequencing with a T7 sequencing kit (Pharmacia). Sequence analyses were performed with the programs HUSAR of the German Cancer Research Center, ANTIGEN and PROSITE. The consensus sequences used for potential modification sites were as follows: N-(P)-(ST)-(P) for Asn glycosylation (N-linked oligosaccharides), (ST)-×-(RK) for protein kinase C, (ST)×(2)-(DE), for casein kinase 2 (CKII), and G-(EDKRHPYFW)-×(2)-(STAGCN) for myristylation.

Antipeptide antisera

The peptide RQPREDGNEEDKEN was synthesized and coupled to KLH (keyhole lympet hemocyanine). Rabbits were immunized once with peptide-KLH in complete Freunds adjuvant and twice with peptide-KLH in incomplete Freunds adjuvant. A more detailed description of this procedure will be published elsewhere (24).

Immunoblotting with anti peptide antisera

Immunoblots were prepared as for binding site blotting without a guanidinium denature renature cycle. For imunodetection of YB-1 the blocked membranes were incubated with a 1:200 fold dilution of the anti-peptide serum in 0.5% dry milk. For the detection of bound antibodies a second antibody coupled to alkaline phosphatase (Dianova) was used at a 1:5000 fold dilution.

Interspecific backcross mapping

Interspecific backcross progeny were generated by mating (C57BL/6J X *M.spretus*) F_1 females and C57BL/6J males as described (25). A total of 205 N_2 progeny were obtained; a random subset of these N_2 mice were used to map the *Yb-1* loci (see text for details). DNA isolation, restriction enzyme digestion, agarose gel electrophoresis, Southern blot transfer and hybridization were performed essentially as described (26). All blots were prepared with Zetabind nylon membranes (AMF-Cuno). The YB-1 probe, a 500 bp EcoRI fragment of cDNA corresponding to the 3'end of the coding sequence, was labeled with (alpha ^{32}P) dCTP using a random prime labeling kit (Amersham). Washing was done to a final stringency of 0.2×SSCP, 0.1% SDS, 65°C. Major fragments of 8.0, 7.4, and 3.5 kb were detected in *SphI* digested C57BL/6J DNA; major fragments of 7.2, 6.2, 5.2, 3.7, 3.6 and 2.7 kb were detected in *SphI* digested *M. spretus* DNA. A description of the probes and RFLP's for the loci used to position the *Yb-1* loci in the interspecific backcross have been reported. These loci include: ecotropic viral integration site-1 (*Evi-1*), fibroblast growth factor basic (*Fgfb*), and fibrinogen gamma polypeptide (*Fgg*) on chromosome 3 (27); transforming growth factor beta-1 (*Tgfb-1*) and glucose phosphate isomerase-1 (*Gpi-1*) on chromosome 7 (28, 29); metallothionein-1 (*Mt-1*), E-cadherin (*Ecad*) and haptoglobin (*Hp*) on chromosome 8 (30); and E26 avian leukemia virus oncogene 5' domain (*Ets-1*) and thymus cell antigen-1 (*Thy-1*) on chromosome 9 (31).

Recombination distances were calculated as described (32) using the computer program SPRETUS MADNESS. Gene order was determined by minimizing the number of recombination events required to explain the allele distribution patterns.

RESULTS

Identification of a YB-1 recognition site in the enhancer of human papillomavirus type 18 by binding site screening

The enhancer of human papillomavirus type 18 (HPV18) (33) is characterized by the presence of multiple octamer related sequence elements binding Oct-1 and a novel octamer binding protein p92 (21, 34). The oligonucleotide RP3 strongly binding to p92 was used for binding site screening experiments in an effort to clone p92 cDNA. A screen of 2 million plaques of a Hela-lambda gt11 expression library yielded 36 positives which were plaque purified. All 36 clones were related as was shown by cDNA insert analysis and crosshybridization. Two representative clones N4 and N19 encode β-galactosidase fusion proteins with a molecular weight of approximately 160kD. From the sizes of the N4 and N19 fusion proteins the molecular weights of the cloned proteins could be determined as 44kD because the β-galactosidase protein has a molecular weight of 116kD. The inserts of N4 and N19 were sequenced and a computer analysis revealed that they are 100% identical with the sequence of a cDNA encoding CCAAT binding protein dbpB (8) and identical

