Alternative Promoter Identified between a Hypermethylated Upstream Region of Repetitive Elements and a CpG Island in Human ABO Histo-blood Group Genes*

Yoshihiko Kominato‡§, Yukiko Hata‡, Hisao Takizawa‡, Kayoko Matsumoto‡, Kazuta Yasui‡, Jun-ichi Tsukada¶, and Fumi-ichiro Yamamoto**

From the ‡First Department of Internal Medicine, Toyama Medical and Pharmaceutical University, Faculty of Medicine, Toyama 930-0194, Japan, the ¶Osaka Red Cross Blood Center, Osaka 536-8505, Japan, the (First Department of Internal Medicine, University of Occupational and Environmental Health, School of Medicine, Kitakyushu 807-8555, Japan, and the **Burnham Institute, La Jolla, California 92037.

Received for publication, May 1, 2002, and in revised form, July 8, 2002 Published, JBC Papers in Press, July 31, 2002, DOI 10.1074/jbc.M204239200

We have studied the expression of human histo-blood group ABO genes during erythroid differentiation, using an ex vivo culture of AC133+/CD34+ cells obtained from peripheral blood. Rapid amplification of cDNA ends analysis of RNA from those cells revealed a novel transcription start site, which appeared to mark an alternative starting exon (1a) comprising 27 bp at the 5'-end of a CpG island in ABO genes. Results from reverse transcription-PCR specific to exon 1a indicated that the cells of both erythroid and epithelial lineages utilize this exon as the transcription starting exon. Transient transfection experiments showed that the region just upstream from the transcription start site possesses promoter activity in a cell type-specific manner when placed 5' adjacent to the reporter luciferase gene. Results from bisulfite genomic sequencing and reverse transcription-PCR analysis indicated that hypermethylation of the distal promoter region correlated with the absence of transcripts containing exon 1a, whereas hypermethylation in the interspersed repeats 5' adjacent to the distal promoter was commonly observed in all of the cell lines examined. These results suggest that a functional alternative promoter is located between the hypermethylated region of repetitive elements and the CpG island in the ABO genes.

In 1900 Karl Landsteiner discovered the ABO blood group system, which is important in blood transfusions and personal identification in criminal investigations (1). Two carbohydrate antigens, A and B, and their antibodies constitute this system. The functional A and B alleles at the ABO genetic locus encode glycosyltransferases A1→3GalNAc transferase (A-transferase) and A1→3Gal transferase (B-transferase), respectively. A-transferase transfers a GalNAc residue from UDP-GalNAc to the precursor H substrate, producing A antigens as defined by the trisaccharide determinant structure GalNAc1→3(Fucα1→2)Galβ1→R. Similarly, B-transferase catalyzes the transfer of a Gal from UDP-Gal to the same H substrate, producing B antigens defined by Galα1→3(Fucα1→2)-Galβ1→R (2–5). Molecular genetic studies of human ABO genes have demonstrated that ABO genes consist of at least seven exons spanning over 18 kb of genomic DNA and that two critical single base substitutions in the last coding exon result in amino acid substitutions responsible for the different donor nucleotide sugar substrate specificity between A- and B-transferases. A single base deletion in exon 6 was ascribed to shift the reading frame of codons and to abolish the transferase activity of A-transferase in most O alleles (6–9).

The ABO antigens are expressed in a cell type-specific manner; the isoantigens A, B, and H of blood groups A, B, and O are not confined to red cells but are also found in many secretions and on some epithelial cells. However, they are absent in connective tissues, muscles, and the central nervous system (10). Moreover, ABH antigens are known to undergo drastic changes during development, differentiation, and maturation of cells in epithelial lineage as well as erythroid lineage. Immature cells in the basolateral layers in nonkeratinized stratified squamous epithelia, for example, were characterized by the expression of sialylated or unsubstituted precursor peripheral cores, whereas differentiated and mature cells in the upper layers sequentially expressed α1→2 fucosylated H structures and A/B antigens, depending on the ABO genotype of the individual (11). Similar to ABH antigen expression in epithelia, studies of A antigen expression during maturation of erythroid progenitors in a two phase liquid culture system showed that A-positive cells gradually increased during erythroid maturation (12, 13). Fluorescence-activated cell sorter (FACS)1 analysis using monoclonal antibodies demonstrated the expression of A antigens on colony cells derived from BFU-E and CFU-E (14). The changes of ABH antigen expression have also been documented in pathological processes such as tumorigenesis. Reduction or complete deletion of A/B antigen expression in carcinomas has been reported (15–17). In addition, the loss of ABH antigens has been correlated with tumor progression of various carcinomas, including those in lung and bladder (18–21).

To elucidate the molecular basis of how ABO genes are controlled in cell type-specific expression, during normal cell differentiation, and in cancer cells lacking A/B antigens with

---

* This work was supported in part by a grant-in-aid for scientific research and from the Ministry of Education, Science, Sports, 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) U22302.

‡ To whom correspondence should be addressed: Dept. of Legal Medicine, Faculty of Medicine, Toyama Medical and Pharmaceutical University, 2630 Sugitani, Toyama 930-0194 Japan. Tel.: 81-76-434-7281; Fax: 81-76-434-5024; E-mail: ykomilm@ms.toyama-mpu.ac.jp.

---

1 The abbreviations used are: FACS, fluorescence-activated cell sorter; GPDH, glyceraldehyde-3-phosphate dehydrogenase; MBP, methyl-CpG binding domain; MEM, minimum Eagle’s medium; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription-PCR; SV40, simian virus 40.
invasive and metastatic potential, it is essential to understand the regulatory mechanism of ABO gene transcription. Previously we determined two transcription start sites just upstream from the initiation codon by the 5'-rapid amplification of cDNA ends (5'-RACE) technique using human pancreatic cDNA as a template (8). Transient transfection experiments demonstrated that the ABO gene promoter was located between -117 and +31, relative to the upstream transcription start site in both epithelial and erythroid lineages (22, 23). Like many housekeeping genes, the ABO gene contains a typical CpG island that extends from 0.7 kb upstream to 0.6 kb downstream from the transcription start site in exon 1 (24). Expression of the ABO genes was shown to be repressed upon DNA methylation of the CpG island in the promoter region (24, 25). Neoplastic cells simultaneously harbor widespread genomic hypomethylation and regional areas of hypermethylation (26). The CpG island located in the gene promoter region is usually the main target of this regional hypermethylation. DNA methylation of CpG islands spanning the promoter is often strongly associated with transcriptional silencing. A group of the methyl-CpG binding domain proteins (MBDs) bind preferentially to methylated CpGs (27–30), and several of these MBD proteins associate with histone deacetylases or Mi-2, a member of the SWI2/SNF2 family of ATP-dependent chromatin-remodeling proteins (31–34). The direct linkages among the MBDs, histone deacetylases, and chromatin remodeling machinery have provided a basis for understanding how DNA methylation can be related to a transcriptionally incompetent chromatin state. Further characterization of the upstream region of the ABO gene is necessary to obtain a better understanding of the underlying mechanisms that result in the cell type-specific expression of the ABO genes and the changes in ABO gene expression during cell differentiation. Furthermore, it will become also informative for the elucidation of the molecular basis of DNA methylation in the ABO gene promoter in cancer cells. Here we report the identification of an alternative promoter of the ABO genes.

**EXPERIMENTAL PROCEDURES**

**Cells**—Human gastric cancer cell lines MKN1 (JCRB0252), MKN28 (JCRB0253), MKN45 (JCRB0254), human erythroleukemic cell lines K562 (JCRB0019) and HEL (JCRB0062), and T cell line Jurkat (RCB0023), and normal human fibroblasts were grown in RPMI 1640 containing 10% fetal bovine serum (Invitrogen), 50 μg/ml penicillin, and 50 μg/ml streptomycin. A human gastric cancer cell line KATOIII (JCRB0611) was grown in 45% RPMI 1640 plus 45% minimum essential medium (MEM) containing 10% fetal bovine serum, 50 μg/ml penicillin, and 50 μg/ml streptomycin. Human embryonal lung fibroblasts were maintained in MEM containing 10% fetal bovine serum, 50 μg/ml penicillin, and 50 μg/ml streptomycin.

