Human Molecular Genetics, 1997, Vol. 6, No. 12 **2051–2060**

Structural and functional characterization of the human FMR1 promoter reveals similarities with the hnRNP-A2 promoter region

Régen Drouin[1,2], Martin Angers[1,2], Nancy Dallaire[1,2], Timothy M. Rose[3], Edouard W. Khandjian[1,2] and François Rousseau[1,2,*]

[1]Unité de Recherche en Génétique Humaine et Moléculaire,Centre de Recherche, Pavillon Saint-François d'Assise, Centre Hospitalier Universitaire de Québec, 10 rue de l'Espinay, Québec, Québec G1L 3L5, Canada, [2]Département de Biologie Médicale, Faculté de Médecine, Université Laval, Québec, Canada and [3]Department of Pathobiology, School of Public Health and Community Medicine, University of Washington, Seattle, WA 98195, USA

Received May 28, 1997; Revised and Accepted August 8, 1997

Fragile X mental retardation syndrome is associated with an expansion of a CGG repeat within the 5′UTR of the first exon of the FMR1 gene, abnormal methylation of the CpG island in the promoter region, and a transcriptional silencing of this gene. We studied transcriptional regulation of the FMR1 gene using protein footprint analysis of the active and inactive gene *in vivo*. We identified four footprints within the FMR1 promoter region which correspond to consensus binding sites of known transcription factors, α-PAL/NRF1, Sp1, H4TF1/Sp1-like and c-myc. These footprints were present in normal cells with a transcriptionally active FMR1 gene. The same footprints were present in different cell types: primary fibroblasts, lymphoblastoid cells and peripheral lymphocytes. However, for the 1.1 kb region analyzed, no footprints were detected in a variety of cell types derived from patients with fragile X syndrome which have a transcriptionally inactive FMR1 gene. A BLAST nucleotide search identified sequence similarities between the region of the FMR1 gene containing the footprints and an analogous region within the promoter region of the gene for the heterogeneous nuclear ribonucleoprotein (hnRNP) A2, a member of a family of ribonucleoproteins implicated in mRNA processing and nuclear-cytoplasm transport. The nucleotide sequences identified in the hnRNP-A2 promoter region correspond to the same consensus binding sites showing DNA–protein interactions in the FMR1 gene. Our previous functional studies and the studies of others demonstrate that FMR proteins, like hnRNP-A2, are also ribonucleoproteins which appear to be involved in mRNA transport. The results from our footprint studies suggest that the expression of the FMR1 gene is regulated by the binding of specific transcription factors to sequence elements in the 5′ region of the gene and that this expression may be regulated by elements in common with the hnRNP-A2 gene. Common regulation of these two genes might play an important role in the cooperative processing and transport of mRNA from the nucleus to the translation machinery.

INTRODUCTION

The fragile X syndrome, a leading cause of inherited mental retardation (1) is associated with an unstable expansion of a polymorphic CGG trinucleotide repeat array localised in the 5′ untranslated region (UTR) of the first exon of the FMR1 gene (2–4). This gene spans 38 kb and comprises 17 exons (5) and a 5′ CpG island of ~1 kb which encompasses the triplet repeats (2). The expansion of the CGG repeats is associated with abnormal DNA methylation of the CpG island and correlates with the transcriptional silencing of FMR1 expression in fragile X patients (2,3,6–11). Although it has been shown that FMR1 is widely expressed in different tissues during embryogenesis (12,13), as well as in young and adult tissues (9,11,14,15), little is known about the control elements in the promoter region. In a transgenic experiment with a reporter gene, a 2.8 kb fragment spanning the region 5′ of exon 1 appeared sufficient to produce a tissue distribution of expression of the reporter gene similar to the endogenous murine Fmr1 gene (16). Primer extension studies have identified a transcription initiation site upstream of the CGG repeat array and 26 bp downstream of a TATA-like sequence (TTACA) (17). In this same study, deletion analysis detected high CAT promoter activity associated with a 466 bp *Pst*I–*Xho*I fragment containing part of the CpG island with 272 bp upstream and 193 bp downstream of the transcription initiation site. Deletion of the transcription initiation site and the 193 bp downstream sequences from this gene fragment inhibited all

*To whom correspondence should be addressed. Tel: +1 418 525 4470; Fax: +1 418 525 4481; Email: francois.rousseau@crsfa.ulaval.ca

promoter activity. Promoter activity was also inhibited after *in vitro* DNA methylation of cytosines in all CpG dinucleotides in the 466 bp promoter fragment.

Clues to understanding both the regulation of FMR1 gene expression and the functions of its encoded protein (FMRP) can be obtained from the study of the factors that modulate FMR1 transcription. We have identified *cis*-controlling elements in the FMR1 promoter by *in vivo* DMS (dimethylsulfate) footprinting. Protein binding to these elements was detected in normal cells (lymphocytes, fibroblasts and lymphoblastoid cells) but not in those with fragile X mutations. Computer-aided analysis identified sequences in the promoter region of the gene coding for heterogenous ribonucleoprotein (hnRNP)-A2 which had striking similarity to the transcriptional regulatory elements showing protein–DNA interactions in the FMR1 promoter.

RESULTS

Functional analysis of the FMR1 promoter

Principle of in vivo *DMS footprinting by ligation-mediated polymerase chain reaction (LM-PCR).* When applied to living cells, DMS diffuses across the plasma and nuclear membranes into the nucleus where it preferentially methylates guanine residues through the major groove at the N-7 position. Guanine residues in contact with sequence-specific DNA-binding proteins display a different degree of reactivity with DMS compared with guanine residues not in contact with binding proteins. Proteins in contact with DNA either decrease accessibility of specific guanines to DMS (protection) or, often at the edges of a footprint, increase reactivity (hyperreactivity) (18). Hyperreactivity can also indicate a greater DMS accessibility allowed by special *in vivo* DNA structures. The glycosylic bond of methylated guanines, as well as the DNA phosphate backbone at the apurinic sites, can be cleaved by hot piperidine leaving a 5′-phosphate (19). These 5′-phosphate single-strand breaks can be quantitatively mapped at the resolution of a single base by the LM-PCR technique coupled with analysis on a DNA sequencing gel. The unique aspect of LM-PCR is the blunt-end ligation of an asymmetric double-stranded linker onto the 5′-end of each nicked DNA molecule (20). The blunt end is created by the extension of a gene-specific primer #1 until a strand break is reached. This linker provides a common sequence at all 5′-ends. A PCR amplification is then performed using the linker primer and the gene specific primer #2 which is nested 3′ of the first gene specific primer (#1). The size of amplification products then corresponds to the position of strand breaks in relation to the position of gene specific primer #2 in the promoter region. The PCR products are size-fractionated on a sequencing gel. LM-PCR amplification with 20 thermal cycles preserves the quantitative representation of each DMS-induced strand break in the original DNA sample. From our experiments and those of others, LM-PCR gives a reproducible quantification of the DMS-induced strand breaks within a given DNA sample (21–24). *In vivo* footprinting is considered to be the most accurate predictor of the state of transcriptional activity of genes because it more accurately represents the DNA–protein interactions encountered in living cells. *In vitro* assays using purified naked DNA are compromised because many of the specific DNA structures present *in vivo* have been lost during the DNA purification steps (18,25,26).

In vivo *DMS protein footprints of the FMR1 promoter.* Transcription initiation of the FMR1 gene has been detected at position +1 (Fig. 3) within a 466 bp fragment which has been shown to have promoter activity (17). We have analyzed both strands of a 1.1 kb sequence in the 5′ region of the FMR1 gene for *in vivo* protein footprints. The region studied corresponds to ~620 bp upstream of the major transcription start site (located at position 13719 in the GenBank sequence #L29074), the first exon including the 5′UTR (273 bp in GenBank) and 51 bp of the initial coding sequence, and ~170 bp of the first intron. As indicated above, the 5′UTR containing the CGG repeats is expanded in DNA samples from patients with fragile X syndrome. We were unable to obtain interpretable results across the G-C rich CGG repeat region probably due to secondary structure problems in the amplification and gel analysis steps.

