Human Molecular Genetics, 2004, Vol. 13, No. 5 **543–549**
DOI: 10.1093/hmg/ddh053
Advance Access published on January 13, 2004

Redistribution of transcription start sites within the *FMR1* promoter region with expansion of the downstream CGG-repeat element

Alexandra Beilina[1], Flora Tassone[1], Philip H. Schwartz[2], Parminder Sahota[1] and Paul J. Hagerman[1,*]

[1]Department of Biological Chemistry, University of California, Davis, School of Medicine, Davis, CA, USA and [2]Children's Hospital of Orange County Research Institute, Orange, CA, USA

Received November 5, 2003; Revised and Accepted December 26, 2003

Fragile X syndrome, the most common form of mental impairment, is caused by expansion of a $(CGG)_n$ trinucleotide repeat element located in the 5′ untranslated region of the fragile X mental retardation 1 (*FMR1*) gene. Repeat expansion is known to influence both transcription and translation; however, the mechanisms by which the CGG element exerts its effects are not known. In the current work, we have utilized 5′-RLM-RACE to examine the influence of CGG repeat number on the utilization of transcription start sites in normal ($n < 55$) and premutation ($54 < n < 200$) cell lines of both non-neural (lymphoblastoid) and neural (primary astrocyte) origin. Our results demonstrate that, in both neural and non-neural cells, transcription of the *FMR1* gene is initiated from several transcription start sites within a ~50 nt region that lies ~130 nt upstream of the CGG repeat element. For normal alleles, most transcripts initiate from the downstream-most start site, close to the single position identified previously. Surprisingly, as the size of the CGG repeat expands into the premutation range, initiation shifts to the upstream sites, suggesting that the CGG element may act as a downstream enhancer/modulator of transcription. The shift in start site selection for both neural and non-neural cells indicates that the effect is general. Furthermore, the correspondence between start site utilization and the degree of elevation of *FMR1* mRNA suggests that a substantial fraction of the increased message in the premutation range may derive from the upstream start sites.

INTRODUCTION

Fragile X syndrome, the most common inherited form of mental impairment, is a trinucleotide repeat expansion disorder in which expansion beyond 200 repeats (full mutation) of a $(CGG)_n$ element in the 5′ untranslated (UTR) region of the fragile X mental retardation 1 (*FMR1*) gene generally leads to transcriptional silencing (1,2). Reduced levels, or absence of *FMR1* protein (FMRP), lead to fragile X syndrome (1,3,4).

The *FMR1* promoter region possesses a CpG-rich island, but lacks a canonical TATA box (5). Previously, primer extension studies have identified a transcription initiation site 134 bp upstream of the CGG repeat (6), which is generally regarded as being the predominant if not the only point of initiation of *FMR1* transcript synthesis. However, no studies have further characterized the promoter region of the *FMR1* gene for the presence of additional transcription start sites, nor has there been any determination as to whether the initiation site may be altered for expanded alleles.

Two major consensus sites, USF1/USF2 and alpha-Pal/Nrf-1, within the *FMR1* promoter have been shown to be involved in positive regulation of *FMR1* expression (7). A mutation in either one of these regions results in a more than 5-fold reduction in promoter activity, whereas mutation of both of these regions completely abolishes promoter activity (5,7,8); mutations within a TATA-like element had essentially no effect on transcriptional activity. Thus, USF1/2 and alpha-Pal/Nrf 1 appear to be important, positive *cis* elements for *FMR1* alpha-Pal/Nrf1 expression.

For CGG repeat expansions in the premutation range (55–200 CGG repeats), *FMR1* mRNA levels are elevated by as much as 5-fold in the upper premutation range (>100 CGG repeats) (9–11). The basis for the increased *FMR1* mRNA levels is not known at present, although the higher levels do not

*To whom correspondence should be addressed at: Department of Biological Chemistry, University of California, Davis, School of Medicine, One Shields Avenue, Davis, CA 95616, USA. Tel: +1 5307547266; Fax: +1 5307547269; Email: pjhagerman@ucdavis.edu

appear to be due to increased mRNA stability (10). Moreover, in spite of the increased levels of *FMR1* mRNA, FMRP levels are slightly to moderately reduced in the premutation range (9,11). A translational deficit directly correlating with CGG repeat number has been reported for premutation alleles (12). These findings suggest that the expanded CGG repeat is capable of influencing both transcription and translation of the *FMR1* gene. However, all of these investigations have implicitly assumed that there is a single site for the initiation of transcription, and that this site remains unchanged for expanded alleles.

To address the question of number and position(s) of sites of transcriptional initiation, we examined transcription initiation start sites in normal and premutation lymphoblastoid cell lines, and in primary astrocyte culture derived from post-mortem tissue, using RNA ligase-mediated rapid amplification of the 5′ cDNA end (RLM-RACE). Our analysis reveals that transcription of the *FMR1* gene is initiated from several start sites, and that start site selection—in both neural and non-neural cells—is modulated by the number of CGG repeats in the downstream $(CGG)_n$ element.

