Transcription of Human ABO Histo-blood Group Genes Is Dependent upon Binding of Transcription Factor CBF/NF-Y to Minisatellite Sequence*

Yoshihiko Kominato‡§, Terumasa Tsuchiya‡, Nobuhide Hata‡, Hisao Takizawa‡, and Fumi-ichiro Yamamoto‡

From the ‡Toyama Medical and Pharmaceutical University, Faculty of Medicine, Department of Legal Medicine, Toyama, 930-01 Japan, the ¶Tokyo Medical and Dental University, Medical Research Institute, Tokyo, 113 Japan, and the ¶The Burnham Institute, La Jolla, California 92037

We have studied the transcriptional regulatory mechanism of the human histo-blood group ABO genes, and identified DNA cis-elements and trans-activating protein that control the expression of these genes which are important in blood transfusion and organ transplantation. We introduced the 5'-upstream sequence of ABO genes into the promoterless reporter vector and characterized the promoter activity of deletion constructs using transient transfection assays with gastric cancer cell line KATO III cells. The sequence just upstream of the transcription start site (cap site), and an enhancer element, which is located further upstream (between −3899 and −3618 base pairs (bp) from the transcription initiation site) and contains 4 tandem copies of a 43-bp repeat unit, were shown in gastric cancer cells to be responsible for the transcriptional activity of the ABO genes. DNA binding studies have demonstrated that a transcription factor, CBF/NF-Y, bound to the 43-bp repeat unit in the minisatellite. Functional importance of these CBF/NF-Y-binding sites in enhancer activity was confirmed by transfection experiments using reporter plasmids with mutated binding sites. Thus, transcriptional regulation of the human ABO genes is dependent upon binding of CBF/NF-Y to the minisatellite.

Histo-blood group ABH(O) antigens, the major alloantigens in humans (1), have been characterized as defined trisaccharide determinants GalNAca1−→3(Fucb1−→2)Galb1−→R, Galα1−→3(Fucb1−→2)Galb1−→R, and disaccharide determinant Fucb1−→2Galb1−→R for A, B, and H, respectively (2, 3). These structures represent the secondary gene products which are synthesized from the precursor H substrate by a1−→3GalNAc (A transferase) and α1−→3Gal transferase (B transferase), the primary gene products coded by the functional alleles at the ABO locus (4, 5). Molecular genetic studies of the ABO genotypes have identified two critical single-base substitutions between A and B genes, the resultant 2-amino acid substitutions being responsible for the different donor nucleotide-sugar substratespecificity between A and B transferases. A single base deletion, which shifts the codon reading frame and abolishes the function of A transferase, has been identified in O allelic cDNAs (6, 7).

ABH antigens are known to undergo drastic changes during development, differentiation, and maturation. Studies of these antigens in stratified squamous epithelia provided one of the clearest examples of differential expression during cell maturation (8). In non-keratinized stratified squamous epithelia, the immature cells in the basal layers are characterized by the expression of sialylated or unsubstituted precursor peripheral cores, while differentiated and mature cells in the upper layers sequentially express a1−→2-fucosylated H structures, and A and B antigens depending on the ABO genotype of the individual. This sequential expression of carbohydrate antigens is associated with the differentiation pattern of the epithelium. An interesting question is how these changes are controlled during cell differentiation. Since keratinocytes are known to greatly change their gene expression during terminal cell differentiation (9), the switch-on of the ABO genes during the maturation may be governed by the same factor(s). To fill in the gap between the expression of the ABO genes and the appearance of the ABO phenotypes in the terminal differentiation of epithelial cells, it is essential to understand the transcriptional regulatory mechanism of the ABO genes. In addition to the normal cell differentiation process, the changes of ABH antigen expression have also been documented in abnormal processes such as tumorigenesis. Reduction or complete deletion of A/B antigen expression in primary lung, bladder, and colorectal carcinomas have been reported. This phenotypic change was well correlated with the invasive and metastatic potentials of the tumors, and with 5- or 10-year mortality rates of the patients (10, 11). Expression of H/Leα, the A/B precursor, was inversely correlated with the survival rate of the patients with lung carcinoma after surgery (12). Disappearance of A/B antigens was ascribed to the absence of A or B transferase gene expression rather than the loss of these genes in human bladder tumors (13). Delineation of transcriptional regulation of the ABO genes, therefore, may provide clues as to the underlying mechanisms resulting in A/B antigen disappearance in cancer cells with invasive and metastatic potentials.

In an initial attempt to elucidate the molecular mechanisms controlling the expression of the human ABO genes, we have previously isolated five overlapping genomic clones to cover the entire coding sequence as well as some 5'-upstream and 3'-downstream sequences (14). In this paper, we report the identification of a DNA cis-element and trans-activating protein which accounts for the expression of the ABO genes in cells from gastrointestinal origin.

* This work was supported in part by grants-in-aid for scientific research from the Ministry of Education, Science and Culture, Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF014105.

‡ To whom correspondence should be addressed: Dept. of Legal Medicine, Faculty of Medicine, Toyama Medical and Pharmaceutical University, 2630 Sugitani, Toyama-shi, 930-01 Japan. Tel.: 81-0764-34-2281; Fax: 81-0764-34-5024; E-mail: ykomilm@ms.toyama-mpu.ac.jp.