In primary fibroblasts, lymphoblasts and peripheral lymphocytes from healthy male donors which contain a transcriptionally active FMR1 gene, only four DMS footprints were observed (Figs 1, 2 and 3). These footprints were present within a 91 bp region 5′ of the major transcription start site (Fig. 3). They were localised at nucleotides –131 to –123 (5′-GCGCATGCG-3′), nt –95 to –90 (5′-GGGCGG-3′), nt –68 to –60 (5′-GGGGGAGGG-3′) and nt –50 to –40 (5′-GATCACGT-GAC-3′). The 5′ and 3′ boundaries of these four sequences involved in DNA–protein interactions were easily identified (Figs 1 and 2). The sequences covering nt –620 to –135 and the proximal 130 bp of the first intron did not show any evidence for protein–DNA interactions (data not shown). None of these four footprints was detected in any of the fibroblasts or lymphoblasts from male patients with fragile X syndrome who have a transcriptionally inactive FMR1 gene (Fig. 1). In addition, the studied promoter region in the fragile X cells showed no other indication of protein binding or special DNA structures. For these patients, the FMR1 gene has a large expansion of >230 CGG repeats in the 5′UTR. This expansion is associated with an abnormal methylation of the CpG residues as indicated by the inability to cleave the classical *Eag*I site in this region (27). The presence of this *Eag*I restriction site for each sample was confirmed by genomic sequencing (provided by the LM-PCR procedure). Furthermore, methylation-sensitive LM-PCR analysis (18) of CpGs within the 1.1 kb region studied confirmed that all CpGs were methylated in these cells (unpublished results).

In the footprint analysis of the transcriptionally active FMR1 gene in normal cells, we also observed a hyperreactive guanine residue at position +14 (Fig. 2). Hyperreactivities can indicate protein binding, some unusual DNA structure not present *in vitro* such as a transcription bubble or RNA polymerase pause site. This DNA region showed no protein binding as detected by DMS footprinting. However, DNA from fragile X patients with a transcriptionally inactive FMR1 gene showed no evidence for a hyperreactive guanine at this position. These results suggest that the normal actively transcribed FMR1 gene contains a singular DNA structure at position +14 which is absent in the mutated FMR1 gene of patients with the fragile X syndrome. The lack of this DNA structure may correlate with the absence of transcription of the FMR1 gene or with the presence of expanded CGG repeats and abnormal DNA methylation which occurs in this region in the fragile X syndrome. Deletion studies have indicated that this region is important for *in vitro* promoter activity as detected in CAT assays (17).

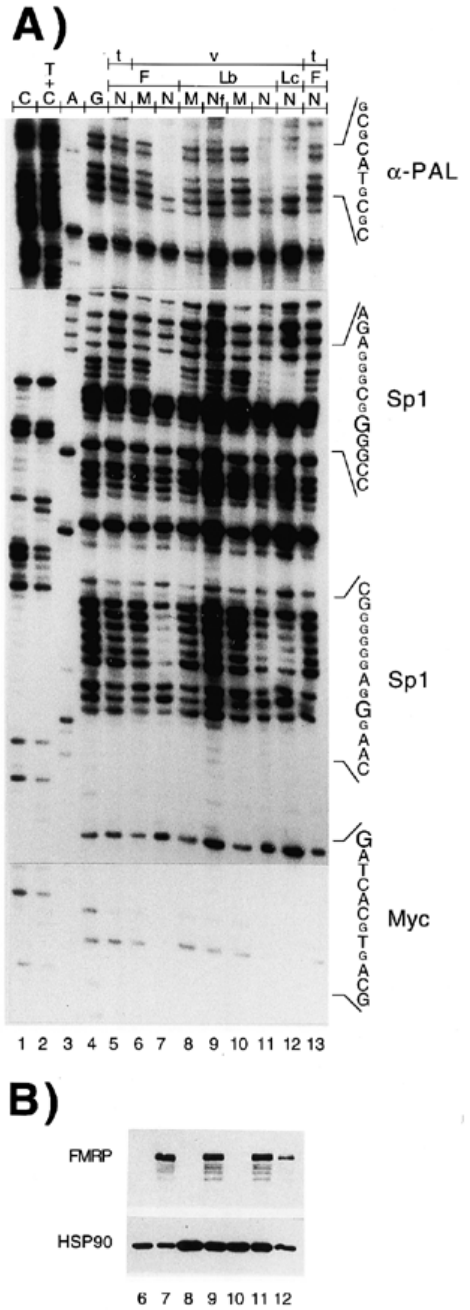

Figure 1. (**A**) Genomic footprinting of the FMR1 gene promoter. The region shown was analyzed with primer set X (primers X1, X2 and X3, Table 2) to reveal upper strand sequences from nt –133 to –35 relative to the major transcription initiation site. Lanes 1–4, LM-PCR of DNA treated with standard Maxam–Gilbert cleavage reactions; lanes 5 and 13, LM-PCR of normal (N) fibroblast (F) DNA treated with DMS *in vitro* (t); LM-PCR of DNA isolated from normal male fibroblasts (lane 7), lymphoblasts (Lb, lane 11) and lymphocytes (Lc, lane 12), from normal female (Nf) lymphoblasts (lane 9) and from fully mutated (M) fibroblasts (lane 6) and lymphoblasts (lanes 8 and 10) treated with DMS *in vivo* (v). As reference, a small portion of the Maxam–Gilbert-derived sequence is shown on the right of the autoradiograms, smaller Gs indicate those Gs which display *in vivo* protection against DMS-induced modification and larger Gs indicate those showing *in vivo* hyperreactivity. From top to bottom, the consensus binding sites for transcription factor α-PAL, Sp1, H4TF1/Sp1-like and Myc are indicated. (**B**) Immunoblot analysis of FMRP expressed in the different cells and cell lines studied. FMRP was detected using 1C3 monoclonal antibody while heat shock protein 90 (HSP90) used as an internal control was detected with affinity-purified rabbit polyclonal IgGs. Lanes 6–12 correspond to the same samples as in (A).

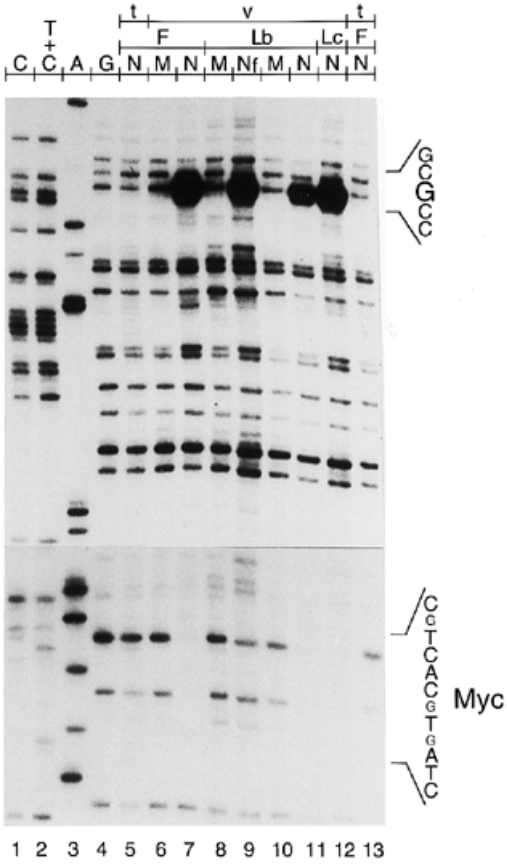

Figure 2. Genomic footprinting of the FMR1 gene promoter. The region shown was analyzed with primer set Y (primers Y1, Y2 and Y3, Table 2) to reveal bottom strand sequences from nt +23 to –58 relative to the major transcription initiation site. Lanes 1–14 are as described in the legend to Figure 1A. As reference, a small portion of the Maxam–Gilbert-derived sequence is shown on the right of the autoradiograms, smaller Gs indicate the Gs displaying *in vivo* protection against DMS-induced modification and larger Gs display *in vivo* hyperreactivity. From top to bottom, the very hyperreactive guanine residue at position +14 and the consensus binding site for transcription factor Myc are indicated.

