Analysis of Human UDP-Glucose Dehydrogenase Gene Promoter: Identification of an Sp1 Binding Site Crucial for the Expression of the Large Transcript

Jaya Vatsyayan¹, Hwei-Ling Peng² and Hwan-You Chang¹,*

¹Institute of Molecular Medicine, National Tsing Hua University; and ²Department of Biological Science and Technology, National Chiao Tung University, Hsin Chu, 300, Taiwan

Received November 18, 2004; accepted March 24, 2005

UDP-glucose dehydrogenase (UGDH) catalyzes the conversion of UDP-glucose to UDP-glucuronic acid, which is required in liver for the excretion of toxic compounds, and for the biosynthesis of complex carbohydrates, such as hyaluronan, in many cell types. Analysis of a human EST database, as well as the results of a 5′-RACE experiment, have revealed the presence of two transcription start sites approximately 160 bp apart in the human UGDH gene confirming previous Northern hybridization results. To delineate the regions in the UGDH promoter required for regulating the expression of the gene, in particular the synthesis of the large transcript, serial deletions of the 2.1-kb UGDH promoter region were constructed and their activities determined by the firefly luciferase reporter gene assay. Our results indicate that the region from nucleotide position –486 to –632 relative to the start of the small transcript contains positive regulatory elements that contribute to gene expression. Mitomycin A, an inhibitor of transcription factor Sp1, abrogates the promoter activity, suggesting the involvement of this specific protein in UGDH expression. By using site-directed mutagenesis, we analyzed the functional contribution of three putative Sp1 binding elements within this region. A mutation at position –564 demonstrated that this site serves as an enhancing element in both HepG2 and HeLa cells. The complex formation pattern revealed by an electrophoretic mobility shift assay as well as an anti–Sp1 antibody–mediated supershift assay confirmed the identity of this GC box as an Sp1 binding motif. Our results thus identify an alternative transcription start site on the UGDH promoter, and locate the cis-element that greatly enhances the basal transcriptional activity of UGDH gene

Key words: EMSA, promoter analysis, Sp-1, transcription start site.

The prominence of UDP-glucose dehydrogenase (UGDH, EC 1.1.1.22) is in its unique and pivotal role in catalyzing the oxidation of UDP-glucose to UDP-glucuronic acid accompanied with the reduction of two molecules of NAD (1). UDP-glucuronic acid participates in detoxification in the liver by conjugating with a variety of xeno- and endobiotic compounds to aid in their solubilization (2, 3). In addition, the compound is an integral component in the biosynthesis of glycosaminoglycans, such as hyaluronan, which serves a variety of functions within the extracellular matrix of nearly all tissues, thus directly influencing cell behavior and developmental processes (4, 5). Mutations in the UGDH gene in Drosophila, known as sugarless, cause disruption in the wingless signaling pathway due to the inability of the mutants to put heparan side chain on proteoglycans (6, 7), while mutations in the same locus in zebrafish lead to a deficiency in the initiation of heart valve formation (8).

The primary structure of the UGDH enzyme predicted as a 468–amino acid protein, was elucidated in 1994 (9) after which obtaining a cDNA clone encoding UGDH became a possibility (10, 11). The human UGDH gene cloned in our laboratory as well as by other groups (11, 12) spans over 26 kb, and contains 12 exons. Although the function of UGDH is well investigated, the knowledge of its transcriptional regulation has just begun to unravel. We have previously shown that two major UGDH transcripts, 2.4 and 2.7 kb in length, can be observed by Northern hybridization (13). The transcription start site together with the critical specificity protein-1 (Sp1) sites responsible for regulating the synthesis of the small UGDH transcript have been determined recently (12). However, the location and contribution of cis-acting elements contained in functional regions on the UGDH promoter for the large transcript have not yet been explored.

The aim of this study was to define the elements responsible for enhancing the basal transcriptional activity of the UGDH gene, particularly those contributing to the synthesis of the large UGDH transcript. As Sp1 bind-
ing GC boxes are one of the most common regulatory elements distributed in promoters of numerous housekeeping as well as tissue specific genes, this study emphasizes the identification of GC boxes crucial to the synthesis of the large $UGDH$ transcript.

