Structural and functional analysis of the human Y-box binding protein (YB-1) gene promoter

Yoshinari Makino, Takefumi Ohga, Satoshi Toh, Koji Koike, Katsuzumi Okumura¹, Morimasa Wada, Michihiko Kuwano and Kimitoshi Kohno*

Department of Biochemistry, Kyushu University School of Medicine, Maidashi, Fukuoka 812-82, Japan and ¹Department of Biochemistry, Mie University of Medicine, Edobashi, Tsu-shi 514, Japan

Received February 6, 1996; Revised and Accepted April 1, 1996 EMBL accession no. X96666

ABSTRACT

We have isolated three overlapping genomic clones containing the 5’ portion of the human YB-1 gene. These clones span ~25 kb of contiguous DNA containing 10 kb of 5’ flanking sequence and 15 kb of the gene. The nucleotide sequence of the first exon and of 2000 upstream base pairs (bp) was determined. The first exon is unusually large and contains a 166 bp coding sequence and a 331 bp untranslated region. CpG sequences cover the 5’-end of the YB-1 gene including its first exon and intron as well as the upstream regions. The GC content around the first exon is ~70% and a CpG-free region was located in the untranslated segment. The segment preceding the major transcription initiation site does not contain a TATA box, CCAAT box and the binding sequence for known transcription factors. A transient expression assay using the chloramphenicol acetyltransferase (CAT) gene showed that the sequence from +24 to +281 is critical for CAT expression. Fluorescence in situ hybridization demonstrated the chromosomal locus of YB-1 gene on chromosome 1p34. Polymerase chain reaction analysis demonstrated the chromosomal locus of YB-1 gene on chromosome 1p34. Polymerase chain reaction analysis on other genomic phage DNAs showed that several clones were derived from pseudogenes.

INTRODUCTION

The human Y-Binding protein (YB-1) is a member of a DNA-binding protein family with a structurally and functionally conserved cold shock domain. YB-1 has been shown to regulate gene expression through binding to Y-box sequences (1,2). Several eukaryotic genes contain the consensus Y-box sequence, 5’-CTGATTGG-3’, in their cis-regulatory elements, including major histocompatibility complex class II (3), thymidine kinase, proliferating cell nuclear antigen, DNA polymerase α, epidermal growth factor receptor and multidrug resistance 1 (MDR1) (4). The multidrug resistance gene encodes an energy-dependent drug efflux pump that is overexpressed in cancer cells resistant to multiple anticancer agents (5). We have previously shown that MDR1 is a stress responsive gene (6–11). Transfection of an mdr1 promoter/ Cat reporter plasmid has shown that an inverted CCAAT box (Y-box) is critical for CAT induction by various environmental stimuli (10,11). Binding activity of MDR NF-1 to this Y-box element is augmented in nuclear extracts prepared from cells treated with either UV irradiation or anticancer agents. It has also been shown that the deletion of the Y-box region leads to markedly reduced expression of the mdr1 promoter/CAT construct (12). We have previously cloned a cDNA for MDR NF-1 by Southwestern screening of a human HeLa cell cDNA expression library using an mdr1 promoter probe. Sequence analysis of the MDR NF-1 cDNA showed that it was identical to the Y-Binding protein, YB-1 (11). Using this cDNA, we have demonstrated that YB-1 mRNA accumulation increased when cells were treated with UV irradiation or anticancer agents. Isolation of cis-regulatory elements should help clarify the complex regulation of YB-1 expression.

We report here the isolation and characterization of several overlapping genomic clones containing the 5’-end of the human YB-1 gene.

MATERIALS AND METHODS

Cloning and sequencing of a human MDR-NF1 cDNA

A human colon cDNA library constructed in the λgt11 expression vector was obtained from Stratagene. The cDNA library was plated on LB agar containing ampicillin and overlaid with nitrocellulose membrane saturated with 10 mM IPTG. A total of 2 × 10⁶ plaques were screened with the end labelled 140 bp XhoI–TagI mdr1 promoter region. A 384 bp partial cDNA of human origin (originally designated as MDR-NF1) encodes a 128 amino acid polypeptide with 99% identity to hYB-1. Plaques (1 × 10⁶) from same λgt 11 cDNA library were screened by DNA hybridization using this clone to isolate full length cDNA. One clone had an insert 1060 bp. Sequence analysis showed 99% identity to the hYB-1/dbpB cDNA and almost covered the coding region.

