MOLECULAR AND CELLULAR BIOLOGY, Sept. 1990, p. 4957–4960
0270-7306/90/094957-04$02.00/0

Vol. 10, No. 9

Highly Repeated Sites in the Apolipoprotein(a) Gene Recognized by Methylated DNA-Binding Protein, a Sequence-Specific DNA-Binding Protein

KENNETH C. EHRLICH[1] AND MELANIE EHRLICH[2]*

Southern Regional Research Center, U.S. Department of Agriculture, New Orleans, Louisiana 70179,[1] and Department of Biochemistry, Tulane Medical School, New Orleans, Louisiana 70112[2]

Received 13 April 1990/Accepted 6 June 1990

Methylated DNA-binding protein (MDBP), a sequence-specific DNA-binding protein, was found to recognize more than 30 sites within an allele of the human apolipoprotein(a) gene. High plasma levels of apolipoprotein(a), a risk factor for atherosclerosis, have been correlated with genetically inherited lower-molecular-mass isoforms of this protein. MDBP might help down modulate the expression of the apolipoprotein(a) gene in a manner dependent on the length of a given allele of the gene and the number of MDBP sites in it.

Elevated concentrations of lipoprotein(a) in human plasma have been correlated with increased risk of cardiovascular disease (24). The most specific protein component of lipoprotein(a) is apolipoprotein(a). Apolipoprotein(a) has a unique structure characterized by many tandem copies of kringle-4, a conserved motif of approximately 80 amino acids within a 114-amino-acid repeat (13, 16). The corresponding intragenic repeat unit is highly conserved (16). The size heterogeneity of human apolipoprotein(a) (M_r ~300,000 to 700,000; 13) is genetically controlled and probably is partially the result of different numbers of kringle-4 units in the apolipoprotein(a) gene (1, 7, 12, 24). In a recent study, it was estimated that the number of repeats of the kringle-4 domain in apolipoprotein(a) can vary from ~15 to ~37 (12). Individuals with lower-molecular-mass isoforms typically have higher (as much as 10-fold or more) plasma concentrations of lipoprotein(a) than do those with the larger isoforms (24), and these isoform levels in plasma correlate with the abundance of the apolipoprotein(a) mRNA (12). The basis of this phenomenon remains unexplained. In this report, we present a possible explanation for it.

In contrast to the tandem intragenic repetition of kringle-4 units in apolipoprotein(a), a single kringle-4 domain is present in human plasminogen (3, 15), a zymogen participating in fibrinolysis, whose binding to endothelial cells is inhibited by lipoprotein(a) (6, 17). By a computer-assisted search of the mammalian DNA data base, we discovered that the single kringle-4 unit in the human plasminogen gene contains a 14-base-pair (bp) site closely resembling binding sites for the sequence-specific protein MDBP (methylated DNA-binding protein) (8, 26). Analogously, most of the kringle-4 units in the human apolipoprotein(a) gene contain binding sites for the same protein.

MDBP, a ubiquitous mammalian protein, binds to sites in several viral enhancers, including to one directly implicated in enhancer function (20, 21; X.-Y. Zhang et al., submitted for publication). The highly degenerate, 14-bp consensus sequence for binding to MDBP is 5′-RTm^{5}YRYYAm^{5}Y RGm^{5}YRAY-3′ (R, A or G; m^{5}Y, 5-methylcytosine or T; Y, C or T). As indicated by this sequence, some sites require methylation of cytosine residues for binding; however, others, like the viral enhancer MDBP sites, do not (11, 20, 21, 26). 5-Methylcytosine, the only genetically determined, modified base detected in mammalian DNA (4), is found predominantly at CpG dinucleotides (2); therefore, many methylated MDBP sites matching the consensus sequence might be present in mammalian DNA. We observed that the human apolipoprotein(a) gene has three different types of oligonucleotide sequences matching the MDBP consensus sequence (Table 1). We predicted that two of these repeated sequences, sites apo(a)1 and apo(a)2, would bind to MDBP in a methylation-independent fashion and that one, site apo(a)3, would bond in a CpG methylation-dependent manner. The apo(a)1 site is identical to the above-mentioned sequence in the human plasminogen gene.

