A Methylation-responsive MDBP/RFX Site Is in the First Exon of the Collagen α2(I) Promoter*

Pritam K. Sengupta‡, Melanie Ehrlich§, and Barbara D. Smith‡§

From the ‡Department of Biochemistry, Boston University School of Medicine, Boston, Massachusetts 02118, the §Boston Department of Veterans Affairs Medical Center, Boston, Massachusetts 02118, and the ¶Human Genetics Program and Department of Biochemistry, Tulane Medical School, New Orleans, Louisiana 70112

DNA methylation inhibits transcription driven by the collagen α2(I) promoter and the 5' end of the gene in transient transfection and in vitro transcription assays. DNA-binding proteins in a unique family of ubiquitously expressed proteins, methylated DNA-binding protein (MDBP)/regulatory factor for X box (RFX), form specific complexes with a sequence overlapping the transcription start site of the collagen α2(I) gene. Complex formation increased when the CpG site at +7 base pairs from the transcription start site was methylated. The identity of the protein was demonstrated by co-migration and cross-competition for a characteristic slowly migrating doublet complex formed on MDBP/RFX recognition sequences and the collagen sequences by band shift assays. A RFX1-specific antibody supershifted the collagen DNA-protein complexes. Furthermore, in vitro translated RFX1 protein formed a specific complex with the collagen sequence that was also supershifted with the RFX1 antibody. MDBP/RFX displayed a higher affinity binding to the collagen sequence if the CpG at +7 was mutated in a manner similar to TpG. This same mutation within reporter constructs inhibited transcription in transfection and in vitro transcription assay. These results support the hypothesis that DNA methylation-induced inactivation of collagen α2(I) gene transcription is mediated, in part, by increased binding of MDBP/RFX to the first exon in response to methylation in this region.

Type I collagen, the most abundant collagen molecule within the collagen family, normally consists of a heterotrimer of two α1(I) chains and one α2(I) chain. Synthesis of these genes is often down-regulated upon oncogenic transformation of cells in culture (1–4). We have previously demonstrated down-regulation of the collagen α2(I) gene, encoding one of the subunits of Type I collagen in an epithelial-like cell line from rat liver, K16 cells upon their conversion to a tumorigenic line, W8, after treatment with the carcinogen 2-N-acetoxyacetyl)-aminofluorine (3). The promoter-5' region of the α2 gene was methylated in the nonexpressing W8 cells and not in the expressing K16 cells (5). Furthermore, reporter gene expression downstream of the 218-bp1 promoter and the 54-bp 5' region of the rat and human collagen α2(I) genes was inactivated by in vitro DNA methylation in transient transfection experiments, whereas an analogous expression plasmid with the SV40 early promoter/enhancer driving expression was not (6). We also demonstrated that a minimal collagen α2(I) promoter containing the preinitiation region (−41 to +54) driving expression of the luciferase reporter gene was inactivated by DNA methylation (7). The inhibition of reporter gene expression was attributable to CpG methylation, specifically of collagen α2(I) sequences. However, all the methylation sites were located in the first exon, not in the promoter.

DNA methylation in the promoter and 5' region of genes often correlates with decreased transcription of vertebrate genes, and many studies indicate that this methylation is often causally involved in down-regulation of gene expression (8–14). Different mechanisms have been hypothesized to explain the inactivation due to methylation. In certain cases, methylation inhibits binding of positive transactivating factors. Alternatively, DNA methylation can induce the binding of the nonspecific methyl-sensitive proteins, such as MeCP1 or MeCP2, that bind to methylated cytosine regardless of surrounding bases and act as global repressors by condensing chromatin. In addition, a family of closely related proteins called methylated DNA-binding protein (MDBP) or regulatory factor for X box (RFX) promoter for X box (RFX) 1–4 (15–18) can bind methylated DNA sequences with higher affinity within a sequence-specific 14-bp consensus sequence, 5'-RT(m5C/T)RYYA(m5C/T)RG(m5C/T)RAY-3' (where (m5C/T) indicates 5-methylcytosine or T, R indicates G or A, and Y indicates C or T). Methylation-dependent binding sites were located for this protein at the beginning of the human genes for hypoxanthine phosphoribosyltransferase; α-galactosidase A; human leukocyte antigens (HLA)-A2, -A3, and -A25 antigens; and the apoferritin H gene (19). For the first two X-linked genes, DNA methylation may help down-regulate gene expression on the inactive X chromosome by increasing binding of MDBP at the three MDBP sites in the hypoxanthine phosphoribosyltransferase promoter/5' region (20). The MDBP sites at the beginning of the first exon of the α-galactosidase A gene are at least partially methylated on the inactive X chromosome but completely unmethylated on the active X chromosome (21).