© 1997 by The American Society for Biochemistry and Molecular Biology, Inc.
CBF/NF-Y Enhances Transcription of the ABO Genes

MATERIALS AND METHODS

Cell Culture, Transfection, and Luciferase Assay—A human gastric cancer cell line, KATO III (JCRB0611), was grown in 45% RPMI 1640 with 45% minimum essential medium containing 10% fetal bovine serum and 0.5% penicillin-streptomycin. Cells were transfected by electroporation. Briefly, the cells were inoculated at 3 × 10^6/ml 16–24 h prior to transfection. They were harvested by centrifugation at 500 × g for 5 min at room temperature, washed once with RPMI 1640, and resuspended at 5 × 10^6/ml in RPMI 1640. An aliquot (0.8 ml) of cell suspension was placed into a 0.4-cm electroporation cuvette (Bio-Rad). Twelve micrograms of supercoiled plasmid DNA was added to the cell suspension. Cells were incubated with DNA for 5 min at room temperature and then electroporated by electroporation at 250 Volts. Immediately after transfection, the cuvettes were transferred to an ice bath and kept for 15 min. The cells were then transferred to 10 ml of RPMI 1640 plus minimal essential medium containing 10% fetal bovine serum and 2 mM L-glutamine, incubated for 48 h, and harvested for luciferase and β-galactosidase assays.

Cell lysis and luciferase assays were performed following the manufacturer’s protocols using the Luciferase Assay System (Promega, Madison, WI). Light emission was measured in a model 1253 luminometer (BioOrbit), and the values were obtained in relative light units. The amount of cell lysate used for each luciferase assay was adjusted so that the observable light emission would fall within the linear range of the instrument. Variations in transfection efficiency were corrected by normalizing the activities of β-galactosidase expressed from the construct into the polylinker site just upstream of the luciferase gene. This activity was measured as described elsewhere (15). Relative activity of each construct in different experiments was obtained by arbitrarily assigning the activity of the pGL3-promoter vector containing the SV40 promoter to be 1.0.

Plasmids—A DNA fragment containing the 5′-upstream sequence of the human ABO gene was subcloned from the genomic clone HG-1 (14) into luciferase reporter vectors, pGL3-basic vector (Promega), pGL3-promoter vector (Promega), pGL2-control vector (Promega), and pTK-luc (16). The Smal site of the pGL3-basic vector was converted to the EcoRI site to facilitate the subcloning of the EcoRI/NcoI genomic fragment into the polylinker site just upstream of the luciferase gene. This construct was then used for the generation of several progressive series of upstream end deletion constructs. Since an NcoI site is present at the translation initiation codon in both ABO and luciferase genes, the linkage between the upstream region and translated region of the ABO gene was conserved in the reporter plasmids containing the progressive series of upstream end deletions. The PGL3-promoter vector contains the SV40 viral promoter, while pTK-luc contains a thymidine kinase promoter; the pGL2-control vector contains both the SV40 promoter and enhancer. The Smal site of the pGL3-promoter vector was converted to EcoRI site to facilitate the subcloning of the genomic fragments or PCR-amplified fragments into the polylinker site just upstream of the SV40 promoter. Nomenclature of the various ABO gene constructs is based on the nature of the inserted fragment. Letter symbols reflect the restriction enzyme cleavage sites used for the construction of these plasmids, while numerals indicate the end points of the primers used for the polymerase chain reaction (PCR). For example, EN construct contains the EcoRI/NcoI fragment (between −4661 and −31), and −3899H construct contains the fragment bordered with PCR primer sequence starting at −3899 on one end and HindIII site on the other. All the DNA fragments were generated by either restriction endonuclease digestion or PCR. Construction −3899–3618 was produced by overlapping PCR mutagenesis. The genomic DNA fragment between the XmnI site at −3252 and the PstI site at −2371 was inserted downstream of the SV40 enhancer in the PGL2-control vector, in the same orientation as in the ABO genes. PLG3-control vector (Promega), containing the SV40 promoter and enhancer, was used in promoter assays.

The −3899 to −3618 fragment containing 4 tandem copies of a 43-bp repeat unit was introduced into the upstream of promoter sequence in some constructs (−3899: −3618 (→)SN), and in the downstream of the luciferase gene in other constructs (SN: −3899: −3618 (→)SN, −3899: −3618 (→)TK, and −3899: −3618 (→)SV), and into the downstream of the luciferase gene in other constructs (SN: −3899: −3618 (→)SN, −3899: −3618 (→)SV, −3899: −3618 (→), and SV: −3899: −3618 (→)). Arrows to the right indicated that the fragment was inserted in the same orientation as appears in the ABO gene; the fragment was cloned in the same orientation as the luciferase genes in the reporter plasmids are transcribed. The fragment (−3899: −3618) was PCR-amplified and cloned first into pCR2 (Invitrogen) to generate construct pCR: −3899: −3618. The plasmid (−3899: −3618) was subsequently digested with EcoRI, modified with the HindIII site, and directionally ligated to the XhoI-digested pGL3-promoter vector and SalI-digested TK-luc, respectively. A fragment containing the −3899 to −3618 sequence was obtained by cleaving the plasmid −3899: −3618 (→)SV with Smal and BglII. One of the 5′ and 3′ ends of the −3899: −3618 fragment was modified to SalI site whereas the other end was oriented in either orientation. These fragments were then directionally ligated into the SalI/BamHI sites of pGL3-promoter vector and SN to construct four plasmids, SV: −3899: −3618 (→)SV, −3899: −3618 (→)SN, −3899: −3618 (→)SN, and −3899: −3618 (→). The plasmid −3899: −3618 (→)SN was prepared by digesting −3899: −3618 (→)SV with BglII, modifying the end to EcoRI site, cleaving with Nhel, and directionally ligating the fragment with Nhel/EcoRI-digested SN. The double-stranded oligonucleotides, TR23 and mTR23, were inserted into the SalI/BamHI sites of pGL3-promoter vector containing the 3′-untranslated region of the SV40 promoter to yield construct pGL3-promoter vector.

Plasmid DNA was purified by alcohol precipitation and recrystallized onto two successive CsCl-ethidium bromide gradients. Orientation, and the 5′ and 3′ boundaries of the insert of all the constructs used in this study were verified by detailed restriction enzyme mapping and DNA sequence analysis using Cycle Sequencing Kit with AmpliTaq DNA Polymerase, FS (Perkin-Elmer). For all the constructs containing PCR-amplified fragments, sequencing was performed over the entire region of the amplified sequences.