**MATERIALS AND METHODS**

**Mapping of the Transcription Start Site of Human $UGDH$ Gene**—The start site of the large $UGDH$ transcript was mapped by 5′-Rapid Amplification of cDNA Ends (5′-RACE). In brief, total RNA was isolated from 80% confluent HepG2 cells cultured in a 10-cm dish using the total RNA extraction kit (Qiagen, Netherlands) and first strand cDNA synthesis using the SMART™ RACE cDNA Amplification Kit (BD Biosciences Clontech, Palo Alto, CA) following the manufacturers’ protocols. Two antisense gene-specific primers (GSP) were designed for the 5′-RACE experiment (Fig. 1B). The primer GSP1 (5′-TCCCAGCAGCAAGCGCAGGGACCGCTCC-3′) was designed downstream of the previously published transcription start site (marked as position +1) of human $UGDH$ at +152 nucleotide. The touchdown PCR conditions were 4 cycles of 95°C for 2 min and 72°C for 1 min followed by 5 cycles of 95°C for 45 s, 65°C for 45 s, 72°C for 1 min, then 30 cycles of 95°C for 45 s, 55°C for 30 s, 72°C for 1 min, and finally 72°C for 5 min. The GSP2 (5′-AGGCTTTTCGCCTGCGAAGGCGGGCCCTCC-3′) was designed downstream of the previously published transcription start site (marked as position +1) of human $UGDH$ at +152 nucleotide. The touchdown PCR conditions were the same as for GSP1 except an annealing temperature of 60°C was used before reducing the temperature to 55°C. Each of the amplicons from the PCR with the UPM-GSP2 primer set was cloned into pGEMT-Easy vector (Promega Inc., Madison, WI) for sequencing.

**Plasmid Construction and Site-Directed Mutagenesis**—The 5.0-kb upstream control region of human $UGDH$ was subcloned from a recombinant lambda phage containing the human $UGDH$ gene into the pUC18 vector. The latter was used as a template to create a series of progressive deletions by PCR. The amplified promoter fragments were cloned into the KpnI and HindIII sites in the pGL2-Basic reporter vector (Promega Inc., Madison, WI) to result in plasmids from pJR001, pJR003, pJR005, and pJR007. Plasmid pJR006 was constructed similarly except that XhoI and HindIII sites were chosen. The internal deletion construct pJR021 was created by using pJR001 as the template for PCR amplification and subcloning the amplified fragment (−1945 to −632) into pJR005 using the KpnI and XhoI sites. Plasmid pJR005 was used as a template to introduce mutations in the Sp1 binding sites (Table 1) following the strategy provided in Stratagene’s site directed mutagenesis instruction manual. All promoter-reporter constructs were confirmed by nucleotide sequencing.

**Cell Culture, Transient Transfection, and Luciferase Assay**—Human hepatoma cell line HepG2 (ATCC No. HB-8065) and human cervical epithelial carcinoma cell line HeLa (ATCC No. CCL-2) cells were grown in DMEM with 10% fetal bovine serum and 10 µg/ml of penicillin/streptomycin, at 37°C and 5% CO2. All cell culture reagents were purchased from Life Technologies, Inc., USA. Lipofectamine Plus reagent (Life Technologies, Inc.) was used for transfection following the manufacturer’s instructions. Cells were seeded at a density of 5 × 10³ cells per well in a 6-well plate 24 h prior to transfection in medium without antibiotics. Approximately 2 µg of the...
sequences were designed in accordance with information provided by Santa Cruz Biotechnology (Table 1). Probe annealing was performed by heating 100 µM of each complementary strand of the oligonucleotide to 95°C for 2 min, and then cooling gradually to 25°C over a period of 45 min in a thermal cycler. Four picomoles of the annealed oligonucleotides were phosphorylated with [γ-32P]-ATP in a reaction catalyzed by T4 polynucleotide kinase (Promega) for 1 h at 37°C. The enzyme was then inactivated by heating at 68°C for 10 min. The unincorporated label was removed by chromatography through a Sephadex G-25 spin column equilibrated with TE buffer (50 mM Tris-HCl, 10 mM EDTA, pH = 8.0).