Cytosine methylation-independent sites have been identified that contain T residues replacing 5-methylcytosine residues (22) in hepatitis B virus, polio virus enhancers, cytomegalovirus (CMV) enhancers and c-Myc intron (19, 23–27). Therefore, MDBP family proteins, consisting of homo- or hetero-

* This work was supported in part by National Institutes of Health Grants R01-CA23540 and F590-HL56386 and by the Department of Veterans Affairs merit review program. The costs of publication of this article were defrayed in part by the payment of page charges. This advertisement must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Dept. of Biochemistry, Boston University School of Medicine, 715 Albany St., Boston, MA 02118. Tel.: 617-638-4159; Fax: 617-638-5339; E-mail: smith@biochem.bumc.bu.edu.

1 The abbreviations used are: bp, base pair(s); MDBP, methylated DNA-binding protein; RFX, regulatory factor for X box; CMV, cytomegalovirus; Gal, galactosidase; hPer, human apoferritin H; MHC, major histocompatibility complex; HLA, human leukocyte antigens; EP, hepatitis B viral enhancer; Py1, polyomavirus enhancer; pB1, methylation-dependent site in pBR322.
DNA Methylation and Collagen α2(I) Transcription

erodimers of RFX1–4 subunits (17), can bind in a cytokine methylation-independent or -dependent fashion to their cognate sites, depending on the sequence of these sites. These constitutively expressed proteins can act as repressors or activators in a context-dependent fashion (28, 29). An activation domain containing a glutamine-rich region is found in the N-terminal half of RFX1, whereas a region with repressor activity overlaps the C-terminal dimerization domain.

In our previous studies (7), we demonstrated that methylation sites within the first exon, which inactivated the α2(I) collagen promoter, bind to a sequence-specific methylation-responsive protein. Also, there was decreased formation of a TATA binding complex on methylated DNA ligands in gel shift experiments. This report demonstrates that MDPB/RFX1 binds to the first 20 bases of the first exon. When this sequence, which matches the consensus sequence for MDPB at 10 out of 14 positions in the center of the oligonucleotide, is methylated at its one CpG dinucleotide pair or is mutated to TpG at this site, there is increased binding of MDPB/RFX and inhibition of collagen gene transcription. Therefore, MDPB/RFX protein is likely to contribute to down-regulation of collagen gene repression and might do so in response to increased methylation associated with oncogenic transformation.

MATERIALS AND METHODS

Cell Culture—Rat skin fibroblasts cells (FR) (CRL-1213, American Type Culture Collection) were grown in Dulbecco’s modified Eagle’s medium supplemented with 10% fetal bovine serum, 1% penicillin G/streptomycin sulfate, 1% sodium pyruvate, and 1% l-glutamine.

Electrophoretic Mobility Gel Shift Assay—Nuclear extracts were prepared essentially according to Dignam et al. (30), with some modifications. Extractions of protein from isolated nuclei were performed at higher salt conditions than normal using 500 mM NaCl or 420 mM NaCl rather than 350 mM NaCl in Buffer C. All buffers contained the protease inhibitors leupeptin (40 μg/ml), aprotinin (200 μg/ml), pepstatin A (40 μg/ml), and phenylmethylsulfonyl fluoride (0.5 mM) as well as the phosphatase inhibitor orthovanadate (1 mM). Protein concentration of the extracts was determined by the Bradford reagent using bovine serum albumin as a standard. Collagen sequences (Table I) or MDPB/RFX consensus sequences (Table II) with HindIII overhangs were synthesized (Oligo Etc. and Integrated DNA Technology) as complementary strands, annealed to make double stranded oligonucleotides and radiolabeled using the [α-32P]dATP and the Klenow fragment to fill in the HindIII overhang. For the DNA mobility shift assay, the binding reaction was performed for 30 min at room temperature in 20 μl of binding buffer containing 90,000–100,000 cpm/200 fmol of labeled probe, 1 μg of poly(dI-dC)·poly(dI-dC), and nuclear extract containing 4.5–5.0 μg of protein. Double-stranded radiolabeled complementary oligonucleotides (Oligo Etc. and Integrated DNA Technology) were used as competitors (Tables I and II). Separation of free radiolabeled DNA from DNA-protein complexes was carried out on a 4–5% nondenaturing polyacrylamide gel with a standard Tris-borate electrophoresis buffer at 300 V in the cold (4 °C). Autoradiography was performed by overexposure to Kodak Biomax film (Eastman Kodak Co.). The intensities of the bands were quantified using Instant Imager (Packard Instrument Co.). In the antibody experiment, the nuclear extract and antisera were preincubated for 20 min at room temperature before the radiolabeled probe was added, followed by another 20 min incubation with the probe. The antibody (kindly supplied by Dr. W. Reith to Dr. M. Ehrlich) is a polyclonal antibody to recombinant RFX1 (31), and its specificity for other family members has been described (18).

In Vitro Transcription and Translation—The RFX1 cDNA in the pBK-RSV vector (Stratagene) was transcribed and translated in vitro using a rabbit reticulocyte lysate (Promega; TNT translation kit) following the manufacturer’s protocol. The in vitro translated proteins were used in electrophoretic mobility gel shift assay.