Preparation of Nuclear Extracts and Electrophoretic Mobility Shift Assay (EMSA)—Nuclear extracts and probes were prepared as previously reported (17, 18). 1 × 10^6 to 3 × 10^6 cells were harvested, washed with phosphate-buffered saline, and incubated for 5 min in 5 ml of ice-cold buffer A (10 mM Hepes pH 7.9, 5 mM MgCl_2, 10 mM NaCl, 0.3 mM sucrose, 0.1 mM EDTA, 0.5 mM dithiothreitol, 0.5 mM phenylmethylsulfonyl fluoride) containing 1 μg/ml each of protease inhibitors, antipain, aprotinin, chymostatin, leupeptin, and pepstatin A. After centrifugation, the cells were resuspended in 1 ml of buffer A with protease inhibitors and then Dounce homogenized. The homogenate was microcentrifuged for 30 s. Nuclei were resuspended in 0.8 ml of buffer B (20 mM Hepes pH 7.9, 5 mM MgCl_2, 300 mM KCl, 0.2 mM EDTA, 0.5 mM dithiothreitol, 0.5 mM phenylmethylsulfonyl fluoride) with protease inhibitors, and gently rocked on a plate at 4°C for 30 min. After 30 min of microcentrifugation at 4°C, the supernatant was dialyzed against 50 volumes of buffer D (20 mM Hepes pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0.2% glycerol, 0.5 mM dithiothreitol, 0.5 mM phenylmethylsulfonyl fluoride) at 4°C overnight. After 30 min of microcentrifugation at 4°C, aliquots of the supernatant were tractioned at −70°C, and the protein concentration was determined using a Bio-Rad protein assay kit.

We have employed two separate protocols for EMSAs. Results shown in Fig. 4, A and B, were obtained by the protocol previously described by Shirakawa et al. (18), using 4% polyacrylamide gel in 0.2× TAE buffer (6.67 mM Tris, pH 7.5, 3.3 mM sodium acetate, 1 mM EDTA). Binding reactions were performed in 15 μl of a buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 40 mM NaCl, 0.1 mM β-mercaptoethanol, 4% glycerol) containing 32P-labeled probe (0.2 μg, 5,000–20,000 cpm), 0.4 μg of poly(dI-dC), and 5 μg of nuclear extracts at room temperature for 15 min. The gels were electrophoresed at room temperature. The EMSAs performed in Fig. 5, A and B, followed the protocol described by Countryside et al. (19), using 4% polyacrylamide gel in 0.5× TBE buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA). Binding reactions were performed in 15 μl of a buffer (25 mM Hepes pH 7.9, 75 mM KCl, 1 mM EDTA, 0.5 mM phenylmethylsulfonyl fluoride, 0.5 mM dithiothreitol, 10% glycerol) containing 32P-labeled probe (0.2 μg, 5,000–20,000 cpm), 0.4 μg of poly(dI-dC), and 5 μg of nuclear extracts at room temperature for 15 min. The gels were electrophoresed at room temperature. One hundred-fold molar excess of unlabeled competitors over the labeled probe were used for competition analyses except those shown in Fig. 4.

For supershift experiments, 2 μl of polyclonal rabbit anti-NF-Y A subunit antibody (Rockland, Gilbertsville, PA), 2 μl of polyclonal rabbit anti-Egr-1 antibody (Santa Cruz Biotechnology) raised against a C-terminal Egr-1 peptide, or 2 μl of normal rabbit serum was added to the nuclear extract, and preincubated for 15 min on ice prior to addition of radio-labeled probes.

The abbreviations used are: PCR, polymerase chain reaction; bp, base pair(s); EMSA, electrophoretic mobility shift assay; kb, kilobase pair(s).
The different double-stranded oligonucleotides were obtained by annealing two chemically synthesized strands: EcoRI, 5'-GAATTCCTTCTGATGATTTTCTGATGGTCCTTATTGGCTATTTGGAAGGAA-3'; NF-I, 5'-GATCCCGTTTTCTGATGATTTTCTGATGGTCCTTATTGGCTATTTGGAAGGAA-3'; mTR23, 5'-CCACCCCAGCCAATAGGGGAAGGACGGATCCCG-3'.

The Eco-23 oligonucleotide sequence corresponds to nucleotides –66 to –45 in the major histocompatibility complex class II promoter (Eco-promoter) (20). The NF-I oligonucleotide contains NF-1 consensus sequence as described by Chodosh et al. (21). The underlines in the mTR23 oligonucleotide indicate the sites of specific mutations designed to abrogate the transcriptional factor CBF/NF-Y binding. We substituted the nucleotide sequences following Dorn et al. (20).

RESULTS

Identification of an Upstream Region Required for ABO Blood Group Gene Transcription—We have previously isolated several genomic clones of the human ABO genes (14). A genomic clone, designated as HG-1, was shown by restriction enzyme analysis and Southern hybridization to contain the 5'-upstream sequence of the coding region. A 7.8-kb EcoRI fragment, which encompassed 4.7 kb of the 5'-flanking region and the first exon as well as the first intron, was subcloned from this clone and sequenced. The complete nucleotide sequence of the 5'-flanking region has been deposited to GenBank (accession number U22309). Two possible transcription initiation sites were mapped by the 5'-rapid amplification of cDNA ends technique using human pancreatic cDNA as a template. Sequence analysis of the proximal region to these sites demonstrated the presence of several GC boxes just upstream of these possible transcription initiation sites, whereas neither TATA nor CAAT boxes are found close to these sites (14).