Band shift assays. To visualize complexes of MDBP with oligonucleotide duplexes containing these sequences, we used a band shift (gel retardation) assay (20). Band shift assays were performed with 20 fmol of ^{32}P-labeled oligonucleotide duplexes containing the putative MDBP sites in the apolipoprotein(a) gene and electrophoresed on 5% polyacrylamide gels as described previously (20). The complexes formed between MDBP and these duplexes were compared with those from our standard ligand for MDBP assays, pB site 1 (Table 1). The latter is the high-affinity, methylation-dependent MDBP-binding site of in vitro-methylated pBR322 DNA (11). Approximately 1 U of MDBP was used per assay; 1 U complexes 1 fmol of the pB site 1 duplex under standard conditions (20). The apo(a)1, apo(a)2, and methylated apo(a)3 [apo(a)3m] oligonucleotide duplexes gave DNA-protein complexes that coelectrophoresed with those of pB site 1 (Fig. 1). Very much lower amounts of these complexes formed with the analogous unmethylated apo(a)3u oligonucleotide duplex (Table 1; Fig. 1). That the binding to these apo(a)1, -2, and -3m sites was specific was shown in DNA competition experiments using either MDBP partially purified from human placenta or a crude nuclear extract from a human cell line (HL-60) cells). Addition of excess unlabeled, MDBP-specific oligonucleotide duplexes completely or almost completely abolished detectable binding by ^{32}P-labeled apo(a)1, apo(a)2, or methylated apo(a)3 oligonucleotide duplexes (Fig. 1 and 2 and data not shown). Heterologous sequences gave no detectable competition (Fig. 2). Only slight competition was observed with a high concentration of unmethylated pB site 1.

* Corresponding author.

TABLE 1. Binding to MDBP by human apolipoprotein(a) gene sequences

Site[a]	No. of repeats[b]	Sequence[c]	Relative binding[d] (%)
		1 5 10 14	
pB site 1		5′–c t a g A T M G T C A M G G M G A T–3′ 3′–T A G M A G T G M C G M T A g a t c–5′	100
apo(a)1	4	5′–a t t G C T A C C A T G G T G A T g g a–3′ 3′–t a a C G A T G G T A C C A C T A c c t–5′	154
apo(a)2	~25	5′–a g t G C T A C C A T G G T A A T g g a–3′ 3′–t c a C G A T G G T A C C A T T A c c t–5′	116
apo(a)3m		5′–a g t G C T A C C A M G G T A A T g g a–3′ 3′–t c a C G A T G G T G M C A T T A c c t–5′	91
apo(a)3u	5	5′–a g t G C T A C C A C G G T A A T g g a–3′ 3′–t c a C G A T G G T G C C A T T A c c t–5′	6

[a] The pB site 1 duplex containing a methylated sequence from pBR322 DNA (11) serves as a standard for MDBP-specific binding. The other sites are from the human apolipoprotein(a) cDNA sequence determined by McLean and co-workers (16).

[b] The number of copies of a given 14-bp MDBP site within a sequenced apolipoprotein(a) cDNA (16) is indicated. Each of these copies is in a separate kringle-4 unit. Relative to the first base of that cDNA (16), the positions of the MDBP apo(a) sites are as follows: apo(a)1, +128, 11072, 11390, and 12074 (kringle-4 units 1, 33, 34, and 36, respectively); apo(a)2, one site in each of ~22 tandemly repeated, 342-bp kringle-4 units (2 to 23) beginning at +470 as well as at +8336, 8678, and 12416 (kringle-4 units 25, 26, and 37, respectively); apo(a)3, four sites in tandemly repeated kringle-4 units (kringle-4 units 27 to 30) beginning at +9020 and one at +7994 (kringle-4 unit 24).