**AC133**—CD34**+** cells were prepared as reported previously with some modifications (35). Mononuclear cells isolated from peripheral blood with Ficoll-Paque PLUS (Amersham Biosciences) were first subjected to immunomagnetic separation using a MACS AC133 Cell Isolation Kit (Miltenyi Biotech, Auburn, CA). The cells in the flow-through fraction were then incubated with AC133 microbeads prior to the application to the second column. The cells in the flow-through fraction were subjected to immunomagnetic separation using a MACS CD34 Progenitor Cell Isolation Kit (Miltenyi Biotech). The trapped cells were collected and designated as AC133**+** CD34**+** cells. As the mononuclear cells isolated from blood were applied to the CD34 immunoisolated cells, the cells in the flow-through fraction from the column were also collected and designated as the flow-through fraction.

**Ex Vivo Expansion**—Ex vivo expansion of the AC133**+** CD34**+** cells was carried out according to a method reported previously (35). The AC133**+** CD34**+** cells were suspended in StemPro-34 serum-free medium (Invitrogen) supplemented with 2 mM L-glutamine, 50 units/ml penicillin, 50 μg/ml streptomycin, 5 ng/ml Flt3 ligand (DIA-CLONE Research, Besancon, Cedex, France), 10 ng/ml thrombopoietin (Kirin Brewery Co. Ltd., Maebashi, Japan), and 50 ng/ml stem cell factor (PeproTech, Inc., Rocky Hill, NJ). The cultures were initiated at 1 or 2 × 10⁶ cells in nontreated 24-well plates and were maintained at 37 °C in a humidified atmosphere of 5% CO₂ for 7 days. On day 7, half of the medium was changed, and 2 units/ml erythropoietin (Kirin Brewery Co. Ltd., Maebashi, Japan) was added. At days 7 and 14, the cells were harvested for RT-PCR and FACS analysis.

**Reverse Transcription-PCR (RT-PCR)**—Total RNA was isolated from cultured cells using the acid guanidium thiocyanate/acid phenol method. cDNA was synthesized from total RNA (2 μg) using random hexamers and Superscript II (Invitrogen). One and 2 μl of the resulting single strand cDNA reaction were used as templates for RT-PCR and quantification, respectively. Five-μl PCR products with bone marrow Marathon-Ready cDNA (Clontech, Palo Alto, CA) was also used as a template for RT-PCR.

**Amplication of the ABO gene message was performed using primers ABO+116 and ABO+802 corresponding to the sequences within exon 3 and 7 of the ABO gene, respectively, as described previously (34). The starting exon-specific RT-PCR was carried out using each of these starting exon-specific primer and the reverse primer ABO+802. The sequences of the starting exon-specific primers were 5’-GGGCGAGGT- GTTGGCGGAGGCT-3’ (ABO+3) and 5’-GAGGTTCTCTCGAGGCAGGC- CAA-3’ (ABO-U678), corresponding to the sequence in exon 1 which was reported previously (8, 9) and the sequence in the alternative starting exon 1a reported in this paper, respectively. Conditions for the first strand RT-PCR were 95 °C for 1 min, 35 cycles of 94 °C for 30 s, and 72 °C for 1 min, followed by incubation at 72 °C for 10 min. Another starting exon-specific RT-PCR was performed using either one of the starting exon-specific primers or the reverse primer ABO+98, of which the sequence was 5’-CCAAACAAGACCAAGACACATTATTAGG-3’, complementary to exon 2 of the ABO gene. Conditions for these amplifications were 95 °C for 9 min, 35 cycles of 94 °C for 1 min, 67 °C for 1 min, and 72 °C for 2 min. These combinations of primers were used in quantitative real time RT-PCR. RT-PCR was performed to amplify the GATA-1, β-globin, and glyceraldehyde-3-phosphate dehydrogenase (G3PDH) gene messages according to the methods by Kosugi et al. (36).

PCR amplifications were performed in a 100-μl reaction mixture containing 20 pmol of each primer, 2.5 units of AmpliTaq Gold (Applied Biosystems, Foster City, CA), 1.5 mM MgCl₂, 150 μM dNTP, and 1× buffer (Applied Biosystems). The products were resolved on a 2% agarose gel. After cloning the PCR-amplified products into a pCR2.1 plasmid vector (Invitrogen), the nucleotide sequences of the amplified fragments were determined, using the ABI PRISM dRhodamine Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA polymerase F8 (Applied Biosystems) with both M13 forward and reverse primers.

**5’-RACE**—5’-RACE was performed using the 5’-RACE system, version 2.0 (Invitrogen) according to the manufacturer’s instruction. The first strand cDNA was synthesized using 5 μg of total RNA and the ABO gene-specific primer, the sequence of which was 5’-TGCGCCAC- CATGAAATGTGCTT-3’ (primer ABO+433). After homopolymeric dC tails were added to the 3’-ends of the cDNA, the second strand DNA was amplified using 5’-RACE Abridged Universal Amplification Primer provided by the supplier. The products were purified, followed by PCR amplification using the Abridged Universal Amplification Primer provided by supplier and the nested primer fy-123 that was reported previously (8). Conditions for the second strand DNA amplification were 95 °C for 9 min, 40 cycles of 94 °C for 1 min, 55 °C for 1 min, and 72 °C for 2 min. These combinations of primers were used in quantitative real time RT-PCR. RT-PCR was performed to amplify the GATA-1, β-globin, and glyceraldehyde-3-phosphate dehydrogenase (G3PDH) gene messages according to the methods by Kosugi et al. (36).

PCR amplifications were performed in a 100-μl reaction mixture containing 20 pmol of each primer, 2.5 units of AmpliTaq Gold (Applied Biosystems, Foster City, CA), 1.5 mM MgCl₂, 150 μM dNTP, and 1× buffer (Applied Biosystems). The products were resolved on a 2% agarose gel. After cloning the PCR-amplified products into a pCR2.1 plasmid vector (Invitrogen), the nucleotide sequences of the amplified fragments were determined, using the ABI PRISM dRhodamine Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA polymerase F8 (Applied Biosystems) with both M13 forward and reverse primers.
cDNA starting from exon 1 and exon 1α, respectively, and plasmid pG3PDH containing the entire G3PDH cDNA were used to generate a calibration curve. Construction protocols of these plasmids will be described later. Amplification of a control plasmid with an unrelated primer set did not generate an increase in reporter fluorescence, and PCR products were barely detected in gel electrophoresis after ethidium bromide staining (data not shown). The plasmid templates were measured using a spectrophotometer, and copy numbers were calculated from the absorbance at 260 nm. For each assay, a standard curve was prepared using serial dilutions of template plasmid DNA with known copy numbers in log steps from 10⁷ copies down to 1 copy in a 2× prepared using serial dilutions of template plasmid DNA with known copy numbers in log steps from 10⁷ copies down to 1 copy in a 2×

| PCR target | Position | 5'-primer | 3'-primer |
|------------|----------|-----------|-----------|
| V          | −1497    | −1017     | TAACTCATACCTACAATCTCTAACA |
| III        | −784     | −414      | CAACCTTACACCGAAGAGAGAGG |
| I+II       | −434     | +47       | CCAACCTTACACCGAAGAGAGG |
| I−II       | −234     | −188      | CAACCTTACACCGAAGAGAGG |

Table I

**Oligonucleotide primer pairs used for amplification of bisulfite-treated genomic DNA**

and self-igation. Plasmid XhN-Δ275→-118 was prepared by directional ligation of the double strand oligonucleotide, corresponding to the −335 to −276 sequence, with the Smal, SacI-digested, blunt ended XhN, Orientation and 5'- and 3'-boundaries of the insert of all of the constructs used in this study were verified by restriction enzyme mapping and by DNA sequencing. For all of the constructs containing PCR-amplified fragments and chemically synthesized oligonucleotides, sequencing was performed over the entire region of the amplified sequences and the oligonucleotides. Plasmid DNA was purified by applying alkaline lysed samples onto two successive CsCl-ethidium bromide gradients.