We have recently identified a minisatellite in the 5' upstream region. This minisatellite sequence is composed of 4 tandem copies of a 43-bp repeat unit, and is located at positions –3843 to –3672 relative to the upstream transcription start site. Minisatellite sequences have been shown to serve as regulatory regions for cellular transcription in certain genes (22–27). Nucleotide sequence between –3950 and –3470 of the human ABO gene is shown in Fig. 1. In this region around the minisatellite, several potential transcription factor-binding site motifs were identified. Therefore, it seemed important and interesting to assess the functional roles of this minisatellite sequence in the transcription of ABO genes.

As a means to examine the promoter/enhancer activity of the 5'-upstream sequences of the human ABO gene, we have employed reporter and transfection systems. We first obtained the EN construct by introducing the 4.7-kb EcoRI/NcoI genomic fragment 5'-flanking the coding sequence of the human ABO gene into the promoterless pGL3-basic vector at the upstream of the luciferase coding sequence (see Fig. 2). This plasmid was transiently transfected into the human gastric cancer cell line KATO III. This cell line was derived from the tumor of a blood group B individual.2 We have observed large amounts of B antigens on its cell surface, as well as high activity of B transferase in cell extracts (1.28 nmol/mg/h). Therefore, we have chosen these cells as candidates for recipients of DNA transfection assuming that enough amounts of B transferase in cell extracts were determined. The promoter activities in cell extracts were determined. The promoter activity of EN construct was at least 20-fold higher than that of pGL3-basic vector, and 3-fold lower than that of pGL3-promoter vector, which demonstrated the promoter activity of the 5'-upstream region of the human ABO genes (Fig. 2).

To better locate the sequences essential for the transcription, we have recently isolated several genomic clones of the human ABO genes (14). A genomic clone, designated as HG-1, was shown by restriction enzyme analysis and Southern hybridization to contain the 5'-upstream sequence of the coding region. A 7.8-kb EcoRI fragment, which encompassed 4.7 kb of the 5'-flanking region and the first exon as well as the first intron, was subcloned from this clone and sequenced. The complete nucleotide sequence of the 5'-flanking region has been deposited to GenBank (accession number U22309). Two possible transcription initiation sites were mapped by the 5'-rapid amplification of cDNA ends technique using human pancreatic cDNA as a template. Sequence analysis of the proximal region to these sites demonstrated the presence of several GC boxes just upstream of these possible transcription initiation sites, whereas neither TATA nor CAAT boxes are found close to these sites (14).

We have recently identified a minisatellite in the 5' upstream region. This minisatellite sequence is composed of 4 tandem copies of a 43-bp repeat unit, and is located at positions –3843 to –3672 relative to the upstream transcription start site. Minisatellite sequences have been shown to serve as regulatory regions for cellular transcription in certain genes (22–27). Nucleotide sequence between –3950 and –3470 of the human ABO gene is shown in Fig. 1. In this region around the minisatellite, several potential transcription factor-binding site motifs were identified. Therefore, it seemed important and interesting to assess the functional roles of this minisatellite sequence in the transcription of ABO genes.

As a means to examine the promoter/enhancer activity of the 5'-upstream sequences of the human ABO gene, we have employed reporter and transfection systems. We first obtained the EN construct by introducing the 4.7-kb EcoRI/NcoI genomic fragment 5'-flanking the coding sequence of the human ABO gene into the promoterless pGL3-basic vector at the upstream of the luciferase coding sequence (see Fig. 2). This plasmid was transiently transfected into the human gastric cancer cell line KATO III. This cell line was derived from the tumor of a blood group B individual.2 We have observed large amounts of B antigens on its cell surface, as well as high activity of B transferase in cell extracts (1.28 nmol/mg/h). Therefore, we have chosen these cells as candidates for recipients of DNA transfection assuming that enough amounts of trans-activating factors are present to support the expression of the introduced reporter constructs. We have used the electroporation method to obtain the higher transfection efficiency with the KATO III cells. pGL3-promoter vector containing the SV40 promoter, which exhibits strong activity in mammalian cells, and pGL3-basic vector without the promoter sequence were used as positive and negative controls, respectively. Forty-eight hours after transfection, the cells were harvested and the luciferase activities in cell extracts were determined. The promoter activity of EN construct was at least 20-fold higher than that of pGL3-basic vector, and 3-fold lower than that of pGL3-promoter vector, which demonstrated the promoter activity of the 5'-upstream region of the human ABO genes (Fig. 2).

To better locate the sequences essential for the transcription, we have recently isolated several genomic clones of the human ABO genes (14). A genomic clone, designated as HG-1, was shown by restriction enzyme analysis and Southern hybridization to contain the 5'-upstream sequence of the coding region. A 7.8-kb EcoRI fragment, which encompassed 4.7 kb of the 5'-flanking region and the first exon as well as the first intron, was subcloned from this clone and sequenced. The complete nucleotide sequence of the 5'-flanking region has been deposited to GenBank (accession number U22309). Two possible transcription initiation sites were mapped by the 5'-rapid amplification of cDNA ends technique using human pancreatic cDNA as a template. Sequence analysis of the proximal region to these sites demonstrated the presence of several GC boxes just upstream of these possible transcription initiation sites, whereas neither TATA nor CAAT boxes are found close to these sites (14).