Isolation of genomic clones and DNA sequencing

Clones were isolated by screening a human placenta genomic library in EMBL3 with a 1060 bp probe from the YB-1 cDNA (13,14). Positive phages were mapped with EcoRI and SalI and hybridized with a 5’ probe, a 350 bp EcoRI–Styl fragment, and a 3’ probe, a 300 bp SalI–EcoRI fragment. Several restriction fragments around the first exon were subcloned in pUC18 and sequenced. All plasmid DNAs were sequenced from both ends using a Shimazu DSQ 1000 automated sequencer. The primer for

* To whom correspondence should be addressed
primer extension analysis, 5'-GCTCATGGTGCAGTGATGG-3' was derived from the sequence of the first exon. The primers for PCR analysis C1, 5'-CAACAGGAATGCACCAAGG-3' and C2, 5'-GATGGTAGAGATGTAACGG-3' were derived from cdNA sequences. This primer pair was used to analyze genomic clones containing pseudogenes. Polymerase chain reaction was performed in 5 μl reactions containing 100 ng DNA, 1μM of each primer pair, 0.2 mM each of the four deoxyribonucleotides (Pharmacia), and 1 U Taq polymerase (Boehringer Mannheim) in PCR buffer [30 mM Tris–HCl (pH 8.5), 2 mM MgCl2, 5 mM β-mercaptoethanol, 0.01% gelatin and 0.1% Thesit] (18–20). Initial denaturation at 94°C for 10 min was followed by 35 amplification cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s in a Gene Amp PCR System 9600 (Perkin Elmer). The product of 773 bp was of the expected size from the cdNA sequence.

**Construction of CAT reporter plasmids**

The vectors, SVOOCA T and pCH110 (pSV-β-Gal), were purchased from Nippon gene (Tokyo) and Pharmacia (Sweden), respectively. Different truncations and deletions of the 5' region of the YB-1 gene were ligated into the HindIII site of pSVOCAT. The fragment from HindIII (~1384) to PvuII (+295) was obtained from a plasmid subclone. From this, fragments from NcoI (~483), Apal (~119), Sall (+24) or BafI (+127) to PvuII (+295) were isolated and filled with the Klenow fragment of DNA polymerase I. After HindIII linker ligation, these fragments were ligated into pSVOCAT and transformed. The resulting constructs were named pYB-CAT1, pYB-CAT2, pYB-CAT3, pYB-CAT4 and pYB-CAT5, respectively. pYB-CAT3R and pYB-CAT5R indicate plasmid with the promoter fragment inserted in reverse orientation. To construct pYB-CAT3Δ1, pYB-CAT3Δ2 and pYB-CAT3Δ3, pYB-CAT3 was digested with Sall–AvrII, AvrII–SmaI and Sall–SmaI, respectively, blunt ended and ligated. Plasmid DNAs were purified by alkaline lysis followed by two cycles of equilibrium centrifugation in a CsCl–ethidium bromide gradient.

**Transfection and assays of cell extracts**

T24, a human bladder carcinoma cell line, was grown as previously described (21). The cells were maintained at 37°C with 5% CO2. Exponentially growing cells were trypsinized, seeded at 3× 105 cells in 60 mm diameter plates and incubated overnight prior to transfection. DNA transfection of T24 cells was then carried out by calcium phosphate precipitation with 5 μg pCH110 (pSV-β-gal) and 10 μg of a reporter construct. The precipitate was added to the cells and after an 8 h transfection, the cells were washed twice, replenished with fresh medium and incubated for another 48 h. At the end of this period, cell extracts were prepared and assayed for both CAT and β-galactosidase activities according to the standard procedure (22,23). Normalized CAT activities were expressed as the ratio of that seen with pYB-CAT1. At least four independent experiments were performed for each construct and all measurements were determined in duplicate.

**Primer extension by reverse transcriptase**

The synthetic primer described above was labelled at its 5'-end and hybridized to poly(A) RNA in 80% formamide, 0.4 M NaCl, 40 mM PIPES (pH 6.4) and 1 mM EDTA for 4 h at 50°C. After precipitation of nucleic acids, the pellet was dissolved in reverse transcriptase buffer (BRL). The primer was extended with 20 U MMTV reverse transcriptase (BRL) using 1 mM each of the four deoxyribonucleotides. After 1 h at 37°C, the reaction was terminated with 20 mM EDTA. RNA was hydrolyzed with 0.125 M NaOH for 1 h at 65°C. The reaction was neutralized and the DNA was collected. Sequencing reactions using the same primer were analyzed on a 7 M urea–6% polyacrylamide gel to determine the size of the extended product (13,14).