**Transfection and Luciferase Assay**—Transient transfection experiments into KATOIII cells, MKN28 cells, and HEL cells were performed as reported previously (22–24). KATOIII cells were cotransfected by using 10 ng of luciferase reporter plasmids and 4 µg of the control plasmid containing the Rous sarcoma virus long terminal repeat directing *Escherichia coli* β-galactosidase expression. MKN28 cells were transfected with Lipofectin reagent (Invitrogen); 1.4 µg of luciferase reporter and 0.6 µg of β-galactosidase control vector were used for each analysis. HEL cells were transfected with DMRIE-C reagent (Invitrogen); 4 µg of luciferase reporter and 2 µg of β-galactosidase control vector were used for each analysis. MKN1 cells were transfected with Luciferin according to the same protocol used in the transfection of MKN28 cells. Human embryonal lung fibroblasts were transfected with Lipofectin; 1.4 µg of luciferase reporter and 0.6 µg of β-galactosidase control vector were used for each analysis. The fibroblasts were split, 18–24 h prior to transfection, into a six-well tissue culture plate (BD Biosciences) at 1 × 10⁶/ml. At the time of transfection, cells were washed once with MEM containing neither fetal bovine serum nor 1-glutamine. Two µg of supercoiled plasmid DNA were suspended in 100 µl of Opti-MEM I reduced serum medium (Invitrogen). Ten µl of Lipofectin reagent was diluted in 100 µl of Opti-MEM I reduced serum medium. The two solutions were combined at room temperature for 10–15 min followed by the addition of 0.8 ml of Opti-MEM I reduced serum medium. The mixture was then overlaid onto the cells. The cells were incubated for 8 h before the DNA-containing medium was replaced with the 2 ml of growth medium containing serum. The cells were harvested for luciferase and β-galactosidase assays at 48 h after replacement of the medium. Cell lysis and luciferase assays were performed using the Luciferase Assay System (Promega). Light emission was measured by the excitation of 560 nm and 540 nm, allowing for conversion of cytosine, but not 5-methylcytosine, to uracil. Methylation of cytosine residue in CpG dinucleotide appears symmetric on either strand of DNA, the upper strand of the bisulfite-modified genomic fragments or PCR-amplified fragments into the expression vector, the pGL3-basic vector (Promega, Madison, WI), of which the *SacI/SacI* site was located downstream from the enhancer/ promoter sequences from the immediate early gene of human cytomegalovirus and upstream from the *A-transferase* cDNA sequence. The *SacII* site was located in the last coding exon of the human *A-transferase* cDNA. These *HindIII/SacII* sites were used to facilitate the subcloning of PCR-amplified fragments for construction of *A-transferase* expression vectors. Using fragment L obtained from exon 1α-specific PCR-amplified fragments for construction of *A-transferase* expression vectors. Using fragment L obtained from exon 1α-specific

**Bisulfite Modification and Genomic Sequencing**—Bisulfite reactions were performed as described by Clark et al. (38) under conditions that allowed for conversion of cytosine, but not 5-methylcytosine, to uracil. In brief, genomic DNA was digested with *EcoRI* followed by phenol/ chloroform extraction and ethanol precipitation. DNA modification and purification were carried out, as described previously (24). Because methylation of cytosine residue in CpG dinucleotide appears symmetric on either strand of DNA, the upper strand of the bisulfite-modified ABO upstream region was amplified with ABO gene-specific primers for the modified sequence, as shown in Table I. The conditions used for PCR I were 95 °C for 9 min, 40 cycles of 94 °C for 2 min, 60 °C for 2 min, 72 °C for 3 min, and finally 10 min at 72 °C. The conditions for other PCRs were the same except for PCR I except for annealing temperatures were 60 °C, 60 °C, 61 °C, 59 °C, and 54 °C in PCR II, I+II, III, IV, and V, respectively. PCR amplification was performed in a 100 µl reaction mixture containing 100 pmol of each primer, 2.5 units of AmpliTaq Gold, 1.5 mM MgCl₂, 150 µM dNTP, and 1× buffer. Amplified DNA was electrophoresed, gel purified, and ligated with pCR2.1. Sequencing was then performed with double strand plasmid templates.
Alternative Promoter for ABO Gene

37939

TABLE II

The number of individual clones sequenced from each cell line and ex vivo culture of AC133-CD34+ cells is indicated. The number in parentheses is the number of clones with distinct methylation patterns in each cell line and ex vivo culture of AC133-CD34+ cells. The data from the clones with asterisks were used in calculation of the percentage of the methylated cytosine residue at each CpG dinucleotide in each PCR target.

| Cells                        | PCR target | V       | IV     | III    | II     | I      |
|------------------------------|------------|---------|--------|--------|--------|--------|
| KATO III                     |            | 14 (11*)| 8 (8*) | 8* (2) | 10* (1)| 10* (1)|
| MKN45                        |            | 12 (11*)| 8 (8*) | 8 (8*) | 10* (5)| 8* (1) |
| MKN1                         |            | 10 (9*) | 8 (8*) | 8 (8*) | 8* (8*)| 8* (1) |
| MKN28                        |            | 10 (10*)| 8 (8*) | 8 (8*) | 8* (8*)| 15 (15*)|
| KG562                        |            | 8 (8*)  | 8 (8*) | 8 (8*) | 8* (2) | 8* (1) |
| HEL                          |            | 14 (9*) | 8 (8*) | 8 (8*) | 8 (8*) | 20 (11*)|
| AC133-CD34+ cells, day 7     |            | 15 (8*) | 8 (8*) | 9* (2) | 8* (1) | 17* (5) |
| AC133-CD34+ cells, day 14    |            | 16 (8*) | 14 (10*)| 9* (2) | 8* (1) | 12* (3) |

The number of clones sequenced at individual PCR targets in each cell culture are indicated in Table II.

RESULTS

Expression of the ABO Genes during Maturation of Peripheral Blood-derived AC133-CD34+ Cells in ex Vivo Culture—We investigated the expression of the ABO genes during erythroid differentiation using AC133-CD34+ cells isolated from peripheral blood. A selective two phase liquid culture system, based on a well defined culture medium supplemented with recombinant growth factors, was utilized for the maturation of erythroid progenitors. The AC133-CD34+ cells are rich in erythroid-committed progenitors, and more than 90% of the colonies produced from those cells are pure erythroid colonies (35). The AC133-CD34+ cells were cultured primarily in serum-free medium supplemented with thrombopoietin, Flt3 ligand, and stem cell factor for 7 days followed by a secondary culture with the addition of erythropoietin for the next 7 days. The FACs profiles of AC133 and CD34 antigen expression revealed that more than 95% of the cells that we isolated from peripheral blood mononuclear cells were AC133-CD34+ (data not shown). The AC133-CD34+ cells proliferated and differentiated into a large number of blasts, and enrichment of BFU-E occurred during the first phase of the culture. After stimulation with erythropoietin, erythroid progenitors proliferated and matured into orthochromatid erythroblasts. The time course profile of erythropoietin receptor during the culture showed a progressive increase of the receptor on erythroid progenitor cells.2

We initially examined by RT-PCR the expression pattern of genes encoding GATA-1 and β-globin in AC133-CD34+ cell differentiation. The important regulatory elements of the β-globin gene were demonstrated to have GATA-1 sites. Neither transcript was detectable in the cells of the flow-through fraction from CD34 immunomagnetic separation nor in the freshly purified AC133-CD34+ cells, although both transcripts became apparent at days 7 and 14 of culture (Fig. 1). The validity of these RT-PCR analyses was confirmed by the presence of both transcripts in the bone marrow and in the erythroleukemia cell line K562 and the absence of the transcripts in the T cell line Jurkat. These results indicated that the AC133-CD34+ peripheral blood mononuclear cells differentiated into erythroid cells during the ex vivo culture. To monitor the expression of the ABO genes during differentiation of the erythroid progenitors, RT-PCR analysis was carried out with ABO transcripts. The ABO gene transcript was barely detectable in the freshly purified AC133-CD34+ cells and the cells in flow-through fraction. However, the ABO transcripts were apparent at day 7 (Fig. 1) but obscure at day 14. The control G3PDH transcripts were detectable before and after the culture. Nucleotide sequence determination using the dRhodamine Terminator Cycle Sequencing Ready Reaction Kit demonstrated that the PCR-amplified products of different sizes in the cells at day 7 consisted of a major full-length transcript and a minor transcript with removal of exon 6, similar to the RT-PCR products observed in human bone marrow. RT-PCR analysis for the ABO gene expression was performed using primers ABO+116 and ABO+802 corresponding to the sequences within exons 3 and 7 of the ABO gene, respectively. PCR products were electrophoresed through a 2% agarose gel followed by ethidium bromide staining. The PCR products of different sizes (A–F) observed in K562 cells and bone marrow are the result of alternative splicing as reported previously (24). Sizes of the PCR products are 334 bp for GATA-1, 244 bp for β-globin, and 452 bp for G3PDH. A 1Kb PLUS DNA ladder was used as a molecular size marker.