We have recently identified a minisatellite in the 5' upstream region. This minisatellite sequence is composed of 4 tandem copies of a 43-bp repeat unit, and is located at positions –3843 to –3672 relative to the upstream transcription start site. Minisatellite sequences have been shown to serve as regulatory regions for cellular transcription in certain genes (22–27). Nucleotide sequence between –3950 and –3470 of the human ABO gene is shown in Fig. 1. In this region around the minisatellite, several potential transcription factor-binding site motifs were identified. Therefore, it seemed important and interesting to assess the functional roles of this minisatellite sequence in the transcription of ABO genes.
we have constructed a series of the human ABO promoter-luciferase chimeric plasmids in which 5'-flanking regions of the ABO gene of different lengths were fused with the coding sequence of the luciferase gene. The left panel of Fig. 2 depicts the 5'-flanking region of the human ABO gene and various restriction enzyme cleavage sites used to generate the fragments subcloned into the promoterless luciferase vector. Deletion of the human ABO gene upstream region from position −3950 to −3252, −2984, −2109, −1685, −1493, or −666 resulted in a large decrease in luciferase activity, demonstrating that the important sequences for the ABO gene transcription were contained within the deleted regions. The strongest decreasing effect appeared to be related to the deletion of the sequence between −3950 and −3252 containing 4 tandem copies of a 43-bp repeat unit. Surprisingly, however, deletion of the sequence from −3252 to either −202 or −117 resulted in an increase in the luciferase activity, suggesting that negative element(s) for the ABO gene transcription were present within this deleted region (−666 to −117). Furthermore, deletion of the sequence from −117 to −35 resulted in a large decrease in luciferase activity, showing the importance of promoter sequences immediately upstream of the transcription start site (cap site) in ABO gene expression. Sequence inspection of the region proximal to the cap site revealed potential binding sites for the regulatory protein Sp1. Taken together, two regions (−3950 to −3253 and −117 to +31) appear to function as positive regulatory cis-elements for the transcription of ABO genes in KATO III cells, whereas negative elements seem to exist in the region between −3252 and −117. To further examine the promoter activity of the region between −3950 and +31, an internal deletion downstream of the XmnI site at −3252 was introduced. The −3950 upstream terminus was chosen, because the maximum reporter activity was observed with the construct PN containing the region between the PstI site at −3950 and the NcoI site at +31 as shown in Fig. 2. An internal deletion located between the XmnI site at −3251 and the SacII site at −118 (construct PXm/SN) was constructed by introducing the PstI/XmnI fragment in SN construct. Transfection of construct PXm/SN into KATO III cells showed a 21-fold higher luciferase activity than did that of construct PN (Fig. 3). This result with the internal deletion seemed consistent with those obtained with upstream end deletions confirming the presence of positive elements for transcription in the PstI/XmnI fragment.

The sequence immediately upstream of the cap site (−666 to −117) seemed to be sufficient for the repression of the ABO cap-site-proximal promoter activity in constructs Xhn and KN. To examine whether additional regions are involved in the repression of the ABO cap-site-proximal promoter activity in the construct XmN, the fragment between the XmnI site at −3252 and the PstI site at −2371 was inserted into downstream of the SV40 enhancer in construct SV/SV-XmPst. Addition of this fragment resulted in a loss of 30% in luciferase activity (data not shown). Thus, it is likely that other negative cis-element(s) are present in the region between the XmnI and the SacII sites.

We have obtained another result to support the presence of positive enhancer elements within the −3950 to −3252 sequence. When the PstI-to-XmnI (PXm) fragment was separated from the ABO promoter-proximal sequence and linked directly to the SV40 promoter of pGL3-promoter vector (PXm/SV construct in Fig. 3), this construct expressed an 18-fold increase in luciferase activity when compared with the
original vector without this fragment. Several additional constructs were prepared from the PXm/SV plasmid and used for the further characterization of the ABO enhancer region. The sequence located between −3899 and −3618 appeared to be a significant functional component. Presence of additional sequence between −3617 and −3470 in construct −3899H/SV yielded a reduced activity, which implicated the presence of a negative regulatory element in the sequence. In agreement with this suggestion, a similar reduction was observed in the constructs PH/SV and P-3618/SV. The −3899 to −3618 sequence was cloned upstream of the ABO gene promoter, thymidine kinase promoter, and SV40 promoter in the same orientation as it appears in the ABO gene. Arrow to the right indicates 5′-CCAAT-3′ sequence located on the upper strand. The v-shaped segment represents deleted sequences.

The −3899 to −3618 sequence was cloned upstream of the ABO gene promoter, thymidine kinase promoter, and SV40 promoter in the same orientation as it appears in the ABO gene. Arrow to the right indicates 5′-CCAAT-3′ on upper strand.

**TABLE I**

| Reporter construct | Promoter | Enhancer | Relative activity | Fold activity over control |
|--------------------|----------|----------|-------------------|---------------------------|
| Test plasmids       |          |          |                   |                           |
| −3899: −3618 (→)/SN | ABO      | −3899: −3618 | 16.23 ± 2.55 | 77                        |
| −3899: −3618 (→)/TK | Thymidine kinase | −3899: −3618 | 4.07 ± 0.10 | 17                        |
| −3899: −3618 (→)/SV | SV40     | −3899: −3618 | 47.22 ± 6.44 | 47                        |
| Controls            |          |          |                   |                           |
| SN                 | ABO      | None     | 0.21 ± 0.04       |                           |
| pTK-luc            | Thymidine kinase | None | 0.23 ± 0.05 |                           |
| pGL3-promoter      | SV40     | None     | 1.00              |                           |

The results are expressed as an average relative activity compared to that observed for the pGL3-promoter vector. Standard deviations are indicated for a minimum of three repetitions.

The constructs as depicted above were transiently transfected into KATO III cells and assessed for functional importance; 10 µg of luciferase reporter and 4 µg of β-galactosidase expression vector were used for each analysis.

Identification of Protein-binding Sites within the Tandem

...
Repeat DNA Sequence—We have performed electrophoretic mobility shift assays (EMSAs) to investigate the nature of specific DNA-protein interaction(s) within the minisatellite. The locations of gene fragments used either as radiolabeled EMSA probes or as specific unlabeled competitors are shown schematically in Fig. 1. The results shown in Fig. 4A demonstrate that oligonucleotide TR43, containing the entire repeat unit in the minisatellite, did bind a KATO III cell nuclear factor (lane 2), and that this binding was inhibited by nonlabeled competitor TR43 (lanes 3 and 4), suggesting that the formation of the DNA-protein complex was specific. This complex, however, was not affected by the addition of either unlabeled TR38 (lanes 5 and 6) or TR39 oligonucleotide (lanes 7 and 8) containing a deletion of 5 or 4 bp, respectively, from the repeat unit, indicating that TR43 bound with a KATO III cell nuclear factor through the 5’-half of the repeat unit. Actually, the DNA-protein complex was competed for by the addition of the TR23 oligonucleotide (lanes 9 and 10) which contained only the 5’-half of the repeat unit. When either TR38 or TR39 was radiolabeled as a probe, definite binding of nuclear proteins was not observed (lanes 11 and 12). These results also support that the 5’-half-sequence is important for the binding.