Nuclear extracts were prepared from confluent dishes of HepG2 and HeLa cells as described by Dignam and coworkers (15), and the protein concentration was determined by the Bradford method (16). The labeled probes were incubated with 4 µg of the nuclear extracts in gel shift binding buffer [20% glycerol, 5 mM MgCl₂, 2.5 mM dithiothreitol, 250 mM NaCl, 50 mM Tris-HCl, pH = 7.5 and 0.25 mg/ml poly (dI-dC)-poly (dI-dC)] for 15 min in a total reaction volume of 20 µl. Then, labeled probes were added, and the reactions were allowed to continue at room temperature for 30 min before the samples were resolved in a 10-cm 5% polyacrylamide gel at 4°C with a 20 mA current for 1.5 h. In competition assays, the unlabeled competitor oligonucleotides were preincubated with the nuclear extract prior to the addition of labeled probe. In the antibody-mediated supershift assay, approximately 2 µg of Sp1-specific antiserum (SC-420 X, Santa Cruz Biotechnology, USA) was incubated with 8 µg of nuclear extract in gel shift binding buffer for 1 h on ice, and then incubated further with the labeled probes for another 30 min before resolving the reaction products in gels. The gels were dried and exposed to superRX film (Fujifilm, Japan) at –80°C overnight.

RESULTS AND DISCUSSION

Identification of Possible Start Sites for UGDH Transcripts—It has been previously shown that the human UGDH gene is capable of producing two major transcripts 2.4 and 2.7 kb in length. The start site for the smaller 2.4-kb transcript has been mapped using primer extension to an A nucleotide 165 bp upstream from the
first in-frame ATG codon (11). To identify the start site of the large transcript, we searched the human expression sequence tag (EST) database provided by the National Center for Biotechnology Information using the BLASTN program for those files that match UGDH and have the longest 5′ extensions. As shown in Fig. 1, two major sets of EST could be clearly identified. As expected, a large number of ESTs have their 5′ ends clustered at approximately 165 bp upstream of the translation initiation codon that matches the reported start site for the small UGDH transcript (11). In addition, approximately 30 ESTs have their 5′ ends extending further upstream as far as 325 bp from the translation start. The entire 325-bp sequence aligned perfectly with the 5′ noncoding region of the human UGDH gene on chromosome 4, suggesting that it is indeed a part of the UGDH transcript. The full length mRNA with a 325 bp 5′-noncoding region is estimated to be 2.7 kb in size, and thus is in good agreement with the size of the large transcript observed on Northern blots. To make subsequent descriptions consistent with the previous report (11), the transcription start site of the human UGDH gene reported previously was designated as +1. The bioinformatics finding was further verified using 5′-RACE technology. When used with the upstream universal primer, the gene-specific primer GSP2, located upstream of the start site for the smaller transcript from −30, was able to amplify two DNA fragments approximately 450 bp and 170 bp in length (Fig. 1C, lane 2). While the sequence of the 450-bp fragment was found to be unrelated to UGDH, the −170-bp fragment contained, other than 45 nucleotides of the UPm primer sequences, a 129-bp 5′ region of UGDH extending to a “G” nucleotide at position −158. The 5′-RACE result is thus consistent with the prediction that an alternative transcription start site exists at a position around −160. The primer GSP1, corresponding to the sequence downstream of the small transcript, was capable of amplifying an expected 152-bp PCR product along with a smear as seen in lane 3 of Fig. 1C, indicating a possible upstream-extended region.