Structural analysis of the FMR1 promoter

We used TESS (Transcription Element Search Software) to search for potential transcription factor binding sites within the proximal FMR1 promoter sequence. The sequence of the FMR1 promoter is shown in Figure 3 with the positions corresponding to some of the potential transcription factor binding sites indicated. In general, the FMR1 promoter is very GC rich and multiple copies of the consensus binding site for the transcription activator Sp1 (28) are present. Sequences matching other transcription factor binding sites were also identified including AP2, UBP1, AGP/EBP, Myc and Zeste as indicated in Figure 3. Moreover, several transcription factor binding sites corresponded to the sequences within the four footprints identified above. As shown in Table 2 and Figure 3, the sequence across the most 5′ footprint in the 1.1 kb fragment at position –131 to –122 (CGCGCATGCGCG) is identical to the palindromic consensus binding site (YGCGCAYGCGCR) for α-PAL/NRF1. α-PAL

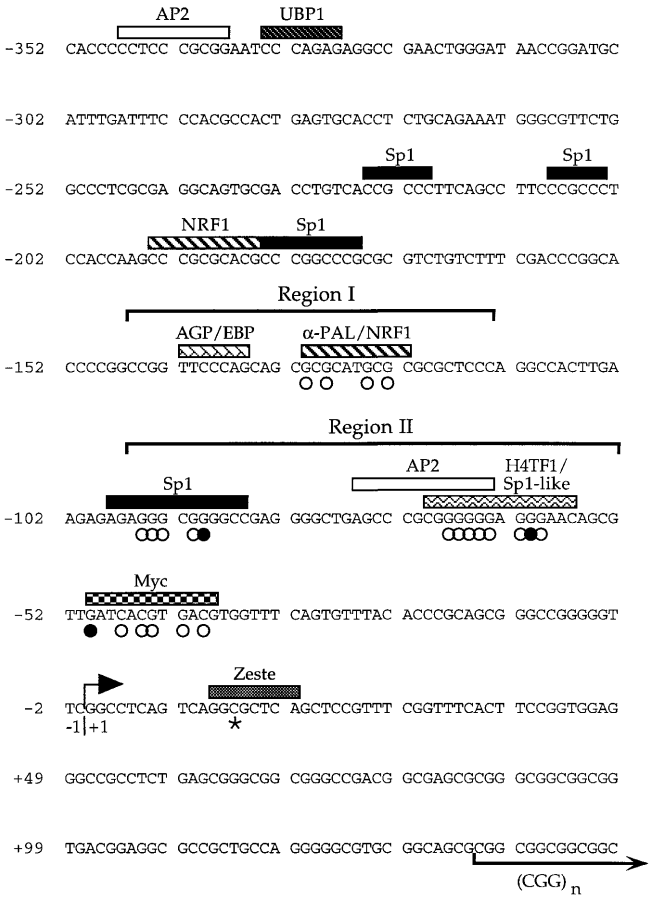

Figure 3. Summary of *in vivo* DMS footprinting data and the consensus binding sequences for transcription factors of the FMR1 gene, from nt –352 to +148. Only the sequence of the upper strand is displayed. Below the sequence line, protection of specific guanine residues is indicated by open circles, and hyperreactive guanine residues are marked by closed circles. Circles below cytosine residues indicate a protected or hyperreactive guanine residue on the bottom strand. Unmarked guanine residues react similarly in cells and in naked DNA. The asterisk at position +14 marks the strongly hyperreactive G residue detected on the lower strand. Putative binding sites as identified by computer searches are shown with horizontal bars on top of the sequence line. Out of the many putative consensus binding sequences for transcription factors identified by the TESS search, only four appear to effectively bind proteins in the living cells studied. Regions I and II which contain sequence similarities to the hnRNP-A2 promoter region are indicated (see also Fig. 4 and Table 1).

Table 1. Consensus sequences recognized by transcriptional factors

α-PAL CONS. (R01727)	G C G C A T G C G C
NRF-1 CONS.	Y G C G C A Y G C G C R
IBR CONS. (Gomez et al. 1995)	R C G C R Y G C G Y
FMR1 REGION I (-132 to -121)	C G C G C A T G C G C G
hnRNP-A2 REGION I (-846 to -857)	G G C G C G C G C G C G
H4TF1/Sp1-like CONS. (R02168)	G G G G G A G G G
SP1 CONS. (M00196)	G G G G G C G G G G Y
FMR1 REGION II (-97 to -87)	G A G G G C G G G G C
FMR1 REGION II (-68 to -58)	G G G G G A G G G A A
hnRNP-A2 REGION II (-897 to -887)	A A G G G C G G G G C
hnRNP-A2 REGION II (-868 to -858)	G G G G G A G G G G C
AP2 CONS. (M00189)	M K C C C S C N G G C G
FMR1 REGION II (-76 to -65)	A G C C C G C G G G G G
hnRNP-A2 REGION II (-876 to -865)	C G C C C G C G G G G G
MYC CONS. (R02207)	C A C G T G
MYC CONS. (Papoulas et al. 1992)	R A C C A C G T G C T C
FMR1 (-50 to -41)	G A T C A C G T G A

M = A,C; N = A,C,G,T; K = G,T; R = A,G; S = G,C; Y = C,T.

was initially detected as a transcription factor involved in the regulation of the expression of the eukaryotic Initiation Factor 2 alpha (eIF-2α), a translation initiation factor (29). Cloning studies have shown this factor to be identical to nuclear respiratory factor (NRF)-1 which stimulates the transcription of nuclear genes whose products function in the mitochondria (30). All of the protected guanines in the –131 to –122 footprint were present in the α-PAL/NRF1 consensus binding site (see Fig. 3) indicating a close correlation between the experimentally determined footprint and the predicted α-PAL/NRF1 binding site.

The sequence across the second *in vivo* footprint (–98, GAGGGCGGGGC, –88), as shown in Figure 3, has 10/11 identical nucleotides with the consensus binding site (GGGGG-CGGGGY) for Sp1, a ubiquitous transcription activator involved in cell development and differentiation (Table 1) (28). The sequence across the third *in vivo* footprint (–68, GGGGGAGGG, –60) is completely contained within the consensus binding site for the histone H4 gene-specific transcription activator (H4TF1; GGGGGAGGG) and Sp1-like transcription factors (GGG-GGCGGGGY; 31), and is partially contained within the consensus binding site for the transactivator AP-2 (MKCCCSCNGGCG; 32) (see Fig. 3 and Table 1). The sequence across the fourth *in vivo* footprint (–50, GATCACGTGACG, –38) is identical to the core binding site (CACGTG) for c-MYC and MYC-like transcription factors and matches well with an extended consensus binding site for c-MYC (RACCACGTGCTC; 33).

Analysis of the sequence (+12, GGCGCTCAGCTCC, +24) across the singular DNA structure which was identified at nt +14 in the LM-PCR study above, revealed the presence of a palindromic structure, centered at nt +18. TESS analysis also revealed similarities between this sequence and the binding site (CGCTCA) for the Zeste transcription factor in *Drosophila*. As indicated above, however, no protein footprint was associated with this DNA region in either actively transcribed or silent FMR1 genes.