**Fluorescence in situ hybridization (FISH)**

Chromosome spreads were obtained from phytohaemagglutinin-stimulated blood lymphocytes of a healthy donor after thymidine synchronization and bromodeoxyuridine incorporation by the method of Takahashi et al. (15,16). Genomic DNA fragments were labelled with biotin-16-dUTP (Boehringer Mannheim) by nick translation. In situ hybridization was performed according to Lichter et al. (17) in the presence of COT-1 DNA (Gibco-BRL, Gaithersburg, MD) as a competitor. Hybridized probes were detected with FITC-conjugated avidin (Boehringer Mannheim). Chromosomes were counterstained with 0.2 mg/ml propidium iodide (PI) for R-banding. Fluorescence signals were imaged using a Zeiss Axioskop epifluorescence microscope equipped with a cooled Charge Coupled Device (CCD) camera (Photometrics, PXL 1400). Image acquisition was performed on a Macintosh Quadra 840AV computer with the software program IPLabTM (Signal Analytics Co.). The images were then pseudocoloured and merged using Adobe Photoshop TM 2.51 (Adobe Systems Inc.). FITC and propidium iodide images were shown in white and grey, respectively. The merged images of FITC and PI were directly printed by Fuji Pictography 3000 from a Macintosh computer.

**RESULTS**

We have previously isolated and characterized a YB-1 cdNA clone (11) that was then used as a hybridization probe to screen a HeLa cell cdNA library (Clontech). Several positive clones were isolated, one of which was a nearly full-length cdNA for YB-1, pYB-1-22 (data not shown). To isolate genomic clones encoding the 5' region of the YB-1 gene, a human genomic library was screened using the 1060 bp cdNA insert of pYB-1-22. More than 50 positive clones were isolated. Twenty-four clones were characterized by Southern hybridization using two probes derived from the cdNA as described in Materials and Methods. Three clones contained non-identical inserts that hybridized with 5' cdNA fragments. The restriction maps for these non-identical clones, EMBL-YBG1, YBG2 and YBG3, are shown in Figure 1. Confirmation that these clones encoded YB-1 was obtained by sequence analysis as described later.

**Sequence analysis of the promoter region**

We determined the nucleotide sequence of the promoter region of the YB-1 gene. A 9 kb EcoRI fragment of EMBL-YBG1 was subcloned into pUC18. To localize the first exon more exactly, this plasmid was digested with restriction enzymes and analyzed by Southern blotting using a 350 bp fragment of the cdNA containing the 5'-end of the cdNA as a probe. Several fragments hybridized with the probe and were subcloned in plasmid pUC18 (data not shown). The nucleotide sequence of the first exon and its 5' flanking region was then determined. This fragment contained exon(s) with sequences identical to those previously
Figure 1. Restriction endonuclease cleavage map of three overlapping genomic clones. Restriction enzyme cleavage sites are indicated as follows: E, EcoRI, S, SalI. Overlapping regions within these clones were confirmed by the Southern blotting (data not shown).

determined from the cDNA (3). Sequence comparison with the
cDNA also indicated that the last codon of this exon is split at the
second base. A similar division is also found in the first exon of
human dbpA gene (24). The dbpA gene is closely related to the
YB-1 gene. The first exon of the dbpA gene is 474 bp long. We
therefore assumed that we had identified the first exon of the YB-1
gene.

We performed primer extension to define precisely the start site
for transcription. The cDNA products extended from the primer
were analyzed by electrophoresis beside sequencing reactions using
the same primer. Three major and four minor transcription initiation
sites were observed. About 60% of the transcripts were initiated at
the three major sites indicated in Figure 2. Sequence analysis of the
region upstream of exon 1 failed to locate any sequence motifs such
as TATA and CCAAT boxes, for the transcription factor binding but
identified a highly GC-rich element. There are two GT boxes and
three GC boxes in the untranslated region that are binding sites for
Sp1 family transcription factors. Multiple E-boxes are found in the
promoter region. Characteristics of the promoter sequences are
summarized in Table 1. The GC content is ~70% around the first
exon. Since ‘housekeeping’ genes have a CpG island overlapping
the transcription start site, we analyzed the 5′ region of the YB-1
gene for the presence of a CpG island. We plotted the frequency of
both CpG and GpC dinucleotides against the position in the
sequence. As shown in Figure 3A, CpG dinucleotides are enriched
at the 5′-end of the YB-1 gene including its first exon.