Fig. 1. Expression of the ABO genes during ex vivo culture of the AC133-CD34+ cells prepared from peripheral blood mononuclear cells. Using RT-PCR, the expression of four genes (ABO, GATA-1, β-globin, and G3PDH) was determined in the flow-through cells from CD34 immunomagnetic separation (FT). The AC133-CD34+ cells were isolated from peripheral blood, ex vivo culture of the AC133-CD34+ cells at days 7 and 14, erythroleukemia cell line K562 cells, T lymphocyte Jurkat cells, and human bone marrow. RT-PCR analysis for the ABO gene expression was performed using primers ABO+116 and ABO+802 corresponding to the sequences within exons 3 and 7 of the ABO gene, respectively. PCR products were electrophoresed through a 2% agarose gel followed by ethidium bromide staining. The PCR products of different sizes (A–F) observed in K562 cells and bone marrow are the result of alternative splicing as reported previously (24). Sizes of the PCR products are 334 bp for GATA-1, 244 bp for β-globin, and 452 bp for G3PDH. A 1Kb PLUS DNA ladder was used as a molecular size marker.

2 K. Matsumoto and K. Yasui, unpublished data.
genes by the 5′-RACE technique using human pancreatic cDNA as a template (8). To examine the transcription start site of the ABO genes in human erythroid cells, 5′-RACE was performed using cDNA synthesized from RNA of the AC133 CD34+ cells cultured at day 7. Agarose gel electrophoresis of the 5′-RACE products showed a major band migrating slower and several faint bands migrating faster. Considering the alternative splicing of the ABO gene transcripts, RNA sequences were purified from the major band and cloned into a sequencing vector. DNA sequences were determined for 20 transformant clones. Except for one clone, the 5′-ends of the 5′-RACE product were located around the transcription initiation sites, as determined previously using human pancreas cDNA. That single clone contained a 371-bp 5′-RACE DNA product that appeared to be a hybrid between exon 2 and the upstream genomic DNA of the ABO gene. That product was 100% identical to exons 2–6 from the 5′-end of the reverse primer fy-123 to the 5′-end of the invariant region of exon 2. Beyond that point, however, the sequence showed 100% identity with the upstream genomic DNA starting at position −656 and running to position −682 relative to the transcription start site in exon 1 (the underlined sequence in Fig. 2). More importantly, the product lacked exon 1. This comparison with the upstream genomic sequence of the ABO gene suggested the presence of an alternative exon, which we named exon 1a. The donor splice site between exon 1a and the subsequent intron had GC, whereas the acceptor site between the subsequent intron and exon 2 had AG. The noncanonical GC-AG splicing site pair was reported at a ratio of 0.56% in mammalian splice site sequences (40). Therefore, the 5′-splice site seems to be compatible with splicing junction.

Inspection of the human ABO gene indicated that repetitive elements, including two Alu repeats and long interspersed nuclear elements, are located between 0.87 and 2.18 kb upstream from the transcription start site in exon 1 in addition to a typical CpG island extending from 0.7 kb upstream to 0.6 kb downstream (see Fig. 5). The newly identified transcription initiation site seems to be located at the 5′-end of the CpG island of the ABO gene.

To confirm the splicing junction between exon 2 and the upstream DNA and to examine whether exon 1a is also used as starting exon in gastric cancer cell lines MKN45 and KATOIII and the erythroid progenitors, RT-PCR was carried out using primers specific for each of two distinct starting exons and a common reverse primer complementary to the sequence in exon 2. The relative abundance of the transcripts starting from exon 1a in the cells examined. Therefore, it is likely that only a small portion of the ABO transcripts starts from alternative starting exon 1a.

**Analysis of Promoter Activity in the 5′-Flanking Region of Exon 1a in the ABO Gene**—Although alternative exon splicing of the ABO gene transcripts has been demonstrated (8, 9, 24),
alternative promoter usage has never been reported with the ABO genes. Distinct promoters and alternative 5'-ends have been reported with some other genes encoding glycosyltransferases, such as α1,2-fucosyltransferase and α1,3-fucosyltransferase (41, 42). Therefore, it is important to characterize the promoter that regulates the transcription of the ABO messages containing exon 1a.

Inspection of the sequence around the distal transcription initiation site revealed putative binding sites for several transcription factors as shown in Fig. 2. As a means to examine the promoter activity in the 5'-flanking region of exon 1a in the ABO gene, we first obtained the −832Xh construct by introducing the −832 to −667 upstream region fragment of the ABO gene into the XhoI site of the pCMV-neo vector. The promoter activities of the −832Xh construct were analyzed in the MKN45, KATOIII, and AC133/CD34+ cells at day 7 by using reporter gene constructs containing either the −832 or −667 fragments upstream of the reporter gene in the pGL2-Basic vector. The results are shown in Fig. 3A and Table III.

**Figure 3.** RT-PCR analysis for detection of the ABO gene starting exon in human gastric cancer cell lines MKN45 and KATOIII and ex vivo culture of the AC133/CD34+ cells at day 7. A, RT-PCR analysis. Total RNA prepared from MKN45 cells, KATOIII cells, and ex vivo culture of the AC133/CD34+ cells at day 7 was reverse transcribed with random primer, and the resulting single-strand cDNA was used as a template for PCR analysis. The ABO gene amplification was performed using either distinct starting exon-specific primer ABO+82 or ABOU-678 and a common reverse primer ABO+802 complementary to exon 7 of the ABO gene. PCR products were electrophoresed through a 2% agarose gel and were stained with ethidium bromide. The amplified fragments were named G to M. A 1Kb PLUS DNA ladder was used as a molecular size marker. B, splicing patterns of the amplified fragments G–M. Nucleotide sequences of these fragments were determined and then compared. Schematically represented ABO genes were aligned with the RT-PCR products amplified, using a set of each starting exon-specific primer (ABOU-678 or ABOU+82) and the ABO+802 primer, which are represented by arrows. The open boxes represent the ABO gene exons, and the thick straight lines represent the intron sequence. Dashed V-shaped lines in RT-PCR-amplified fragments G–M indicate regions that are removed by splicing. The number at the right of each RT-PCR product represents the length of the product. The figure also shows the locations of the primers fy-123 and ABOU+433 used in 5'-RACE, the location of the primer ABOU+116 used in RT-PCR of Fig. 1, and the location of the reverse primer ABO+98 used in another starting exon-specific PCR and quantitative real time RT-PCR. Alternative splicing transcripts I–K shifts the reading frame of codons, whereas stop codons occur in transcripts H and M at the splicing junctions.

**Table III**

Comparison of the levels of the ABO splice variants containing exon 1 with those containing exon 1a in cultured cells

| Cells          | Exon 1-containing transcripts | Exon 1a-containing transcripts | Ratio  |
|----------------|--------------------------------|--------------------------------|--------|
| MKN45          | 5,376 (961)                    | 10 (2)                         | 0.002  |
| KATO III       | 3,494 (453)                    | 217 (55)                       | 0.062  |
| AC133 CD34+, day 7 | 1,525 (76)                   | 29 (4)                         | 0.019  |
gene into the promoterless pGL3-basic vector upstream from the luciferase coding sequence. The −832 upstream terminus was chosen because the vast majority of CpG dinucleotide are methylated and reside within repetitive elements (43), and the locations of the recognition sites for the putative transcription factors were taken into account. The reporter plasmid was transfected into the KATOIII cells. 48 h after transfection, the cells were harvested, and the luciferase activities in cell extracts were analyzed (Fig. 4). The pGL3-promoter vector containing the SV40 promoter and pGL3-basic vector without the promoter sequence were used as positive and negative controls, respectively. The relative luciferase activity of the −832Xh construct was at least 8-fold higher than that of pGL3-basic vector and was 6-fold lower than that of pGL3-promoter vector. This indicated the promoter activity of the 5′-flanking region of exon 1a in the ABO gene. The luciferase activity of the construct −832Xh was similar to that of the construct SN containing the −117/+31 proximal promoter sequence. Deletion of the upstream end from position −832 to −781 resulted in a loss of one-third of the activity, suggesting that the important elements for the distal promoter function were contained within the deleted region. Moreover, the presence of an additional sequence between −666 and −336 in construct −832Sma yielded an elevated activity, which implicated the presence of positive regulatory element(s) in the sequence. These results suggest that the region just upstream from the proximal transcription start site acts as a promoter. Thus, it appears that the alternative promoter is present at the 5′-end of the CpG island in the ABO gene and that the exon 1a is transcribed from this promoter.