In order to determine whether sequences similar to those of the FMR1 promoter exist in the promoter regions of previously characterized genes, the 1100 bp fragment of the FMR1 gene was used as a probe in a search of the non-redundant nucleotide sequence database using the BLAST search algorithm. This type of search is constrained by the defined word length of 12 used in the BLAST search which only allows sequences containing an initial match of 12 identical nucleotides in a row to be further considered for similarity scores. Most identified promoter elements have sequence lengths <12 nt. Unexpectedly, a 36 bp sequence within the promoter region of the hnRNP-A2 gene (from –896 to –854; GenBank accession no. U09120) was identified as having an 83% match with a sequence between –96 and –60 of the FMR1

promoter region (Fig. 5b). This corresponds to the region in FMR1 containing the footprints for the Sp1, AP2 and H4TF1/Sp1-like binding sites, and is herein denoted as region II (Figs 3 and 4). While the BLAST score for the nucleotide match (0.94) would not normally be considered significant for gene to gene comparisons, such a match between small promoter elements could. Therefore, a comparison of the two DNA sequences was undertaken. This analysis demonstrated complete conservation across the region of the footprints in the FMR1 promoter region with the corresponding sequence in the hnRNP-A2 promoter region (Fig. 5b). In addition, the sequence of the hnRNP-A2 gene also corresponded to the consensus binding sites for the Sp1, AP2 and H4TF1/Sp1-like transcription factors, as seen in FMR1 (see Table 1). Subsequent analysis of this region in both genes revealed a second region of nucleotide similarity upstream of region II. This region, called region I, contained α-PAL/NRF1 and AGP/EBP consensus binding sites in both promoters, although the hnRNP-A2 sequence was inverted in relationship to the FMR1 sequence (Figs 3 and 5a). Both promoter regions contained comparable palindromic sequences within the predicted transcription factor binding sites, and the guanines protected in the FMR1 footprint were conserved in the hnRNP-A2 sequence (Fig. 5a). Recently, a chicken homolog of α-PAL/NRF-1 was identified (34). This factor which represses transcription of histone H5 has high affinity binding to the consensus site 'RCGCRYGCGY' which is identical to the potential α-PAL/NRF-1 binding site in hnRNP-A2 (Table 1).

The hnRNP-A2 promoter sequence was also analyzed using the TESS software and a number of Sp1 and H4TF1/Sp1-like binding sites were identified (Fig. 4).

Table 2. Synthetic oligomer primers for LM-PCR analysis of the human FMR1 gene

Primer	Sequence (5′→3′)	Position[a]	T_m (°C)[b]
Non-transcribed strand or upper strand			
W1	TGGTTTGGTCCGGTTC	–385 to –400	54.1
W2	TTTGGTCCGGTTCAAAGTAGCGCAG	–388 to –412	63.6
W3	GTAGCGCAGTCTGACTGAGCGGG	–404 to –426	67.3
Z1	CGAAAGACAGACGCGC	–161 to –176	56.6
Z2	TGCGCGGGCTTGGTGGAGGGCGG	–187 to –209	72.6
Z3	CGGGAAGGCTGAAGGGCGG	–207 to –225	66.1
X1	AAGTGAAACCGAAACGGAGC	+39 to +20	57.7
X2	GAGCTGAGCGCCTGACTGAGGCCG	+23 to –1	70.5
X3	GCCGAACCCCCGGCCCGCTGCGGG	+3 to –21	77.3
T1	CTGGTCTCTCATTTCGAT	+538 to +521	52.7
T2	CTCATTTCGATAGGCGCTAGGGCC	+531 to +508	65.3
T3	CTCTCGGAGTCGGAGAGGGGC	+508 to +488	67.7
Transcribed strand or bottom strand			
P1	GAAAATGAGTTGAGGAAAGGCG	–695 to –674	58.2
P2	GAGTACGTGGGTCAAAGCTGGGTCTGAGG	–674 to –646	69.0
P3	TGAGGAAAGGCTCACATTTTGAGATCCCG	–650 to –622	64.8
R1	TTCACCCCTATTCTCGCC	–457 to –440	57.2
R2	TTCTCGCCTTCCACTCCACCTCCCGC	–447 to –422	69.8
R3	ACTCCACCTCCCGCTCAGTCAG	–435 to –414	65.6
M1	TGTCACCGCCCTTCAGC	–230 to –214	59.4
M2	CCTTCCCGCCCTCCACCAAGCCCGC	–214 to –190	73.4
Y1	TCCACCAAGCCCGCGCA	–203 to –187	61.8
Y2	TCTGTCTTTCGACCCGGCACCCCG	–171 to –148	68.7
Y3	TTCCCAGCAGCGCGCATGCGCGCGC	–142 to –118	73.4
A1	GGAACAGCGTTGATCACGT	–61 to –43	57.5
A2	CGTGACGTGGTTTCAGTGTTTACACCCG	–45 to –18	66.4
S1	CCACCTCTCGGGGGC	+218 to +232	61.7
S2	GGCTGAAGAGAAGATGGAGGAGCTGGTGG	+252 to +280	69.0
S3	GGAGCTGGTGGTGGAAGTGCGGG	+270 to +292	69.0

[a]Primer positions are given relative to the transcription start site (see Fig. 3).
[b]T_m determined by GeneJockey software program.

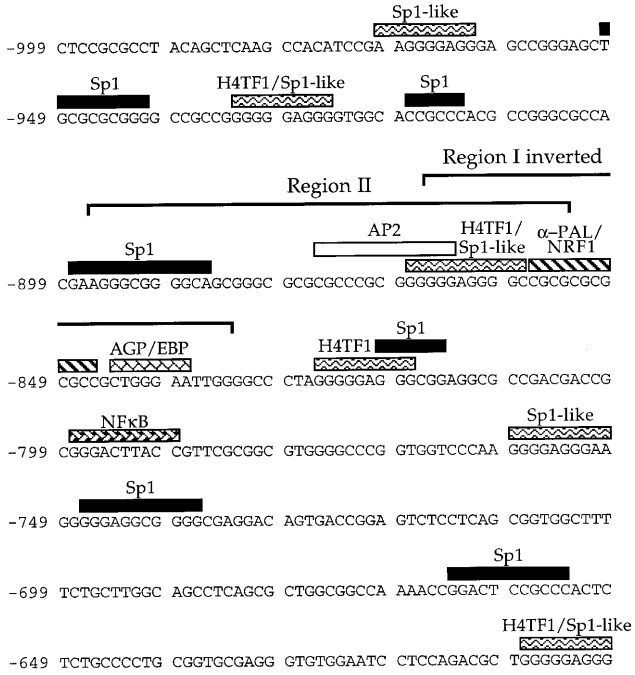

Figure 4. Summary of the consensus binding sequences for transcription factors of the hnRNP-A2 gene, from nt –999 to –600. Only the sequence of the upper strand is displayed. Regions with sequence similarity to the FMR1 promoter are shown.