Table 1. Features of the sequence

<table>
<thead>
<tr>
<th>Feature</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promoter</td>
<td>1</td>
<td>1855</td>
</tr>
<tr>
<td>Exon 1</td>
<td>1856*</td>
<td>2352</td>
</tr>
<tr>
<td>E-box</td>
<td>48</td>
<td>53</td>
</tr>
<tr>
<td>E-box</td>
<td>353</td>
<td>358</td>
</tr>
<tr>
<td>E-box</td>
<td>458</td>
<td>463</td>
</tr>
<tr>
<td>E-box</td>
<td>531</td>
<td>536</td>
</tr>
<tr>
<td>E-box</td>
<td>1147</td>
<td>1152</td>
</tr>
<tr>
<td>E-box</td>
<td>1201</td>
<td>1206</td>
</tr>
<tr>
<td>GT box</td>
<td>1957</td>
<td>1962</td>
</tr>
<tr>
<td>GC box</td>
<td>2001</td>
<td>2006</td>
</tr>
<tr>
<td>GC box</td>
<td>2024</td>
<td>2029</td>
</tr>
<tr>
<td>GC box</td>
<td>2033</td>
<td>2038</td>
</tr>
<tr>
<td>GT box</td>
<td>2100</td>
<td>2105</td>
</tr>
<tr>
<td>Start codon</td>
<td>2187</td>
<td>2189</td>
</tr>
<tr>
<td>Inverted repeat 1</td>
<td>1904</td>
<td>1914</td>
</tr>
<tr>
<td>Inverted repeat 2</td>
<td>2034</td>
<td>2044</td>
</tr>
<tr>
<td>Other initiation sites</td>
<td>1883, 1886, 1938, 1947, 1959, 1972</td>
<td></td>
</tr>
</tbody>
</table>

The most 5′ transcription initiation site is started at 1856 as indicated by an asterisk (*) and designated as +1 in the text and Figure 3.

Identification of a YB-1 promoter by analysis of transcriptional activity

To determine whether the region upstream of the putative transcriptional start sites can activate transcription, available restriction sites were utilized to construct a series of deletion plasmids (Fig. 3B). These constructs were then tested by transient transfection in human bladder carcinoma T24 cells. DNA to only –119 yielded full promoter activity. Deletion of –119 to +24 diminished promoter activity by ~45%, but this construct was still able to promote some CAT activity. Further deletion of +24 to +127 or all internal deletions of the untranslated region almost completely abolished promoter activity. Surprisingly, no CAT activity was found when cells were transfected with pYBCAT-3Δ2.
Figure 3. (A) Structure and deletion analysis of the YB-1 promoter. Relevant restriction enzyme cleavage site are as indicated: H, HindIII; S, SalI; N, NcoI; A, ApaI; B, BalI; Sm, SmaI; Av, AvrII; P, PstI. The small black triangles indicate transcriptional start sites determined by primer extension analysis. The diagram shows the distribution of GpC and CpG dinucleotides. Bold line indicates the first exon. (B) Constructs containing various portion or deletions of the 5′ region were subclones in front of a CAT reporter gene, and transiently transfected into T24 cells. The β-galactosidase reporter gene was co-transfected as an internal control. CAT activity was normalized to β-galactosidase activity. At least four independent experiments were done and all observations were performed in duplicate. The results are all normalized to the value of the CAT activity of pYBCAT-1 and are indicated as %.

which still contains both the promoter region and all transcription initiation sites.

Chromosome mapping and analysis of other genomic clones for YB-1

FISH was used to determine the chromosome location of YB-1. Comparison of the fluorescence signals and the banding patterns of the chromosomes indicated that this region was located on chromosome band 1p34 (Fig. 4). These results were based on observations made on >10 prometaphase chromosomes. Fluorescence signals were not consistently observed on the other chromosomes. We mentioned before that >50 clones were isolated in the first screening. This might indicate the existence either of a gene family or of pseudogenes. To test for pseudogenes, all genomic clones were analyzed by a PCR assay using a primer pair derived from the cDNA sequence. The PCR product includes half of the mRNA. We used human placenta DNA and cDNAs transcribed from total RNA of KB and T24 cells as templates (Fig. 5, lanes P, R1 and R2), and a 773 bp PCR product was consistently observed. A similar product was also detected when some phage genomic DNA clones were used as the template (lanes 1 and 3).