[c] The base pairs corresponding to the 14-bp consensus sequence for MDBP are capitalized and numbered in pB site 1 and the apolipoprotein(a) sites. M, 5-Methylcytosine.

[d] Determined by cutting bands from gels and quantitating Cerenkov radiation.

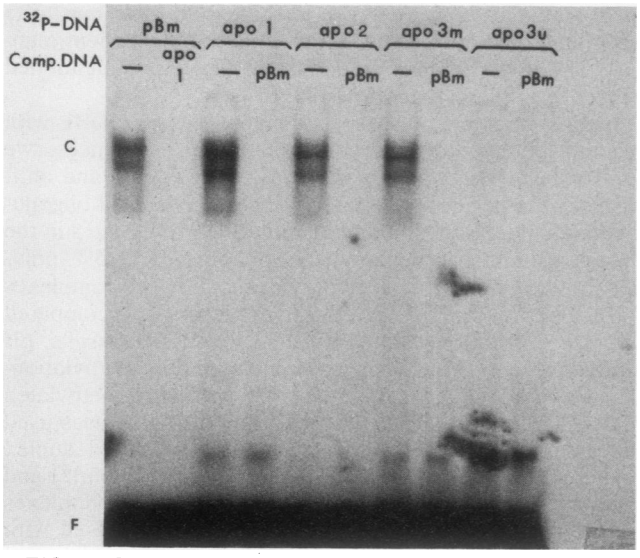

FIG. 1. Complex formation between the hydroxylapatite fraction of placental MDBP (8) and oligonucleotide duplexes containing MDBP sites in the apolipoprotein(a) gene. These ^{32}P-labeled duplexes, whose sequences are given in Table 1, were apo(a)1, apo(a)2, CpG-methylated apo(a)3 [apo3m], unmethylated apo(a)3 [apo3u], or pB site 1 (pBm), as a standard CpG-methylated MDBP ligand. First, MDBP was preincubated for 10 min at room temperature with 0.5 μg of poly(dI) · poly(dC) and, where indicated, with 400 fmol of a specific competitor, pBm or apo(a)1 oligonucleotide duplexes in 10 mM Tris hydrochloride (pH 7.6)–80 mM NaCl–5 mM $MgCl_2$–10% glycerol–0.2 mM sodium EDTA–2 mM dithiothreitol–80 μg of bovine serum albumin per ml. Then the radiolabeled ligand (20 fmol) was added, and incubation continued for 15 min. C and F denote the closely related (21) family of MDBP-specific complexes and the free (uncomplexed) fragment, respectively.