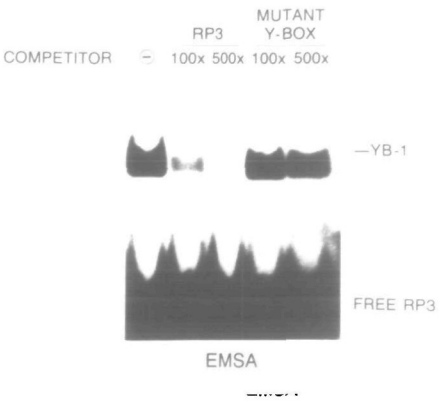

Fig. 1. Sequence specificity of YB-1 DNA binding. EMSA competition analysis with the HPV18 YB-1 recognition site present in RP3 and a mutant Y-Box oligonucleotide. Three μg of an E.coli lysate 1 hour after IPTG induction were incubated with RP3 (lane 1), and in the presence of 100×(lane 2) and 500×molar excess of unlabeled RP3 (lane 3). Competition with 100×(lane 4) and 500×molar excess (lane 5) of double stranded mutant Y-box oligonucleotide. Retarded RP3:YB-1 DNA:protein complexes and free RP3 are indicated.

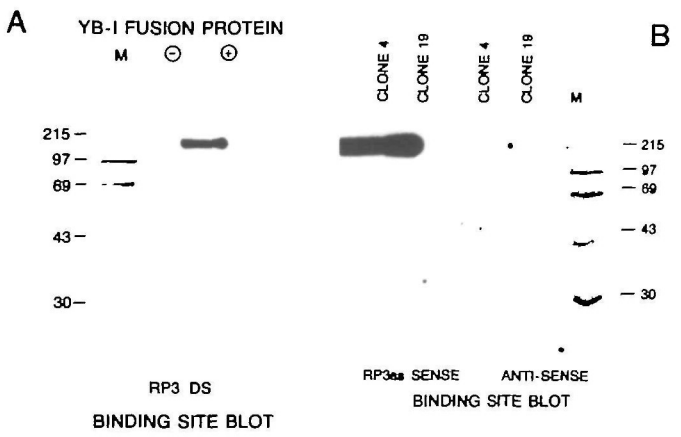

Fig. 2. Selective binding of YB-1 fusion protein to the sense strand of a YB-1 recognition sequence. (A) Binding site blot of YB-1 fusion protein. Bacterial lysate of clone N4 without induction (−) and 1 hour after induction with IPTG (+). In each lane 60 μg bacterial lysate was used. The blot was probed with double stranded end labeled RP3. (B) Binding site blots as in (A) probed with single stranded end labeled sense RP3 or end labeled anti-sense RP3.