To examine whether the 5’-half of the repeat unit is sufficient for binding with a nuclear factor, labeled TR23 oligonucleotide was used as an EMSA probe. Fig. 4B shows that this probe bound to a KATO III cell nuclear factor (lane 2). The complex formation was effectively inhibited by the addition of either the TR23 (lane 3) or the TR43 (lane 4) oligonucleotide as...
CBF/NF-Y Enhances Transcription of the ABO Genes

CBF/NF-Y Specifically Recognizes the 43-bp Repeat Unit of the Minisatellite—To examine whether CBF/NF-Y binds to the minisatellite, the competition assay was carried out using unlabeled oligonucleotide E\textsubscript{a}-23 containing the strong CBF/NF-Y-binding site in Y boxes of MHC class II promoter (lane 3). Since the TR23 oligonucleotide contained a half-palindromic sequence TGG(C/A) of the NF-I-binding site (31, 32), we have examined oligonucleotide NF-I (lane 4) which contained the NF-I consensus sequence (21), however, no competition was observed. In addition, anti-NF-I antibody failed to cause a supershift of the DNA-protein complex (data not shown). To more precisely identify the binding location, we mutated the CCAAT site as reported for the mouse E\textsubscript{a}-23 probe, which had the same relative migration as that observed with TR23 probe (compare lanes 1 and 7). Formation of this complex was inhibited by competition with unlabeled E\textsubscript{a}-23 and TR23 oligonucleotides (lanes 8 and 10) but not with NF-I or mutated TR23 oligonucleotide (lanes 9 and 11). Therefore, we concluded that CBF/NF-Y in KATO III cells nuclear extracts bound to the minisatellite sequence at the 23-bp sequence. In Fig. 4A, we have shown that the nucleotide sequences in the TR38 or TR39 oligonucleotide were not sufficient to form DNA-protein complex similar to that obtained with TR43 oligonucleotide. Neither TR38 nor TR39 oligonucleotides contained all the nucleotide sequence necessary for the contact with CBF/NF-Y, which was reported by Chodosh (21).

Furthermore, the motif flanking the 3' end of the CCAAT sequence was reported to be important for the high-affinity binding of CBF/NF-Y to the C1 element in the grp78/Bip promoter (33). Thus, it was unlikely that CBF/NF-Y could bind strongly to TR38 or TR39 oligonucleotides in the absence of an important nucleotide sequence. Since the E\textsubscript{a}-23 competitor inhibited the binding more efficiently than the TR23 oligonucleotide, binding of CBF/NF-Y to the 5' half of the minisatellite may be weaker than that to the mouse E\textsubscript{a} promoter CCAAT box.

To further investigate whether CBF/NF-Y binds to the repeat unit of the minisatellite, additional DNA binding studies were carried out using an anti-NF-Y A-subunit-specific antibody raised against an N-terminal NF-Y A-subunit peptide. The results of those experiments are shown in Fig. 5B. The anti-NF-Y A-subunit antibody supershifted the DNA-protein complexes between TR23 probe and the nuclear factor from KATO III cells (lane 4). In addition, the supershifted DNA-protein complex was also competed for by the E\textsubscript{a}-23 oligo containing the strong CBF/NF-Y-binding site (lane 5). Furthermore, DNA binding experiments performed with E\textsubscript{a}-23 probe, KATO III cell nuclear extracts, and anti-NF-Y antibody showed similar results to the ones obtained with TR23 probe (lanes 7–12). The supershifted DNA-protein complex was also competed for by the addition of TR23 oligonucleotide (lane 11). An
FIG. 6. Effects on enhancer activity of nucleotide substitutions in CBF/NF-Y-binding sites of the minisatellite. The structures of the reporter plasmids are shown schematically in the left panel. The reporter constructs contain the SV40 promoter alone, four copies of the 5′-half wild-type (TR23) or mutated (mTR23) oligonucleotide of the repeat unit cloned 5′ of the SV40 promoter region. Arrow to the right indicates the 5′-CCAAT-3′ on the upper strand. The open circles represent mutated sequences. The constructs as depicted at the left were transiently transfected into KATO III cells and assessed for functional activity; 10 μg of luciferase reporter and 4 μg of β-galactosidase expression vector were used for each analysis. The results are presented by the average fold of the activity observed with pGL3-promoter vector. Standard deviations were calculated from a minimum of three experiments.

CBF/NF-Y Enhances Transcription of the ABO Genes

To ascertain the functional importance of CBF/NF-Y sites in the minisatellite, four copies of the 5′-wild-type or mutated oligonucleotide were inserted upstream of the SV40 promoter. Transfection of the wild-type construct into KATO III cells showed an 8-fold increase in luciferase activity compared with that of the SV40 promoter vector (Fig. 6, construct 4×23/SV). Mutation of the CBF/NF-Y-binding site, which abrogated CBF/NF-Y binding as shown in Fig. 5a, resulted in a drastic loss of the luciferase activity and approached within a factor of 2 the expression level of pGL3-promoter vector (construct 4×m23/SV). This result suggests that CBF/NF-Y sites are important for the enhancer function and that these sites in the minisatellite may play a key role in the human ABO blood group gene expression. The wild-type construct exhibited 8-fold reduction in relative luciferase activity when compared with −3899:−3618/SV. This may imply that the minisatellite and its flanking regions function to exert a synergistic effect, although another explanation may be that the decreased activity may reflect the altered pitch of the CBF/NF-Y-binding sites relative to the SV40 promoter.