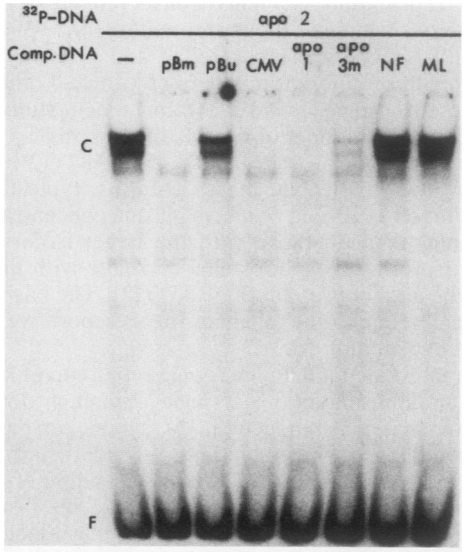

FIG. 2. Complex formation between MDBP in crude nuclear extracts and site apo(a)2. Band shift assays were performed as for Fig. 1 with 20 fmol of ^{32}P-labeled apo(a)2 duplex except using a crude nuclear extract (4 μg of protein) from HL-60 cells, a promyelocytic leukemic human cell line, as the source of MDBP and various unlabeled DNA competitors (200 fmol of oligonucleotide duplex) in the reaction. The MDBP-specific oligonucleotide duplexes used as competitors were methylated pB site 1 (pBm), apo(a)1, apo(a)3m, and a 20-bp duplex containing cytomegalovirus (CMV) site 1, one of the two MDBP sites from the major immediate-early enhancer of human cytomegalovirus DNA (Zhang et al., submitted). The nonspecific competitors were the unmethylated form of pB site 1 (pBu), a 34-bp duplex containing a site for the mammalian sequence-specific DNA-binding protein NF1 (25) (NF), and 0.1 μg of *Taq*I-digested *Micrococcus lysodeikticus* DNA (ML). ^{32}P-labeled pB site 1 formed complexes with MDBP in this nuclear extract that coelectrophoresed with those of ^{32}P-labeled apo(a)1, -2, and -3m oligonucleotide duplexes (data not shown).

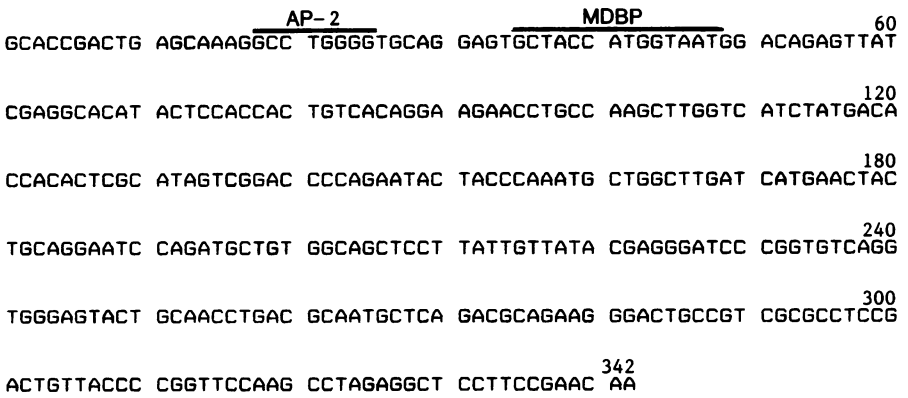

FIG. 3. The 342-bp sequence of the exactly conserved and most highly repeated of the three variants of the kringle-4 unit found in the previously sequenced apo(a) cDNA (16). The MDBP site, apo(a) site 2, and the site with the consensus sequence for AP-2 are indicated. The ~15 to ~37 repeats of the kringle-4 units occur in tandem and constitute most of the cDNA (12, 16).

We do not know whether MDBP will bind well in vivo to the several genomic copies of the methylation-dependent apo(a)3 site because the methylation status of CpG dinucleotides in the apolipoprotein(a) gene is unknown. However, the other >20 copies of the MDBP site in the cDNA sequenced by McLean et al. (16) should be available in genomic DNA for binding because they do not require cytosine methylation for recognition. Furthermore, it is unlikely that these 14-bp sites are interrupted by introns because this site is uninterrupted in the highly homologous exon of the human plasminogen gene although the kringle-4 DNA domain in the plasminogen gene is interrupted by intron sequences 33 bp in front of and 113 bp after the MDBP site (15). An important biological role for these sites is consistent with our finding that MDBP complexes with these sites were long-lived. The half-time for dissociation was ~50 min, as determined by incubation of ^{32}P-labeled apo(a) oligonucleotide duplexes with MDBP for 60 min, followed by addition of 1 pmol of unlabeled pB site 1 and removal of samples at various time intervals for band shift assays (21).

Conclusions. Such multiple intragenic repeats of sequence-specific DNA-binding sites have not been reported previously for any gene. In addition to the one human apolipoprotein(a) cDNA (16), the only apolipoprotein(a)-encoding DNA whose sequence has been reported is that of rhesus monkeys. This cDNA shows a similar high density of MDBP sites and kringle-4 units (23). Rhesus plasminogen cDNA, as well as the analogous human cDNA, contains a single kringle-4 unit (15, 23), in which one MDBP-specific apo(a)1 site is found. Indeed, it has been proposed that the apolipoprotein(a) gene evolved from the nearby plasminogen gene (1, 14) during primate evolution via rearrangements including multiple iterations of the kringle-4 domain (13, 16, 24).