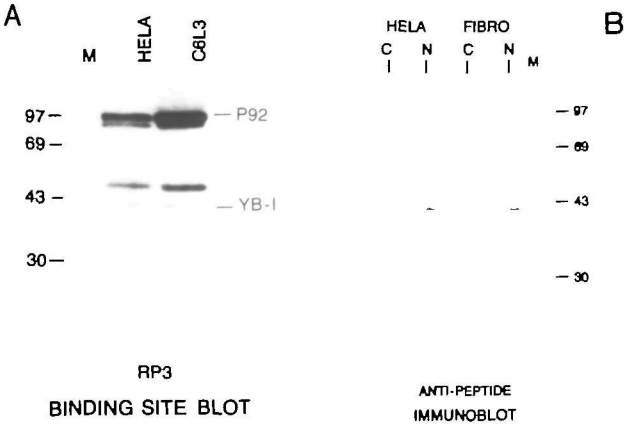

Fig. 3. Identification of a 42kD nuclear YB-1 protein in Hela cells and fibroblasts. (A) Binding site blot of nuclear extracts made from Hela cells and Cgl3 cells using end labeled RP3. Nuclear RP3 binding proteins p92 and YB-1 are indicated. (B) Detection of a nuclear 42kD protein with a peptide antiserum selected from the YB-1 sequence RQPREDGNEEDKEN. Hela and fibroblast cytoplasmic (C) and nuclear (N) extracts were analyzed by immunoblotting with a second antibody coupled to alkaline phosphatase. The molecular weights of ^{14}C labeled protein size markers are indicated.

with the sequence of a cDNA encoding the Y-box binding protein YB-1 (7), where a few sequence differences were observed. DNA binding of YB-1 is considered as being sequence specific because a mutation of the CCAAT element present in the Y-box to a CCC-AG motif abolished binding to IPTG induced YB-1 lambda gt11 phage plaques (7). We confirmed sequence specific DNA binding of the β-galactosidase-YB-1 fusion protein by EMSA competition analysis with the mutated long Y-box oligonucleotide (7). In this assay a 100 fold and 500 fold molar excess of unlabeled RP3 oligonucleotide efficiently competed for YB-1 binding (Fig. 1, lanes 2 and 3). In contrast, however, the long Y-box mutant oligonucleotide (7) did not compete in 100 fold and 500 fold molar excess with YB-1 DNA binding (Fig. 1, lanes 4 and 5). Therefore, we use the term YB-1 for this protein rather than dbpB because sequence specific DNA binding activity was established with a Y box motif.

The enhancer of HPV18 consists of two functionally redundant domains, one of which has been analyzed for DNA:protein interactions (21). The β-galactosidase-YB-1 fusion protein bound to HPV18 enhancer oligonucleotide RP3 (21) but not to other oligonucleotides from this enhancer domain suggesting that one YB-1 binding site exists in this domain.

YB-1 binds selectively to the sense strand of a YB-1 recognition site from the HPV18 enhancer

We have observed in binding site blotting experiments using a variety of synthetic binding sites that some nuclear proteins bound to double as well as to single stranded recognition sites. We sought to determine whether YB-1 fusion protein would also bind to single stranded recognition sites in HPV18 enhancer oligonucleotide RP3. Expression of clone N4 was induced for 1 hr with IPTG, the fusion protein size fractionated on a polyacrylamide gel and transferred to nitrocellulose. The membranes were probed with double stranded RP3 (Fig. 2A), single stranded RP3 sense strand and the RP3 anti-sense strand (Fig. 2B). RP3 double strand and sense strand bound to YB-1 fusion protein, in contrast, no binding was detected using the anti-sense strand. In this experiment equal amounts of fusion protein were run in parallel on the same gel and the nitrocellulose membrane was cut after transfer prior to incubation with sense and anti-sense oligonucleotides. A similar result has been reported for estrogen receptor DNA binding activities were the coding or sense strand of the estrogen responsive element (ERE) binds the estrogen receptor with 60 fold higher affinity than the double stranded ERE and the anti-sense ERE shows no estrogen receptor binding activity (35).