In Vitro Mutagenesis—Mutation at the +7 and +23 sites of collagen α2(I) gene (all positions given relative to the transcription start site) were performed by site-directed mutagenesis based upon Kunkel’s method (Muta-Gen e phagemid mutagenesis kit, Bio-Rad) following the manufacturer’s protocol. The mutated constructs were then cloned into pH 20 (−25 to +54 of the collagen α2(I) fused to the luciferase coding sequence) at SmaI/HindIII sites. The DNA sequence of the mutated constructs were confirmed by DNA sequencing (U. S. Biochemical Corp.) prior to their use in transfection and in vitro transcription assays.

Transient Transfection and Luciferase Assays—Plasmid DNA was transfected by lipofection (LipofectAMINE, Life Technologies, Inc.) into rat fibroblasts 24 h after plating cells. Plasmids containing the wild type or mutated bp in −220 to +54 of collagen α2(I) promoter/5′ region driving expression of the luciferase coding sequence were co-transfected with a reference plasmid, pCMV-green fluorescent protein (CLONTECH) containing the CMV immediate early promoter driving expression of the gene encoding the green fluorescent protein. CMV-green fluorescent protein was used to normalize the transfection efficiency.

Luciferase assays were performed under standard conditions (Luciferase kit; Promega Corp.). Briefly, the cells were washed twice with phosphate-buffered saline buffer and scraped with lysis reagent. The cells and solution were centrifuged at 12,000 × g to pellet the debris. The cell extract was mixed with the luciferase assay reagent, and light emission was measured in a scintillation counter. The luciferase activity was assayed in duplicate within the linear range of the instrument. Ten readings at 60-s intervals were averaged in each assay. Values were normalized to fluorescence of the green fluorescent protein.

In Vitro Transcription Assay—The primer mixture for in vitro transcription contained 50–90 μg of nuclear extract, 1 μg of super-coiled template DNA (purified on a CsCl gradient), 20 mM HEPES, pH 7.9, 4 mM MgCl2, 60 mM KCl, 2 mM EDTA, 0.5 mM dithiothreitol, 12% glycerol, 600 μM of each of rNTP, in a final volume of 25 μl. Reaction was carried out at 30 °C for 1 h and terminated by the addition of 175 μl of stop solution, which contained 0.3 mM sodium acetate, 0.5% SDS, 3 μg/ml RNase A, pH 5.2. After extraction of protein with phenol/chloroform, the RNA was precipitated by ethanol. To detect the newly synthesized transcript, antisense oligonucleotide primer corresponding to a sequence in the luciferase gene was generated (Table I). The primer was end labeled with polynucleotide kinase and [γ-32P]ATP, hybridized to in vitro transcription products and extended using Moloney murine leukemia virus reverse transcriptase. The primer-extended products were analyzed by 5% polyacrylamide gel electrophoresis containing 7 M urea. Transcription reaction and primer extension reactions always included an RNase inhibitor (RNase inhibitor protein, cloned human pancreatic RNase A lytic enzyme inhibitor, Ambion, Inc.). Gels were dried and autoradiographed at −80 °C with an intensifying screen.

RESULTS

In our earlier study (7), we demonstrated that the −25 to +30 sequence of α2(I) promoter could bind sequence-specific nuclear proteins preferentially when CpGs were methylated. This sequence contains two CpG sites, at +7 and +23, respectively, relative to the transcription start site. In order to investigate which sites are important for the methylation responsiveness, a gel shift experiment was performed using the wild type and mutated probes (both unmethylated and methylated) with rat fibroblast nuclear extracts (Fig. 1). The DNA fragment with a CpG to TpG transition at position +23 specifically complexed with proteins in the extract in a similar manner as the wild type DNA fragment. There was a 3-fold increase in

FIG. 1. Methylation and mutation at the +7 CpG site increases the protein-DNA complex formation on the α2(I) initiator probe as judged by electrophoretic mobility shift assay. Duplex oligonucleotide sequences corresponding to positions −25 to +30 of the α2(I) gene containing C to T mutations at +7, +23, or both were incubated with a reference plasmid, pCMV-green fluorescent protein (CLONTECH) containing the CMV immediate early promoter driving expression of the gene encoding the green fluorescent protein. CMV-green fluorescent protein was used to normalize the transfection efficiency.

Luciferase assays were performed under standard conditions (Luciferase kit; Promega Corp.). Briefly, the cells were washed twice with phosphate-buffered saline buffer and scraped with lysis reagent. The cells and solution were centrifuged at 12,000 × g to pellet the debris. The cell extract was mixed with the luciferase assay reagent, and light emission was measured in a scintillation counter. The luciferase activity was assayed in duplicate within the linear range of the instrument. Ten readings at 60-s intervals were averaged in each assay. Values were normalized to fluorescence of the green fluorescent protein.

