ANALYSIS OF FMRP, THE PROTEIN DEFICIENT IN FRAGILE X SYNDROME

Kersten Small and Stephen T. Warren
Howard Hughes Medical Institute and Departments of Biochemistry and Pediatrics, Emory University School of Medicine,
Atlanta, Georgia

Fragile X syndrome results from a massive trinucleotide repeat expansion in the 5’ untranslated region of the gene FMR1. Methylation of FMR1 occurs as a result of this expansion, causing transcriptional silencing. Therefore, the absence of FMR protein (FMRP) appears to result in fragile X syndrome. Characterization of the gene and its expression, as well as the cellular localization and functional properties of the protein, provide insight into how the absence of FMRP results in mental retardation and the related phenotype. Alternative splicing of FMR1 transcripts leads to protein isoforms and suggests that this cell is functionally diverse. Among vertebrates, FMRP is highly conserved at the amino acid level and localizes mainly to the cytoplasm. The presence of two protein sequence motifs, the RGG box and the KH domain, suggests that FMRP is an RNA-binding protein. Direct experimentation shows that FMRP is capable of binding its own RNA in vitro, as well as a subset of brain-derived messages. Further understanding of how these interactions take place and identification of those genes having messages that interact will provide insight into the mechanisms of cognitive function in humans. © 1995 Wiley-Liss, Inc.

Key Words: trinucleotide repeat expansion, gene expression, protein function, RNA-binding proteins, mental retardation, X chromosome

Fragile X syndrome, the most common form of inherited mental retardation, is associated with a fragile site on the X chromosome at position Xq27.3 [Lubs, 1969]. This chromosomal abnormality, observed as a gap in up to 50% of metaphase chromosomes, results from in vitro cell culture conditions that effect deoxyribonucleotide synthesis [Sutherland, 1977]. In 1991, the molecular basis for this fragile site was elucidated and shown to be associated with a massive trinucleotide repeat expansion within the gene fragile X mental retardation-1 (FMR1) [Kremer et al., 1991; Oberlé et al., 1991; Verkerk et al., 1991; Yu et al., 1991].

FMR1 is an evolutionarily conserved gene that spans approximately 38 kilobases (kb) of DNA on the X chromosome at the position of the fragile site (Fig. 1) [Verkerk et al., 1991; Eichler et al., 1993]. A CGG trinucleotide repeat is located within the 5’ untranslated region (UTR) of the first exon of this gene [Ashley et al., 1993a]. In normal human this repeat is highly polymorphic and ranges in size from 6 to 52 repeats, with the most common size being 30 repeats [Fu et al., 1991]. In patients with fragile X syndrome, this repeat has undergone a tremendous expansion, ranging from hundreds to more than a thousand repeats [Snow et al., 1993; Rousseau et al., 1993]. Fragile X carriers have an FMR1 allele with an intermediate level of expansion, typically between 60 and 200 CGG repeats, known as a premutation. Current evidence indicates that this expanded region of trinucleotide repeats is the putative fragile site itself, and possibly is the result of the late replication of the FMR1 region when the repeat is expanded [Hansen et al., 1993a].

As a result of the repeat expansion, abnormal methylation of both a nearby CpG island and the repeat itself takes place, causing transcriptional silencing of FMR1 [Bell et al., 1991; Oberlé et al., 1991; Pieretti et al., 1991; Sutcliffe et al., 1992; Hansen et al., 1993b]. Methylation does not occur in premutation alleles with normal transcription and normal FMR1 protein levels [Feng et al., 1995]. The absence of FMR1 mRNA, however, has been reported in the majority of fragile X patients analyzed. It is consistent with these results that antibodies fail to detect FMR protein (FMRP) in the lymphocytes of fragile X patients (Fig. 2) [Devys et al., 1993; Verheij et al., 1993]. Several deletion mutations as well as intragenic loss of function mutations in FMR1 have been characterized in rare patients with the fragile X phenotype [Gedeon et al., 1992; Wöhrl et al., 1992; Melier et al., 1994; Hirsh et al., 1995; Ludden et al., in press]. In addition, FMR1 knockout mice display a mild phenotype consistent with fragile X syndrome [The Dutch-Belgian Fragile X Consortium, 1994]. Thus, fragile X syndrome occurs as a consequence of the absence of FMRP. Understanding the normal cellular functions of this protein will be the key to deciphering the molecular pathways involved in this syndrome and may lead to possible therapeutic approaches for treating fragile X patients. Furthermore, the elucidation of FMRP function should provide insight into biochemical aspects of cognitive function in humans.