To compare the proximal and distal promoters, we prepared the −832N construct by introducing the −832 to +31 sequence incorporating both the promoter regions of the ABO gene into the promoterless pGL3-basic vector. The luciferase activity of the plasmid −832N was less than 2-fold of the activity for the control pGL3-basic vector, suggesting negative element(s) located between those two promoters. Deletion of the upstream end from position −832 to −670, −548, −409, or −275 did not result in any significant changes. Thus, the proximal promoter does not seem to be affected by the distal promoter activity in transient transfection experiments. Deletion of the sequence from −275 to either −202 or −117 elicited an increase in the luciferase activity. In addition, an internal deletion of the sequence between −275 and −118 in construct XhNΔ−275/−118 resulted in a 7-fold or 2-fold increase in luciferase activity compared with the construct XhN or SN, respectively. Moreover, another internal deletion of the −335 to −118 sequence in construct XhNΔ−335/−118 resulted in similar increases. These results suggested that negative element(s) for the ABO gene transcription were present in the region between −275 and −118 and that positive regulatory element(s) were present in the sequence between −670 and −336.

**Methylation Profile of the Upstream Region in the ABO Gene from the Repetitive Elements through the Distal Promoter to the Proximal Promoter**—Because the vast majority of CpG dinucleotide are methylated and reside within repetitive elements (43), we examined the methylation status of the upstream region in the ABO gene to define a boundary between methylated and unmethylated domains. The methylation status in the region between −140 and +26 was analyzed by bisulfite genomic sequencing. This region spanned 128 or 127 CpG sites and was analyzed by PCR amplification of five regions (PCR I–V) followed by subcloning of the PCR-amplified fragments into a sequencing vector. PCR fragments IV and V contained polymorphisms, which included variation in the number of \((\text{AAAAAT})_{n-1}\) in the downstream Alu sequence and the presence or absence of 36 bp in a long interspersed nuclear element (44), respectively. Data obtained from genomic sequencing of
bisulfite-treated DNA were compiled in human gastric cancer cell lines KATOIII, MKN45, MKN1, MKN28, human erythroleukemia cell lines K562 and HEL, and the culture of AC133−CD34+ cells at days 7 and 14. The percentage of the methylated cytosine residue in each CpG dinucleotide was calculated for the individual sites in the −1469 to +26 region using more than eight clones. The number of the clones used for calculation of the percentage of the methylated cytosine in each CpG dinucleotide was shown in Table II. Fig. 5 shows the methylation profiles in the −1469 to +26 region in each cell line and the ex vivo culture. Analysis of methylation in KATOIII cells, K562 cells, and the culture of AC133−CD34+ cells at days 7 and 14 demonstrated hypermethylation in the region of repetitive elements and hypomethylation in the region downstream position −832 except for methylation at low ratios in a few CpG sites. The distal promoter region appeared to be located 3′ adjacent to the methylated region of repetitive elements.

In contrast, the methylation profile in MKN28 cells demonstrated that hypermethylation extended from the region of repetitive elements through the distal promoter region to the proximal promoter region. Because methylation of CpG islands spanning the promoter regions is strongly associated with transcriptional silencing, starting exon-specific RT-PCR was performed using each distinct starting exon-specific primer and a common reverse primer complementary to exon 2 to examine transcripts from both exons 1a and 1 (Fig. 6). These exon-specific RT-PCR analyses showed that neither exon 1a- nor exon 1-containing transcript was detectable in MKN28 cells, although the control G3PDH transcript was detectable in MKN28 cells similar to those expressing cells in which the ABO transcripts from both exons were detected.

The lack of any transcript in MKN28 cells by the starting exon-specific amplifications agreed with the previous result that the ABO transcript was not found by RT-PCR using the primers corresponding to the sequences in exons 3 and 7 (24). Thus, hypermethylation of both promoter regions in MKN28 cells might be responsible for the absence of any transcripts from either promoter. The methylation profile in MKN1 cells showed an intermediate pattern between KATOIII cells and MKN28 cells. Hypermethylation extended from the region of repetitive elements through the distal promoter region to around 0.4 kb upstream relative to the transcription start site of the ABO gene exon 1. The distal promoter was hypermethylated, whereas the proximal promoter was hypomethylated. These methylation profiles appeared to correspond to the results of expression analysis that the ABO transcript was barely detectable by exon 1a-specific RT-PCR, whereas the transcript was detected by exon 1-specific RT-PCR. As was observed in MKN28 cells, the methylation profile of HEL cells showed hypermethylation extending from the region of repetitive elements to the proximal promoter region, whereas methylation ratios decreased in the proximal promoter region. Our previous study indicated cellular heterogeneity in DNA methylation of the −117 to +26 sequence in HEL cells (24). To define a cellular heterogeneity in methylation of the −434 to +47 sequence, additional PCR amplification was performed with a distinctive combination of primers in PCR I−II. Fig. 7 shows the methylation profile obtained from individual clones generated from the PCR I−II fragment. The methylation patterns were heterogeneous. Hypomethylation was found in the −117 to +26 sequence in some clones, whereas hypermethylation was found in other clones. These methylation profiles seemed to be consistent with the results that the transcript starting from exon 1a was not detectable by exon 1a-specific RT-PCR, although the transcript starting from exon 1 was detected by exon 1-specific RT-PCR in HEL cells. In MKN45 cells, hypermethylation was demonstrated in the region of repetitive elements, whereas the proximal promoter region was unmethylated. However, clusters of methylated CpG sites were found at low or moderate ratios in an interval of around 0.2 kb in the region between the hypermethylated region of repetitive elements and the hypomethylated proximal promoter region. With regard to the methylation status of the region within a

![Fig. 5. Analysis of in vivo methylation of CpG dinucleotides in the upstream region of the ABO gene using the bisulfite conversion technique. The top diagram indicates the five sequence regions that were amplified. The primers used are represented as open boxes at both ends of individual thick lines. The second diagram from the top indicates two transcription start sites and their upstream region. The sequence located between −1500 and +26 relative to the transcription start site of the ABO gene exon 1 was compiled from GenBank accession number U22302. The schematic shows the locations of proximal and distal transcription start sites as small closed arrows, whereas the location of the CpG island is indicated within broken lines. The open arrow represents the position and orientation of the downstream Alu sequence, and the open squares indicate long interspersed nuclear elements, which were found by computer analysis using RepeatMasker. The second exon is not indicated in the figure because it is located 13 kb downstream from exon 1. The third diagram from the top represents the distribution of CpG dinucleotides, and the vertical lines indicate the position of each CpG dinucleotide in the DNA sequence from −1469 to +26 relative to the transcription start site of the ABO gene exon 1. Dense clustering of CpG sites is shown in the CpG island in which CpG density is 12.9%. Below the CpG dinucleotide plot, eight panels show the percentage of DNA methylation at individual CpG sites in the −1469 to +26 sequence of the ABO genes. The ratios of DNA methylation at each CpG site are represented as the length of the vertical lines at the relative positions of the CpG dinucleotide in human gastric cancer cell lines KATOIII, MKN45, MKN1, MKN28, human erythroleukemia cell lines K562 and HEL, and the ex vivo culture of AC133−CD34+ cells at days 7 and 14. Inspection of the −1469 to +26 sequence reveals 128 or 127 CpG dinucleotides because a polymorphism was reported with regard to the presence or absence of the sequence between −973 and −943, which contains one CpG site. The V-shaped lines in MKN45 cells, MKN28 cells, and the ex vivo culture of AC133−CD34+ cells indicate the regions that were deleted in the ABO gene upstream region. +1 represents the transcription start site in exon 1, which was determined previously.](Image 317x412 to 544x728)
radius of 100 bp around the distal transcription start site, one PCR product showed hypomethylation, and the others demonstrated methylation at a few CpG sites. Although the starting exon-specific RT-PCR analyses demonstrated the presence of transcripts from both exon 1a and exon 1, the relative ratio of the exon 1a-containing transcripts to the exon 1-containing transcripts was smaller than those observed with KATOIII cells and the ex vivo culture of erythroid cells (Table III).