RESULTS

Transcription of the *FMR1* gene is initiated at multiple sites

The transcription start sites of the human *FMR1* gene were determined using 5′ RNA ligase-mediated rapid amplification (RLM-RACE) analysis, which has proven to be a very sensitive and accurate method to obtain full-length 5′ ends of cDNAs, and to ensure that truncated messages are eliminated from the amplification reactions (13,14). 5′-RLM-RACE analysis was used to determine the transcription start sites in a lymphoblastoid cell line with a normal number of CGG repeats (female, AG; 21, 29 CGG repeats). Sequencing of 27 clones for the normal cell line has identified the presence of three distinct clusters (sites) for initiation of *FMR1* transcripts, starting at the positions within −132 to −128, −166 to −162, and −181 to −177 (where +1 for the translation start site with a reference CGG repeat number of 16; Fig. 1); these clusters are designated sites I–III, respectively. Site I is close to the previously reported transcription start site determined by primer extension analysis (−135; Fig. 1) (6). Our results indicate that site I is the major transcription start site, represented in 67% of total clones (Fig. 2). The other two sites (sites II and III) are represented in 15% and 18% of the clones, respectively, and are located ∼40 and 50 bp upstream of the major transcription start site (site I).

The length of the downstream CGG repeat element influences the distribution of transcription start sites

Premutation alleles are known to be associated with elevated levels of *FMR1* mRNA (9–11). However, it is generally assumed that the increased mRNA levels with increasing CGG repeat are simply quantitative; that is, that the same mRNA species is/are being produced, independent of CGG repeat length. To examine this issue in more detail, we determined the relative utilization of each of the three transcription start sites, for three lymphoblastoid cell lines harboring premutation alleles (60, 82, 160 CGG repeats; Table 1), using 5′-RLM-RACE. As for the normal allele, transcription initiated at all three sites; however, the fraction of initiation events from each site differed significantly between normal and premutation alleles (Fig. 2; Table 2).

We sequenced 27 clones for the 21/29 CGG repeat alleles, 32 clones for the 60 CGG repeat allele, 33 clones for the 82 CGG repeat allele and 47 clones for the 160 CGG repeat allele (Table 1). Site I was the major transcription start site for the normal cell line and was represented in 67% of the clones (Fig. 2). In contrast, site III was the major transcription start site for the two largest premutation cell lines (82 and 160 repeats), represented in 55 and 58% of clones, respectively. For the low premutation allele (60 repeats), mRNA generated from site II was the major species, represented in 50% of the clones. As demonstrated by the χ^2 statistics presented in Table 2, start site utilization undergoes a significant shift to upstream start sites as the CGG repeat expands from normal to the low premutation range (normal to 60; $P = 0.017$), and from the low premutation to mid/high premutation range (60 to 82, 60 to 160; $P = 0.0001$). These results indicate that start site selection for transcription of the *FMR1* gene is influenced by the downstream CGG element.

Transcripts initiated from upstream start sites (sites II and III) are proportionally represented within the pool of polyadenylated mRNA

We have investigated the possibility that the RNA species produced from the upstream sites are the products of abortive initiation. This possibility is underscored by the observation that abortive transcripts ∼200 nt in length are produced from upstream sites during HIV virus transcription (15). To address this issue, we repeated the 5′-RLM-RACE experiments with polyA(+) mRNA isolated from lymphoblastoid lines that possess either normal (21 or 29 CGG repeats) or premutation (82 CGG repeats) alleles. PolyA(+) RNA representing all three start sites was found for both lines, and in approximately the same proportions to the fractions initiating at each site in the total RNA pools. For sites I–III, the polyA(+) RNAs yielded 16, 5 and 3 clones, respectively, for the normal line [total versus polyA(+) RNA; $P = 0.88$]; and 3, 6, and 14 clones, respectively, for the 82 CGG allele [total versus polyA(+) RNA; $P = 0.86$]. Thus, transcripts initiated from all three sites are processed into polyA(+) message.