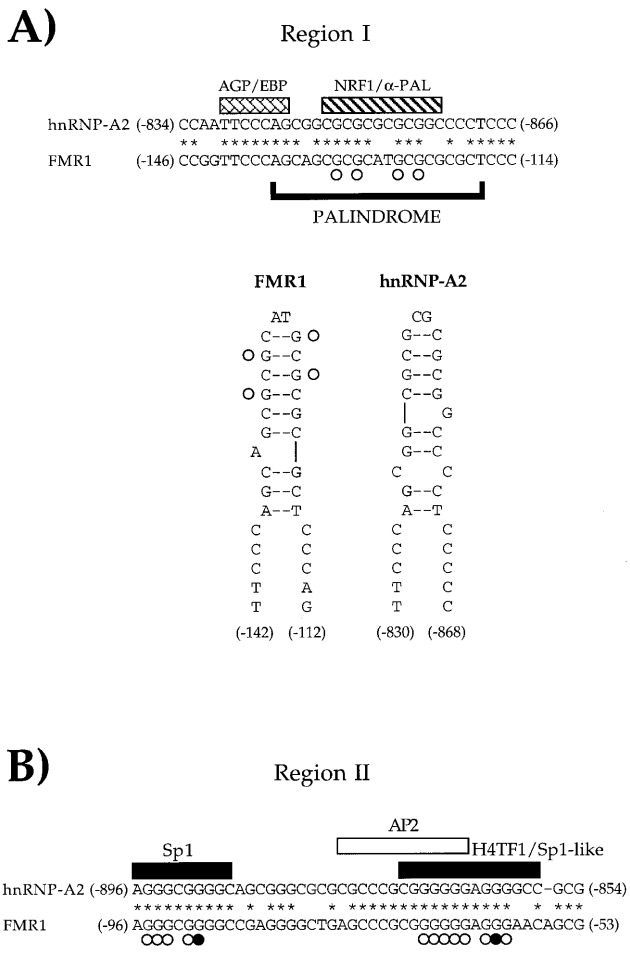

Figure 5. Schematic summary of *in vivo* DMS footprinting data of the FMR1 gene and comparison of the consensus binding sequences for transcription factors of regions I (**A**) and II (**B**) of the FMR1 and hnRNP-A2 genes. The 21 nt that can form a palindromic structure in region I are illustrated. Only the sequence of the upper strand is displayed. Protection of specific guanine residues is indicated by open circles, and hyperreactive guanine residues are marked by closed circles. Circles below cytosine residues indicate a protected or hyperreactive guanine residue on the bottom strand. Unmarked guanine residues react similarly in cells and in naked DNA.

DISCUSSION

We used LM-PCR after DMS treatment of living cells to localise *in vivo* DNA–protein interactions within ~1100 bp of the promoter region for the FMR1 gene. Four specific protein footprints were identified in all cells from normal individuals which contain a transcriptionally active FMR1 gene, including peripheral lymphocytes, primary fibroblast cultures and transformed lymphoblastoid cell lines. These four footprints were all localized within a 91 bp region of the FMR1 promoter (–131 to –40) located 40 bp upstream of the major transcriptional start site (Fig. 3) identifying them as potential *cis*-regulatory elements in FMR1 gene expression. Our sequence comparison studies correlated these footprints with the consensus binding sites for Sp1, AP2, α-PAL/NRF1, Myc and H4TF1/Sp1-like transcription factors. The absence of all footprints in cells obtained from fragile X syndrome patients in which the FMR1 gene is transcriptionally inactive suggests that these, or related transcription factors play an activating role in the modulation of FMR1 expression. It is probable that the expansion of the CGG repeats in the 5′ non-coding region and the abnormal methylation of the CpG rich FMR1 promoter region which accompanies this expansion is preventing binding of the transcription factors to the *cis*-regulatory elements and causing a silencing of FMR1 transcription. The DNA binding of transcription factors which recognize G-C rich binding sites, such as AP2, can be directly inhibited by DNA methylation at cytosine residues (35,36). In our study, we were unable to detect any protein interactions within the 1.1 kb promoter region with DNA from patients with fragile X syndrome that would indicate the presence of a transcription repressor.

At the 3′ extremity of the Myc consensus sequence described above is located a consensus sequence for a cAMP-responsive element (TGACGT) that was recently reported to bind the CRE-binding protein in *in vitro* gel-shift assays (37). The guanine protection hyperreactivity patterns we observed in this region does not suggest *in vivo* binding of a CREB transcription factor to this sequence in the cells studied. This underlines the importance of studying protein–DNA interactions in living cells.

We identified multiple binding sites for the transcription factors Sp1 and AP2 in the FMR1 gene promoter region. Although Sp1 appears to be ubiquitous in nature and AP2 is expressed in neural crest linages, both are implicated in the regulation of many facets of development and differentiation. We have also detected two closely spaced footprints, in region II of the FMR1 promoter, which are predicted to bind Sp1, AP-2 and Sp1-like transcription

factors. Multiple binding sites for Sp1 have been found in a wide variety of promoters (28) and physical interactions between adjacent Sp1 molecules have been proposed to play a role in transcriptional synergism (38). It is possible that the Sp1 and AP2 binding sites in the FMR1 promoter region which were not associated with *in vivo* footprints in this study can functionally bind their respective transcription factor and regulate FMR1 transcription under altered conditions such as during development or differentiation when there is widespread expression of the FMR1 gene (12,13). The Sp1-like element that is associated with an *in vivo* footprint in the FMR1 promoter contains the consensus binding site for the histone 4-specific transcription factor H4TF1. The protected guanines in this footprint correspond exactly to the guanine footprints determined for H4TF1 on the histone 4 promoter (31) suggesting that transcription of FMR1 may be regulated by this factor. It is of interest that the H4TF1 binding site on the histone 4 promoter has a closely associated Sp1 binding site, as is seen in FMR1.

We identified two α-PAL/NRF1 consensus binding sites in the FMR1 promoter region, and the one in region I corresponded to an *in vivo* footprint. The α-PAL/NRF1 transcription factor plays a role in the expression of nuclear respiratory genes and may help coordinate respiratory metabolism with other biosynthetic and degradative pathways. α-Pal/NRF1 may function at the point of translation initiation in these pathways since it has been shown to regulate the expression of the eukaryotic initiation factor 2 alpha (eIF-2α), a translation initiation factor (29). Sequence comparison studies have shown that α-Pal/NRF-1 is closely related to developmental transcription factors in sea urchins and *Drosophila* which appear to function by modulating the transcription of metabolic genes required for cellular growth and proliferation (39).

The functional attributes of the transcription factors which appear to be involved in the regulation of the FMR1 gene mesh closely with the proposed functions for the FMR1 gene product, FMRP, a ribonucleoprotein. The RNA binding properties of FMRP were discovered after the identification of two KH domains and one RGG box that are homologous to RNA-binding sequences in hnRNPs (40,41). FMRP is predominantly a cytoplasmic protein (9,11) that is found in actively translating polyribosomes (42,43) through its association with messenger RNA–protein complexes (mRNPs) (44). Since FMRP contains nuclear localization and nuclear export signals (45–47) and has been detected *in situ* in the nuclei (43), it has been proposed that FMRP would chaperone mRNA from its source in the nucleus to the cytoplasmic translation machinery as part of a complex with mRNPs (43,44).

The sequence similarity that we have detected between the functionally active part of the proximal FMR1 promoter and the upstream region of the promoter for the hnRNP-A2 also appears to tie in with the proposed functions of the two genes. hnRNP-A2 is a member of a family of ribonucleoproteins that are involved in mRNA processing and nuclear-cytoplasm transport (48,49). Unlike FMRP, the hnRNPs are found predominately in the nucleus with only a minor component present in the cytoplasm. Although the coding sequence of hnRNP-A2 is highly homologous to that of the hnRNP-A1 gene, both genes differ in the sequences upstream of their transcription initiation sites (50). However, little is known about the hnRNP-A2 promoter other than the fact that the upstream 500 nt have a stronger promoter activity than SV40 in reporter assays suggesting that other unidentified control elements must ensure a coordinate expression in different cells and tissues (50). Since both proteins appear to be involved with mRNA transport from the nucleus to the cytoplasm, it is possible that, in the perinuclear area, nuclear RNP chaperones (hnRNPs) are exchanged for cytoplasmic RNP chaperones. FMRP now appears to also be one of a family of related RNPs whose other members, such as the closely related FXR1 and FXR2 (51–53) may also play a part in this donor–acceptor relay during mRNA transport. Theoretically, the coordinate regulation of the transcription of donor hnRNPs and cytoplasmic acceptor RNPs (i.e., FMRP and related proteins) may be biologically important. The exact coincidence of *in vivo* footprint sequences in the FMRP promoter regions homologous to the hnRNP-A2 promoter region is suggestive of coordinate expression of these two genes. However, it should be noted that the *in vivo* footprints in the FMR1 gene lie within 150 bp of the transcription initiation start site while the corresponding homologous region in the hnRNP-A2 gene is positioned around 900 bp of its putative start site. Functional *in vivo* footprinting of the hnRNP-A2 promoter region should substantiate whether this region is actually involved in regulation of the hnRNP-A2 gene expression.

