Am. J. Hum. Genet. 60:1354–1362, 1997

Characterization of FMR1 Promoter Elements by In Vivo–Footprinting Analysis

Sabine Schwemmle,[1] Esther de Graaff,[2] Heidrun Deissler,[3] Dieter Gläser,[1] Doris Wöhrle,[1] Ingo Kennerknecht,[1] Walter Just,[1] Ben A. Oostra,[2] Walter Dörfler,[3] Walther Vogel,[1] and Peter Steinbach[1]

[1]Abteilung Medizinische Genetik, Universität Ulm, Ulm, Germany; [2]Department of Clinical Genetics, Erasmus University, Rotterdam; and [3]Institut für Genetik, Universität zu Köln, Köln, Germany

Summary

Fragile X syndrome is associated with silencing of the FMR1 gene. We studied the transcriptional regulation, by analysis of the FMR1 promoter region for the presence of in vivo protein/DNA interactions and for cytosine methylation at the single-nucleotide level. Four protein-binding sites were present in the unmethylated promoter of the active FMR1 gene. In the methylated promoter of inactive genes no footprints were detected, and no evidence of active repression was found in the region investigated. We propose that the silencing of FMR1 gene transcription results from a lack of transcription-factor binding.

Introduction

The transcriptional activity of a gene is regulated by numerous transacting protein factors that interact with specific *cis*-acting elements (La Thangue and Rigby 1988; Wasylyk 1988; Latchman 1991). Changes in these interactions occur during the transition from an inactive to an active state and often are accompanied by alterations in cytosine methylation (Ehrlich and Ehrlich 1993; Graessmann and Graessmann 1993). Therefore, methylation of control regions in the genome is thought to play a critical role in the regulation of gene expression.

In fragile X syndrome, full expansion of the FMR1 CGG repeat usually is associated with methylation and gene inactivation (Bell et al. 1991; Oberlé et al. 1991; Pieretti et al. 1991; Verkerk et al. 1991; Sutcliffe et al. 1992; Verheij et al. 1993). Since this phenomenon is found in affected individuals only, hypermethylation of the FMR1 CpG island may represent an alteration of normal regulation. Analysis of the active gene is a prerequisite to the understanding of the dysfunctional regulation in the fragile X patient. A first definition of the 5′ FMR1 regulatory region was obtained on the basis of reporter-gene analysis (Hwu et al. 1993) and from a transgenic-model system (Hegersberg et al. 1995). Proteins binding to synthetic CGG repeat sequences were detected and were characterized, by in vitro gel-shift assays (Richards et al. 1993), but the functional relevance of these findings is not clear. Data on methylation of the FMR1 promoter still are limited, since only single recognition sites of methylation-sensitive restriction endonucleases have been analyzed (Bell et al. 1991; Oberlé et al. 1991; Pieretti et al. 1991; Vincent et al. 1991; Hansen et al. 1992; Sutcliffe et al. 1992; Richards et al. 1993). Thus, an interpretation of the methylation status of the complete promoter region is not possible. Previous methylation analysis at the single-nucleotide level concerned the repeat itself and its surrounding region but included only a small segment of the promoter (Hornstra et al. 1993).

The aim of the present work was to study the transcriptional regulation of the FMR1 gene and the interferences in regulation by mutation and/or by methylation, in the intact living cell. Therefore, we characterized the in vivo interaction of transcription factors with the FMR1 promoter and evaluated the DNA methylation of the corresponding region. Our data demonstrate the involvement of methylation in the process of FMR1 gene regulation. Footprinting analysis of the methylated promoter suggests that gene silencing is characterized by lack of protein binding.

Material and Methods

Cell Lines

Human fibroblast cell lines were established from skin biopsies taken from adult normal males, transmitting males, and fragile X patients. Embryonic cell lines were derived from a 24-wk-old full-mutation embryo. All cell cultures were maintained in Dulbecco's Modified Eagle's Medium (DMEM) or in RPMI 1640, supplemented with 10% FCS and 2 mM of L-glutamine.

Received June 26, 1996; accepted for publication March 19, 1997.
Address for correspondence and reprints: Dr. Sabine Schwemmle, Abteilung Medizinische Genetik, Universität Ulm, Parkstrasse 11, 89073 Ulm, Germany.
© 1997 by The American Society of Human Genetics. All rights reserved.
0002-9297/97/6006-0013$02.00

Southern Blot Analysis and Reverse Transcriptase–PCR (RT-PCR) Analysis

Genomic DNA was extracted from cultured cells, was cleaved with restriction endonuclease *Eag*I, *Eco*RI, or *Pst*I (New England Biolabs), was separated on 0.8% agarose gels, was blotted onto Hybond N^+ membranes, and was hybridized with DNA probe Ox1.9 or Ox0.55 (Nakahori et al. 1991). Probes were radiolabeled by use of the random-priming method (Feinberg and Vogelstein 1983). Poly$(A)^+$ RNA was purified by oligo(dT) cellulose chromatography, was transcribed into cDNA, and subsequently was amplified by use of primers 4924 and 4925 (for FMR1) and primers 243 and 244 (for the hypoxanthine guanine phosphoribosyltransferase [HPRT] gene), according to the protocol described in the study by Pieretti et al. (1991). Aliquots of the PCR reaction were added to a loading-dye mix and were electrophoresed in an 80-V constant field, in 2% agarose gels.