Identification of a 42kD nuclear YB-1 protein in HeLa cells and fibroblasts

A synthetic peptide approach was initiated aimed at identifying YB-1 protein in HeLa cells with the help of an anti YB-1 antiserum (24). This approach is based on the selection of antigenic regions of YB-1 protein primary sequence using the computer program ANTIGEN. The program ANTIGEN identifies the hydrophilicity of proteins. It is assumed that high hydrophilicity corresponds to high antigenicity. YB-1 protein displayed one region of very high hydrophilicity which was selected for peptide synthesis. We synthesized the peptide RQPREDGNEEDKEN from this region. The peptide was coupled to keyhole lympett hemocyanine (KLH) and rabbits were immunized. In binding site blots HPV18 enhancer oligonucleotide RP3 containing the YB-1 site binds to a protein with a molecular weight of 92kD which is a novel octamer binding factor (21) and to two additional proteins p50 and p42 (Fig. 3A). We analyzed nuclear and cytoplasmic extracts of Hela cells and fibroblasts in immunoblots using the anti-peptide antiserum in combination with a second antibody coupled to alkaline phosphatase. A 42kD protein was recognized by the anti peptide serum in nuclear extracts of Hela cells and human fibroblasts (Fig. 3B). Therefore the 42kD protein which was detected by binding site blotting (Fig. 3A) corresponds most likely to YB-1 protein indicating that YB-1 protein has an apparent molecular weight of 42kD. In cell extracts from tissue culture cells YB-1 protein was exclusively localized to the nucleus (Fig. 3B). Preimmune serum did not react with a 42kD protein (24).

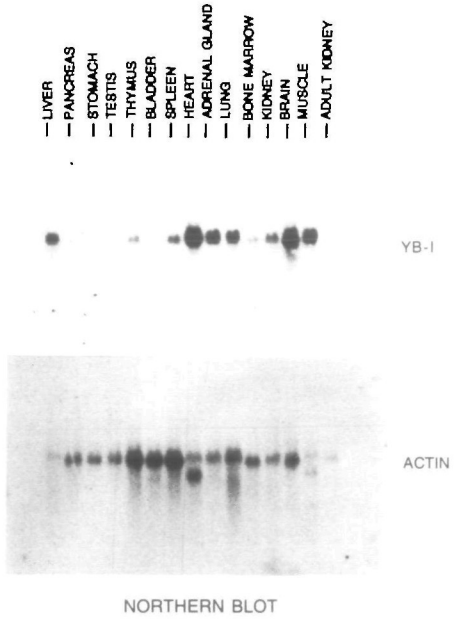

Fig. 4. Tissue restricted expression of the *Yb-1* gene in a 24 week old human fetus. Northern hybridization of total RNAs (5 μg) isolated from different tissues indicated on top of each lane. Brain RNA was from the cerebrum. Adult kidney RNA was from a kidney of a 24 year old male. For RNA concentration control the blot was hybridized to a human β-actin probe.

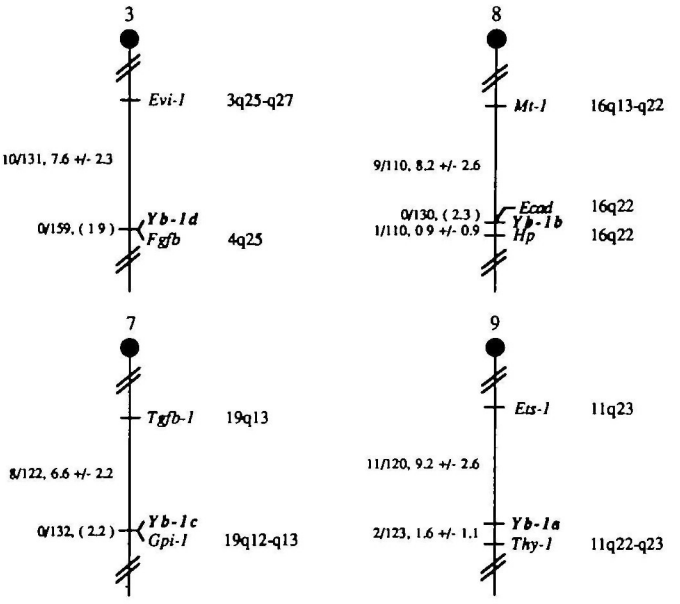

Fig. 5. Linkage maps showing the chromosomal locations of four *Yb-1* loci in the mouse genome. The *Yb-1* loci were mapped by interspecific backcross analysis. The number of recombinant N_2 animals over the total number of N_2 animals typed plus the recombination frequencies, expressed as genetic distance in centimorgans (+/− one standard error), is shown for each pair of loci on the left of the chromosome maps. Where no recombinants were found between loci the 95% confidence limit of the recombination distance is given in parentheses. The positions of loci in human chromosomes are shown to the right of the chromosome maps.