The CGG element also influences start site selection in human primary astrocytes

To determine whether the distribution of start sites, and their utilization, is similar in neural cells, we performed 5′-RLM-RACE experiments with primary astrocytes derived from both normal (15 CGG) and premutation (87 CGG) alleles (16). In accord with the results from lymphoblastoid lines, the same three transcription initiation sites are found in both lines (Fig. 1). Moreover, as with the lymphoblastoid lines, start site utilization underwent a significant shift toward the upstream sites for expanded CGG alleles (87 repeats; Fig. 3; Table 2). The partitioning among the three sites in the astrocyte line with 87 CGG repeats is closer to the lymphoblastoid cell line with

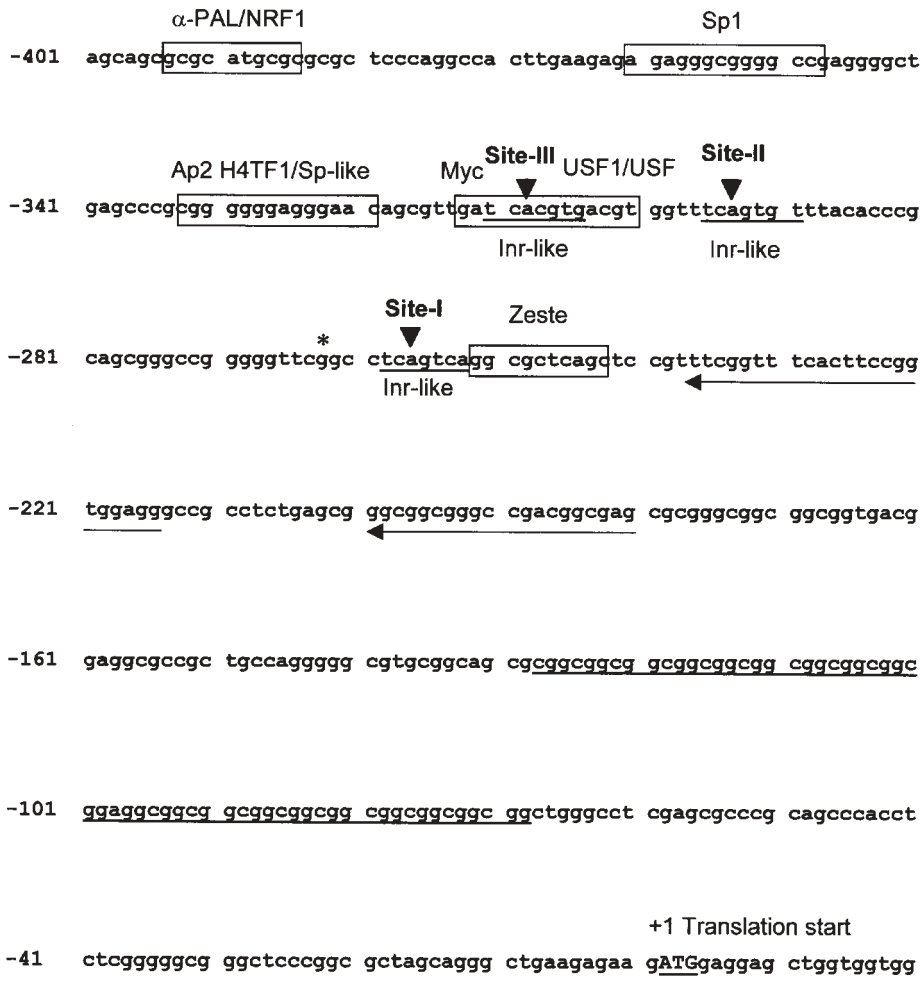

Figure 1. Region of the *FMR1* gene including the CGG element (16 CGG repeats), CpG island, transcription factor binding sites, and translation start sites. Nucleotide positions are numbered to the right with respect to the start of translation at +1 (ATG). The multiple transcription start sites deduced from the current 5′RLM-RACE experiments are noted by the solid triangles, and are designated as site I, site II and site III. The asterisk marks the previously identified transcription start site (6). Transcription factor binding sites are indicated by boxes. Initiator-like (Inr-like) sequences are underlined. *FMR1*-specific primer positions are designated by arrows.

60 CGG repeats than to the lymphoblastoid cell line with 82 CGG repeats, which may explain the absence of elevated *FMR1* mRNA in the premutation (astrocyte) line. In this regard, the fractional utilization of site III is unchanged from the normal control, suggesting that the most upstream site may have a significant influence on translation. Taken together, these observations demonstrate that the number of CGG repeats, *per se*, is responsible for differences in transcription initiation, and that either the level or specificity of some additional factors are likely to play a role in *FMR1* transcription initiation.

DISCUSSION

In the current study, we have utilized 5′-RLM-RACE to demonstrate that transcription of the *FMR1* gene is initiated at multiple transcription start sites for both neural (primary cultured astrocytes) and non-neuronal (lymphoblastoid) cells. We have also observed that the relative utilization of the three principal start sites is significantly altered for expanded-CGG alleles in the premutation range compared with normal alleles, indicating that the downstream CGG element has a direct influence on transcriptional initiation. The similar distributions of site utilization for cultured astrocytes and lymphoblastoid lines establish that the mechanism of start site selection for the *FMR1* gene is not cell-type-specific. Furthermore, the elevated levels of total *FMR1* mRNA in the premutation alleles, relative to the normal control, are consistent with an absolute increase in the utilization of sites II and III, not simply a shift away from site I. This latter observation suggests that the CGG element may be acting as a downstream enhancer for initiation at the more upstream sites (sites II and III).