DISCUSSION

In this study, we have identified sequences essential for the transcriptional control of the human ABO blood group genes. Transient transfection of KATO III cells with luciferase reporter constructs demonstrated the importance of the promoter sequence immediately upstream of the transcription start site (cap site) in regulating the ABO gene expression. In addition, the far-upstream sequences located between −3899 and −3618 was shown to be crucial for the ABO gene transcription. The region contains a minisatellite composed of 4 tandem copies of a 43-bp repeat unit which exhibits enhancer activity. EMSA analysis demonstrated that the transcription factor CBF/NF-Y binds the minisatellite. A mutant construct, 4×m23/SV, containing the mTR23 sequence with 1-bp substitution which abrogates the binding of this factor, showed almost complete loss of activity in KATO III cells. This finding demonstrates that the CBF/NF-Y-binding sites in the minisatellite have a functional role. Thus, transcription of the human ABO blood group gene may be regulated, at least in part, by the binding of CBF/NF-Y to the tandem copies of repeats in the minisatellite. Our finding that the minisatellite has enhancer activity offers additional clues as to how other repetitive sequences may play functional roles.

Minisatellites are highly repetitive DNA sequences typically found in mammalian genomes. They vary in length from a few base pairs to several thousands. They also vary in complexity from simple di- and trinucleotide repeats (microsatellites) to more complicated repetitive elements. Repetitive DNA, most notably trinucleotide repeats, have been implicated in several human diseases such as fragile X syndrome, myotonic dystrophy, and Kennedy's disease (34, 35). Although the mechanisms by which repetitive DNA is associated with these diseases have been intensively studied in the past years, they are still unknown. However, the studies on the biological functions of minisatellites have brought new findings. For example, recent studies on trinucleotide repeats have underscored the importance of repetitive DNA in a variety of biological processes, ranging from recombination to generation of nucleosome positioning signals (36, 37). Other observations on tandem repeats have demonstrated that some human minisatellites might serve as regulatory elements for cellular transcription. Indeed, the 28-bp repeat unit of a minisatellite 1-kb downstream from the human HRAS1 gene was shown to bind several members of the rel/NF-kB family of transcriptional regulatory factors (24). This minisatellite was later reported to possess transcriptional regulatory activity that is dependent on the promoter and the cell lines used in the transient transfection experiments (25). A member of the myc/helix-loop-helix family closely related to upstream factor/major late transcription factor was shown to bind to the 50-bp repeat unit of a minisatellite within the D12-I4 interval of the human immunoglobulin heavy chain locus. This minisatellite was demonstrated to compete for upstream factor/major late transcription factor, resulting in suppression of the activation of the adenovirus major late promoter (26). More recently, transcriptional function was recognized for a polymorphic minisatellite in the 5′-flanking insulin-linked polymorphic region of the human insulin gene. Numerous high-affinity binding sites for the transcription factor Pur-1 were identified in the minisatellite, and minisatellite-dependent enhancement in transcription was observed with both the native human insulin promoter and a heterologous promoter linked to the reporter gene (27). Moreover, transcriptional activity was dependent on the length and the number of repeats in the minisatellite, with the longer minisatellites (up to ~2.5 kb) being more active than the short ones (~0.5–1 kb). Interestingly, a subgroup of short minisatellite seemed to be associated with susceptibility to type I diabetes (38). Our results, together with others, strongly indicate that minisatellites influence the behaviors of the nearby genes.

The results that transcriptional regulation of the human ABO gene is dependent upon the binding of transcription factor CBF/NF-Y to the minisatellite may provide clues as to how the expression of ABO genes is controlled during differentiation of epithelial cells. Although CBF/NF-Y is a well known constitutive and ubiquitous factor, a possibility remains that CBF/NF-Y may be modulated during cellular differentiation. CBF/NF-Y activity was shown to be serum dependent in IMR-90 diploid fibroblasts (39). It is of interest that the serum-induced enhancement of CBF/NF-Y binding to the distal CCAAT box of the human TK gene was mediated by the enhanced expression of NF-YA subunit, but not NF-YB. Recently, another example of enhanced CBF/NF-Y activity by elevated levels of NF-YA subunit has been reported in the heme-treated Friend leukemia cells and during the monocyte-to-macrophage diferenti-
tion process (40). The relevance of this enhanced binding of CBF/NF-Y to the CCAAT box has also been reported in the promoter region of the ferritin H gene in differentiated Caco-2 cells, a human colon cell line capable of differentiating to enterocyte-like cells (41). Therefore, it seems plausible to assume that the ABO gene expression is influenced by CBF/NF-Y binding activity through cell differentiation.

Various human cancers are known to frequently lose blood group A/B antigens. Recently, a complete correlation between the absence of the ABO mRNA transcript and the absence of group A/B antigens. Recently, a complete correlation between blood group A/B antigens has been reported (13). The authors also showed that growth stimulation with the cholera toxin B or epidermal growth factor led to a total loss of the ABO mRNA formed. Reduced levels of transcription can be caused several ways, including inhibition of an active transcription factor or appearance of repressors binding to the ABO gene sequence. This down-regulation may be due to both decreased transcription and increased breakdown of mRNA transcripts. This down-regulation may be due to both decreased transcription and increased breakdown of mRNA formed. Reduced levels of transcription can be caused several ways, including inhibition of an active transcription factor or appearance of repressors binding to the ABO gene sequence. Our results help delineate the mechanism of A/B antigen disappearance which correlates with invasive and metastatic potentials of cancer cells. Further investigation is required to elucidate the inhibitory mechanisms.