Genomic Sequencing of DNA by Ligation-Mediated PCR (LMPCR)

Genomic DNA was isolated according to standard procedures. After cleavage with *Eco*RI, aliquots of 25–50 μg of DNA were subjected to Maxam-Gilbert sequencing reactions. Chemical sequencing was performed as described in the study by Pfeifer and Riggs (1993), and LMPCR was performed according to the protocol described in the studies by Mueller and Wold (1989) and Pfeifer et al. (1989). For first-strand synthesis with Sequenase, 3 μg of cleaved genomic DNA, 0.3 pmol of primer I, and 3 μl of 5× Sequenase buffer were mixed, and water was added, bringing the total volume to 15 μl. Denaturation at 95°C for 5 min was followed by primer annealing at 45°C for 30 min. Then the probes were cooled on ice, and 7.5 μl of a freshly prepared Mg-dNTP mix (20 mM of $MgCl_2$, 20 mM of DTT, and 0.25 mM of each dNTP) and 1.5 μl of a 1:4 dilution of Sequenase (version 2; United States Biochemical) were added. The mixture was incubated at 46°C for 15 min. After the reaction was stopped (by cooling, the addition of 6 μl of 310 mM Tris-HCl [pH 7.7], and incubation at 67°C for 15 min), 45 μl of freshly prepared ligation mix (13 mM of $MgCl_2$, 30 mM of DTT, 1.7 mM of ATP, 80 μg of BSA/ml, and 100 pmol of linker/reaction) and 3 Weiss units of T4 DNA ligase (Promega) were added. Tubes were incubated overnight at 17°C. After ligation, the reaction was stopped by heating of the mixture to 70°C for 10 min, and the samples were precipitated with 10 μg of carrier tRNA. For PCR amplification, the sample was suspended in 50 μl of water, and 50 μl of *Taq*-polymerase mix was added (final concentrations in a 100-μl PCR reaction were as follows: 1 × *Taq* buffer, 2 mM of $MgCl_2$, 0.25 mM of dNTP, 10 pmol of primer II, 10 pmol of the longer oligomer of the linker, and 3 units of *Taq* polymerase [Perkin Elmer Cetus]). Samples were covered with mineral oil, and PCR was performed. After initial denaturation at 95°C for 3 min, the samples were denatured at 95°C for 1 min, were annealed at 66°C (all the different primers II that were used worked at this temperature), and were extended at 76°C for 3 min. With each cycle the extension time was increased for 5 s. After 20 cycles, the sample was incubated at 76°C, and 5 μl of booster solution (containing diluted *Taq*-polymerase mix plus 1 unit of fresh *Taq* polymerase per sample) was added. After 10 min of incubation, samples were placed on ice, and 1–5 pmol of end-labeled primer III, 2.5 units of *Taq* polymerase, and 20 nmol of each dNTP were added. After initial denaturation at 94°C for 3 min, samples were cycled twice as follows: 94°C for 1 min; 69°C–72°C, depending on the melting temperature of primer III, for 2 min; and 76°C for 10 min. Then, polymerase activity was stopped, and the samples were extracted with phenol/chloroform, were precipitated, and were resuspended in loading dye. Half the reaction was separated on a sequencing gel and was visualized by autoradiography. The control reactions were done with the cloned 5.2-kb *Eco*RI fragment of the FMR1 gene (Nakahori et al. 1991). For analysis of the upper DNA strand, three primer sets, A, B, and C, were used. These sets included the following sequences: AI, 5′-CGCCCGCTCAGAGGC-3′; AII, 5′ CAGAGG-CGGCCCTCCACCGGAA-3′; AIII, 5′-AGTGAA-ACCGAAACGGAGCTGAGCGCCT-3′; BI, 5′-ACCGAAACGGAGCTGAG-3′; BII, 5′-GAGCTG-AGCGCCTGACTAGGGCCGAA-3′; BIII, 5′-ACC-ACGTCACGTGATCAACGCTGTTCCCTC3′; CI, 5′-TCTCTCTTCAAGTGGCCTGG-3′; CII, 5′-GCA-TGCGCGCTGCTGGGAACC-3′; and CIII, 5′-CCG-GGGTGCCGGGTCGAAAGACAGACG-3′. For analysis of the lower strand, primer sets D, E, and F were used. These sets included the following sequences: DI, 5′-GAATCCCAGAGAGGCCGAACTG-3′; DII, 5′-GGCCGAACTGGGATAACCGGATGCA-3′; DIII, 5′-GGATGCATTTGATTTCCCACGCCACTG-AGTG-3′; EI, 5′-CCGCCCTCCACCAAG-3′; EII, 5′-CGCGTCTGTCTTTCGACCCGGCA-3′; EIII, 5′-GGCCGGTTCCCAGCAGCGCGCATG-3′; FI, 5′-TCAGTGTTTACACCCGCAGC-3′; FII, 5′-CCT-AGTCAGGCGCTCAGCTCCGTTTC-3′; and FIII, 5′-CAGCTCCGTTTCGGTTTCACTTCCGGTGG-3′.

In Vivo Dimethyl Sulfate (DMS)–Footprinting Analysis

The cultured cells were washed twice with PBS and subsequently were treated with 0.1% DMS in serum-free DMEM medium for 8 min at room temperature. Then, the DMS-containing medium was removed quickly, and the cells were washed with ice-cold PBS. Subsequently, DNA was extracted by standard proce-

dures, was treated with piperidine, and, finally, was subjected to LMPCR as described above.