Deletion Analysis of the Human UGDH Promoter—To delineate the core promoter and other regulatory elements controlling the transcription of the UGDH gene, a 2.1-kb genomic fragment comprising the possible 5′ control region of the UGDH gene was dissected into a series of progressive 5′ and 3′ deletions. The plasmid containing the entire 2.1-kb fragment was named pJR001, while the three other successive 5′ deletion constructs chosen for this study were named pJR005 to pJR007. Since the UGDH gene is expressed at high levels in the liver, we used hepatoma cell line HepG2 as the model for this study. In addition, the cervical carcinoma cell line HeLa, which also expresses the UGDH gene according to our real-time quantitative PCR results, and shows a Ct value of ~20 for UGDH mRNA levels, was used for comparison in order to demarcate the minimal and essential regions in the UGDH promoter responsible for its expression.

The luciferase activities of the 5′ deletion constructs have been presented as percentage activity of the 2.1-kb promoter construct pJR001 (Fig. 2). The region from +183 to −59 in pJR007 displays negligible promoter activity. Consistent with this finding, deletion of this region in the 2.1-kb fragment (pJR008) also did not result in a significant effect on the UGDH promoter activity. The results indicate the presence of an additional region with promoter activity, presumably the promoter for the expression of the large UGDH transcript. As the region extends to −290 (pJR006), the luciferase activity rises 140% in HepG2 cells and 100% in HeLa cells. Since the start of the large transcript is around nucleotide position −159, this region can thus be thought to contain the core promoter for transcribing the large UGDH transcript. For plasmid pJR005, which contains the segment from position +183 to −632, the luciferase activity rises 150% and 200% higher than pJR001 in HepG2 and HeLa cells, respectively. This result suggests the presence of a strong cis-regulatory element in the region from position −290 to −632. The promoter activity of the next construct, pJR003 (+183 to −1340), dropped sharply in HeLa cells, while in HepG2 cells it remained the same as the full length promoter construct. This finding suggests the presence of an element that explains the differential expression of the UGDH gene between liver and epithelial cells. Finally, as observed from the promoter activity of the 2.123 kb construct, pJR001, this full length promoter region shows a much less promoter activity than pJR005. The distal region was not investigated further in the present study, but holds great promise for exploring the causes behind the suppression of promoter activity in the region upstream of −632.

An internal deletion of region −632 to −486 was made in construct pJR021, resulting in a significant decrease in promoter activity to approximately 14% in HepG2 cells and 20% in HeLa cells in comparison with pJR001 (Fig. 2). The transient transfection data suggest that the region from −632 to −486 is important for UGDH promoter activity, and is indicative of the presence of important positive regulatory elements in this proximal region of –159, this region can thus be thought to contain the core promoter for transcribing the large UGDH transcript.

Fig. 2. Deletion analysis of promoter activity of the human UGDH gene. Schematic diagram of the reporter constructs with a deletion of the 5′ control region of the UGDH gene is depicted on the left. The regions deleted are shown as blank spaces while the solid lines are the regions cloned upstream of the luciferase gene in the pGL-2 Basic vector. The luciferase activities normalized with EGFP and total protein concentration are shown on the right as bars and are relative to the percentage activity of the 2.1-kb full-length promoter construct, pJR001. The promoter activity in HeLa cells is depicted as black and in HepG2 as gray bars. The data represent the means of at least three independent experiments done in triplicate.
of the promoter, because its removal leads to a severe reduction in luciferase reporter activity in both cell lines.

Identification of Sp1 Binding Motifs That Contribute to UGDH Promoter Activity—The NIH Proscan program was used to predict a promoter region on the forward strand in –570 to –260 with a score of 60.39 (cutoff = 53.00). CpG island prediction revealed a large CpG island expanding from –250 to –570 (Fig. 3), along with as many as fourteen GC boxes in the 2.1-kb long 5′-flanking sequence of the UGDH gene. Our results of deletion analysis of the UGDH control region are consistent with the computer prediction that the region from –570 to –260 is important for UGDH gene expression. This region is most likely the core promoter region for transcribing the 2.7-kb mRNA, since it is in proximity to the transcription start site of this messenger RNA located at –159.