Analysis of FMR1 sequence and expression has provided the initial clues to how FMRP might function within the cell. FMR1 is highly conserved throughout evolution, showing cross-hybridization with DNA from a wide range of species [Verkerk et al., 1991]. In humans, FMR1 is composed of 17 exons that code for an approximately 4.4 kb cDNA (Fig. 1)

© 1995 Wiley-Liss, Inc.
[Eichler et al., 1993]. The CGG repeat is located within the noncoding portion of the first exon, and a relatively long (2.3 kb) 3' UTR has been described. Alternative splicing has been demonstrated in the 3' portion of this gene, creating the potential for individual protein isoforms with novel C-termini that may have important functional significance [Ashley et al., 1993a; Verkerk et al., 1993].

Both mouse and chicken FMR1 cDNAs have been characterized; they show strong sequence conservation within the coding regions, respectively having 98% and 91% amino acid identity with humans [Ashley et al., 1993a; Price et al., in press]. Significant homology between the 5' and 3' UTRs has also been noted. But whereas the mouse cDNA contains nine CGG repeats, the chicken cDNA contains a TCC repeat flanked by two short dinucleotide repeats in the 5' untranslated region. Also, exons 11 and 12 are not present in the chicken message. Although both mouse and chicken FMR proteins have been shown to bind RNA in vitro [Ashley et al., 1993b; Price et al., in press], the additional exonic information in the mammalian FMR1 genes may reflect enhancement of function in higher species. Also, the second KH domain found in the mouse and human genes is interrupted by an intron that is absent from the chicken gene. This is consistent with a functional protein unit, in this instance the KH domain, evolving from single exonic units. The strong homology between chicken, mouse, and human cDNAs supports the functional significance of FMR1 in normal individuals.

In order to determine how loss of FMR1 gene expression and subsequent protein expression results in the clinical manifestations of fragile X syndrome, the normal expression pattern for this gene has been examined. Northern analysis of normal human tissues shows high levels of FMR1 mRNA in brain and testis, which is consistent with the fragile X phenotype. Relatively high levels of message were also detected in heart, lung, kidney, and placenta. Low to negligible levels were found in liver, pancreas, and skeletal muscle [Hinds et al., 1993].

In situ detection of mRNA in sections of mouse tissues gives a more informative picture of the distribution of FMR1 transcripts [Hinds et al., 1993]. In agreement with the Northern results obtained from human tissues, FMR1 mRNA shows widespread distribution in murine tissues as well. In the adult mouse brain, expression appears to be strongest in neuronal cells of the granular layers of the hippocampus and cerebellum and in discrete cell types of the cerebral cortex and habenula. Interestingly, very high levels of FMR1 expression, having an almost uniform distribution, appear during the early stages of the mouse embryo, suggesting that FMRP may play an important role during embryonic development. The results of studies by Abitbol et al. [1993], in which in situ hybridization of human FMR1 message was characterized in human fetal brain, are consistent with the expression patterns observed in the mouse. FMR1 is also strongly expressed in the follicular cells of the ovary and in the seminiferous tubules of the testis, although a decline in expression levels occurs early in adult life [Bachner et al., 1993]. No consistent ovarian involvement has been described in fragile X patients, but the macroorchidism displayed in most postpubescent fragile X males may reflect connective tissue dysplasia in the testis.