No clones were identified that correspond precisely to the published transcription start site (6), although the principal start site in the control cell lines (site I) lies within 5 nt of the

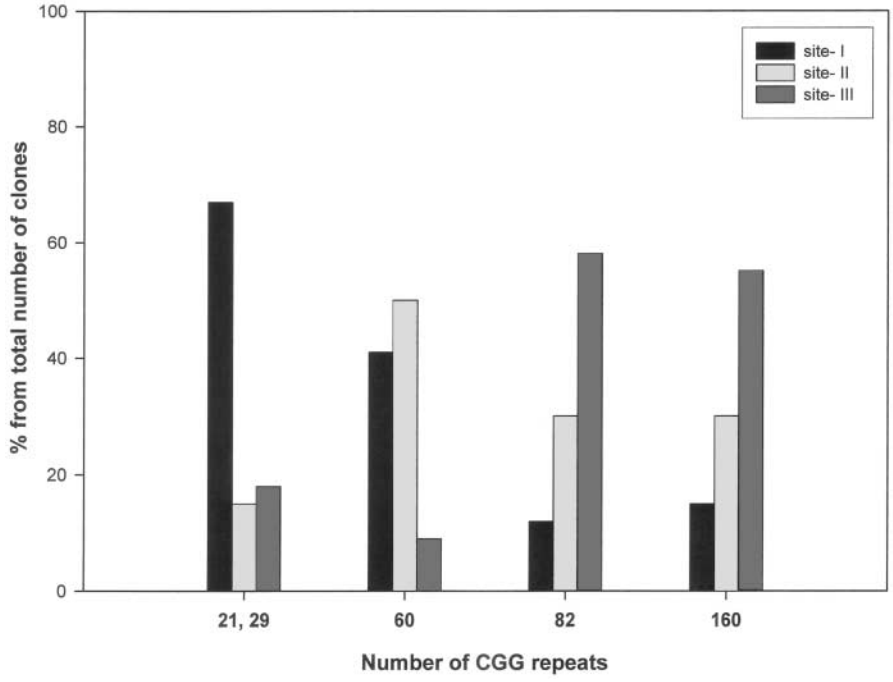

Figure 2. Distribution of transcription start sites for normal (21 or 29 CGG), and premutation (60, 82 or 160 CGG repeats) lymphoblastoid cell lines; 100% is the total number of sequenced clones for each lymphoblastoid cell line. Site I (black); site II (light gray); site III (dark gray).

Table 1. Summary of cell lines used in 5′-RLM-RACE experiments

Cell line	Cell type	CGG repeat[a]	mRNA level[b]	No. clones[c]
AG (normal, female)	Lymphoblastoid	21, 29	1.24 ± 0.19	27
SR (premutation)	Lymphoblastoid	60	2.50 ± 0.14	32
EB (premutation)	Lymphoblastoid	82	2.70 ± 0.19	33
MM (premutation)	Lymphoblastoid	160	4.48 ± 0.44	47
SC21 (normal)	Astrocyte	15	1.61 ± 0.21	31
SC17 (premutation)	Astrocyte	87	1.43 ± 0.04	28

[a]CGG repeat number determined by PCR.
[b]*FMR1* mRNA levels reported as relative value compared to normal controls (9,10).
[c]Number of sequenced clones in 5′RLM-RACE experiments using total RNA.

Table 2. χ^2 and *P*-values for differences in start site utilization for various pairs of alleles

CGG transition	χ^2	*P*
Human lymphoblastoid lines		
Normal to 60	8.14	0.017
Normal to 82	19.2	0.0001
Normal to 160	20.7	0.0000
82 to 160	0.128	0.94
60 to 82	17.8	0.0001
60 to 160	18.0	0.0001
Primary (human) astrocyte		
Normal to 87	7.71	0.021
Lymphoblastoid versus astrocyte		
Normal versus normal	0.839	0.66
60 (lymph) versus 87 (astrocyte)	0.806	0.67
82 (lymph) versus 87 (astrocyte)	14.3	0.0008

previously reported start site (6). It is possible that this residual difference is due to the lower precision of the primer extension method used previously, which was based on gel sizing. In addition, it is likely that the primer extension method lacked sufficient sensitivity to detect the minor transcription start sites (sites II and III) in a cell harboring normal alleles. An important advantage of 5′-RLM-RACE, relative to primer extension analysis as a means for accurate determination of the legitimate 5′ ends of mRNAs, is that all non-capped RNAs are eliminated as candidates for RACE by calf intestinal protein (CIP) dephosphorylation (13,14).