Results

In Vivo Protein-Binding Studies

Analysis of in vivo protein/DNA interaction was performed by use of DMS footprinting by ligation-mediated PCR (Mueller and Wold 1989; Pfeifer et al. 1989). Adult human fibroblast cell lines and a fetal fibroblast cell line were investigated. All cell lines were characterized for a fragile X mutation, by Southern blot analysis, and for the transcriptional activity of the genes FMR1 and HPRT, by RT-PCR, as illustrated in figures 1 and 2, respectively. The lines were derived from adult male controls (fig. 1*a* and *b*, lanes 2 and 3), from two transmitting males—one showing expansions in the pre- and full-mutation ranges, with partial methylation of the largely expanded fragments at the *Eag*I restriction site (fig. 1*a* and *b*, lane 5), and the other presenting only an unmethylated premutation (fig. 1*a* and *b*, lane 4)—from two adult full-mutation males (fig. 1*a* and *b*, lanes 7 and 8), and from one male fetus (fig. 1, lane 6) with largely expanded and methylated fragments. Analysis of the cell lines is summarized in table 1.

In order to define the promoter and upstream-promoter elements of the FMR1 gene, we investigated the 300 nucleotides upstream of the cap site, for the presence of in vivo protein/DNA interactions. In the normal FMR1 gene, four footprints, designated "I," "II," "III," and "IV" (fig. 3), were present. The positions of the protein-binding sites are given in figure 4. The position numbers of figure 4 correspond to file HSFXDNA of GenBank (accession X61378), which contains the sequence originally reported in the study by Fu et al. (1991). Site I is close to the presumed TATA box–like sequence element (Hwu et al. 1993) and shows a palindromic sequence motif. In this palindrome, two G residues are protected on either strand. The sequences of the protein-binding site are 5′-CACGTG-3′ and 3′-GTGCAC-5′, with the protected G's indicated by underlining.

The second protein-binding site, 5′-GGGGGAGG-3′, shows protection of a cluster of six adjacent guanosines (positions 2517–2521 and 2523) at the upper DNA strand. A potential binding site for the transcription factor Sp1 is given with footprint III, showing the consensus sequence 5′-GGGCGG-3′ at the upper strand (positions 2490–2495). The most upstream-located and also palindromic *cis*-acting element shows protected G residues on both DNA strands. The sequences are 5′-GCGCATGCGC-3′ and 3′-CGCGTACGCG-5′. In this case, the exact 5′ boundaries of protection could not be identified on either strand. However, the corresponding

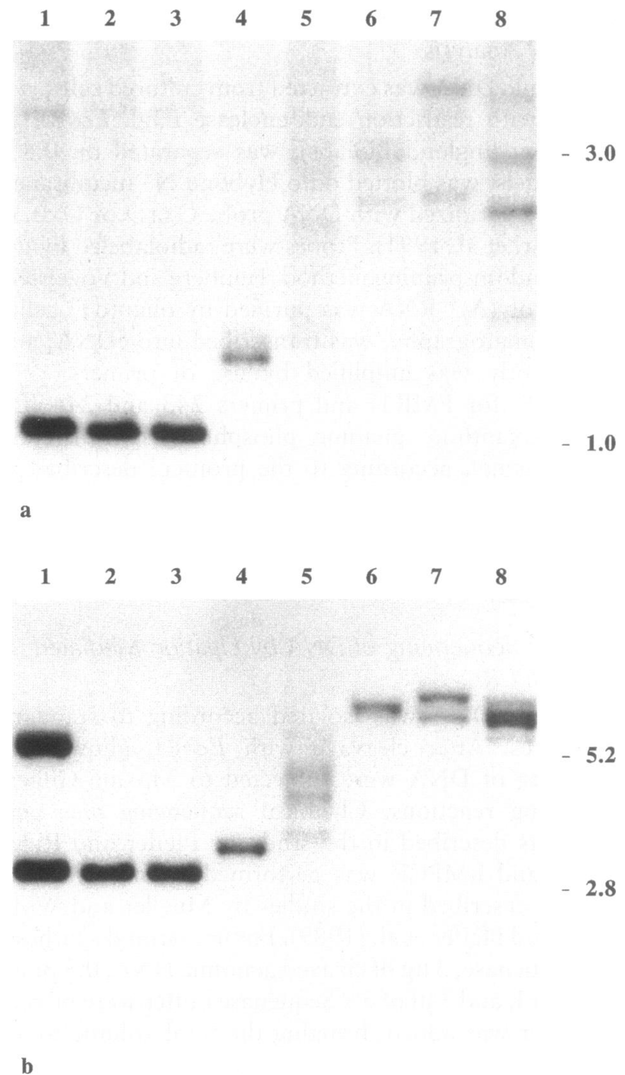

Figure 1 Southern blot analysis of DNA samples from a control female and from normal, transmitting, and full-mutation males, digested with *Pst*I (*a*) and *Eco*RI + *Eag*I (*b*). The filters were hybridized with probes Ox0.55 (*a*) and Ox1.9 (*b*). Lane 1, Control female with fragments within the normal size range (*a*) and with a methylated inactive X (5.2-kb band) and an unmethylated active X (2.8-kb band), at the *Eag*I site (*b*). Lanes 2 and 3, Normal males with unmethylated fragments of normal size. Lanes 4 and 5, Transmitting males. Lane 6, Embryonic cell line from a male fetus, showing a full mutation. Lanes 7 and 8, Cell lines from adult full-mutation males.