The structural features of the FMR1 and hnRNP-A2 promoters region, such as high G-C content, lack of a TATA box, and the presence of multiple Sp1 sites, are characteristic of typical housekeeping genes. This is confirmed by their widespread patterns of expression (15,54). However, high levels of expression of both hnRNP-A2 and FMRP have been detected in neuronal cells (9,55). This neuronal expression coincides with a high level of expression of the AP-2 transcription factor (56) which is postulated from our studies to be involved in the transcriptional regulation of the FMR1 and hnRNP-A2 genes. In our study, protein binding to the AP2, Sp1, α-PAL/NRF1, H4FT1/Sp1-like and Myc binding sites on the FMR1 promoter region correlated with the transcriptional activity of this gene. The absence of transcription factor binding to the FMR1 gene in patients with fragile X syndrome, as revealed in our studies, suggests that expansion of the CGG repeats and hypermethylation of the FMR1 promoter region prevents transcriptional activation of the gene. The absence of FMRP early in embryogenesis, especially in neuronal cells, could alter the transport and expression of important mRNAs leading to the syndrome of mental retardation.

Since submission of this manuscript, Schwemmle *et al.* (57) have reported *in vivo* footprinting analyses of the FMR1 gene in established fibroblast cell lines from normal and fragile X carrier individuals. Although we describe DNA–protein interactions for the same four sites, our findings provide clearly delimited footprints that are significantly larger in some instances. Also the present study involves three other cell types, namely, two untransformed cell types: peripheral lymphocytes and primary fibroblasts and (EBV-transformed) lymphoblastoid cells. The presence of similar DNA–protein interactions in four different normal cell types suggests that the various regulatory elements identified are likely to be involved in control of FMR1 gene expression in several tissues. Furthermore, we have identified a previously undetected hyperreactive guanine at position +14, the possible significance of which has been discussed above. Finally, our combination of functional and structural analyses of the FMR1 promoter region allowed us to identify the hnRNP-A2 gene as being possibly under similar transcription control to the FMR1 gene.

MATERIALS AND METHODS

Cell culture and *in vivo* footprinting

Human skin fibroblasts from healthy male donors and the fibroblasts from a male with the fragile X syndrome were grown in DMEM supplemented with 10% calf serum (FCS). For LM-PCR (see below) analysis, the culture medium was replaced with fresh unsupplemented medium containing 0.2% dimethylsulfate (DMS, Aldrich) (22). After incubation for 6 min at room temperature, the cell monolayer was washed with Ca-Mg-free HBSS (Hanks' Balanced Salt Solution) medium and detached by trypsinization. After cell lysis, nuclei were isolated and the DNA purified as described (20,22). Peripheral blood lymphocytes were isolated from heparinized blood by centrifugation on Ficoll-Hypaque (Pharmacia) gradients. Epstein–Barr-virus-transformed lymphoblastoid cell lines originating from healthy female and male donors as well as from fragile X patients were maintained in RPMI medium supplemented with 15% FCS. Lymphocytes and lymphoblastoid cell lines were treated with 0.2% DMS for 6 min prior to isolation of the nuclei. DNA concentration was measured by fluorometry after staining with 4′-6-diamidino-2-phenylindole (DAPI) (20). For comparison, DNA purified from lymphocytes was treated with 0.5% DMS (*in vitro* treatment) as described (23). In the present study, six different types of cells were studied: fragile X mutated primary fibroblasts NIGMS #GM04024B, fragile X mutated lymphoblastoid cell line RMGA#Q.691, normal lymphoblastoid cells from a male RMGA#Q.001 and a female RMGA#Q.772 as well as normal low-passage foreskin fibroblasts cultures and fresh lymphocytes from two different normal male donors.

Chemical cleavage of 7-methylguanine

DMS-exposed DNA was treated with hot piperidine (Fluka) to convert methylated bases to DNA strand breaks (23). The single-strand break frequency in the total genomic DNA was determined by alkaline agarose (1.5%) gel electrophoresis (58).

Ligation-mediated polymerase chain reaction

Details of the LM-PCR protocol used for this work have already been published (20). The procedure can be divided into six steps: (i) primer extension of an annealed gene-specific oligonucleotide (primer 1) to generate blunt ends from nicked genomic DNA; (ii) ligation of a universal asymmetric double-strand linker; (iii) PCR amplification using a second gene-specific oligonucleotide (primer 2); (iv) separation of the DNA fragments on a sequencing polyacrylamide gel; (v) transfer of the DNA to a nylon membrane by electroblotting; (vi) hybridization of a radiolabeled probe prepared by repeated primer extension using a third gene-specific oligonucleotide (primer 3).

Approximately 1100 base pairs of the human FMR1 gene (GenBank accession no. L29074: from bp 13080 to 14180), including the promoter region, part of the first exon and proximal part of the first intron, were studied (see Results) on both strands using the primer sets described in Table 2. Primer extension was initiated with 0.5–1.0 μg of DNA in duplicate. Initially, one half of each duplicate sample was analyzed as described (20). If there was no significant variation between duplicate samples, the remainder of the two samples were pooled and a combined gel was produced. Thus, each analysis represents the sum of both duplicate samples. The nylon membranes were exposed to a phosphor-sensitive imaging plate (Type III-s) and the band intensities were quantified by Phosphoimager, Fuji BAS 1000 (Fuji Medical Systems USA Inc., Stanford, CT).

DNA analysis by Maxam–Gilbert cleavage reactions was carried out as described (23), except DNA for the 'adenine' reactions was treated with K_2PdCl_4 at pH 2.0 followed by piperidine treatment (59). Chemically cleaved G, A, T+C and C samples were included along with the other samples in the LM-PCR assays in order to provide base markers in the sequencing gels.

Western blots

Total protein extracts were prepared by lysing the cells in SDS sample buffer (68 mM Tris-HCl, 2%SDS, 2% β-mercapto-ethanol, 6% glycerol) followed by sonication and heat denaturation in a boiling bath for 3 min. Immunoblot analyses were performed using the ECL (DuPont) system as described (15). FMRP was detected using 1C3 monoclonal antibody (9) and hsp90 was detected with affinity-purified rabbit polyclonal antibodies (60).

Sequence analysis

To determine whether similarities exist between the FMR1 promoter and other characterized nucleotide sequences, we performed a BLASTN search against the non-redundant DNA database, as implemented in the Baylor College of Medicine Search Launcher (http://dot.imgen.bcm.tmc.edu:9331/seq-search/gene-search.html). To identify potential transcription regulatory sites in the FMR1 and hnRNP-A2 promoters, we used TESS (Transcription Element Search Software) as implemented on the Baylor College of Medicine Search Launcher (as above). This software accesses the TRANSFAC MATRIX database Release 3.1 which contains the consensus binding sites for a variety of transcription factors.

ACKNOWLEDGEMENTS

We thank S. Tremblay for cell culture, J.-P. Therrien for assistance with the Phosphorimager and M. Vaillancourt for his help in the initial steps of this work. We thank D. Devys and J.-L. Mandel for 1C3 antibody and R. Tanguay for the affinity-purified hsp90 antibody. This project was supported by the Canadian Genetic Diseases Network (MRC/NSERC NCE Program) and by the Medical Research Council (MRC) of Canada. Lymphoblastoid cell lines of fragile X patients were established by the 'Banque Lymphoblastique du Réseau de Médecine Génétique Appliquée du FRSQ'. R.D. is a research scholar of the Cancer Research Society Inc./MRC program. M.A. holds a studentship from the Department of Medical Biology, Faculty of Medicine, Université Laval. F.R. is an MRC Scientist.