Because transcriptional repression by DNA methylation was reported to appear when the density of methyl-CpGs approach to approximately 1 in 100 bp (45, 46), the ABO gene distal promoter could be repressed only moderately by DNA methylation in MKN45 cells.

Both promoter regions were hypomethylated in the culture of AC133+/CD34+ cells at day 14. However, the transcript starting from exon 1a was barely detectable, whereas the transcript starting from exon 1 was detectable (Fig. 6). These results may suggest that negative regulatory mechanism(s) other than DNA methylation might play a role in the down-regulation of transcription from the distal promoter during differentiation of erythroid progenitors in the ex vivo culture.

Effects of DNA Demethylation on Expression of the ABO Gene

Exon 1a-containing Transcripts—To address the question of whether methylation of the distal promoter region is itself inhibiting expression from the distal promoter, we treated MKN1 cells, MKN28 cells, and HEL cells with 5-aza-2'-deoxycytidine, an inhibitor of DNA methyltransferase which causes the demethylation of DNA. We then monitored the expression of the ABO transcripts by the starting exon-specific RT-PCR. Results are shown in Fig. 8. DNA fragments of the expected size were amplified by the exon 1a-specific RT-PCR of RNA from MKN1 cells and HEL cells treated with 5 μM concentrations of 5-aza-2'-deoxycytidine for 3 days. These bands were confirmed to be derived from the ABO gene message by nucleotide sequencing.

Demethylation analysis of the distal promoter in those cells after the 5-aza-2'-deoxycytidine treatment was carried out by PCR of the bisulfite-treated DNA using the primers for the PCR III target corresponding to the sequence from −413 to +26 relative to the transcription start site of exon 1. Below the CpG dinucleotide plot, the methylation profiles of 15 individual clones with distinct patterns are shown. The presence of the methylated cytosine residue in each CpG dinucleotide is represented as a vertical line at the individual position of the CpG dinucleotide. The percentage of the methylated cytosine residue in each CpG dinucleotide was calculated for the individual sites in the −413 to +26 region using the results from the clones shown in this figure. At the bottom, the ratios of DNA methylation at each CpG site are represented as the lengths of vertical lines at the relative positions of the CpG dinucleotide.

FIG. 6. RT-PCR analysis for detection of the ABO gene starting exons in various tumor cells. ABO amplification was performed using distinct starting exon-specific primers ABOU-678 and ABO+3 in combination with a common reverse primer ABO+98 complementary to the sequence in exon 2 of the ABO gene. PCR products were electrophoresed through a 2% agarose gel and were stained with ethidium bromide. Sizes of the PCR products are 96 bp for starting exon 1-specific PCR and 93 bp for starting exon 1a-specific PCR. A 1Kb PLUS DNA ladder was used as a molecular size marker.

FIG. 7. Methylation analysis of individual clones from HEL cells. The top diagram indicates the region of the ABO gene which was amplified by PCR I+II and which corresponds to positions −434 to +47. Sodium bisulfite-modified genomic DNA from HEL cells was used as the PCR template. The sequenced region and the locations of the primers are represented as closed and open boxes of the thick line, respectively. The second diagram from the top indicates the restriction enzyme sites and the transcription start site in the ABO gene exon 1. The next portion represents the distribution of CpG dinucleotides from position −413 to +26 relative to the transcription start site of exon 1. Below the CpG dinucleotide plot, the methylation profiles of 15 individual clones with distinct patterns are shown. The presence of the methylated cytosine residue in each CpG dinucleotide is represented as a vertical line at the individual position of the CpG dinucleotide. The percentage of the methylated cytosine residue in each CpG dinucleotide was calculated for the individual sites in the −413 to +26 region using the results from the clones shown in this figure. At the bottom, the ratios of DNA methylation at each CpG site are represented as the lengths of vertical lines at the relative positions of the CpG dinucleotide.
from unmethylated DNA. Only the methylated sequences were cleaved by the enzyme, yielding bands at 255 and 121 bp. By comparing the bands derived from the methylated (255 bp) and unmethylated (376 bp) alleles before and after the treatment, the proportions of the demethylated sequences seemed to increase in both cell lines after the DNA methyltransferase inhibitor treatment. Because the demethylation of the promoter could reactivate the distal promoter, it strongly supports the idea that DNA methylation is responsible for the absence of the transcripts from the distal promoter in MKN28 cells.

No amplification of the exon 1a-specific transcripts was observed by the RT-PCR of RNA from MKN28 cells treated with 5-aza-2'-deoxycytidine despite the repeated attempts, although the demethylation of the distal promoter region was demonstrated by PCR of bisulfite-treated DNA, followed by TaqI digestion. In contrast, the exon 1-specific RT-PCR confirmed the appearance of transcripts containing exon 1. These results suggest that negative regulatory mechanisms other than DNA methylation might play a role in the down-regulation of transcription from the distal promoter in MKN28 cells.

Cell Type-specific Activity of Transfected ABO Gene Distal Promoter—By performing transient transfection experiments using KATOIII cells, the sequence located between −832 and −667 was found to be responsible for transcriptional activity. To examine whether the ABO gene distal promoter exhibited cell type-specific activity correlating with endogenous expression of the ABO genes, transient transfection studies were performed using MKN1 cells, MKN28 cells, HEL cells, and fibroblasts. In fibroblasts, the transcripts starting from the ABO gene exon 1 and exon 1a were barely detectable, suggesting that the distal promoter functioned more efficiently in ABO gene-expressing cells than fibroblasts. It is expected that the distal promoter might be more active in ABO gene-expressing cells than fibroblasts.

| Cell         | Reporter construct | Ratio of promoter activity between SN and −832Xh | relative luciferase activity |
|--------------|--------------------|-----------------------------------------------|-----------------------------|
| KATOIIIa     | pGL3-basic vector  | 0.018 ± 0.003 | 0.014 ± 0.021 | 0.015 ± 0.025 | 0.15 |
| MKN1b        | SN                 | 0.016 ± 0.002 | 0.134 ± 0.014 | 0.139 ± 0.015 | 1.03 |
| MKN28c       | −832Xh             | 0.028 ± 0.009 | 0.269 ± 0.048 | 0.113 ± 0.037 | 0.42 |
| HELd         | 0.041 ± 0.026      | 0.316 ± 0.045 | 0.297 ± 0.078 | 0.93 |
| Fibroblasta  | 0.011 ± 0.003      | 0.234 ± 0.019 | 0.037 ± 0.007 | 0.15 |

a The ABO proximal promoter used sequences located between −117 and +31.
b The ABO distal promoter used sequences located between −832 and −667.
c The results are expressed as an average relative activity compared with that observed for the pGL3-promoter vector containing the SV40 promoter.
d KATOIII cells were transfected with electroporation; 10 µg of luciferase reporter and 4 µg of β-galactosidase control vector were used for each analysis.
e MKN1 cells, MKN28 cells, and fibroblasts were transfected with Lipofectin; 1.4 µg of luciferase reporter and 0.6 µg of β-galactosidase control vector were used for each analysis.
f HEL cells were transfected with DMRIE-C reagent; 4 µg of luciferase reporter and 2 µg of β-galactosidase control vector were used for each analysis.

The expression of the ABO genes was regulated by the distal promoter. The results showed that the distal promoter was more active in ABO gene-expressing cells than fibroblasts. It is suggested that DNA methylation is responsible for the absence of the transcripts from the distal promoter in MKN28 cells.