3′ end on the opposite DNA strand could be defined. With repeated experiments, a region of approximately eight bases adjacent to the boundaries of this footprint appeared very faint on one DNA strand and thus did not permit evaluation of the degree of protection of the G residues. This also is true for the other Maxam-Gilbert base-specific control reactions (G, G/A, C/T, and C). Therefore, although located within or near a protein-binding site, this phenomenon cannot be due to protein binding. A similar observation of the 5′

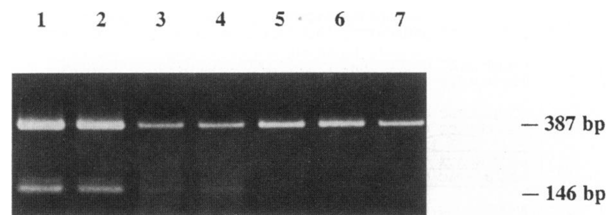

Figure 2 RT-PCR products from normal males (lanes 1 and 2), from carriers (lanes 3 and 4), and from fragile X patients (lanes 5 and 6, adult tissue; lane 7, fetal tissue). The larger band (387 bp) corresponds to the amplification product of the HPRT gene. The smaller band (146 bp) corresponds to amplification of the FMR1 transcript.

flanking region upstream of the CGG repeat has been described by Hornstra et al. (1993), who discussed methodical problems and particular structural alterations in this gene region as possible causes. No indications of further protein/DNA interaction were detected in our analysis of the neighboring upstream DNA region. A scheme summarizing the footprinting data is presented in figure 4.

Footprinting data of the two transmitting males who were investigated showed some individual differences. One of these nonretarded males was typical in that he presented, on *Pst*I and *Eco*RI + *Eag*I digests, restriction fragments in only the premutation size range and no evidence of *Eag*I-site methylation (fig. 1*a* and *b,* lane 4). This male showed protein/DNA interactions representing the same binding pattern that was seen in the normal FMR1 gene (fig. 5, lane 3). The other transmitting male presented a complex somatic mosaic consisting of repeats with length variation reaching from the premutation range to the full-mutation range. *Eag*I site–methylation analysis of this male revealed that probably the majority of expansions were unmethylated (fig. 1*a* and *b,* lane 5). In this particular case, footprint data were not that clear and may represent incomplete protection.

Footprints were absent in the full-mutation males—two adults and one fragile X fetus—who were investigated (fig. 3, lanes 3 and 4; fig. 5, lanes 1 and 2). In these cases, RT-PCR confirmed the absence of FMR1 transcripts.

Table 1

Summary of Results Showing the Correlation between in Vivo Protein Binding and Methylation Status

			FMR1 Gene Methylation, as Determined by		
Patient and Sample	Mutation Range[a]	FMR1 Gene Expression[b]	*Eag*I-Site Analysis	Genomic Sequencing	In Vivo Footprints
Male fragile X fetus:					
Fibroblast cell line	Full	−	+	+	−
Fragile X male:					
Fibroblast cell line	Full	−	+	+	−
Fragile X male:					
Fibroblast cell line	Full	−	+	+[c]	−
Transmitting male:[d]					
Fibroblast cell line	Pre/full	+	−/+	No data	+
Transmitting male:					
Fibroblast cell line	Pre	+	−	−	+
Normal male:					
Fibroblast cell line	...	+	−	−	+
Normal male:					
Fibroblast cell line	...	+	−	−	+
Fragile X male:[e]					
Different tissues	Full	−	+	No data	No data
Tissue of lung tumor	Pre	+	+	+	No data

Note.—A plus sign (+) indicates the presence of gene expression, methylation, and/or in vivo footprints; and a minus sign (−) indicates the absence of gene expression, methylation, and/or in vivo footprints.

[a] Data from Southern blot analysis.

[b] Transcriptional activity assayed by RT-PCR.

[c] Most CpG dinucleotides were methylated, and single CpGs remained unmethylated.

[d] Presented with a mosaic pattern of pre- and fully mutated, methylated and unmethylated fragments.

[e] Mutation, expression (FMR1 protein detection by immunohistochemistry), and *Eag*I and *Bss*HII methylation data according to Smeets et al. (1995); note that only native tissues (no cell lines) were investigated.

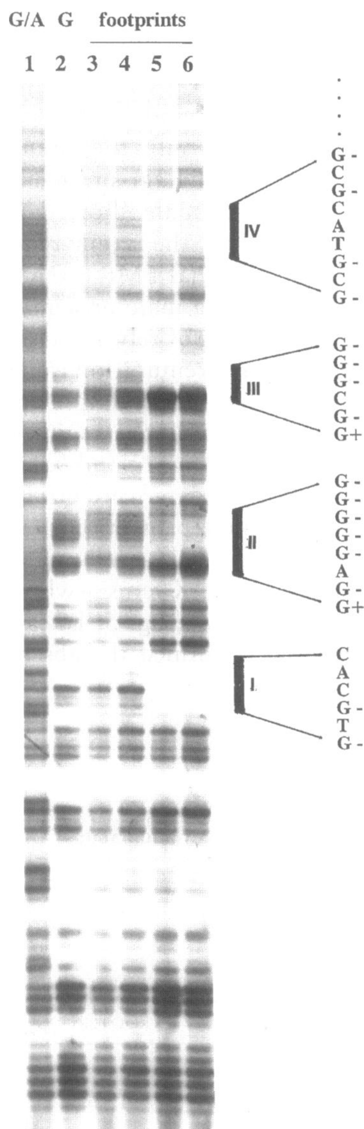

Figure 3 In vivo DMS footprinting of the upper DNA strand of the FMR1 promoter region. Lanes 1 (G/A) and 2 (G), Control reactions of naked DNA. Lane 3, Fibroblast cell line of an adult fragile X patient. Lane 4, Fibroblast cell line of a fetus with full mutation. Lane 5, Fibroblast cell line of a normal male. Lane 6, Cell line of another phenotypically normal male. The positions of the footprints (lanes 5 and 6) are indicated by vertical bars, and the corresponding sequences are shown to the right of each bar (a minus sign [−] denotes a protected G residue and a plus sign [+] denotes a visible G residue). Primer set A was used.