The *FMR1* promoter does not possess a typical TATA element, and deletion of the putative TATA-like sequence and an initiator-like (Inr) sequence (corresponding to site I only) does not dramatically reduce promoter activity (7). These observations support the idea that other transcription start sites as well as other important regulatory sequences are operating during transcriptional initiation from the *FMR1* gene. In this regard, our observation of multiple start sites is not surprising, since many TATA-less promoters transcribed by RNA polymerase II initiate transcription at multiple start sites (17).

The *FMR1* promoter region possesses three Inr sequences that correspond to sites I–III. The Inr-like sequences were identified by sequence similarity search using the Inr *consensus* sequence [YYAN(T/A)YY] (18,19) (Fig. 1). Inr sequences are usually located near transcription start sites and have been implicated in transcription initiation from TATA-less promoters (18,20). For the TATA-less promoters, binding of Inr-binding proteins to the Inr-elements on the promoter is thought to facilitate pol II complex formation either via interaction with TBP (TFIID) and/or other general and specific transcription factors, or by

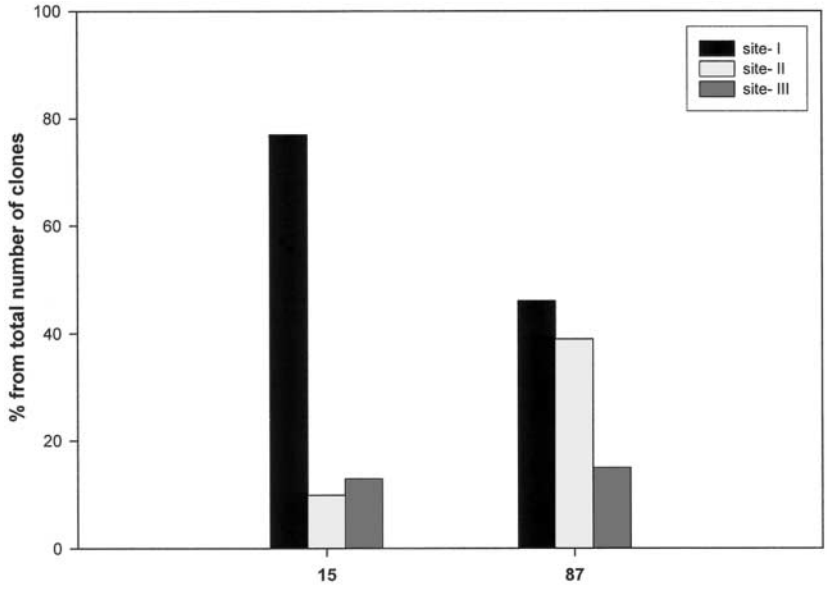

Figure 3. Distribution of transcription start sites for normal (15 CGG), and premutation (87 CGG) primary cultured astrocytes; 100% is the total number of sequenced clones for each primary astrocyte cell culture. Bars are as represented in Figure 2.

interacting with the CTD of RNA pol II (18). Among the several putative *cis*-regulatory elements within the promoter region of the *FMR1* gene, USF1/USF2 and alpha-Pal/Nrf-1 appear to be major regulatory factors that are involved in activation of the *FMR1* gene (7). Interestingly, USFs have been reported to be Inr-element-binding proteins (21,22), suggesting that these factors may be playing a role, perhaps via the Inr-like sequences associated with sites I–III, in the regulation of *FMR1* transcription.

Although we have demonstrated that transcriptional initiation from the *FMR1* gene is modulated by the downstream CGG repeat, the mechanism by which the CGG element influences start site selection is not known. One intriguing aspect of this coupling of CGG repeat size and start site utilization is that repeat expansion appears to direct the activation of the more upstream start sites. A precedent for such a modulator is the downstream element associated with the TATA-less promoter of the rat xanthine dehydrogenase/oxidase gene (*XDH*/*XO*), which contains four functional initiator sequences (Inr1, Inr2, Inr3 and Inr4) (20). In the absence of the downstream element, Inr 3 and Inr 4 are the major start sites; however, in the presence of the downstream element, the start site usage is shifted to Inr 1 and Inr 2, owing to YY-1 binding to the downstream element (23).

One possible mechanism by which the downstream CGG element could influence start site utilization would be through the interaction of a CGG binding protein with both the CGG element itself and components of the transcriptional regulatory machinery (e.g. USFs); such an interaction could be altered as a consequence of the increased length of the CGG repeat element. It is conceivable that the binding of additional CGG-BPs to the expanded repeat would lead to enhanced recruitment of pol II and increased production of *FMR1* mRNA, as is observed among premutation carriers (9,11).