Differential expression of the *Yb-1* gene in organs and tissues of a 24 week old human fetus

Transcription factors and sequence specific DNA binding proteins can be divided into two groups, the first beeing the group of ubiquitously occurring proteins and the second beeing those with a restricted pattern of expression or a cell type specific expression. We sought to establish whether *Yb-1* gene expression is regulated in a tissue restricted manner or whether YB-1 is a ubiquitous factor. In order to clarify this issue we have extracted RNAs from fourteen different tissues of a 24 week old human fetus and determined *Yb-1* gene transcription by Northern hybridization (Fig. 4). High levels of YB-1 mRNA were present in six different tissues: cerebrum, heart, muscle, adrenal gland, lung and liver. Low levels of YB-1 mRNA were detected in four different tissues: kidney, bone marow, spleen and thymus. In four additional tissues: bladder, testis, stomach and pancreas, the *YB-1* gene was either not transcribed or at very low levels which could not be detected by Northern hybridization. Further experiments will be necessary to determine whether the *Yb-1* gene is silent in the latter four tissues. In adult human kidney YB-1 mRNA was present in low amounts (Fig. 4). The *Yb-1* gene is also expressed in human placenta and B-cells because YB-1 cDNA clones have been be isolated from a human placenta and a B-cell cDNA expression library (7, 8). In tissue culture cells YB-1 protein was present in Hela and in human fibroblasts (Fig. 3B). In different cell lines from T-cells, B-cells and monocytes *Yb-1* gene expression was observed (7). We conclude that *Yb-1* gene expression is not ubiquitous and that the levels of transcripts vary considerably in different tissues.

The *Yb-1* gene is a member of a dispersed gene family

The mouse chromosomal locations of four *Yb-1* related genes were determined by interspecific backcross analysis using progeny derived from matings of ((C57BL/6J×*Mus spretus*) F_1×C57bL/6J) mice. This interspecific backcross mapping panel has been typed for over 800 loci that are well distributed among all the autosomes as well as the X chromosome (25). C57BL/6J and *M. spretus* DNAs were digested with several enzymes and analyzed by Southern blot hybridization for informative restriction fragment length polymorphisms (RFLP's) using the YB-1 probe. *M. spretus*-specific polymorphisms were followed in these studies since the interspecific backcross was to C57BL/6J and C57BL/6J-specific polymorphisms could thus only be typed by hybridzation intensity. SphI digestion generated six major fragments in *M. spretus* DNA of 7.2, 6.2, 5.2, 3.7, 3.6 and 2.7 kb, five of which (6.2, 5.2, 3.7, 3.6 and 2.7 kb) could be followed in backcross mice. The strain distribution pattern (SDP) of each RFLP in the interspecific backcross mice was then determined and used to position the YB-1 related sequences on the interspecific map (Fig. 5).

The 6.2 kb M. spretus-specific fragment mapped to mouse chromosome 9 and defined the *Yb-1a* locus, which is 1.6 cM proximal of *Thy-1*. The 5.2 kb fragment (*Yb-1b*) was placed on chromosome 8 and did not recombine with *Ecad*. No recombinants were detected among the two loci in 130 animals suggesting that they are within 2.3 cM of each other (upper 95% confidence interval). The *Yb-1c* locus, characterized by the 3.7 and 3.6 kb *SphI* fragments, was located on chromosome 7. *Yb-1c* did not recombine with *Gpi-1* in 132 animals typed for both loci, suggesting the 2 loci are within 2.2 cM of each other. The 2.7 kb *SphI* fragment defined *Yb-1d* and was located on chromosome 3 (Fig. 5). *Yb-1d* did not recombine with *Fgfb* in 159 animals typed for both loci suggesting that the 2 loci are within 1.9 cM

of each other. These studies suggest there are at least four loci in the mouse that are related to YB-1 sequences. It remains to be determined whether these loci correspond to functional genes or pseudogenes.