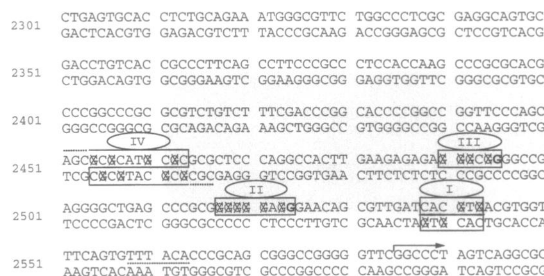

Figure 4 Sequence of the investigated promoter region of the FMR1 gene. The position numbers correspond to sequence file HSFXDNA of GenBank (accession X61378). The sites of protein protection, footprints I, II, III, and IV, are indicated by boxes (crossed-out G's denote protected G residues and boldface G's denote visible G residues). The exact 5′ boundaries (indicated by dotted lines) of footprint IV could not be determined on either strand. The TATA-like box (position 2558–2563) described in the study by Hwu et al. (1993) is indicated by a dotted line, and the putative transcription start site is indicated by an arrow.

DNA Methylation Analysis

With methylation analysis by LMPCR-aided genomic sequencing, the methylation status of each individual cytosine residue can be defined. Compared with unmethylated control samples (plasmid-cloned DNA containing the region of interest), methylated cytosines result in the absence of single bands within the cytosine-specific sequencing ladder. We primarily were interested in the methylation status of the FMR1 control region. Genomic DNA from normal males, transmitting males, and affected males (from adult as well as from fetal tissues) was subjected to methylation analysis. The cytosine-specific sequencing lanes of normal males and normal transmitting males (fig. 6, lane 4) displayed a continuous

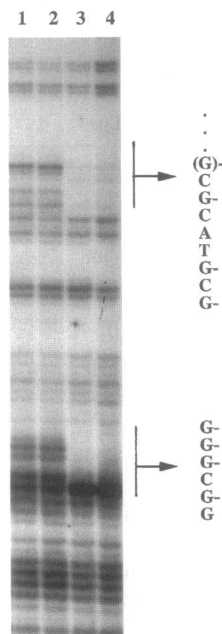

Figure 5 In vivo DMS footprinting of the FMR1 promoter region of the upper DNA strand, including footprint III (*bottom*) and footprint IV (*top*). G− = protected G residues; G = visible G residues; and (G) = G residues located in the region of faint bands. Lane 1, Fibroblast cell line of an adult fragile X patient. Lane 2, Fibroblast cell line of a fetus with full mutation. Lanes 3 and 4, Fibroblast cell lines of two normal males. Primer set B was used.

1 2 3 4

GGGGCTGAGCCCGCGGGGGGAGGGAACAGCGTTGATCACGTGACGTGGTTTCAGTGTTTACACC

Figure 6 Genomic sequencing and methylation analysis of part of the FMR1 promoter region (nucleotide positions 2502–2565). The methylation state of five CpGs (indicated by blackened circles) in the C tracks of different genomic DNA samples are shown (lanes 1–4). Lanes 1 and 2, Lung tumor of a male fragile X patient. Lane 3, Fibroblast cells of a male with full mutation, showing both methylated and unmethylated CpGs. Lane 4, Fibroblast cells of a normal transmitting male. Unblackened circles indicate other C residues. Vertical lines to the right of the sequence indicate footprint I (*bottom*) and footprint II (*top*). Primer set A was used.

ladder of single bands. This holds true for the entire region that was investigated and for both the upper and lower DNA strands, suggesting that all CpGs were unmethylated in these phenotypically normal males. In two fragile X patients, there was an absence of signals in all 5′-CG-3′ dinucleotide positions of the FMR1 promoter on either DNA strand, suggesting the presence of methylcytosine residues in all 5′-CG-3′ positions. No difference in the methylation pattern was detected between the DNA samples from the adult and the fetal fragile X patients. However, a fibroblast cell line of one adult, affected male (fig. 6, lane 3) revealed a combination of methylated and unmethylated CpG residues. Interestingly, during proliferation of this cell line, changes were observed in both the mutation and the methylation pattern, which is a phenomenon that is investigated further. The methylation data of the investigated FMR1 promoter region are summarized in figure 7.

Data on methylation and expression of the FMR1 gene in neoplastic tissue were first reported in a male fragile X patient with lung cancer (De Graaff et al. 1995). The premutated FMR1 gene of this patient was expressed in the tumor tissue instead of methylation at the *Eag*I and *Bss*HII sites. Since methylation at these two sites does not necessarily imply methylation of the complete promoter region, we studied the entire region by genomic sequencing, using the DNA sample originally isolated from the same piece of tumor tissue that was characterized previously by Southern blot analysis and by immunohistochemistry. We had to restrict our investigation to methylation analysis of the upper DNA strand, because of the limited amount of tumor DNA available. All CpG dinucleotides were methylated (fig. 6, lanes 1 and 2).