Acknowledgments—We thank Naoko Taneso for providing anti-NF-I antibody. We also thank K. Shiraki and K. Hiraga for helpful suggestions and criticism, and R. J. Connor and S. Ramey for comments on the manuscript.

REFERENCES

1. Landsteiner, K. (1900) Zentralbl. Bakteriol. 27, 337–363
2. Morgan, W. T. J., and Watkins, W. M. (1969) Br. Med. Bull. 25, 30–34
3. Kabat, E. A. (1973) in Carbohydrates in Solution (Advanced Chemistry Society 117) (Isbell, H., ed) pp. 334–361, American Chemical Society, Washington, D. C.
4. Watkins, W. M. (1980) in Advances in Human Genetics (Harris, H., and Hirschhorn, K., eds) Vol. 10, pp. 1–136, Plenum Publishing Co., New York
5. Ginsburg, V. (1972) Adv. Enzymol. 36, 131–149
6. Yamamoto, F., Clausen, H., White, T., Marken, J., and Hakomori, S. (1990) Nature 345, 229–233
7. Yamamoto, F., and Hakomori, S. (1990) J. Biol. Chem. 265, 19257–19262
8. Dabelsteen, E., Vedtofte, P., Hakomori, S., and Young, W. W. (1982) J. Invest. Dermatol. 79, 3–7
9. Teumer, J., Zeeulak, K., and Green, H. (1994) in The Keratinocyte Handbook (Leigh, I. M., Lane, E. B., and Watt, F. M., ed) pp. 235–242, Cambridge University Press, Cambridge
10. Lee, J. S., Ro, J. Y., Sahin, A. A., Hong, W. K., Brown, B. W., Mountain, C. F., and Hittelman, W. N. (1991) N. Engl. J. Med. 324, 1084–1090
11. Matsumoto, H., Muramatsu, H., Shimotakahara, T., Yanagi, M., Nishijima, H., Mitani, N., Baba, K., Muramatsu, T., and Shimazu, H. (1993) Cancer 72, 75–81
12. Miyake, M., Taki, T., Hitomi, S., and Hakomori, S. (1992) N. Engl. J. Med. 327, 14–18
13. Ørntoft, T. F., Meldgaard, P., Pedersen, B., and Wolf, H. (1996) Cancer Res. 56, 1031–1036
14. Yamamoto, F., McNill, P. D., and Hakomori, S. (1995) Glycobiology 5, 51–58
15. Edlund, T., Walker, M. D., Barr, P. J., and Rutter, W. J. (1985) Science 229, 912–916
16. McKnight, S. L., Gavis, E. R., and Kingsbury, R. (1981) Cell 25, 385–398
17. Kominato, Y., Galson, D. L., Waterman, W. R., Webb, A. C., and Aaron, P. E. (1995) Mol. Cell. Biol. 15, 58–68
18. Shirakawa, F., Saito, K., Bonagura, C. A., Galsen, D. L., Fenton, M. J., Webb, A. C., and Aaron, P. E. (1995) Mol. Cell. Biol. 13, 1332–1344
19. Country, F., Maity, S. N., and de Crombrugghe, B. (1995) J. Biol. Chem. 270, 468–475
20. Dorn, A., Bollekens, J., Staub, A., Benoist, C., and Mathis, D. (1985) Cell 50, 363–372
21. Chodosh, L. A., Baldwin, A. S., Carthew, R. W., and Sharp, P. A. (1988) Cell 53, 11–24
22. Cohen, J. B., Walter, M. V., and Levinson, A. D. (1987) J. Cell. Physiol. 5, suppl. 75–81
23. Spanidou, D. A., and Holmes, L. (1987) FEBS Lett. 218, 41–46
24. Trepechio, W. L., and Krontris, T. G. (1992) Nucleic Acids Res. 20, 2427–2434
25. Green, M., and Krontris, T. G. (1993) Genomics 17, 429–434
26. Trepechio, W. L., and Krontris, T. G. (1993) Nucleic Acids Res. 21, 977–985
27. Kennedy, G. C., German, M. S., and Rutter, W. J. (1995) Nature Genet. 9, 293–298
28. Maity, S. N., Gelumbek, P. T., Karsenty, G., and Crombrugghe, B. (1988) Science 241, 582–585
29. Maire, P., Wuarin, J., and Schibler, U. (1989) Science 246, 343–346
30. Sinha, S., Maity, S. N., Lu, J., and Crombrugghe, B. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 1624–1628
31. Jones, K. A., Kadenaga, J. T., Rosenfeld, P. J., Kelly, T. J., and Tjian, R. (1987) Cell 48, 79–89
32. Santoro, C., Mermod, N., Andrews, P. C., and Tjian, R. (1988) Nature 334, 218–224
33. Roy, B., and Lee, A. S. (1995) Mol. Cell. Biol. 15, 2263–2274
34. Maddox, J. (1994) Nature 366, 685
35. Mandel, J.-L. (1997) Nature 386, 767–769
36. Wahl, W. P., Wallace, L. J., and Moore, P. D. (1990) Cell 60, 95–103
37. Wang, Y.-H., Amirhaeri, S., Kang, S., Wells, R. D., and Griffith, J. D. (1994) Science 265, 669–671
38. Owerbach, D., and Gublay, K. H. (1993) Diabetes 42, 1708–1714
39. Chang Z.-F., and Liu, C.-J. (1994) J. Biol. Chem. 269, 17893–17898
40. Marziali, G., Perrotti, E., Ilari, R., Testa, U., Coccia, E. M., and Battistini, A. (1997) Mol. Cell. Biol. 17, 1387–1395
41. Bevilacqua, M. A., Faniello, M. C., D'Agostino, P., Quaresima, B., Tiano, M. T., Pignata, S., Russo, T., Cimino, F., and Costanzo, F. (1995) Biochem. J. 311, 769–773