The functional significance of the apolipoprotein(a) gene having so many intragenic copies of MDBP sites remains to be determined. However, a role for these sites in transcription control is strongly suggested by the occurrence of a consensus sequence for another enhancer-binding protein, AP-2, in 29 of the human apolipoprotein(a) kringle-4 units that contain MDBP sites in the sequenced cDNA. This sequence, 5′-GCCTGGGG-3′, which has been identified as a binding site (18) for the phorbol ester- and cyclic AMP-inducible AP-2 (9, 10), is found only 9 bp, or approximately one turn of the B-DNA helix, from the MDBP site (Fig. 3). If MDBP helps control apolipoprotein(a) gene expression, it might be acting from a distance, as in silencers or enhancers, on the transcription complex in concert with other sequence-specific DNA-binding proteins, such as AP-2. Alternatively, it might help organize chromatin structure. Some transcription factors control transcription positively from certain DNA regions and negatively from others (5, 19, 22). In the latter case, they might be engaged in nonproductive interactions with positively acting factors. If MDBP can down modulate transcription from within the apolipoprotein(a) gene, more MDBP molecules binding over a longer region of chromatin in the larger alleles could help explain the observed inverse relationship between the molecular mass of apolipoprotein(a) and plasma levels of this glycoprotein.

This research was supported in part by Public Health Service grant GM33999 from the National Institutes of Health.

LITERATURE CITED

1. **Drayna, D. T., R. A. Hegele, P. E. Hass, M. Emi, L. L. Wu, D. Eaton, R. M. Lawn, R. R. Williams, R. L. White, and J. M. Lalouel.** 1988. Genetic linkage between lipoprotein(a) phenotype and a DNA polymorphism in the plasminogen gene. Genomics **3**:230–236.
2. **Ehrlich, M., and R. Y.-H. Wang.** 1981. 5-Methylcytosine in eukaryotic DNA. Science **212**:1350–1357.
3. **Forsgren, M., B. Raden, M. Israelsson, K. Larsson, and L.-O. Heden.** 1987. Molecular cloning and characterization of a full-length cDNA clone for human plasminogen. FEBS Lett. **213**: 254–260.
4. **Gama-Sosa, M. A., R. M. Midgett, V. A. Slagel, S. Githens, K. C. Kuo, C. W. Gehrke, and M. Ehrlich.** 1983. Tissue-specific differences in DNA methylation in various mammals. Biochim. Biophys. Acta **740**:212–219.
5. **Glass, C. K., J. M. Holloway, O. V. Devary, and M. G. Rosenfeld.** 1988. The thyroid hormone receptor binds with opposite transcriptional effects to a common sequence motif in thyroid hormone and estrogen response elements. Cell **54**:313–323.
6. **Hajjar, K. A., D. Gavish, J. L. Breslow, and R. L. Nachman.** 1989. Lipoprotein(a) modulation of endothelial cell surface fibrinolysis and its potential role in atherosclerosis. Nature (London) **339**:303–304.
7. **Hixson, J. E., M. L. Britten, G. S. Manis, and D. L. Rainwater.** 1989. Apolipoprotein(a) (apo(a)) glycoprotein isoforms result from size differences in apo(a) mRNA in baboons. J. Biol. Chem. **264**:6013–6016.
8. **Huang, L.-H., R. Wang, M. A. Gama-Sosa, S. Shenoy, and M. Ehrlich.** 1984. A protein from human placental nuclei binds preferentially to 5-methylcytosine-rich DNA. Nature (London) **308**:293–295.