Discussion

The phenotype of fragile X syndrome is assumed to be caused by absence of the FMR1 protein. A large

```
2251  CCCAGAGAGG CCGAACTGGG ATAACCGGAT GCATTTGATT TCCCACGCCA
      GGGTCTCTCC GGCTTGACCC TATTGGCCTA CGTAAACTAA AGGGTGCGGT

2301  CTGAGTGCAC CTCTGCAGAA ATGGGCGTTC TGGCCCTCGC GAGGCAGTGC
      GACTCACGTG GAGACGTCTT TACCCGCAAG ACCGGGAGCG CTCCGTCACG

2351  GACCTGTCAC CGCCCTTCAG CCTTCCCGCC CTCCACCAAG CCCGCGCACG
      CTGGACAGTG GCGGGAAGTC GGAAGGGCGG GAGGTGGTTC GGGCGCGTGC
                                            E
2401  CCCGGCCCGC GCGTCTGTCT TTCGACCCGG CACCCCGGCC GGTTCCCAGC
      GGGCCGGGCG CGCAGACAGA AAGCTGGGCC GTGGGGCCGG CCAAGGGTCG
        B          B
2451  AGCGCGCATG CGCGCGCTCC CAGGCCACTT GAAGAGAGAG GGCGGGGCCG
      TCGCGCGTAC GCGCGCGAGG GTCCGGTGAA CTTCTCTCTC CCGCCCCGGC
                    S
2501  AGGGGCTGAG CCCGCGGGGG GAGGGAACAG CGTTGATCAC GTGACGTGGT
      TCCCCGACTC GGGCGCCCCC CTCCCTTGTC GCAACTAGTG CACTGCACCA

2551  TTCAGTGTTT ACACCCGCAG CGGGCCGGGG GTTCGGCCCT AGTCAGGCGC
      AAGTCACAAA TGTGGGCGTC GCCCGGCCCC CAAGCCGGGA TCAGTCCGCG

2601  TCAGCTCCGT TTCGGTTTCA CTTCCGGTGG AGGGCCGCCT CTGAGCGGGC
      AGTCGAGGCA AAGCCAAAGT GAAGGCCACC TCCCGGCGGA GACTCGCCCG

2651  GGCGGGCCGA CGGCGAGCGC GGGCGGCGGC GGTGACGGAG GCGCCGCTGC
      CCGCCCGGCT GCCGCTCGCG CCCGCCGCCG CCACTGCCTC CGCGGCGACG

2701  CAGGGGGCGT GCGGCAGCGC GGCGGCGGCG GCGGCGGCGG CGGCGGCGGC
      GTCCCCCGCA CGCCGTCGCG CCGCCGCCGC CGCCGCCGCC GCCGCCGCCG
```

Figure 7 Summary of DNA cytosine-methylation data of the investigated FMR1 region (nucleotides 2270–2681). The position numbers correspond to sequence file HSFXDNA of GenBank (accession X61378). Shaded C's indicate methylated CpG sites detected in the adult patients and in the fetus with full mutation. Boxed C's indicate CpG sites that could not be analyzed because of their localization in the so-called faint region surrounding footprint IV. An asterisk (*) indicates methylated CpG sites, as previously described in the study by Hornstra et al. (1993). The boxed area (nucleotides 2301–2590) indicates nucleotides for which methylation analysis of both DNA strands was performed. S = *Sac*II restriction site; B = *Bss*HII restriction site; and E = *Eag*I restriction site.

majority of affected individuals show large expansions of the CGG repeat, methylation of the CpG island, and a lack of FMR1 transcripts (Bell et al. 1991; Oberlé et al. 1991; Steinbach et al. 1993). This indicates that the major control point for the lack of gene expression lies at the level of transcriptional regulation. Therefore, we have studied methylation and in vivo protein binding in the FMR1 promoter.

Promoter activity previously has been assigned to the DNA segment 5′ to the CGG repeat, by CAT (chloramphenicol acetyltransferase gene) assay and transient gene expression (Hwu et al. 1993; Hegersberg et al. 1995). Nuclear proteins have been found to bind synthetic p(CGG)n oligonucleotides and a 430-bp DNA fragment of the CpG island containing 30 CGG repeats (Richards et al. 1993; Zhong et al. 1995), but the functional relevance of these findings remained unclear. We detected four actual protein-binding sites in the FMR1 promoter, using the in vivo–footprinting technique. These protein/DNA interactions were found in transcriptionally active genes of normal individuals and were absent from inactive genes of affected males. Hence, the binding sites represent *cis*-regulatory elements of the FMR1 promoter, and their sequences now can be used to identify the corresponding transcription factors. The DNA segments covered by the four observed footprints could be identified in transcription-factor databases. An imperfect Sp1-binding motif is identified in site II, followed by a perfect Sp1 consensus site (footprint III). Control element I in the FMR1 promoter presented the palindromic sequence CACGTG, which is a potential binding site for transcription factor c-myc (Ariga et al. 1989; Blackwell et al. 1990; Blackwood and Eisenman 1991; Dang et al. 1991). Binding site IV could be occupied by alpha pal or by Nrf1 (Jacob et al. 1989; Chan et al. 1993*a*, 1993*b*). There have been only limited data on in vitro gel-shift experiments with DNA fragments that include the four in vivo–defined binding sites. In a published abstract, Zhong et al. (1995) reported in vitro binding of transcription factors Sp1, Ap1, and CREB. As shown by our in vivo–footprinting experiments, however, only the binding of Sp1, most probably to the perfect consensus sequence GGGCGG (site III), is functionally relevant. Preliminary data from our own gel-shift experiments suggest that the imperfect Sp1-binding motif is not used by Sp1 (data not shown). A CREB consensus-sequence motif (TGACGT) is located adjacent to element I but did not contribute to the protein-binding pattern at this site. The same holds true for the Ap1-binding motif.