The widespread but low levels of FMR1 expression in most tissues examined may suggest a general housekeeping role for FMRP, whereas the higher levels of expression observed in brain and testis may be indicative of a special function for FMRP in these tissues [Bachner et al., 1993]. The mental retardation, as well as the neurological and behavioral disorders of the fragile X syndrome, may be associated with regional brain dysfunction: a general correlation exists between these regions and regions of FMR1 expression [Reiss and Freund, 1990a, b; Reiss et al., 1991]. Indeed, the physical features of this syndrome may reflect inadequate FMR1 expression during critical stages of development [Mintz and Russell, 1957].

Immunohistochemical studies also show the widespread expression of FMRP in adult human tissues [Devys et al., 1993]. In patterns of expression similar to those previously described for mRNA, FMRP was also detected in testes and in neuronal cells of the brain. In the cerebellum, the granular layer and the Purkinje cells at the interface of the granular and molecular layers show the highest levels of FMRP, although neurons in the cortex also show FMRP. In almost all cells examined, FMRP localization appears to be cytoplasmic (Fig. 3).
Epithelial cells show less abundant levels of protein than are observed in neurons. Despite the presence of FMR1 message, tissues of mesodermal origin show no significant levels of protein [Devys et al., 1993].

Western analysis using monoclonal antibodies directed against the N-terminal portion of the protein [Verheij et al., 1993] reveals multiple FMR1-specific protein bands at approximately 80 kD (Fig. 2). These bands are likely to correspond to different isoforms of FMRP. The apparent molecular mass determined by SDS-polyacrylamide gel electrophoresis (PAGE) is larger than expected from the predicted amino acid sequence (69 kD). This discrepancy...
Fig. 5. Binding of RNA to FMRP. A: Competition of $^{35}$S FMRP-binding by the addition of unlabeled FMRP. FMR1 RNA was transcribed in vitro in the presence of biotin-uridine 5' triphosphate (UTP), after which 80 ng of biotinylated RNA was mixed with a constant amount of in vitro-translated $^{35}$S FMRP (2 μl) and increasing amounts (0.5–4 μl) of unlabeled competitor FMRP. RNA-protein complexes were captured with streptavidin-linked magnetic beads. After several washes, the bound protein was released by denaturation and analyzed by SDS-PAGE and fluorography. Densitometry was used to estimate amounts of bound protein. Bound $^{35}$S FMRP (y axis) was plotted against amounts of unlabeled competitor FMRP. Squares denote individual data points. B: Selectivity of FMRP binding to random human fetal brain messages. Plasmid DNA from individual colonies in a human fetal brain library (Invitrogen, San Diego, CA) was used as a template for in vitro transcription in the presence of biotin UTP. The binding of in vitro-translated $^{35}$S FMRP to 80 ng biotinylated RNA from three representative brain clones is shown (lane 2, clone 19.1, 0.7 kb; lane 3, clone 19.2, 1.0 kb; lane 4, clone 19.3, 1.0 kb) as compared to biotinylated FMR1 RNA alone (lane 1). RNA transcribed from clone 19.1 (lane 2) demonstrates binding equivalent to that of FMR1 RNA, whereas RNA produced from the other clones demonstrated little or no interaction with FMRP.

between molecular mass and mobility has been observed in proteins that contain stretches of acid-rich amino acids [Hope and Struhl, 1986; Swanson et al., 1987] that are also present in FMRP. Alternatively, the slower mobility on SDS-PAGE may reflect posttranslational modifications of the protein that have yet to be determined.

Alternative splicing of the FMR1 gene predicts at least 12 potential isoforms, suggesting that FMRP is functionally diverse [Verkerk et al., 1993]. Alternative splicing of exon 12 in both mouse and human FMR1 transcripts has been demonstrated, as has alternative splicing of exon 14, which alters the transcript reading frame, thereby creating potential FMRP isoforms with novel C-termini [Ashley et al., 1993a; Eichler et al., 1993]. Alternative use of splice acceptor sites has been shown in exons 10, 15, and 17. Significant differences between the hydrophobicity profiles of different FMRP isoforms further supports the concept that this protein is functionally diverse. Although FMRP localization has been shown to be primarily cytoplasmic [Verheij et al., 1993], some nuclear localization has been observed. Also, transfection studies have shown that when the N-terminal portion of the protein is translated in COS cells, it localizes to the nucleus [Devys et al., 1993]. Therefore, it is possible that alternatively spliced FMR1 products may have function in the nucleus as well as in the cytoplasm. Protein isoforms with novel C-termini may have tissue-specific or developmental stage-specific functions, and may be partitioned to distinct regions within tissues such as those of the brain.