In Vitro Transcription Assay—The reaction mixture for in vitro transcription contained 50–90 μg of nuclear extract, 1 μg of super-coiled template DNA (purified on a CsCl gradient), 20 mM HEPES, pH 7.9, 4 mM MgCl2, 60 mM KCl, 2 mM EDTA, 0.5 mM dithiothreitol, 12% glycerol, 600 μM of each of rNTP, in a final volume of 25 μl. Reaction was carried out at 30 °C for 1 h and terminated by the addition of 175 μl of stop solution, which contained 0.3 mM sodium acetate, 0.5% SDS, 3 μg/ml RNase A, pH 5.2. After extraction of protein with phenol/chloroform, the RNA was precipitated by ethanol. To detect the newly synthesized transcript, antisense oligonucleotide primer corresponding to a sequence in the luciferase gene was generated (Table I). The primer was end labeled with polynucleotide kinase and [γ-32P]ATP, hybridized to in vitro transcription products and extended using Moloney murine leukemia virus reverse transcriptase. The primer-extended products were analyzed by 5% polyacrylamide gel electrophoresis containing 7 M urea. Transcription reaction and primer extension reactions always included an RNase inhibitor (RNase inhibitor protein, cloned human pancreatic RNase A lytic enzyme inhibitor, Ambion, Inc.). Gels were dried and autoradiographed at −80 °C with an intensifying screen.

RESULTS

In our earlier study (7), we demonstrated that the −25 to +30 sequence of α2(I) promoter could bind sequence-specific nuclear proteins preferentially when CpGs were methylated. This sequence contains two CpG sites, at +7 and +23, respectively, relative to the transcription start site. In order to investigate which sites are important for the methylation responsiveness, a gel shift experiment was performed using the wild type and mutated probes (both unmethylated and methylated) with rat fibroblast nuclear extracts (Fig. 1). The DNA fragment with a CpG to TpG transition at position +23 specifically complexed with proteins in the extract in a similar manner as the wild type DNA fragment. There was a 3-fold increase in
binding when the CpG at position +7 was methylated whether or not the +23 site was mutated. This result suggests that the +23 CpG site is not important for the increased binding upon cytosine methylation. On the other hand, there was no difference in binding between unmethylated and methylated DNA fragments when the ligand had a CpG toTpG mutation at the +7 position, and only the +23 CpG was differentially methylated. This indicates that only the +7 site is important for increasing protein binding on methylated constructs. In addition, the amount of complex formation with the +7 mutated DNA fragment was approximately 3 times more than with the analogous wild type unmethylated fragment.

Ten base pairs out of 14 in the CpG methylated sequence from position –1 to +13 of the collagen α2(1) 5′ region match the consensus sequence for the transcription regulatory protein MDBP/RFX (Fig. 2A). Furthermore, the CpG of this sequence is in the same position as the CpG of several previously described methylation-dependent MDBP/RFX sites (19). Lastly, CpG to TpG transitions in the methylation-dependent sites have been shown to increase binding by MDBP/RFX in a similar fashion to cytosine methylation at these sites. Therefore, we suspected that the collagen α2(1) sequence in this region is an MDBP/RFX site.

Four short oligonucleotides from different parts of the 55-bp α2(1) initiator region DNA fragment (position –25 to +30) initially were used as ligands in electrophoretic mobility shift assays to determine whether there was methylation sensitive binding of a nuclear protein to these shorter probes. Nuclear extracts from rat skin fibroblasts were used in these assays with oligonucleotide duplexes containing α2(1) sequences –20 to +1, –14 to +6, +10 to +30, or –1 to +20 bordered by a HindIII site (AAGCTT) added to ends of both strands (see Table I). The sequence in the beginning of the first exon of the α2(1) gene containing the region homologous to MDBP/RFX was the only sequence that resulted in a methylation responsive formation of a specific complex when used as a radiolabeled ligand (data not shown). Two specific, slowly migrating DNA-protein complexes formed with the –1 to +20 oligonucleotide that migrated only slightly faster (commensurate with the small size of the ligand) than the methylation-responsive complexes formed from the longer –25 to +30 oligonucleotide. Furthermore, just as methylation increased the amount of this specific complex formation from the –25 to +30 oligonucleotide, methylation increased binding by the smaller –1 to +20 oligonucleotide 2.5–3-fold (Fig. 2B, lanes 1 and 2). This complex formation was specific, as it was competed by an excess of the identical unlabeled sequence (Fig. 2, lanes 3 and 7) but not a similar excess of other tested α2(1) sequences (–14 to +6, +10 to +30, and –20 to +1) (Fig. 2B, lanes 4–6 and 8). Competition for binding to methylated sequence (–1 to +20) probe increased when the specific competitor was methylated (Fig. 2B, lanes 3 and 7), whereas methylation of the +10 to 30 competitor did not visibly change the amount of labeled complex formation (Fig. 2B, lanes 6 and 8).