With genomic sequencing, we have analyzed DNA methylation of the CpG-rich promoter region, bridging the gap between the single, upstream-located restriction-enzyme recognition sites and the immediately adjacent repeat area. Our data seem to confirm that methylation of the FMR1 promoter region in males is restricted to full-mutation alleles. In two full-mutation patients, there was evidence for methylation of each individual CpG that was investigated. However, single CpG sites were not methylated in another patient with full mutation. Some of these unmethylated CpGs were located within a defined in vivo protein-binding site. This partial lack of methylation obviously did not allow either for normal protein binding or for transcriptional activity. Further genomic-sequencing analysis of partially methylated FMR1 promoters will help to detect common patterns of methylation and to identify those sites of CpG methylation that are functionally relevant. The fact that methylation is associated with full mutations and with inactivation of the FMR1 gene already has been known, and our data do not change this perception. We report that the absence of transcription-factor binding in the investigated promoter region of inactive genes could be related to methylation. The restriction sites of *Bss*HII, *Eag*I, and *Sac*II are located either within (*Bss*HII) or immediately adjacent to (*Eag*I and *Sac*II) regulatory element IV. They frequently are used in diagnostic methylation approaches but may not always be representative for the complete promoter.

In normal somatic cells, protein/DNA interaction at the identified regulatory elements was shown to be sensitive to methylation, since there were no in vivo footprints detectable in the methylated promoter regions. Thus, the repression of FMR1 transcription in fragile X syndrome and the inhibition of transcription-factor binding probably are related to DNA methylation. There were no alternative footprints found in the methylated promoter region that was investigated. Repression, therefore, probably does not result from binding of a particular repressor protein to the methylated sequences that were investigated here. Missing footprints in the methylated promoter might indicate methylation sensitivity of single, critical, or even most transcription factors involved in the regulation of FMR1 gene transcription. A possible exception is Sp1, which is known to be methylation insensitive (Harrington et al. 1988; Höller et al. 1988).

FMR1 gene regulation in the lung tumor could be different from the situation in the normal, nontransformed cell. Despite methylation at the *Eag*I and *Bss*HII sites, gene expression was detected in the tumor, containing a premutation of the FMR1 gene (De Graaff et al. 1995). Our genomic-sequencing data seem to support the evidence that this promoter is completely methylated, but we cannot be absolutely sure that our methylation data do represent the tumor cells expressing the premutated FMR1 gene. Since no culture of lung-tumor cells was available, footprint analysis could not be done and further questions concerning FMR1 gene regulation in this methylated tissue could not be addressed. Methyl-

ation insensitive isoforms or different sets of transacting proteins may be involved in transcriptional regulation in the transformed lung cells. The transcriptional activity of certain protein factors could be increased in tumor tissue, as a consequence of transformation. Further experimental analysis of cancer tissue in general will be necessary.

The major mutation affecting the silencing of the FMR1 gene is full expansion of the CGG repeat, but this is not a sufficient cause of promoter methylation, as indicated by the normally transcribed unmethylated full mutations found in some chorionic-villi samples (Sutcliffe et al. 1992), as well as in some nonretarded fragile X males (Smeets et al. 1995). The biological relevance of nuclear protein binding, in vitro, to p(CGG)n oligonucleotides is not clear yet. There also has been no in vivo–footprinting analysis demonstrating actual binding of protein to the FMR1 CGG repeat in the living cell. Our further experiments, therefore, focus on different functional aspects of methylation and on in vivo protein binding of the FMR1 CGG repeat.

Acknowledgments

The authors are grateful to Dr. G. Pfeifer (Department of Biology, Beckman Research Institute, City of Hope, Duarte, CA) for his valuable discussions and substantial help, during the revision of this paper. The technical assistance of Christine Klett, Heike Kozma, and Renate Weber also is gratefully acknowledged. This study was supported by the Deutsche Forschungsgemeinschaft.

References

Ariga H, Imamura Y, Iguchi-Ariga SM (1989) DNA replication origin and transcriptional enhancer in c-myc gene share the c-myc protein binding sequences. EMBO J 8: 4273–4279

Bell MV, Hirst MC, Nakahori Y, MacKinnon RN, Roche A, Flint TJ, Jacobs PA, et al (1991) Physical mapping across the fragile X: hypermethylation and clinical expression of the fragile X syndrome. Cell 64:861–866

Blackwell TK, Kretzner L, Blackwood EM, Eisenman RN, Weintraub H (1990) Sequence-specific DNA binding by the c-Myc protein. Science 250:1149–1151

Blackwood EM, Eisenman RN (1991) Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. Science 251:1211–1217

Chan JY, Han XL, Kan YW (1993*a*) Cloning of Nrf1, an NF-E2 related transcription factor, by genetic selection in yeast. Proc Natl Acad Sci USA 90:11371–11375

——— (1993*b*) Isolation of the cDNA encoding the human NF-E2 protein. Proc Natl Acad Sci USA 90:11366–11371

Dang CV, Barrett J, Villa-Garcia M, Resar LMS, Kato GJ, Fearon ER (1991) Intracellular leucine zipper interactions suggest c-Myc hetero-oligomerization. Mol Cell Biol 11: 954–962