Based on the result of our 5′-end studies as well as the internal deletion analysis of the UGDH promoter, focus was placed on the region from –632 to –486 that possesses positive regulatory elements for the promoter activity. No consensus TATA or CAAT sequence could be identified upstream of the start site of the large transcript. Since the Sp1 transcription factor commonly plays an important role in regulating TATA-less genes, serving as the critical determinant of promoter activity and positioning the start site of transcription (17). Therefore, we then focused our research on the Sp1 binding site in the region from –632 to –486. By searching the TRANSFAC database, three putative Sp1 binding GC boxes, named SP#1, SP#2, and SP#3 (Fig. 3), were found in the important region and analyzed for this study. To verify the possible contribution of Sp1 on UGDH promoter activity, we treated cells transfected with the construct pJR005 with mithramycin A, a compound that competes with the transcription factor Sp1 for binding to GC boxes. HeLa cells were taken as a control for the presence of Sp1 transcription factors, and, therefore, were transfected together with HepG2 cells. At a concentration of 10^{-6} M, mithramycin A treatment reduced the activity of pJR005 to almost 15% of its original activity, suggesting the involvement of Sp1 (Fig. 4). This result, together with that of pJR0021 (Fig. 2B), clearly implicates the involvement of these putative Sp1 binding motifs in the –632 to –486 bp region in the regulation of UGDH gene expression.

Effects of Mutation of the Putative Sp1 Binding Sites on UGDH Promoter Activity—To examine further the individual contribution of these putative Sp1 binding sites to the transcriptional regulation of the UGDH gene, mutations were introduced in their core sequences (Table 1) in plasmid pJR005 (–632 to +183), which exhibits the highest promoter activity. The promoter activities of the mutations introduced in the three GC boxes in plasmid pJR005 were determined and are shown in Fig. 4. The mutation of sequence CGG to AAA in the SP#1 site led to a slight reduction in promoter activity. The mutation of GCGG to AAAA in SP#2 resulted in a 10% and 25% increase in promoter activity over the wild type in HepG2 and HeLa cells, respectively. In contrast, a mutation of SP#3 where GCGG was changed to AAAA caused a sig-

---

Received 11 June 2004; accepted 16 October 2004. Published online 22 March 2005.

Vol. 137, No. 6, 2005

Fig. 3. Nucleotide sequence of the upstream control region of human UGDH. The 650-bp sequence of the genomic fragment upstream of the human UGDH coding region is shown. The transcription start for the small transcript, an A nucleotide, that has been reported previously (11) is indicated by a gray background. The longest alternative transcription start identified from EST analysis as well as 5′-RACE is marked with a # sign. The Proscan predicted core promoter in the –158 to –567 region is underlined. The putative Sp1 binding GC boxes under study are shown in bold type letters and are named according to the order in which the respective GC boxes occur on the human UGDH promoter. The numbers to the left of the sequence are relative to the start site of the small UGDH transcript, which is designated as +1.

Fig. 4. Effects of site-specific mutation of three putative Sp1 binding sites on UGDH promoter activity. Two micrograms of pJR005 (wild type) and Sp1 binding site mutants were cotransfected along with 0.5 μg of EGFP control plasmid into HepG2 (gray bar) and HeLa cells (black bar). The activity of pJR005 has been set at 100%. The location of these Sp1 sites is depicted as circles, among which solid circles represent the wild type GC box and open ones indicate mutated sites. Treatment with mithramycin A (1.8 μM) of the cells transfected with pJR005 is a control for the suppression of Sp1 activity. The data represent the means of three independent experiments done in duplicate.
significant suppression of promoter activity in both HeLa and HepG2 cell lines. Thus site directed mutagenesis of the three GC boxes revealed the important role of SP#3 as a strong enhancer in both HepG2 and HeLa cells.