DISCUSSION

YB-1 is a sequence specific DNA binding protein (7) because a mutation of the core CCAAT to a CCCAG motif in the Y-box element abolished YB-1 binding activity (7), which was confirmed by EMSA competition analysis (Fig. 1, lanes 4 and 5). We have found that YB-1 selectively bound to the sense strand of HPV18 enhancer oligonucleotide RP3. This kind of binding specificity was previously described for estrogen receptor binding to the estrogen responsive element (35). DNA dependent RNA polymerases II and III require auxiliary proteins for accurate and selective initiation of transcription (36, 37, 38). Some of these factors are sequence specific DNA binding proteins and transcription factors. The function of YB-1 as a transcription factor remains to be established, in Xenopus, however, highly similar Y-box (CCAAT) binding proteins exist (FRG Y1 and FRG Y2) which are transcriptional activators (39). The double strand binding specificity of YB-1 may be needed to concentrate YB-1 at a promoter or enhancer and the single strand binding specificity would keep YB-1 DNA bound during transitory strand separation in an initiation complex.

The open reading frame of full length YB-1 cDNA is 972 (8) bases long and the predicted molecular weight of YB-1 protein is 35.6kD. Antiserum which was raised against a synthetic YB-1 peptide, however, detected a 42kD nuclear protein in immunobloting experiments. Therefore we conclude that cellular YB-1 protein is a 42kD nuclear protein. Some proteins have higher molecular weights in polyacrylamide gels than that predicted by the primary amino acid sequence. One reason for this difference could be the occurrence of postranslational protein modifications such as phosphorylation or glycosylation.

Sequence specific DNA binding proteins and transcription factors can be ubiquitous, tissue restricted or cell type specific. Examples for cell type specific regulatory factors are the Oct-2 protein from B-cells (40) and Myo D1 which is muscle specific (41). Examples of transcription factors which are present in multiple tissues are: AP2, JUNB, JUNC, C/EBP. NFY-A, NFY-B and SP1 (10, 42, 43, 44). *Yb-1* gene expression was high in heart, cerebrum, muscle, adrenal gland, liver and lung. As the *Yb-1* gene is part of a family of related genes, it is possible that the observed expression pattern is not due to the activity of a single gene. To sort out the roles of all these factors in transcriptional regulation and during development it is necessary to define their individual fields of action in the organism and identify the target genes that they regulate within those fields. A first step in this analysis is to determine the expression of the genes encoding these factors in many tissues of an organism. The *Yb-1* gene may negatively regulate *HLA DR alpha* gene expression because its level of expression is inversely correlated with the expression of the *HLA DR alpha* gene (7). The occurrence of CCAAT motifs and Y-box motifs in many gene control regions suggests that Y-box transcription factors may regulate tissue-specific gene expression in completely different contexts (39).