It is also possible that the expanded CGG repeat element exerts a direct effect, in *cis*, by modulating the local chromatin architecture (e.g. nucleosome positioning), thus altering accessibility of transcription factors to regions within the proximal *FMR1* promoter. Previous studies have suggested that there may be a subtle effect of expanded CGG repeats on nucleosome affinity and/or positioning (24,25), and such positioning effects may influence start site selection.

In a more recent study of the effects of expanded CGG repeats on promoter activity, it was found that expanded CGG repeats had a repressive effect on HSV thymidine kinase activity upon injection of a reporter construct into *Xenopus* oocytes (26). Whereas the repressive effect was thought to be mediated by histone deacetylation, it is likely to be a specific phenomenon of the *Xenopus* system, or the use of a heterologous 5′-UTR, since such repressive effects are not observed for premutation alleles in patients (9,11) or in cultured cells that have been transfected with CGG repeats within the *FMR1* 5′-UTR context (27). However, the *Xenopus* study does raise the separate issue of how the CGG element may be contributing to the silencing process.

Finally, the use of post-mortem human brain tissue for the generation of cultured primary astrocytes with premutation CGG expansions represents an important advance in the study of fragile X syndrome. Using cultured astrocytes derived from brains with both normal and premutation alleles, we have shown that start site utilization in neural cells mirrors the pattern of initiation in peripheral tissues (lymphoblastoid cells). These results affirm that the mechanisms responsible for the differential distribution of these transcriptional start sites are functional in the brain and may thus represent a novel target for therapeutic intervention. These data also suggest that the harvesting of neural progenitor cells from these same tissues (16) may provide a preparation with which to investigate the effects of CGG expansions on both neuronal and astrocytic lineages. This latter capability should allow us to investigate the role of the CGG repeats *per se*, or the alternative start sites, in

the pathogenesis of the newly-discovered fragile X-associated tremor/ataxia syndrome (FXTAS) (28–30). These studies are currently underway.

In conclusion, this work extends our understanding of the mechanism of transcription initiation of the *FMR1* gene. Our results indicate that the CGG repeat element in the *FMR1* gene directly influences transcription initiation, raising the possibility that redistribution of the 5′ ends of the *FMR1* message could play a role in the reduced translation efficiency observed for premutation alleles.

MATERIALS AND METHODS

Cell culture

Lymphoblastoid cell lines established by Epstein–Barr virus transformation were maintained in RPMI 1640 medium supplemented with 10% fetal bovine serum (Gibco BRL), glutamine, and penicillin-streptomycin (Sigma) at 37°C in a humidified atmosphere of 5% CO_2.

Primary astrocyte cell cultures were established in accord with UC Davis IRB-approved human subjects protocols from post-mortem normal and premutation human brain tissue following the Schwartz protocol (16). Primary astrocytes were maintained in DMEM medium supplemented with 15% fetal bovine serum (Gibco BRL), glutamine and penicillin–streptomycin (Sigma) at 37°C in a humidified atmosphere of 5% CO_2. All lymphoblastoid cell lines and primary astrocytes cultures used in 5′-RLM-RACE studies are specified in Table 1.

PCR Southern blot hybridization analysis

DNA was extracted from cultured cells (Puregene DNA-isolation kit, Gentra) and analyzed by PCR with oligonucleotide primers specific for a portion of the *FMR1* genomic DNA sequence that includes the CGG repeat region. The PCR primers used were c and f as described by Fu (31). PCR products were separated on a 6% polyacrylamide sequencing gel and transferred to a nylon membrane (Roche). Hybridization was performed with DIG labeled oligonucleotide probe $(CGG)_{10}$ overnight in DIG Easy Hyb solution (Roche) at 42°C. Probe detection utilized anti-DIG antibody-alkaline phosphatase conjugate (Roche) with CDP-Star substrate (Roche). Further details of the PCR Southern blot hybridization method are reported by Tassone and Hagerman (32).