The ability of various other oligonucleotide duplexes to com-

**FIG. 2.** Methylation-dependent sequence-specific binding activity is located at the transcription start site of α2(1) collagen gene (–1 to +13). A, sequence homology between consensus sequence for MDBP/RFX and α2(1) collagen at the transcription start site (–1 to +13). R, purine; m, methylated C or T; Y, pyrimidine. B, methylation-sensitive and sequence-specific binding to the α2(1) consensus sequence and the competitor is indicated at the top. Competitors at 50-fold molar excess over the labeled ligand (200 fmol of labeled ligand) were incubated with nuclear proteins, and then radioactive probes were added. Different regions of the α2(1) collagen gene were used as competitors as follows: lane 1 and 7, –1 to +20; lane 4, –14 to +6; lane 5, –20 to +1; lanes 6 and 8, +10 to +30 (see Table I for sequence information). Competitor sequences are methylated in lanes 7 and 8. The assay conditions were the same as in Fig. 1. The arrows indicate the protein-DNA complexes generated by MDBP/RFX and α2(1) sequence.

**TABLE I**

Sequences of the collagen oligonucleotides used in experiments

| Name          | Position       | 5′–3′ sequence                                      |
|---------------|----------------|----------------------------------------------------|
| α2(1), –25 to +30 | α2(1) promoter/first exon sequence | GGCAGGTCGGCTGCGCTTTATATTGGACCCAAGAGCGAGGTTTCGACTAA |
| α2(1), –1 to +20 | α2(1) first exon sequence | CCGCCAGCCCAGCCGAGAAATAATAAATAATCGTGGTGGCTGGTCTTGGTCGCAAGCTGATActtc |
| α1(1), –1 to +20 | α1(1) first exon | aagcttAAGGAGGTTTGCAATGAA |
| α2(1), –14/+6 | α2(1) promoter/first exon sequence | TCGTGGTGGCGCTGGTCTGCCATCTGaa |
| mTAE         | Mutated TAE in α1(1) promoter at –1624 | CAGAATGAAAATCGTGGTGTGttcag |
| α2(1), –20/+1 | α2(1) promoter sequence | TGTGGGCGCCGCCTACAGGGCCCCG |
| α2(1), +10/+30 | α2(1) first exon sequence | aagcttTCTGGTTGCCTTATATTGGAG |
| Antisense primer | Luciferase | TCCATAGGCCTATATCTGGCTCTCCAGCG |
Fig. 3. The MDBP/RFX recognition sites compete for protein-DNA complex formed on the α2(I) sequence (A), and the α2(I) protein complex co-migrates with the MDBP/RFX protein-DNA complex (B). A, competitors at 50-fold molar excess over the labeled ligand (200 fmol of labeled ligand) were incubated with nuclear proteins, and then radioactive probes were added as in Fig. 2B. MDBP/RFX sites used as competitors are shown in lanes 2–7 (lane 2, EP from hepatitis B virus enhancer 1; lane 3, X box from MHC type II gene; lane 4, Py1 from polyomavirus enhancer B enhancer; lane 5, α-GaLA from the α-galactosidase A gene; lane 6, pBSite1 from pBR322; lane 7, hFer from apoferitin gene) (Table II). The methylation-sensitive competitor sites (α-GaLA, hFer, and pB1) were methylated (lanes 5–7). competitor sequences unrelated to MDBP/RFX binding sites are α1(I) (lane 8) and mTAE (lane 9) (Table I). Lane 1 contains no competitors. The arrows indicate the protein-DNA complexes generated by MDBP/RFX and the α2(I) methylated sequence. B, nuclear extracts were incubated with the same five known MDBP/RFX ligands in addition to the α2(I) sequence position –1 to +20. All probes were labeled at the same specific activity (100,000 cpm/200 fmol). MDBP/RFX sites used as probes are shown in lanes 3–7 (lane 3, EP from hepatitis B virus enhancer 1; lane 4, X box from MHC type II; lane 5, α-GaLA from the α-galactosidase A gene; lane 6, pBSite1 from pBR322; lane 7, hFer from the apoferitin gene) (Table II). Methylation status of probes is indicated at the top. The arrows indicate the protein-DNA complexes generated by MDBP/RFX and the α2(I) methylated sequence.

To test for control of gene expression at this sequence in the beginning of the α2(I) gene, C to T mutations at +7 and/or +23 sites were introduced by site-directed mutagenesis into the α2(I) promoter-luciferase construct, pH 20, containing the −220 to +54 region of the promoter. These expression plasmids were transfected into rat fibroblast cells. The +7 mutation inhibited transcription to a greater extent (80 ± 2.6% (S.D.)) than the +23 mutation (47 ± 3.8%) (Fig. 5A). The same constructs were used in an in vitro transcription assay. The +7 mutation decreased the transcription by 56% but the +23 mutation decreased only 23% (Fig. 5B) in a representative experiment repeated three times. Therefore, a CpG-to-TpG mutation at the +7 site, which increases the formation of specific protein complexes in vitro, also decreases transcription in vivo and in reaction mixtures containing a nuclear extract. This mutation had more of an effect on transcription than did an analogous mutation at position +23, which does not match the MDBP/RFX consensus sequences.
DNA Methylation and Collagen α2(I) Transcription

DISCUSSION

One of the unusual characteristics of MDBP/RFX proteins is their behavior with respect to CpG-containing recognition sites. These sequence-specific DNA-binding proteins show increased binding to these sites in their CpG-methylated form. The sequence specificity of the binding site distinguishes these proteins from methylated DNA-binding proteins, such as MeCP1, MeCP2, DBP-m, and MDBP-2-H1 (22, 32–35). We demonstrated that the MDBP/RFX could bind to a sequence at the very beginning of the α2(I) gene first exon in the rodent and human genomes in a methylation-dependent manner. Similar methylation-responsive MDBP sites are present at the beginning of the human genes for hypoxanthine phosphoribosyl transferase; HLA-A2, -A3, -A25 antigens; and α-galactosidase A (19).