Examination of the predicted protein sequence of FMRP shows that it contains sequence motifs that are present in many RNA-binding proteins (Fig. 4) [Ashley et al., 1993b; Sionni et al., 1993a]. One such domain is a region rich in arginine and glycine that is known as the RGG box [for review, see Burd and Dreyfuss, 1994]. At the carboxyl end, FMRP has two RGG boxes that show a high degree of sequence homology with fibrillarin and pre-mRNA-binding protein A1, two previously described RNA-binding proteins [Aris and Blobel, 1991; Buvoli et al., 1988]. More centrally located within the protein are two amino acid repeats with similarity to the KH motif found in a family of ribonucleoproteins [Sionni et al., 1993b]. KH motif proteins are a diverse group that includes the heterogeneous nuclear ribonucleoprotein (hnRNK), ribosomal S3 protein, bacterial polynucleotide phosphorylase, and a meiosis-specific alternative splicing factor in yeast known as mer-1 [Reigner et al., 1987; Engebrecht and Roeder, 1990; Zheng et al., 1990; Sionni et al., 1993]. All KH domain proteins of known function are associated with RNA and many have been shown to bind RNA in vitro [Burd and Dreyfuss, 1994]. Interestingly, a severe and unique case of fragile X syndrome results from a point mutation (Ile to Asn at position 304) in a highly conserved hydrophobic residue within the second KH domain of FMRP [De Boulle et al., 1993].
Experimental evidence shows that FMRP does bind RNA in vitro (Fig. 5) [Ashley et al., 1993b; Sioni et al., 1993a]. In vitro translated FMRP can bind to its own message as well as to poly(G) and poly(U), synthetic homopolymer RNA molecules. Truncated forms of the protein that delete the RGG box show little or no binding to RNA, suggesting that the RGG box is essential for RNA binding [Sioni et al., 1993a, Warren unpublished data]. Also, experiments by Sioni et al. [1994] reveal that two Ile to Asn mutations within each KH motif (one of which is present in the severely affected fragile X patient described earlier) prevent FMRP from binding to poly(U).

Ashley et al. [1993b] provided evidence that FMRP selectively binds to a subset of brain-derived RNA, with approximately 4% by mass of human fetal brain message specifically binding FMRP (Fig. 5). A strategy developed by Tsi et al. [1992] may help elucidate which RNA sequences are recognized by FMRP. A methodology based on polymerase chain reaction can be used to identify specific RNA epitopes from a library of random RNA sequences. This technique may identify specific binding sites or structural determinants required for FMRP RNA-binding. Determination of the RNA targets of FMRP binding will be an important step toward understanding how the absence of a single protein results in the pleiotropic phenotype observed in fragile X syndrome.

The functions of RNA-binding proteins are quite diverse. Heterogeneous nuclear RNA must be processed in the nucleus, transported to the cytoplasm, and ultimately translated into protein. RNA-binding proteins are involved in all of these processes as well as in the regulation of message stability, potentially providing many levels of post-transcriptional gene regulation. Because of its primarily cytoplasmic localization, FMRP may have a role in the translational regulation of specific messages.

The study of how proteins regulate gene expression via RNA interactions is becoming an increasingly exciting area of research. The discovery of new RNA-binding proteins such as FMRP will provide a challenge for future research. How, when, and where FMRP interacts with RNA, which particular RNA molecules are bound by FMRP, and the consequence of this interaction are the key questions that need to be answered in order to understand fragile X syndrome.

ACKNOWLEDGMENTS
KS is an associate and STW is an investigator with the Howard Hughes Medical Institute. The authors thank Yue Feng, Derek Eberhart, and Fuping Zhang for help with the figures.

REFERENCES
Siomni H, Siomi MC, Nishbaum RL et al. (1993a): The protein product of the fragile X gene,