DISCUSSION

We have described the isolation and characterization of the 5′ region of the YB-1 gene. We isolated overlapping genomic clones encompassing 15 kb of the 5′ part of the YB-1 gene and 10 kb of its 5′ flanking regions. We determined the nucleotide sequence surrounding the 5′-end of the gene and the start site for the transcription.

The dbpA gene is an another member of a protein family containing cold shock domains. The human dbpA gene covers 24 kb and includes 10 exons (24). Characteristic of its exon–intron structure is a large first exon of 474 bp. YB-1 gene may have a similar structure. The first exon of YB-1 is 497 bp long resembling that of the dbpA gene (Table 1). The regulatory regions are highly GC rich and the CpG islands are located at 5′ to the first exon and intron. Thus, the YB-1 promoter has structural features common to those of housekeeping genes. This feature of the YB-1 5′ region is consistent with its ubiquitous mRNA expression in all human tissues (24), and suggests that it encodes a protein assumed with an essential cellular function. No typical TATA and CCAAT boxes exist in the proximal promoter sequences of the YB-1 and dbpA genes. This divergence in the promoter region between YB-1 and dbpA may reflect the presence of different regulatory mechanisms. Functional analysis of the promoter region in a transient expression system demonstrated significant promoter activity in the region upstream of the first exon (Fig. 3B). The promoter activity of the YB-1 gene is strong and comparable with that of the SV40 promoter. Our initial analysis of this promoter revealed
Figure 4. Mapping of the gene region encoding human YB-1 by fluorescence in situ hybridization (FISH). FISH was carried out using phage DNA (EMBL-YBG1) as a probe. Fluorescence signals on the R-handed metaphase chromosomes are indicated by arrow heads. Based on observations made with more than 10 metaphase chromosomes, the band is 1p34.

complex regulation. When a region between −119 and +24 was deleted, CAT activity was diminished to 45%. This result may reflect variation in the amount of transcriptional initiation. Thus, the region between +24 and +127 may act as a promoter of the transcription initiated from minor sites. A unique structure of this region is that it lacks CpG dinucleotides, as shown in Figure 3A. The CAT activity of pYBCAT-3Δ2 was very weak even though pYBCAT-3Δ2 contains the promoter region and all sites for transcription, suggesting that the untranslated region is critical for the expression of CAT activity. The untranslated region may play a role in mRNA stability or may confer altered affinities for the translational initiation and elongation machinery. The DNA sequence analysis revealed the presence of inverted repeat in the 5′ untranslated region. The 11 nucleotides long (+48 to +58) showed high inverse complementarity to the sequence at position +178 to +188. These sequences could form the stem–loop secondary structure, which may affect the translational efficiency. More precise analysis is required to define the molecular mechanism of how YB-1 expression is controlled at the translational level. The EMBL-YBG1 clone was used as a probe for in situ hybridization to human metaphase chromosomes. A high specific signal was obtained on the short arm of chromosome 1. Our result confirmed mapping data published previously (24). Other genomic clones were mapped to various chromosomes (data not shown), and many positive signals were observed in the first screening of a genomic library, suggesting the existence of pseudogenes scattered in the human genome. Primer pairs derived from cDNA sequence generated fragments of 773 bp, the expected size of the cDNA, when human placenta DNA or genomic phage DNA were used as templates (Fig. 5). This indicated that PCR products are derived either from intronless pseudogenes or from a single big exon. If the exon–intron structure of YB-1 is similar to that of the dpbA gene, these results indicate the existence of pseudogenes. Kudo et al. (24) have reported that 16 YB-1 genomic clones contained intronless processed genes. Isolation and characterization of these pseudogenes is now in progress in our laboratory.

We have previously reported that both steady state level of YB-1 mRNA and DNA binding activity of YB-1 protein were increased when cells were treated with either UV irradiation or anticancer agents (10,11). YB-1 protein is also known to involve in cell proliferation (1,2). YB-1 protein is expected to influence stress response to anticancer agents or UV irradiation as well as cell proliferation through signal transduction cascade. The characterization of this promoter region allows further exploration of mechanisms governing the YB-1 expression.

ACKNOWLEDGEMENTS

We thank Miss Yukiko Mine for editorial assistance. This work was supported by a grant-in-aid for cancer research from the Ministry of Education, Science and Culture, Japan, and the Fukuoka Anti-cancer Research Fund.

REFERENCES