MDBP/RFX also binds to many of its sites in a cytosine methylation-independent manner if 1–3 of the CpGs in the consensus sequence are replaced by TpG or TpA. There is a CpG site 7 base pairs downstream from the transcription start site of the α2(I) gene that is within the MDBP/RFX site. This CpG matches the central CpG of the MDBP/RFX consensus sequence (Fig. 2A), which confers methylation-dependent binding on other methylation-responsive binding sites for this family of closely related proteins (19). The location of this MDBP/RFX site in the α2(I) gene was defined by homology to many known MDBP/RFX sites, electrophoretic mobility shift assays with rat fibroblast nuclear extracts and with the in vitro transcription/translation product from an MDBP/RFX1 template, and supershift assays using specific antibody to MDBP/RFX1 (Figs. 2–4). When we replaced this unmethylated CpG within the MDBP/RFX site with a TpG, there was increased specific complex formation comparable to the increase seen upon methylation of this CpG dinucleotide pair (Fig. 1), just as has been observed with other methylation-dependent MDBP/RFX sites (19). When the next downstream CpG in this region of the α2(I) gene, namely, the CpG at position +22, was converted to a TpG, there was no effect on complex formation with MDBP/RFX proteins in a nuclear extract, as expected, because this dinucleotide is outside the MDBP/RFX site.

A reporter containing the α2(I) promoter/5′ region gene driving expression of a luciferase gene was equivalently mutated at the CpG at position +7. The observed decrease in promoter activity in the TpG mutant in both transfection experiments in cultured rat fibroblast and in vitro transcription with fibroblast extracts may be due to increased binding of MDBP/RFX to the mutated sequence (Fig. 5). Although this mutation may also affect formation of the preinitiation complex for transcription in a manner independent of MDBP/RFX binding, a similar mutation in a more downstream methylation-dependent MDBP/RFX site (48 bp downstream of the major transcription start site) in the human α-galactosidase A also decreases gene expression in transient transfection assays. Furthermore, when that MDBP/RFX site was mutated to a yeast GAL4 binding site and yeast GAL4 DNA-binding domain chimeras with mammalian transcription factors were present, a hybrid GAL4 DNA-binding domain-MDBP/RFX down-regulated reporter gene expression, whereas the intact GAL4 transcription factor and a hybrid GAL4 DNA-binding domain-VP16 activation domain up-regulated expression of the reporter gene (1).

All of the homo- and heterodimeric members of the MDBP/RFX1–3 family of DNA-binding protein exhibit methylation-dependent binding to certain of their cognate sites (16, 17). In contrast, RFX5 is a more distant member of this family, which is involved in positive regulation of mammalian histocompatibility type II genes and does not display methylation-dependent binding (17). Yeast proteins involved in cell cycle control and DNA damage are also present in this family (36, 37). MDBP/RFX family members have very similar DNA binding domains but different N-terminal regions. RFX2–4 show appreciable tissue-specificity in their distribution, whereas RFX1 is present at similar levels in a variety of examined tissues (16).

RFX1 homodimer, a large protein with 979-amino acids per subunit, contains both transcription repression and activation domains, the ability of which to positively or negatively modulate transcription may vary depending upon the location of its cognate sites in a given gene region and the other proteins with which it interacts (28, 29). Positive regulation of transcription by MDBP/RFX has been demonstrated for the methylation-independent binding site in the EP sequence of the hepatitis B virus enhancer (38). In contrast, one of us previously demonstrated that a low affinity methylation-independent binding site beginning at position +5 of the CMV IE transcription unit can down-regulate transcription when its binding by MDBP/RFX is increased by mutation (39).

Dual function transcription factors have been described that switch their function by interaction with different co-activators or co-repressors (40–42), different neighboring transacting factors (43–45), or interaction with specific DNA sequences in different locations relative to the transcription start site (46). MDBP/RFX interacts with c-Abl, greatly stimulating its auto-phosphorylation (47). We have preliminary data using antibody supershifting experiment suggesting that c-Abl protein is present in the complex formed with the collagen sequence.3 MDBP/RFX can down-regulate transcription when its binding by MDBP/RFX is increased by mutation (39).

| Table II  |
|-----------------------------------------------|
| **Sequences of the MDBP/RFX consensus oligonucleotides used in experiments** |
| All of the collagen sequences are derived from GenBank™ sequences, shown in uppercase letters. Lowercase letters are Hin dIII overhangs used for labeling. |
| **EP** | Hepatitis B virus |
| **X** | HLA-A2, -A3, -A25 |
| **Polymavirus enhancer/ori region** | human apoferritin +202 |
| **pBR322** | Human apoferritin +202 |
| **α-Gal** | α-Galactosidase A +49 |
| **hFer** | α-Galactosidase A +49 |

4 M. Ehrlich, manuscript in preparation.
3 P. K. Sengupta and B. D. Smith, unpublished data.
4 M. Ehrlich, unpublished data.
lysate (lanes 2–4).