5′-RLM-RACE analysis of capped-RNA

Total RNA was isolated from lymphoblastoid cells and primary astrocytes using the Trizol method (Invitrogen). 5′-RLM-RACE was performed using GeneRacer kits (Invitrogen) according to the manufacturer's instructions. Three micrograms of total mRNA, or 0.5 µg of polyA(+) RNA, were used for each sample. Briefly, total mRNA, or polyA(+) RNA, was treated with calf intestinal phosphatase (CIP) to remove the 5′ phosphates from any truncated (i.e. non-capped) mRNA. Dephosphorylated RNA was treated with tobacco acid pyrophosphatase (TAP) to remove the 5′ cap from full-length mRNA, leaving a 5′ phosphate. The GeneRacer™ RNA oligomer was then ligated to the 5′ end of the mRNA using T7 RNA ligase. First-strand cDNA was produced with AMV-reverse-transcriptase using random hexamer primers. The regions corresponding to the legitimate 5′ ends of the capped-RNA species were PCR amplified from cDNA templates using the GeneRacer 5′ primer (5′-CGACTGGAGCACGAGGACACTGA-3′) and an *FMR1* gene-specific primer (5′-CTCGCCGTCGGCCCGC CGCC-3′). Nested PCR was performed with the GeneRacer 5′ nested primer (5′-GGACACTGACATGGACTGAAGGAGT A-3′) and the nested *FMR1* gene specific primer (5′-CCT CCACCGGAAGTGAAACCGAAA-3′). The position of both *FMR1*-specific primers is indicated in Figure 1. The PCR products of the 5′-RLM-RACE RT-PCR reactions were cloned into pCR 4.1 vector (TA Cloning Kit, Invitrogen), and ligated products were transfected into One-Shot chemically competent cells (Invitrogen). Plasmids were purified using plasmid mini or midi kits (Promega), and were screened for insertions following *Eco*RI digestion. Positive clones for each transfection reaction were sequenced by means of an automated sequencer (ABI PRISM 310 Genetic Analyzer, Perkin Elmer) with T7 or T3 primers. Table 1 lists the number of clones that were sequenced for each lymphoblastoid line and primary astrocyte culture.

Preparation of polyA(+) RNA

PolyA(+) mRNA from lymphoblastoid cell lines with normal (21 or 29 CGG) and premutation (82 CGG) alleles was prepared using the PolyAPure mRNA Isolation System (Ambion), which is based on the elution of polyA mRNA from oligo (dT) cellulose columns, according to the manufacturer's protocol. ~5 µg of PolyA(+) mRNA were isolated from 10^6 cells for each lymphoblastoid cell line.

Statistical analysis

Chi-squared analyses were performed by global comparison of the site distributions (number of clones for each site) between pairs of alleles.

ACKNOWLEDGEMENTS

The authors wish to thank Drs David Bentley and John Hershey for helpful discussions. We thank Lisa Becker for editorial assistance. This work was supported by a grant from the National Institute of Child Health and Development (P.J.H.; HD 40661), by the Boory and Cooper/Kraff/Fishman family funds (P.J.H.), by general laboratory support from the UC Davis MIND Institute and by Children's Hospital of Orange County Foundation for Children (P.H.S).

REFERENCES

1. Pieretti, M., Zhang, F.P., 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.
2. 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. *Nat. Genet.*, **4**, 335–340.
3. 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.