9. **Hyman, S. E., M. Comb, J. Pearlberg, and H. M. Goodman.** 1989. An AP-2 element acts synergistically with the cyclic AMP- and phorbol ester-inducible enhancer of the human proenkephalin gene. Mol. Cell. Biol. **9**:321–324.
10. **Imagawa, M., R. Chiu, and M. Karin.** 1987. Transcription factor AP-2 mediates induction by two different signal-transduction pathways: protein kinase C and cAMP. Cell **51**:251–260.
11. **Khan, R., X.-Y Zhang, P. C. Supakar, K. C. Ehrlich, and M. Ehrlich.** 1988. Human methylated DNA-binding protein: determinants of a pBR322 recognition site. J. Biol. Chem. **263**:14374–14383.
12. **Koschinsky, M. L., U. Beisiegel, D. Henne-Bruns, D. L. Eaton, and R. M. Lawn.** 1990. Apolipoprotein(a) size heterogeneity is related to variable number of repeat sequences in its mRNA. Biochemistry. **29**:640–644.
13. **Kratzin, H., V. W. Armstrong, M. Niehaus, N. Hilschmann, and D. Seidel.** 1987. Structural relationship of an apolipoprotein(a) phenotype (570 kDa) to plasminogen: homologous kringle domains are linked by carbohydrate-rich regions. Biol. Chem. Hoppe-Seyler **368**:1533–1544.
14. **Lindahl, G., E. Gersdorf, H. J. Menzel, C. Duba, H. Cleve, S. Humphries, and G. Utermann.** 1989. The gene for the Lp(a)-specific glycoprotein is closely linked to the gene for plasminogen on chromosome 6. Hum. Genet. **81**:149–152.
15. **Malinowski, D. P., J. E. Sadler, and E. W. Davie.** 1984. Characterization of a complementary deoxyribonucleic acid coding for human and bovine plasminogen. Biochemistry **23**: 4243–4250.
16. **McLean, J. W., J. E. Tomlinson, W.-J Kuang, D. L. Eaton, E. Y. Chen, G. M. Fless, A. M. Scanu, and R. M. Lawn.** 1987. cDNA sequence of human apolipoprotein(a) is homologous to plasminogen. Nature (London) **330**:132–137.
17. **Miles, L. A., G. M. Fless, E. G. Levin, A. M. Scanu, and E. F. Plow.** 1989. A potential basis for the thrombotic risks associated with lipoprotein(a). Nature (London) **339**:301–303.
18. **Mitchell, P. J., C. Wang, and R. Tjian.** 1987. Positive and negative regulation of transcription in vitro: enhancer-binding protein AP-2 is inhibited by SV40 T antigen. Cell **50**:847–861.
19. **Shore, D., and K. Nasmyth.** 1987. Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. Cell **51**:721–732.
20. **Supakar, P. C., D. Weist, D. Zhang, N. Inamdar, X.-Y. Zhang, R. Khan, K. C. Ehrlich, and M. Ehrlich.** 1988. Methylated DNA-binding protein is present in various mammalian cell types. Nucleic Acids Res. **16**:8029–8044.
21. **Supakar, P. C., X.-Y. Zhang, S. Githens, R. Khan, K. C. Ehrlich, and M. Ehrlich.** 1989. How different DNA sequences are recognized by a DNA-binding protein: effects of partial proteolysis. Nucleic Acids Res. **17**:8611–8629.
22. **Thierry, F., and M. Yaniv.** 1987. The BPV-E2 trans-acting protein can be either an activator or a repressor of the HPV18 regulatory region. EMBO J. **6**:3391–3397.
23. **Tomlinson, J. E., J. W. McLean, and R. M. Lawn.** 1989. Rhesus monkey apolipoprotein(a). J. Biol. Chem. **264**:5957–5965.
24. **Utermann, G.** 1989. The mysteries of lipoprotein(a). Science **246**:904–910.
25. **Zhang, D., K. C. Ehrlich, P. C. Supakar, and M. Ehrlich.** 1989. A plant DNA-binding protein that recognizes 5-methylcytosine residues. Mol. Cell. Biol. **9**:1351–1356.
26. **Zhang, X.-Y., P. C. Supakar, R. Khan, K. C. Ehrlich, and M. Ehrlich.** 1989. Related sites in human and herpes virus DNA recognized by methylated DNA-binding protein from human placenta. Nucleic Acids Res. **17**:1459–1474.