**Electrophoretic Mobility Shift Assay (EMSA) of Oligonucleotides Spanning the SP#3 GC Box**—To observe the pattern of protein interaction at the functionally important SP#3 GC box, and to verify whether this box represents Sp1 binding sites, EMSA was performed with HeLa and HepG2 nuclear extracts. The double stranded probes covering the GC box together with two other probes representing the consensus sequence for the binding site of Sp1 and AP-2 were designed (Table 1) and tested in the EMSA. The Sp1 consensus sequence probe formed the three major complexes, C1, C2, and C3 (Fig. 5A, lanes 1 and 3), with nuclear extracts from both cell types that are typically observed in other related studies concerning Sp1 (11). Similar to this finding, probe SP#3 produced three major complexes with the HepG2 nuclear extract that are identical to those found in the Sp1 consensus probe (Fig. 5B, lane 2). The complexes formed by SP#3 could be abolished by unlabeled self-probe as well as by the Sp1 consensus probe (Fig. 5B, lanes 7 and 9). The change of GCGG to AAAA in SP#3, the M3 probe, resulted

![Fig. 5. Verification of Sp1 binding on the GC boxes contained in the region from nucleotide positions -632 to -486 by EMSA. (A) Complex formation of the Sp1 consensus sequence oligonucleotide. Lane 1, with HepG2 nuclear extract; 2, without any nuclear extract; 3, with HeLa nuclear extract. (B) Binding pattern and competition assay for probe SP#3. Except lane 3, HepG2 nuclear extract was used in all cases. Lane 1, SP#3 probe without nuclear extract; 2, SP#3 probe; 3, SP#3 probe with HeLa nuclear extract; 4, SP#3 mutant probe M3; 5, Sp1 consensus probe; 6, AP-2 consensus probe. Lanes 7–10 are SP#3 probe competed by 100× concentration of unlabeled oligonucleotides of Sp1 consensus sequence (lane 7); SP#3 mutant M3 (lane 8); SP#3 (lane 9); AP-2 consensus (lane 10). The three complexes formed are shown by arrows and have been marked C1, C2 and C3. Hep, HepG2; NE, nuclear extract.**

![Fig. 6. Identification of authentic Sp1 binding sites by supershift assay. (A) Complex formation of the Sp1 consensus oligonucleotide probe with HepG2 nuclear extract in the absence (lane 1) and presence of an antibody specific to Sp1 factor (lane 2). (B) The supershift pattern of complexes formed between the SP#3 oligonucleotide probe and HepG2 nuclear extract in the presence of an antibody specific to Sp1 (lane 1) and an antibody specific to an unrelated antigen AP-2 (lane 2). Supershifts are marked with thick arrows. The three complexes are designated C1, C2, and C3.**

J. Biochem.
in the disruption of complex formation (Fig. 5B, lane 4). The unlabeled mutant SP#3 oligonucleotides could not compete away all the complexes formed with the labeled wild type probe. A distinct binding pattern was observed using the AP-2 consensus probe, and competition of the complex formed between SP#3 and the nuclear extract with the unlabeled AP-2 oligonucleotide did not show any effect.

To establish the identity of the protein interacting at the SP#3 GC box we included an antibody specific for Sp1 in the EMSA to generate supershifts. In the case in which the Sp1 consensus sequence was taken as the positive control, a super-shifted band was observed along with a high molecular weight complex (Fig. 6A, lane 2) that did not enter the gel (18). The complex C1 was nearly eliminated and partial shifts at C2 and C3 were also observed. A very similar supershift pattern was observed for the SP#3 probe (Fig. 6B, lane 1). Addition of the antibody against the AP-2 transcription factor did not cause any supershift of complexes formed at the SP#3 probe. Together, these results indicate SP#3 is indeed the binding site for Sp1.

In summary, our deletion analyses have delineated the important Sp1 sites on the UGDH promoter that were not analyzed before, and, at the same time have indicated the presence of inhibitory elements in the region upstream of –632 in the 2.1-kb upstream control region of the UGDH gene. The identification of at least two UGDH transcripts indicates that there are two core promoters typically located within 100 bp upstream of the transcription starts. A CpG island comprising an important Sp1 binding site at –564, which has been verified by site-directed mutagenesis and EMSA, is predicted to be located from –260 to –570. It would also be intriguing to investigate the elements in the distal region of the human UGDH promoter as well as many other factors as yet unexplored that might be responsible for the differential expression of this gene in tissues such as liver.

This work was supported by the National Program of Genomic Medicine of Taiwan, R.O.C. (NSC 92-2321-B-007-001).

REFERENCES