In vitro translation product containing bases 220 to 254 of the collagen gene were transcribed *in vitro* with rat fibroblast nuclear extract. The RNA transcript was detected by primer extension as described under “Materials and Methods.” This is a representative of an experiment performed three times.

The data presented here indicate that MDBP/RFX behaves as transcriptional repressor in the context of methylated collagen α2(I) sequence. We have previously demonstrated that in a rat cell line that has lost expression of this gene, there is methylation in the promoter region and that treatment of this cell line with the DNA demethylating agent 5-azacytidine results in the gain of expression of this gene (43). Because abnormal hypermethylation of the 5’ region of tumor suppressor genes is so frequent in cancer and down-regulation of collagen gene expression has been proposed to contribute to carcinogenesis, abnormal methylation of the 5’ region of the α2(I) gene, including the MDBP/RFX site, might occur in cancers and play a role in down-regulation of the expression of this gene. Future genomic sequencing experiments will allow testing of this hypothesis and of the possibility that the RFX/MDBP is a transcriptional repressor involved in tissue-specific regulation of the collagen α2(I) gene transcription in response to naturally occurring differential methylation.

Fig. 5. Mutation of two CpG sequences to TpG sequences (+7 or +23) in the collagen α2(I) exon inhibited transcription in transient transfection (A) and *in vitro* transcription assays (B). A, rat skin fibroblasts were transfected by the LipofectAMINE method with collagen promoter-luciferase constructs (2 μg) containing C to T mutations at +7 or +23. A CMV promoter driving the green fluorescent protein gene (1 μg) was co-transfected to normalize the transfection efficiency. The graph represents the average luciferase activity/μg of protein with standard error from four different experiments performed in duplicate. B, the control and mutated α2(I) collagen promoter-luciferase constructs (1 μg) containing bases −220 to +54 of the collagen gene were transcribed *in vitro* with rat fibroblast nuclear extract. The RNA transcript was detected by primer extension as described under “Materials and Methods.” This is a representative of an experiment performed three times.

**REFERENCES**

1. Adams, S. L., Sobel, M. E., Howard, B. H., Olden, K., Yamada, K. M., de Crombrugghe, B., and Pastan, I. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 3299–3303.
2. Rowe, D. W., Moen, R. C., Davidson, J. M., Byers, P. H., Bornstein, P., and Palmiter, R. D. (1978) *Biochemistry* 17, 1581–1590.
3. Sandmeyer, S., Smith, R., Kiehn, D., and Bornstein, P. (1981) *Biochemistry* 20, 8258–8264.
4. Guenette, D. K., Ritzenthaler, J. D., Foley, J., Jackson, J. D., and Smith, B. D. (1992) *Biochem. J.* 283, 699–703.
5. Sengupta, P. K., and Smith, B. D. (1998) *Biochem. Biophys. Acta* 1443, 75–89.
6. Umekawa, A., Yamamoto, H., Rhodes, K., Klemm, M. J., Maki, R. A., and Oshima, R. G. (1997) *Mol. Cell. Biol.* 17, 4885–4894.
7. Okuse, K., Matsuzaka, I., and Kurihara, K. (1997) *Brain Res. Mol. Brain Res.* 46, 197–207.
8. Singal, R., Ferris, R., Little, J. A., Wang, S. Z., and Ginder, G. D. (1997) *Proc. Brain Res.* 774, 367–370.
DNA Methylation and Collagen α2(I) Transcription