The *Yb-1* gene belongs to a dispersed gene family

We have determined by interspecific backcross mapping that YB-1 related sequences are present on four different mouse chromosomes. There was an essentially invariant *SphI* band (7.2 kb in *Spretus*; 7.4 kb in B6) that could not be typed. This fragment might represent yet another *Yb-1* related gene. We have compared our interspecific map in which the *Yb-1* related genes have been placed with a composite mouse linkage map that includes many uncloned mutations compiled by M.T.Davisson, T.H.Roderick, A.L.Hillyard, and D.P.Doolittle and provided from GBASE (a computerized database maintained at the Jackson Laboratory, Bar Harbour, ME) and find that several mouse mutations lie in the vicinity of some of these loci (not shown). Assuming these loci encode functional genes, it will be of interest to determine the mouse developmental and tissue specific expression patterns of these genes and whether any of the mutations are candidates for alterations in a *Yb-1* related gene. Finally, it is often possible to predict where a gene will map in humans by determining its location in mouse. Based upon the positions of the *Yb-1* related genes on the interspecific map we can predict that *Yb-1a* will likely reside on human chromosome 11q, *Yb-1b* on 16q, *Yb-1c* on 19q and *Yb-1d* on 4q.

ACKNOWLEDGMENTS

We thank H. zur Hausen for support, C.de Vack and B.Luckow for comments on the manuscript. We thank Peter Angel for the lambda gt11 HeLa expression library.

REFERENCES

1. Chodosh, L. A., Baldwin, A. S., Carthew, R. W. and Sharp, P. A.(1988) Cell 53, 11–24.
2. Dorn, A., Bollekens, J., Staub, A., Benoist, C., and Mathis, D.(1987) Cell 50, 863–872.
3. Chodosh, L. A., Olesen, J., Hahn, S., Baldwin, A. S., Guarente, L. and Sharp, P. A.(1988) Cell 53, 25–35.
4. Jones, K. A., Yamamoto, K. R. and Tjian, R.(1985) Cell 42, 559–572.
5. Graves, B .J., Johnson, P.F. and McKnight, S.L.(1986) Cell 44, 565–576.
6. Lum, B. J., Williams, G. T. and Morimoto, R. I.(1990) Proc. Natl. Acad. Sci. USA 84, 2203–2207.
7. Didier, D. K., Schiffenbauer, J., Woulfe, S. L., Zacheis, M. and Schwartz, B. D.(1988) Proc. Natl. Acad. Sci. USA 85, 7322–7326.
8. Sakura, H., Maekawa, T., Imamoto, F., Yasuda, K. and Ishii, S.(1988) Gene 73, 499–507.
9. Gallinari, P., LaBella, F. and Heintz, N.(1989) Mol. Cell. Biol. 9, 1566–1575.
10. van Huijsduijnen, R. H., Li, X. Y., Black, D., Matthes, H., Benoist, C. and Mathis, D.(1990) EMBO J. 9, 3119–3127.
11. Santoro, C., Mermod, N., Andrews, P. C. and Tjian, R. (1988) Nature 334, 218–224.
12. Landschulz, W. H., Johnson, P. F., Adashi, E. Y., Graves, B. J. and McKnight, S. L.(1988) Genes Dev. 2, 1759–1764.
13. Jones, K. A., Kadonaga, J. T., Rosenfeld, P. J., Kelly, T. J. and Tjian, R.(1987) Cell 48, 79–89.
14. Pruijn, G. J. M., Vvan Driel, W. and van der Vliet, P. C. (1986) Nature 322, 656–659.
15. Birkenmeyer, E. H., Gwynn, B., Howard, S., Jerry, J., Gordon, J. I., Landschulz, W. H. and McKnight, S. L.(1989) Genes Dev. 3, 1146–1156.
16. McKnight, S. L., Lane, M. D. and Gluecksohn-Waelsch, S. (1989) Genes Dev. 3, 2021–2024.
17. Umek, R. M., Friedman, A. D. and McKnight, S. L. (1990) Science 251, 288–292.
18. Kim, Y. K., and Lee, A. S.(1991) Mol. Cell. Biol. 11, 2296–2302.
19. Wu, B. J., Williams, G. T.and Morimoto, R.(1987) Proc. Natl. Acad. Sci. USA 84, 2203–2207.
20. Stanbrige, E. J., Der, C. J., Doersen, C. J., Nishimi, R. Y., Peehl, D. M., Weissman, B. E. and Wilkinson, J. E.(1982) Science 215, 252–259.
21. Royer, H.-D., Freyaldenhoven, M. P. Napierski, I., Spitkovsky, D. D., Bauknecht, T. and Dathan, N.(1991) Nucl. Acids. Res. 19, 2363–2371.
22. Vinson, C. R., LaMarco, K. L., Johnson, P. E. Landschulz, W. H. and McKnight, S. L.(1988) Genes Dev. 2, 801–806.
23. Chirgwin, J. M., Przybyla, A. E., MacDonald, R.J. and Rutter, W. J.(1979) Biochemistry 18, 5294–5299.