4. Hagerman, P.J. (2002) Gene expression and molecular approaches to therapy. In Hagerman, R.J. and Hagerman, P.J. (eds), *Fragile X Syndrome: Diagnosis, Treatment and Research*, 3rd edn. The Johns Hopkins University Press, Baltimore, MD, pp. 465–494.
5. Drouin, R., Angers, M., Dallaire, N., Rose, T.M., Khandjian, W. and Rousseau, F. (1997) Structural and functional characterization of the human FMR1 promoter reveals similarities with the hnRNP-A2 promoter region. *Hum. Mol. Genet.*, **6**, 2051–2060.
6. 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. Commun.*, **193**, 324–329.
7. Kumari, D. and Usdin, K. (2001) Interaction of the transcription factors USF1, USF2, and (alpha)-PAL/Nrf-1 with the *FMR1* promoter. *J. Biol. Chem.*, **276**, 4357–4364.
8. Schwemmle, S., de Graaff, E., Deissler, H., Glaser, D., Wohrle, D., Kennerknecht, I., Just, W., Oostra, B.A., Dorfler, W., Vogel, W. *et al.* (1997) Characterization of *FMR1* promoter elements by *in vivo*-footprinting analysis. *Am. J. Hum. Genet.*, **60**, 1354–1362.
9. Tassone, F., Hagerman, R.J., Chamberlain, W.D. and Hagerman, P.J. (2000) Transcription of the *FMR1* gene in individuals with fragile X syndrome. *Am. J. Med. Genet. (Semin. Med. Genet.)*, **97**, 195–203.
10. Tassone, F., Hagerman, R.J., Taylor, A.K., Gane, L.W., Godfrey, T.E. and Hagerman, P.J. (2000) Elevated levels of *FMR1* mRNA in carrier males: a new mechanism of involvement in fragile X syndrome. *Am. J. Hum. Genet.*, **66**, 6–15.
11. Kenneson, A., Zhang, F., Hagedorn, C.H. and Warren, S.T. (2001) Reduced FMRP and increased *FMR1* transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Hum. Mol. Genet.*, **10**, 1449–1454.
12. Primerano, B., Tassone, F., Hagerman, R.J., Hagerman, P., Amaldi, F. and Bagni, C. (2002) Reduced *FMR1* mRNA translation efficiency in fragile X patients with premutations. *RNA*, **8**, 1482–1488.
13. Volloch, V., Schweitzer, B. and Rits, S. (1994) Ligation-mediated amplification of RNA from murine erythroid cells reveals a novel class of beta globin mRNA with an extended 5′-untranslated region. *Nucl. Acids Res.*, **22**, 2507–2511.
14. Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano, S. (1997) Construction and characterization of a full length-enriched and a 5′-end-enriched cDNA library. *Gene*, **200**, 149–156.
15. Niikura, M., Dornadula, G., Zhang, H., Mukhtar, M., Lingxun, D., Khalili, K., Bagasra, O. and Pomerantz, R.J. (1996) Mechanisms of transcriptional transactivation and restriction of human immunodeficiency virus type I replication in an astrocytic glial cell. *Oncogene*, **13**, 313–322.
16. Schwartz, P.H., Bryant, P.J., Fuja, T.J., Su, H., O'Dowd, D.K. and Klassen, H. (2003) Isolation and characterization of neural progenitor cells from post-mortem human cortex. *J. Neurosci. Res.*, **74**, 838–851.
17. Kawai, Y., Asai, K., Miura, Y., Inoue, Y., Yamamoto, M., Moriyama, A., Yamamoto, N. and Kato, T. (2003) Structure and promoter activity of the human glia maturation factor-gamma gene: a TATA-less, GC-rich and bidirectional promoter. *Biochim. Biophys. Acta*, **1625**, 246–252.
18. Weis, L. and Reinberg, D. (1992) Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. *FASEB J.*, **6**, 3300–3309.
19. Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. and Smale, S.T. (1994) DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell. Biol.*, **14**, 116–127.
20. Chow, C.W., Clark, M.P., Rinaldo, J.E. and Chalkley, R. (1995) Multiple initiators and C/EBP binding sites are involved in transcription from the TATA-less rat XDH/XO basal promoter. *Nucl. Acids Res.*, **23**, 3132–3140.
21. Roy, A.L., Meisterernst, M., Pognonec, P. and Roeder, R.G. (1991) Cooperative interaction of an initiator-binding transcription initiation factor and the helix–loop–helix activator USF. *Nature*, **354**, 245–248.
22. Du, H., Roy, A.L. and Roeder, R.G. (1993) Human transcription factor USF stimulates transcription through the initiator elements of the HIV-1 and the Ad-ML promoters. *EMBO J.*, **12**, 501–511.
23. Clark, M.P., Chow, C.W., Rinaldo, J.E. and Chalkley, R. (1998) Correct usage of multiple transcription initiation sites and C/EBP-dependent transcription activation of the rat XDH/XO TATA-less promoter requires downstream elements located in the coding region of the gene. *Nucl. Acids Res.*, **26**, 1801–1806.
24. Wang, Y.H., Gellibolian, R., Shimizu, M., Wells, R.D. and Griffith, J. (1996) Long CCG triplet repeat blocks exclude nucleosomes: a possible mechanism for the nature of fragile sites in chromosomes. *J. Mol. Biol.*, **263**, 511–516.
25. Godde, J.S., Kass, S.U., Hirst, M.C. and Wolffe, A.P. (1996) Nucleosome assembly on methylated CGG triplet repeats in the fragile X mental retardation gene 1 promoter. *J. Biol. Chem.*, **271**, 24325–24328.
26. Chandler, S., Kansagra, P. and Hirst, M. (2003) Fragile X $(CGG)_n$ repeats induce a transcriptional repression in *cis* upon a linked promoter: evidence for a chromatin mediated effect. *BMC Mol. Biol.*, **4**, 3.
27. Chen, L.S., Tassone, F., Sahota, P. and Hagerman, P.J. (2003) The $(CGG)_n$ repeat element within the 5′ untranslated region of the FMR1 message provides both positive and negative *cis* effects on *in vivo* translation of a downstream reporter. *Hum. Mol. Genet.*, **12**, 3067–3074.
28. Hagerman, R.J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., Grigsby, J., Gage, B. and Hagerman, P.J. (2001) Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology*, **57**, 127–130.
29. Greco, C.M., Hagerman, R.J., Tassone, F., Chudley, A.E., Del Bigio, M.R., Jacquemont, S., Leehey, M. and Hagerman, P.J. (2002) Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain*, **125**, 1760–1771.
30. Jacquemont, S., Hagerman, R.J., Leehey, M., Grigsby, J., Zhang, L., Brunberg, J.A., Greco, C., Des Portes, V., Jardini, T., Levine, R. *et al.* (2003) Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *Am. J. Hum. Genet.*, **72**, 869–878.
31. Fu, Y.H., Kuhl, D.P., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J., Holden, J.J., Fenwick, R.G., Jr., Warren, S.T. *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.
32. Tassone, F. and Hagerman, P.J. (2003) Expression of the *FMR1* gene. *Cytogenet. Genome Res.*, **100**, 124–128.