REFERENCES

1. Nussbaum,R.L. and Ledbetter,D.H. (1995) The Fragile X syndrome. In Scriver,C.R., Beaudet,A.C., Sly,W.S. and Valle,D. (eds) *The Metabolic and Molecular Bases of Inherited Disease*. McGraw-Hill, Inc. New York, 7th ed., vol.1, pp. 795–810.
2. Oberlé,I., Rousseau,F., Heitz,D., Kretz,C., Devys,D., Hanauer,A., Boué,J., Bertheas,M.F. and Mandel,J.-L. (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* **252**, 1097–1102.

3. Verkerk,A.J.M.H., Pieretti,M., Sutcliffe,J.S., Fu,Y.-H., Kuhl,D.P., Pizzuti,A., Reiner,O., Richards,S., Victoria,M.F., Zhang,R., Eussen,B.E., van Ommen,G.J.B., Blonden, L.A.J., Riggins,G.J., Chastain,J.L., Kunst,C.B., Galjaard,H., Caskey,C.T., Nelson,D.L., Oostra,B.A. and Warren,S.T. (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.
4. Kremer,E.J., Pritchard,M., Lynch,M., Yu,S., Holman,K., Baker,E., Warren,S.T., Schlessinger,D., Sutherland,G.R. and Richards,R.I. (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CGG)*n*. *Science* **252**, 1711–1714.
5. Eichler,E.E., Richards,S., Gibbs,R.A. and Nelson,D.L. (1993) Fine structure of the human FMR1 gene. *Hum. Mol. Genet.* **2**, 1147–1153.
6. Pieretti,M., Zhang,F., Fu,Y.-H., Warren,S.T., Oostra,B.A., Caskey,C.T. and Nelson,D.L. (1991) Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* **66**, 817–822.
7. Hansen,R.S., Gartler,S.M., Scott,C.R., Chen,S.H. and Laird,C.D. (1992) Methylation analysis of CGG sites in the CpG island of the human FMR1 gene. *Hum. Mol. Genet.* **1**, 571–578
8. Sutcliffe,J.S., Nelson,D.L., Zhang,F., Pieretti,M., Caskey,C.T., Saxe,D. and Warren,S.T. (1992) DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum. Mol. Genet.* **1**, 397–400.
9. Devys,D., Lutz,Y., Rouyer,N., Bellocq,J.-P. and Mandel,J.-L. (1993) The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. *Nature Genet.* **4**, 335–340.
10. Hornstra, I.K., Nelson, D.L., Warren, S.T. and Yang,T.P. (1993) High resolution methylation analysis of the FMR1 gene trinucleotide repeat region in fragile X syndrome. *Hum. Mol. Genet.* **2**, 1659–1665
11. Verheij,C., Bakker,C.E., de Graaff,E., Keulemans,J., Willemsen,R., Verkerk, A.J., Galjaard,H., Reuser,A.J., Hoogeveen,A.T. and Oostra,B.A. (1993) Characterization and localization of the *FMR-1* gene product associated with fragile X syndrome. *Nature* **363**, 722–724.
12. Hinds,H.L., Ashley,C.T., Sutcliffe,J.S., Nelson,D.L., Warren,S.T., Housman,D.E. and Schalling,M. (1993) Tissue specific expression of FMR-1 provides evidence for a functional role in fragile X syndrome. *Nature Genet.* **3**, 36–43.
13. Bachner,D., Manca,A., Steinbach,P., Wohrle,D., Just,W., Vogel,W., Hameister,H. and Poustka,A. (1993) Enhanced expression of the murine FMR1 gene during germ cell proliferation suggests a special function in both the male and the female gonad. *Hum. Mol. Genet.* **2**, 2043–2050
14. Ashley,C.T.,Jr, Sutcliffe,J.S., Junst,C.B., Leiner,H.A., Eichler,E.E., Nelson,D.L. and Warren,S.T. (1993) Human and murine FMR1: alternative splicing and translational initiation downstream of the CGG-repeat. *Nature Genet.* **4**, 244–251.
15. Khandjian,E.W., Fortin,A., Thibodeau,A., Tremblay,S., Côté,F., Devys,D., Mandel,J.-L. and Rousseau,F. (1995) A heterogeneous set of FMR1 proteins is widely distributed in mouse tissues and is modulated in cell culture. *Hum. Mol. Genet.* **4**, 783–789.
16. Hergersberg,M., Matsuo,K., Gassmann,M., Schaffner,W., Lüscher,B., Rülicke,T. and Aguzzi,A. (1995) Tissue-specific expresssion of a FMR1/beta-galactosidase fusion gene in transgenic mice. *Hum. Mol. Genet.* **4**, 359–366.
17. Hwu,W.-L., Lee,Y.-M., Lee,S.-C. and Wang,T.-R. (1993) In vitro DNA methylation inhibits FMR-1 promoter. *Biochem. Biophys. Res. Comm.* **193**, 324–329.
18. Pfeifer,G.P., Tanguay,R.L., Steigerwald,S.D. and Riggs,A.D. (1990) In vivo footprint and methylation analysis by PCR-aided genomic sequencing: comparison of active and inactive X chromosomal DNA at the CpG island and promoter of human PGK-1. *Genes Dev.* **4**, 1277–1287.
19. Maxam,A.M. and Gilbert,W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**, 499–560.
20. Drouin,R., Rodriguez,H., Holmquist,G.P. and Akman,S.A. (1996) Ligation-mediated PCR for analysis of oxidative DNA damage. In Pfeifer,G.P. (ed.), *Technologies for Detection of DNA Damage and Mutations*, Plenum Press, New York, pp 211–225.
21. Pfeifer,G.P. (1992) Analysis of chromatin structure by ligation-mediated PCR. *PCR Meth. Appl.* **2**, 107–111.
22. Pfeifer,G.P. and Riggs,A.D. (1993) Genomic footprinting by ligation mediated polymerase chain reaction. In White,B. (ed.) *PCR protocols: Current Methods and Applications, Methods in Molecular Biology*, vol. 15, 153–168, Humana Press Inc, Totowa, NJ, USA.
23. Pfeifer,G.P. and Riggs,A.D. (1993) Genomic sequencing. In Griffin,G.H. and Griffin,A.M. (eds) *DNA Sequencing Protocols, Methods in Molecular Biology*, vol. 23, 169–181, Humana Press Inc, Totowa, NJ, USA.
24. Pfeifer,G.P., Singer-Sam,J. and Riggs,A.D. (1993) Analysis of methylation and chromatin structure. *Methods Enzymol.* **225**, 567–583.
25. Pfeifer,G.P. and Riggs,A.D. (1991) Chromatin differences between active and inactive X chromosomes revealed by genomic footprinting of permealized cells using DNase I and ligation-mediated PCR. *Genes Dev.* **5**, 1102–1113.
26. Tornaletti,S. and Pfeifer,G.P. (1995) UV light as a footprinting agent: modulation of UV-induced DNA damage by transcription factors bound at the promoters of three human genes. *J. Mol. Biol.* **249**, 714–728.
27. Rousseau,F, Heitz,D., Biancalana,V., Blumenfeld,S., Kretz,C., Boué,J., Tommerup,N., Van Der Hagen,C., Blanchet-DeLozier,C., Croquette,M.-F., Gilgenkrantz,S., Jalbert,P.,Voelckel,M.-A., Oberlé,I. and Mandel,J.-L. (1991) Direct diagnosis by DNA analysis of the fragile X syndrome of mental retardation. *N. Engl. J. Med.* **325**, 1673–1681.
28. Kadonaga,J.T and Tjian,R. (1986) Affinity purification of sequence-specific DNA binding proteins. *Proc. Natl. Acad. Sci. USA* **83**, 5889–5893.
29. Jacob,W.F., Silverman,T.A., Cohen,R.B. and Safer,B. (1989) Identification and characterization of a novel transcription factor participating in the expression of eukaryotic initiation factor 2a*. *J. Biol. Chem.* **264**, 20372–20384.
30. Evans,M.J. and Scarpulla,R.C. (1990) NRF-1: a trans-activator of nuclear-encoded respiratory genes in animal cells. *Genes Dev.* **4**, 1023–1034.
31. Dailey,L, Boseman,R.S. and Heintz,N. (1988) Purification of the human histone H4 gene-specific transcription factors H4TF-1 and H4TF2. *Genes Dev.* **2**, 1700–1712.
32. Williams,T. and Tjian,R. (1991) Analysis of the DNA-binding and activation properties of the human transcription factor AP-2. *Genes Dev.* **5**, 670–682.
33. Papoulas,W., Williams,N.G. and Kingston,R.E. (1992) DNA binding activities of C-MYC purified from eukaryotic cells. *J. Biol. Chem.* **267**, 10470–10480.
34. Gomez-Cuadrado,A., Martin,M., Noel,M. and Ruiz-Carrillo,A. (1995) Initiation binding repressor (IBR), a factor that binds to the transcription initiation site of the histone H5 gene, is a glycosylated member of a family of cell growth regulators. *Mol. Cell. Biol.* **15**, 6670–6685.
35. Comb,M. and Goodman,H.M. (1990) CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. *Nucleic Acids Res.* **18**, 3975–3982.
36. Hermann,R. and Doerfler,W. (1991) Interference with protein binding at AP2 sites by sequence-specific methylation in the late E2A promoter of adenovirus type 2 DNA. *FEBS Lett.* **281**, 191–195.
37. Hwu,W.L., Wang,T.R. and Lee,Y.M. (1997) FMR1 enhancer is regulated by cAMP through a cAMP-responsive element. *DNA Cell Biol.* **16**, 449–453.
38. Mastrangelo,I.A, Courey,A.J, Wall,J.S, Jackson,S.P. and Hough,P.V. (1991) DNA looping and Sp1 multimer links: a mechanism for transcriptional synergism and enhancement. *Proc. Natl. Acad. Sci. USA* **88**, 5670–5674.
39. Efiok,B.J., Chiorini,J.A. and Safer, B. (1994) A key transcription factor for eukaryotic initiation factor-2 alpha is strongly homologous to developmental transcription factors and may link metabolic genes to cellular growth and development. *J. Biol. Chem.* **269**, 18921–18930.
40. Siomi,H., Siomi,M.C., Nussbaum,R.L. and Dreyfuss,G. (1993) The protein product of the fragile X gene, *FMR1*, has characteristics of an RNA-binding protein. *Cell* **74**, 291–298.
41. Ashley,C.T.,Jr, Wilkinson,K.D., Reines,D. and Warren,S.T. (1993) *FMR-1* protein: conserved RNP family domains and selective RNA binding. *Science* **262**, 563–566.
42. Khandjian,E.W., Corbin,F., Woerly,S. and Rousseau,F. (1996) The fragile X mental retardation protein is associated with ribosomes. *Nature Genet.* **12**, 91–93.
43. Feng,Y, Gutekunst,C.A., Eberhart,D.E., Yi,H., Warren,S.T. and Hersh,S.M. (1997) Fragile X mental retardation protein: nucleocytoplasmic shuttling and association with somatodendritic ribosomes. *J. Neurosci.* **17**, 1539–1547.
44. Corbin,F., Bouillon,M., Fortin,A., Morin,S., Rousseau, F. and Khandjian,E.W. (1997) The fragile X mental retardation protein is associated with poly(A)+ mRNA in actively translating polyribosomes. *Hum. Mol. Genet.* **6**, 1465–1472.
45. Eberhart,D.E, Malter,H.E, Feng,Y. and Warren,S.T. (1996) The fragile X mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear export signals. *Hum. Mol. Genet.* **5**, 1083–1091.
46. Fridell,R.A., Benson,R.E., Hua,J., Bogerd,H.P. and Cullen,B.R. (1996) A nuclear role for the fragile X mental retardation protein. *EMBO J.* **15**, 5408–5414.
47. Sittler,A., Devys,D., Weber,C. and Mandel,J.-L. (1996) Alternative splicing of exon 14 determines nuclear or cytoplasmic localisation of fmr1 protein isoforms. *Hum. Mol. Genet.* **5**, 95–102.
48. Pinol-Roma,S. and Dreyfuss,G. (1992) Shuttling or pre-mRNA binding proteins between nucleus and cytoplasm. *Nature* **355**, 730–732.