**Table IV**

| Cell      | Reporter construct | Ratio of promoter activity between SN and −832Xh | relative luciferase activity |
|-----------|--------------------|-----------------------------------------------|-----------------------------|
| KATOIII   | pGL3-basic vector  | 0.018 ± 0.003 | 0.014 ± 0.021 | 0.015 ± 0.025 | 0.15 |
| MKN1      | SN                 | 0.016 ± 0.002 | 0.134 ± 0.014 | 0.139 ± 0.015 | 1.03 |
| MKN28     | −832Xh             | 0.028 ± 0.009 | 0.269 ± 0.048 | 0.113 ± 0.037 | 0.42 |
| HEL       | 0.041 ± 0.026      | 0.316 ± 0.045 | 0.297 ± 0.078 | 0.93 |
| Fibroblast| 0.011 ± 0.003      | 0.234 ± 0.019 | 0.037 ± 0.007 | 0.15 |

a The ABO proximal promoter used sequences located between −117 and +31.
b The ABO distal promoter used sequences located between −832 and −667.
c The results are expressed as an average relative activity compared with that observed for the pGL3-promoter vector containing the SV40 promoter.
d KATOIII cells were transfected with electroporation; 10 µg of luciferase reporter and 4 µg of β-galactosidase control vector were used for each analysis.
e MKN1 cells, MKN28 cells, and fibroblasts were transfected with Lipofectin; 1.4 µg of luciferase reporter and 0.6 µg of β-galactosidase control vector were used for each analysis.
f HEL cells were transfected with DMRIE-C reagent; 4 µg of luciferase reporter and 2 µg of β-galactosidase control vector were used for each analysis.

The expression of the ABO genes was regulated by the distal promoter. The results showed that the distal promoter was more active in ABO gene-expressing cells than fibroblasts. It is suggested that DNA methylation is responsible for the absence of the transcripts from the distal promoter in MKN28 cells.
likely that the distal promoter activity is dependent upon cell types and that the cell type-specific promoter is located 3' adjacent to the hypermethylated region of repetitive elements that can be recognized by MBDS involved in repressive complexes in association with histone deacetylases. In MKN28 cells, the promoter construct −832Xh revealed half the activity of the construct SN. Comparison of the ratio of promoter activities between constructs SN and −832Xh in MKN28 cells with those ratios in those expressing cells suggested the reduction of the distal promoter activity in MKN28 cells. Thus, the decreased promoter activity may have resulted in the inhibition of transcription from the distal promoter in MKN28 cells. The reduction of the promoter activity was incomplete in the transient transfection experiments, and the exon 1a-containing transcript was barely detectable by the starting exon-specific RT-PCR from MKN28 cells treated with 5-aza-2'-deoxycytidine although demethylation of the distal promoter region was found. Therefore, other generalized mechanisms such as methylation of histone tails may have also participated in the inhibition of transcription from the distal promoter. Moreover, a negative control other than DNA methylation has been suggested to play a role in the down-regulation of transcription from the distal promoter during differentiation of the erythroid progenitors. Although hypermethylation was found in the upstream region, further investigation is required to elucidate other negative regulatory mechanisms in MKN28 cells.

**DISCUSSION**

We have studied the transcriptional regulatory mechanism of the human ABO genes. Our previous characterization of the 5'-upstream sequence of the ABO gene demonstrated that the region between −117 and +31 had promoter activity in both epithelial and erythroid lineages (22, 23). Furthermore, expression of the ABO genes was found to be dependent upon DNA methylation of the proximal constitutive promoter (24). In this paper, we have identified a novel transcription start site at the 5'-end of the CpG island containing the proximal promoter in the ABO gene. This start site appears to mark an alternative starting exon (1a), which is utilized as a transcription starting exon in both erythroid and epithelial lineage cells examined. The region just upstream from the transcription start site is sufficient for the expression of a reporter gene in a cell type-specific manner when placed 5' adjacent to the luciferase gene. Hypermethylation is commonly observed in the region of repetitive elements 5' adjacent to the distal promoter region in all of the cell lines examined. These results indicate that a cell type-specific promoter is located between the hypermethylated region of repetitive elements and the CpG island containing the constitutive promoter in the ABO genes.

The utilization of multiple promoters and transcription start sites is a frequently used mechanism to create diversity and flexibility in the regulation of gene expression (47). The level of transcription initiation can vary between alternative promoters: the turnover or translation efficiency of mRNA isoforms with different leader exons can differ, alternative promoters can have different tissue specificity and can react differently to growth signals, and alternative promoter usage can lead to the generation of protein isoforms differing in amino acid sequence.

The use of an alternative exon 1a seems to be common in the ABO genes because the ABO blood group genotypes of the Ac133−CD34+ cells used in this study were inferred to be AA based on A allele-specific nucleotide substitutions (6), whereas those of KATOIII cells, MKN45 cells, and K562 cells were BO, AA, and OO, respectively (24). Although the nucleotide sequence of exon 1a does not contain an ATG codon, transfection experiments of MKN28 cells with the expression plasmid A(1a) containing the entire cDNA starting from exon 1a demonstrated the appearance of large amounts of A antigens, suggesting that the functional protein involved in expression of A antigens could be synthesized by the usage of an internal preferential translation start site in the transmembrane domain. However, quantitative real time RT-PCR showed that most of the ABO gene transcripts start from exon 1. The low level of the transcripts starting from exon 1a may be caused by different turnover efficiencies of mRNA isoforms with different leader exons or by reduction of the distal promoter activity conferred by corepressor complexes recruited by MBDS that can bind to the hypermethylated interspersed repeats in chromatin. On the other hand, the short open reading frames upstream from the coding sequence may serve to control expression at the translational level, as reported with other genes (49). In the murine acetylcholinesterase gene, an alternative promoter was identified 5'-upstream from the proximal promoter (50). Our analyses of the G+C density and CpG frequency in the nucleotide sequence of the upstream region of the murine acetylcholinesterase gene (GenBank accession no. AF148849) demonstrated that the alternative exon is located 5' adjacent to a CpG island. Based on the histone code hypothesis that distinct histone amino-terminal modifications can generate synergistic or antagonistic interaction affinities for chromatin-associated proteins, which in turn dictate dynamic transitions between transcriptionally active and silent chromatin states (51), it is possible that some activities of the cell type-specific promoter located at the 5'-end of the ABO gene island prevent histone tails from modifications such as deacetylation and methylation leading to silenced chromatin domains. Further studies are needed to define the relative roles of the distal cell type-specific promoter and the proximal constitutive promoter in the regulation of ABO gene expression. Moreover, elucidation of an association among the hypermethylated region of repetitive elements, the distal promoter, the negative element from position −275 to −118, and the proximal promoter may lead to more precise resolution of a molecular basis for the ABO gene expression in a cell type-specific fashion and of the changes during cell differentiation.

CpG islands are almost always maintained in unmethylated states, unlike the CpG sites in the remainder of the genome. However, methylation of CpG islands can occur on an inactive X chromosome, in promoters of imprinted genes, with oncogenesis, and during aging. In all of these cases, methylation of CpG islands spanning the promoter regions is strongly associated with transcriptional silencing. DNA methylation of the distal and proximal promoters in the ABO gene is correlated with the absence of transcripts from the distal and proximal promoter, respectively, in KMN1 cells, MKN28 cells, and HEL cells. However, the possibility still exists that additional factor(s) could function in the down-regulation of ABO gene expression. As of now, there is no clear understanding of what leads to the aberrant methylation of CpG islands commonly seen in cancer. Moreover, it is unclear whether hypermethylation is initiated from the ends of CpG island or at “hypermethylation center(s)” within the CpG island in tumor cells. It has been suggested that maintenance of the unmethylated CpG islands is dependent on continued active transcription and/or the binding of specific proteins that may protect them from being methylated (52). Studies of the adenine phosphoribosyltransferase gene locus have identified a potential mechanism of how CpG islands remain free of methylation in embryonic cells (53–56).

---

3 Y. Hata and Y. Kominato, unpublished data.
S1 binding sites and presumably trans-acting factor(s) that bind to those sites apparently protect the CpG island of this gene from methylation. Thus, the loss of specific proteins or interference with their binding and/or the lack of active transcription may contribute to CpG island promoter methylation in cancer. It has also been proposed that repetitive elements such as Alu repeats and long interspersed nuclear elements might serve as foci for de novo methylation and that methylation may spread from such attractors of modification (57, 58).