24. Spitkovsky, D. D., Royer, H.-D., Mazurenko, N., Michaleva, I., Prudchenko, I., Korbuch, I., Suchova, N. and Kisseljov, F. (1992) Manuscript in preparation.
25. Copeland, N.G. and Jenkins, N.A.(1991) Trends Genet. 7, 113–118.
26. Jenkins, N.A., Copeland, N.G., Taylor, B.A. and Lee, B.K. (1982) J. Virol. 43, 26–36.
27. Cox, R. D., Copeland, N.G., Jenkins, N.A. and Lehrach, H (1991) Genomics 10, 375–384.
28. Dickinson, M.E., Kobrin, M.S., Silan, C.M., Kingsley, D.M., Miller, M.J., Ceci, J.D., Lock, L.F., Lee, A., Buchberg, A.M., Siracusa, L.D., Lyons, K.M., Derynck, R., Hogan, B.L.M., Copeland, N.G. and Jenkins, N.A.(1990) Genomics 6, 505–520
29. Siracusa, L.D., Jenkins N.A. and Copeland, N.G.(1991) Genetics 127, 169–179.
30. Hatta, M., Miyatani, S., Copeland, N.G., Gilbert, D.J., Jenkins, N.A. and Takeichi, M.(1991) Nucl. Acids Res. in press.
31. Kingsley, D.M., Jenkins, N.A. and Copeland, N.G.(1989) Genetics 123, 165–172.
32. Green, E.L.(1981) In Genetics and Probability in Animal Breeding Experiments. Macmillan, New York, pp 77–113.
33. Swift, F. V., Bhat, K., Younghusband, H. B. and Hamada, H.(1987) EMBO J. 6, 1339–1344.
34. Weitz, J., Kopun, M., Stöhr, M., Napierski, I. and Royer, H.-D.(1991) Nucl. Acids Res. 19, 5725–5730.
35. Lannigan, D. A. and Notides, A. C.(1989) Proc. Natl. Acad. Sci. USA 86, 863–867.
36. Matsui, T., Segall, J., Weil, P. A. and Roeder, R.(1980) J. Biol. Chem. 255, 11992–11996.
37. Segall, J., Matsui, T. and Roeder, R.(1980) J. Biol. Chem. 255, 11986–11991.
38. Weil, P. A., Luse, D. S., Segall, J. and Roeder, R.(1979) Cell 18, 469–484.
39. Tafuri, S. R. and Wolffe, A. P.(1990) Proc. Natl. Acad. Sci. USA 87, 9028–9032.
40. Staudt, L. M., Singh, H., Sen, R., Wirth, T., Sharp, P. A. and Baltimore, D.(1986) Nature 323, 640–643.
41. Davis, R. L., Weintraub, H. and Lassar, A. B.(1987) Cell 51, 987–1000.
42. Mitchell, P. J., Timmons, P. M., Hebert, J. M., Rigby, P. W. J. and Tjian, R.(1991) Genes Dev. 5, 105–119.
43. Hirai, S.-I., Ryseck, R.-P., Mechta, F., Bravo R. and Yaniv, M.(1989) EMBO J. 8, 1433–1439.
44. Saffer, J. D., Jackson S. P. and Anarella, M. B.(1991) Mol. Cell. Biol. 11, 2189–2199.