Natl. Acad. Sci. U. S. A. 94, 13724–13729
11. Wolff, G. L., Kodell, R. L., Moore, S. R., and Cooney, C. A. (1998) FASEB J. 12, 949–957
12. Kudo, S., and Fukuda, M. (1995) J. Biol. Chem. 270, 13298–13302
13. Kirillov, A., Kistler, B., Mostoslavsky, R., Cedar, H., Wirth, T., and Bergman, Y. (1996) Nat. Genet. 13, 435–441
14. Iannello, R. C., Young, J., Sumarsono, S., Tymms, M. J., Dahl, H. H., Gould, J., Hedger, M., and Kola, I. (1997) Mol. Cell. Biol. 17, 612–619
15. Huang, L. H., Wang, R., Gama-Sosa, M. A., Shenoy, S., and Ehrlich, M. (1984) Nature 308, 293–295
16. Reith, W., Ucla, C., Barras, E., Gaud, A., Durand, B., Herrero-Sanchez, C., Kobr, M., and Mach, B. (1994) Mol. Cell. Biol. 14, 1230–1244
17. Emery, P., Durand, B., Mach, B., and Reith, W. (1996) Nucleic Acids Res. 24, 803–807
18. Zhang, X. Y., Jabrune-Ferrat, N., Asiedu, C. K., Samac, S., Peterlin, B. M., and Ehrlich, M. (1999) Mol. Cell. Biol. 13, 6810–6818
19. Zhang, X. Y., Asiedu, C. K., Supakar, P. C., Khan, R., Ehrlich, K. C., and Ehrlich, M. (1999) Nucleic Acids Res. 18, 6253–6260
20. Hornstra, I. K., and Yang, T. P. (1994) Mol. Cell. Biol. 14, 1419–1430
21. Samac, S., Rice, J. C., and Ehrlich, M. (1998) Biol. Chem. 379, 541–544
22. Zhang, X. Y., Supakar, P. C., Khan, R., Ehrlich, K. C., and Ehrlich, M. (1989) Nucleic Acids Res. 17, 1459–1474
23. Ghazal, P., Lubon, H., Fleckenstein, B., and Hennighausen, L. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 3658–3662
24. Karpen, S., Banerjee, R., Zeleni, A., Price, P., and Acs, G. (1988) Mol. Cell. Biol. 8, 5159–5165
25. Ostapchuk, P., Scheirle, G., and Hearing, P. (1989) Mol. Cell. Biol. 9, 2787–2797
26. Zhang, X. Y., Supakar, P. C., Wu, K. Z., Ehrlich, K. C., and Ehrlich, M. (1990) Cancer Res. 50, 6865–6869
27. Zhang, X. Y., Inamdar, N. M., Supakar, P. C., Wu, K., Ehrlich, K. C., and Ehrlich, M. (1991) Virology 182, 865–869
28. Katan, Y., Agami, R., and Shaul, Y. (1997) Nucleic Acids Res. 25, 3621–3628
29. Katan-Rhaykovitch, Y., and Shaul, Y. (1998) J. Biol. Chem. 273, 24504–24512
30. Dignam, J. D., Martin, P. L., Shastry, B. S., and Roeder, R. G. (1983) Methods Enzymol. 101, 582–598
31. Herrero Sanchez, C., Reith, W., Silacci, P., and Mach, B. (1992) Mol. Cell. Biol. 12, 4076–4083
32. Nan, X., Cross, S., and Bird, A. (1998) Nucleic Acids Res. 26, 6–16
33. Nan, X., Ng, H. H., Johnsen, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., and Bird, A. (1998) Nature 393, 386–389
34. Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992) Cell 69, 905–914
35. Schwarz, S., Hess, D., and Joist, J. P. (1997) Nucleic Acids Res. 25, 5052–5056
36. Wu, S. Y., and McLeod, M. (1995) Mol. Cell. Biol. 15, 1479–1488
37. Huang, M., Zhou, Z., and Elledge, S. J. (1998) Cell 94, 595–605
38. Siegrist, C. A., Durand, B., Emery, P., David, E., Hearing, P., Mach, B., and Reith, W. (1995) Mol. Cell. Biol. 15, 6375–6384
39. Zhang, X. Y., Ni, Y. S., Saifudeen, Z., Asiedu, C. K., Supakar, P. C., and Ehrlich, M. (1995) Nucleic Acids Res. 23, 3026–3033
40. Shi, Y., Seto, E., Chang, L. S., and Shenk, T. (1991) Cell 67, 377–388
41. Treitel, M. A., and Carlson, M. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 3132–3136
42. Mahseswaran, S., Park, S., Bernard, A., Morris, J. F., Rauscher, F. J., Hill, D. E., and Haber, D. A. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 5109–5104
43. Smith, K. J., Smith, W. J., Hamilton, T., Skelton, H. G., Graham, J. S., Okerberg, C., Moeller, R., and Hackley, B. E., Jr. (1998) Am. J. Dermatopathol. 20, 22–28
44. Ip, Y. T. (1995) Curr. Biol. 5, 1–3
45. Lehning, N., Thoms, D., Brickman, J. M., Ma, J., Maniatis, T., and Ptashne, M. (1994) Nature 371, 175–179
46. Kurokawa, R., Soderstrom, M., Horlein, A., Halachmi, S., Brown, M., Rosenfeld, M. G., and Glass, C. K. (1995) Nature 377, 451–454
47. Agami, R., and Shaul, Y. (1998) Oncogene 16, 1779–1788
48. Liu, M., Lee, B. H., and Mathews, M. B. (1999) J. Biol. Chem. 274, 15433–15439
49. Reinhold, W., Emens, L., Itkes, A., Blake, M., Ichinose, I., and Zajac-Kaye, M. (1995) Mol. Cell. Biol. 15, 3041–3048
50. Lee, B. H., Liu, M., and Mathews, M. B. (1998) J. Virol. 72, 1138–1145
51. Labrie, C., Lee, B. H., and Mathews, M. B. (1995) Nucleic Acids Res. 23, 3732–3741
52. Blake, M., Niklinski, J., and Zajac-Kaye, M. (1996) J. Virol. 70, 6060–6066