In particular, Graff et al. (59) showed that Alu sequences upstream from E-cadherin and von Hippel-Lindall (VHL) genes are methylated in normal tissues, whereas adjacent CpG island sequences are not. In addition, from their studies of transfected DNA they have suggested that methylation may progressively encroach from the methylated Alu sequence regions flanking CpG islands. However, others have suggested that methylation is initiated at “centers” within the islands and progressively spreads (60). Recently, Millar et al. (39) demonstrated that a marked boundary of the methylated and unmethylated domains correlated with an (ATAAA)19 repeated sequence at the 3′-end of hypermethylated Alu sequence of the upstream region in the glutathione S-transferase gene in normal tissues involving prostate tissue. It was also discovered that DNA methylation was present in the core CpG-rich promoter region but did not extend through the 5′-flanking region in two prostate cancer cell lines, whereas the extensive high level of DNA methylation observed in the core promoter region was found to spread through the upstream region to and beyond the boundary in another prostate cancer cell line. Supporting the latter hypothesis, the leukemia-promoting promyelocytes-retinoic acid receptor (PML-RAR) fusion protein has been reported to induce gene hypermethylation and silencing by recruiting DNA methyltransferases to target promoters (48). These authors suggested that the newly methylated CpGs worked as docking sites for methyl-binding proteins, which in turn interacted with both histone deacetylase and DNA methyltransferases, leading to the spreading of hypermethylation to the neighboring regions. In the present study, we found that the region consisting of repetitive elements was methylated in all of the cells examined, although a definite boundary between the methylated and unmethylated domains could not be determined in the ABO gene upstream region. A similar methylation pattern in which the methylated domain is restricted beyond the distal promoter region, was observed in KATOIII cells, K562 cells, and the erythroid cells in 

Acknowledgments—We thank Professor Kiimiyasu Shiraki for providing human embryonal lung fibroblasts, Professor Koichi Hiraga for helpful suggestions, Yoshimi Takata for providing the plasmid pG3PDH, Yoshinori Asano for providing the software to make Figs. 5 and 7, and Trung T. Luong and Ami Yamamoto for editing the manuscript.

REFERENCES

1. Landsteiner, K. (1900) Zentralbl. Bakteriol. 27, 357–363
2. Morgan, W. T. J., and Watkins, W. M. (1969) Br. Med. Bull. 25, 30–34
3. Kabat, E. A. (1977) in Carbohydrates in Solution (Ishib, H., ed.) pp. 334–361, American Chemical Society, Washington, D. C.
4. Watkins, W. M. (1980) Adv. Hum. Genet. 10, 1–136
5. Ginsburg, V. (1972) Adv. Enzymol. 36, 131–149
6. Yamamoto, F., Claussen, 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. Yamamoto, F., McNeill, P. D., and Hakomori, S. (1990) Glycobio 8, 51–58
9. Benet, P. F., Steffenes, R., Claussen, H., Wherbaugh, D. O., and van Kessel, A. G. (1995) Biochem. Biophys. Res. Commun. 206, 318–325
10. Schenkel-Brunner, H. (2000) in Human Blood Groups, 2nd Ed., pp. 54–150, Springer-Verlag, New York.
11. Millar, T., and Bannister, A. J. (1999) Cell 99, 75–81
12. Kominato, Y., Tsuchiya, T., Hata, N., Takizawa, H., and Yamamoto, F. (1997) J. Biol. Chem. 272, 25898–25908
13. Hata, Y., Kominato, Y., Yamamoto, F., and Yamamoto, H. (2002) Vox Sang. 82, 39–46
14. Kominato, Y., Hata, Y., Takizawa, H., Tsuchiya, T., Tanaka, J., and Yamamoto, F. (1999) J. Biol. Chem. 274, 37240–37250
15. Iwamoto, S., Withers, D. A., Handa, K., and Hakomori, S. (1999) Glycoconjug. J. 16, 659–666
16. Meehan, R. R., Bachman, M. K., Herman, J. G., and Baylin, S. B. (2001) Oncogene 20, 3156–3165
17. Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L., and Bird, A. P. (1989) Cell 58, 499–507
18. Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogel, I., Jepsson, P., Klein, F., and Bird, A. P. (1992) Cell 69, 905–914
19. Croce, C. M., Meehan, R. R., Nan, X., and Bird, A. P. (1997) Nat. Genet. 16, 256–259
20. Hendrich, B., and Bird, A. (1998) Mol. Cell. Biol. 18, 6538–6547
21. Jones, P. L., Veenastra, G. J., Wade, P. A., Vermaak, K. H., Kass, S. U., Landsberger, N., Strouboulis, J., and Wolfe, A. P. (1988) Nat. Genet. 187, 189–191
22. Nan, X., Ng, H. H., Johnson, C. A., Loh, M., editors., Bird, A., and Bird, A. P. (1999) Genes Dev. 13, 1924–1935
23. Peden, G. O., Ponder, A. B., and Bird, A. P. (1999) Nat. Genet. 21, 69–73
24. Kominato, Y., Tsuchiya, T., Hata, N., Takizawa, H., and Yamamoto, F. (2000) Stem Cell 18, 196–203
25. Kominato, Y., Tsuchiya, T., Hata, N., and Takizawa, H. (2000) J. Biol. Chem. 275, 24893–24899
26. Burrell, M., Selstedt, I. A., and Soley, S. V. (2001) Nuclear Acids Res. 29, 255–259
27. Koda, Y., Ojio, M., and Kimura, H. (1997) J. Biol. Chem. 272, 7501–7505
28. Cameron, D. H., Stasevskaia, D., and Weston, B. W. (1995) J. Biol. Chem. 270, 20112–20122
29. Yoder, J. D., Walsh, C. P., and Bower, H. K. (2000) Trends Genet. 16, 335–340
30. Shioda, I., Kominato, Y., Hata, N., and Takizawa, H. (1999) Legal Med. 1, 217–225
31. Nan, X., Campoy, P. J., and Bird, A. (1997) Cell 88, 471–481
32. Yesberger, J. S., and You, D. S. (1999) Mol. Cell. Biol. 187, 454–459
33. Ayoubi, T. A. Y., and van den Van, W. J. M. (1996) FASEB J. 10, 453–460
34. Croce, C. M., Raker, F. A., Corsaro, M., Fazi, F., Prelli, M., Faretta, M., Fuks, C., Foci, L., Kozakis, T., Nervi, C., Minucci, S., and Pelicci, P. G. (2002) Science 295, 1079–1082
35. Kozak, M. (1996) Mamm. Genome 7, 563–574
36. Atanasescu, E., Chiappa, S., Wieben, K., and Brimijoin, S. (1999) J. Biol. Chem. 274, 21078–21084
51. Strahl, B. D., and Allis, C. D. (2000) *Nature* **403**, 41–45
52. Cross, S. H., and Bird, A. P. (1995) *Curr. Opin. Genet. Dev.* **5**, 309–314
53. Macleod, D., Charlton, J., Mullins, J., and Bird, A. P. (1994) *Genes Dev.* **8**, 2282–2292
54. Brandeis, M., Fank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., and Cedar, H. (1995) *Nature* **371**, 435–438
55. Mummanei, P., Walker, K. A., Bishop, P. L., and Turker, M. S. (1995) *J. Biol. Chem.* **270**, 788–792
56. Mummanei, P., Yates, P., Simpson, J., Rose, J., and Turker, M. S. (1998) *Nucleic Acids Res.* **26**, 5163–5169
57. Magewu, A. N., and Jones, P. A. (1994) *Mol. Cell. Biol.* **14**, 4225–4232
58. Turker, M. S. (1999) *Semin. Cancer Biol.* **9**, 329–337
59. Graff, J. R., Herman, J. G., Myohanen, S., Baylin, S. B., and Vertino, P. M. (1997) *J. Biol. Chem.* **272**, 22322–22329
60. Toth, M., Lichtenberg, U., and Doerfler, W. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 3728–3732

**Alternative Promoter for ABO Gene**