49. Siomi,H. and Dreyfuss,G. (1995) A nuclear localization domain in the hnRNP A1 protein. *J. Cell. Biol.* **129**, 551–560.
50. Biamonti,G., Ruggiu,M., Saccone,S., Della Valle,G. and Riva,S. (1994) Two homologous genes, originated by duplication, encode the human hnRNP proteins A2 and A1. *Nucleic Acids Res.* **22**, 1996–2002.
51. Siomi,M.C., Siomi,H., Sauer,W.H., Srinivasan,S., Nussbaum,R.L. and Dreyfuss,G. (1995) FXR1, an autosomal homolog of the fragile X mental retardation gene. *EMBO J.* **14**, 2401–2408.
52. Zhang,Y., O'Connor.J.P., Siomi,M.C., Srinivasan,S., Dutra,A., Nussbaum,R.L. and Dreyfuss,G. (1995) The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2. *EMBO J.* **14**, 5358–5366.
53. Siomi,M.C., Zhang,Y., Siomi,H. and Dreyfuss,G. (1996) Specific sequences in the fragile X syndrome protein FMR1 and the FXR proteins mediate their binding to 60S ribosomal subunits and the interactions among them. *Mol. Cell. Biol.* **16**, 3825–3832.
54. Biamonti,G, Bassi,M.T., Cartegni,L., Mechta,F., Buvoli,M., Cobianchi,F. and Riva,S. (1993) Human hnRNP protein A1 gene expression. Structural and functional characterization of the promoter. *J. Mol. Biol.* **230**, 77–89.
55. Kozu,T., Henrich,B. and Schafer,K.P. (1995) Structure and expression of the gene (HNRPA2B1) encoding the human hnRNP protein A2/B1. *Genomics* **25**, 365–371.
56. Mitchell,P.J., Timmons,P.M., Hebert,J.M., Rigby,P.W. and Tjian,R. (1991) Transcription factor AP-2 is expressed in neural crest cell lineages during mouse embryogenesis.*Genes Dev.* **5**, 105–119.
57. Schwemmle,S., de Graff,E., Heidrun,D., Gläser,D., Wöhrle,D., Kennerknecht,I., Just,W., Oostra,B.A., Dörfler,W., Vogel,W. and Steinbach,P. (1997) Characterization of FMR1 promoter elements by in-vivo footprinting analysis. *Am. J. Hum. Genet.* **60**, 1354–1362.
58. Drouin,R., Gao,S. and Holmquist,G.P. (1996) Agarose gel electrophoresis for DNA damage analysis. In Pfeifer,G.P. (ed.) *Technologies for Detection of DNA Damage and Mutations*. Plenum Press, New York, pp 37–43.
59. Iverson,B.L. and Dervan,P.B. (1987) Adenine specific DNA chemical sequencing reaction. *Nucleic Acids Res.* **19**, 7823–7830.
60. Tanguay,R.M., Wu,Y. and Khandjian,E.W. (1993) Tissue-specific expression of heat shock proteins of the mouse in the absence of stress. *Dev. Genet.* **14**, 112–118.