De Graaff E, Willemsen R, Zhong N, de Die-Smulders CEM, Brown WT, Freling G, Oostra B (1995) Instability of the CGG repeat and expression of the *FMR1* protein in a male fragile X patient with a lung tumor. Am J Hum Genet 57: 609–618

Ehrlich M, Ehrlich KC (1993) Effect of DNA methylation on the binding of vertebrate and plant proteins to DNA. In: Jost JP, Saluz HP (eds) DNA methylation: molecular biology and biological significance. Birkhäuser Verlag, Basel, Boston, Berlin, pp 145–168

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Fu YH, Kuhl DPA, Pizzuti A, Pieretti M, Sutcliffe J, Richards S, Verkerk AJMH, et al (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell 67:1047–1058

Graessmann M, Graessmann A (1993) DNA methylation, chromatin structure, and the regulation of gene expression. In: Jost JP, Saluz HP (eds) DNA methylation: molecular biology and biological significance. Birkhäuser Verlag, Basel, Boston, Berlin, pp 404–424

Hansen RS, Gartler SU, Scott CR, Chen SH, Laird CD (1992) Methylation analysis of CGG sites in the CpG island of the human FMR1 gene. Hum Mol Genet 1:571–578

Harrington MA, Jones PA, Imagawa M, Karin M (1988) Cytosine methylation does not affect binding of transcription factor Sp1. Proc Natl Acad Sci USA 85:2066–2070

Hegersberg M, Matsuo K, Gassmann M, Schaffner W, Lüscher B, Rülicke T, Aguzzi A, et al (1995) Tissue-specific expression of a FMR1/beta-galactosidase fusion gene in transgenic mice. Hum Mol Genet 4:359–366

Höller M, Westin G, Jiricny J, Schaffner W (1988) Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. Genes Dev 2: 1127–1135

Hornstra IK, Nelson DL, Warren ST, Yang TP (1993) High resolution methylation analysis of the FMR1 gene trinucleotide repeat region in fragile X syndrome. Hum Mol Genet 2:1659–1665

Hwu WL, Lee YM, Lee SC, Wang TR (1993) In vitro DNA methylation inhibits FMR-1 promoter. Biochem Biophys Res Commun 193:324–329

Jacob WF, Silverman TA, Cohen RB, Safer B (1989) Identification and characterization of a novel transcription factor participating in the expression of eukaryotic initiation factor 2α. J Biol Chem 264:20372–20384

La Thangue NB, Rigby PWJ (1988) Trans-acting protein factors and the regulation of eukaryotic transcription. In: Hames B, Glover D (eds) Transcription and splicing. IRL Press, Oxford, pp 3–42

Latchman DS (1991) Eukaryotic transcription factors. Academic Press, London

Mueller PR, Wold B (1989) In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science 246: 780–786

Nakahori Y, Knight SJL, Holland J, Schwartz C, Roche A, Tarleton J, Wong S, et al (1991) Molecular heterogeneity of the fragile X syndrome. Nucleic Acids Res 19:4355–4359

Oberlé I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer

A, Boue J, et al (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science 252:1097–1102

Pfeifer GP, Riggs AD (1993) Genomic sequencing. In: Griffin H, Griffin A (eds) Methods in molecular biology. Vol 23: DNA sequencing protocols. Humana Press, Totowa, NJ, pp 169–181

Pfeifer GP, Steigerwald SD, Mueller PR, Wold B, Riggs AD (1989) Genomic sequencing and methylation analysis by ligation mediated PCR. Science 246:810–813

Pieretti M, Zhang FP, Fu YH, Warren ST, Oostra BA, Caskey CT, Nelson DL (1991) Absence of expression of the FMR-1 gene in fragile X syndrome. Cell 66:817–822

Richards RI, Holman K, Yu S, Sutherland GR (1993) Fragile X syndrome unstable element, p(CGG)n, and other simple tandem repeat sequences are binding sites for specific nuclear proteins. Hum Mol Genet 2:1429–1435

Smeets HJM, Smits APT, Verheij CE, Theelen JPG, Willemsen R, van de Burgt I, Hoogeven AT, et al (1995) Normal phenotype in two brothers with full FMR1 mutation. Hum Mol Genet 4:2103–2108

Steinbach P, Wöhrle D, Tariverdian G, Kennerknecht I, Barbi G, Edlinger H, Enders H, et al (1993) Molecular analysis of mutations in the FMR-1 segregating in fragile X families. Hum Genet 92:491–498

Sutcliffe JS, Nelson DL, Zhang F, Pieretti M, Caskey CT, Saxe D, Warren ST (1992) DNA methylation represses FMR-1 transcription in fragile X syndrome. Hum Mol Genet 1: 397–400

Verheij C, Bakker CE, de Graaff E, Keulemans J, Willemsen R, Verkerk AJ, Galjaard H, et al (1993) Characterization and localization of the FMR-1 gene product associated with fragile X syndrome. Nature 363:722–724

Verkerk AJMH, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DPA, Pizzuti A, Reiner O, et al (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65:905–914

Vincent A, Heitz D, Petit C, Kretz C, Oberlé I, Mandel JL (1991) Abnormal pattern detected in fragile X patients by pulsed field gel electrophoresis. Nature 349:624–626

Wasylyk B (1988) Enhancers and transcription factors in the control of gene expression. Biochim Biophys Acta 951:17–35

Zhong N, Curley D, Dobkin C, Brown WT (1995) Determination of transcription factors binding to the FMR1 promoter region. Am J Hum Genet Suppl 57:A155