Characterization
of the fragile X syndrome gene products

Mikiko Chihara Siomi

1994

Acknowledgments

The author wishes to express her sincere and heartfelt gratitude to Dr. Gideon Dreyfuss, Investigator of the Howard Hughes Medical Institute and Professor of Biochemistry and Biophysics at the University of Pennsylvania, School of Medicine, for his continuous advice and encouragement.

The author wishes to express the most hearty appreciation to Professor Jun'ichi Oda, Institute for Chemical Research, Kyoto University, for his continuous guidance and support.

The author wishes to thank these colleagues and ex-colleagues in the Dreyfuss labratory: Serafin Piñol-Roma, Michael J. Matunis, Erika L. Matunis, Matthias Görach, Megerditch Kiledjian, Christopher Burd, Joeng-Kook Kim, Mieyoung Choi, Douglas Portman, W. Matthew Michael, Qing Liu, Gloria Kim, and William H. Sauer. The author could not have accomplished this work without their invaluable advice, discussions, kind encouragement and great tolerance.

The author wishes to thank Professor Akira Hasegawa and Professor Makoto Kiso, Gifu University, Associate Professor Takaaki Nishioka, Institute for Chemacal Research, Kyoto University, Professor Yasuo Hotta and Associate Professor Satoshi Tabata, Nagoya University, Professor Akira Kimura, Institute for Food Science, Kyoto University, Associate Professor Yukio Yamamoto, Kyoto University, and Professor Naomichi Baba, Okayama University, for their kind encouragement.

The author is very much grateful to Dr. Hiroaki Kato, Dr. Jun Hiratake, and Dr. Takuji Tanaka, for their kind encouragement and support.

The author thanks the members of the Oda labratory for their kind help.

Finally, the author thanks all of the Chiharas and the Siomis who have been essential during this whole process, especially Dr. Haruhiko Siomi, whose involvment in all aspects of this endeavor has been fundamental.

Mikiko Chihara Siomi

Contents

Chapter 5

General introduction

Fragile X mental is one of the most common human genetic diseases and the most common cause of hereditary mental retardation, affecting approximately 1 in 1200 males and 1 in 2500 females (Sutherland et al., 1985; Nussbaum and Ledbetter, 1986; Richards and Sutherland, 1992; Oostra and Verkerk, 1992; Caskey et al., 1992). The clinical features include variable but generally severe mental retardation, resulting in IQ. of 20-60, a typical facial appearance, and enlarged testicles in adult males. The fragile X syndrome is strongly associated with a chromosomal fragile site, a segmental gap of poorly staining chromatin at Xq27.3 (Sutherland, 1977; Tommerup, 1989), that can be induced to appear in ~50% of metaphases from affected males if cells are starved for precursors of DNA synthesis during the preceding S phase. The genetics of this X-linked disorder are unusual: Approximately 30% of carrier females exhibit mental deficiency while 20% of males with the fragile X chromosome are non-penetrant, phenotypically normal individuals (Sherman et al., 1984; Sherman et al., 1985). These males in turn will transmit the chromosome to daughters, who are consistently unaffected, and they may have affected grandsons. This phenomenon of anticipation, with the risk of mental impairment in fragile X pedigrees appearing to be contingent upon the position of individuals in the pedigrees, has commonly been referred to as the "Sherman paradox".

Fragile X syndrome is caused by the amplification of a simple trinucleotide repeat $(CGG)_n$ located within the 5' untranslated region of the fragile X mental retardation gene, *FMR1* (Verkerk et al., 1991). Two classes of mutations have been described (Oberle et al., 1991; Fu et al., 1991; Rousseau et al., 1991): premutation are characterized by moderate expansions of the trinucleotide repeat (n=54-200) and do not cause mental retardation. They are found in normal transmitting males and in normal female carriers (including all daughters of normal transmitting males); full mutations are present in affected males or females patients and are characterized by larger expansions (n>250) associated with an abnormal methylation of the CpG island which includes the CGG repeat (Oberle et al., 1991; Hansen et al., 1992). The transition from premutation to full mutations occurs only by transmission through a female carrier at a frequency

which depends on the size of the premutation (Fu et al., 1991; Yu et al. 1992; Heitz et al., 1992).

There is strong evidence that defects in *FMR1* play a central role in the clinical syndrome. First, amplification of the trinucleotide repeat is associated with hypermethylation of the CpG island 5' to *FMR1* (Bell et al., 1991; Heitz et al., 1991; Oberle et al., 1991; Vincent et al., 1991) and a marked decrease or extinction of steady-state levels of *FMR1* mRNA (Pieretti et al., 1991; Verheij et al., 1993). Additional evidence for a direct role of FMR1 is that three patients with a partial or complete deletion of *FMR1* exhibit fragile X syndrome in the absence of the fragile site (Gedeon et al., 1992; Wohrle et al., 1992; Meijer et al., 1994). The finding of a single point mutation in the open reading frame of *FMR1* in a patient with very severe fragile X syndrome, but without cytogenetic expression of the fragile site (De Boulle et al., 1993), also suggests that *FMR1* is directly responsible for fragile X syndrome. Recently, fmr1 knockout mice have been reported. These knockout mice lack normal *FMR1* RNA and protein and show enlarged testes, impaired cognitive function, and aberrant behavior. This animal model might serve as a valuable tool in the elucidation of the physiological role of FMR1 and the mechanisms involved in macroorchidism, abnormal behavior, and mental retardation (The Dutch-Belgian Fragile X Consortium, 1994).

Although its apparently ubiquitous expression and extraordinary cross-species conservation (Verkerk et al., 1991) suggest that it may have a 'housekeeping' role, the normal function of FMR1 is not known. The predicted amino acid sequence of the *FMR1* gene product has been published (Verkerk et al., 1991), as no homology to proteins characterized previously has been reported. *FMR1* expression is widespread as shown by northern blot analysis on human tissues and by in situ hybridization in mouse, and appears strong in brain and testis which are involved in the clinical phenotype (Hinds et al., 1993).

To characterize the FMR1 protein, the predicted amino acid sequence of FMR1 is examined. In Chapter 1, it is shown that the predicted amino acid sequence of FMR1 contains two RNA binding motifs that strongly suggest that it is an RNA-binding protein: the K homology (KH) domain (Siomi et al., 1993; Gibson et al., 1993b; Ashley et al., 1993a) and an RGG box (Kiledjian and Dreyfuss, 1992), and FMR1 does indeed bind to RNA (Ashley et al., 1993). The

identification of the KH domains in FMR1 provides a framework for thinking about the nature of the Ile304 to Asn missense mutation in a patient with severe fragile X syndrome (De Boulle et al., 1993). Ile304 at one of the most conserved residues of the KH domain (Siomi et al., 1993; Gibson et al., 1993a), indicating that the conserved residues of the KH domains are required for normal function of FMR1. In Chapter 2, it is demonstrated that KH domains of FMR1 protein play a role in RNA binding and is shown that the RNA binding activity of the human FMR1 Ile304 to Asn mutant found in the severely retarded fragile X patient is strongly impaired.

The diagnosis of the fragile X syndrome was confirmed initially by finding 50% expression of the fragile site in transformed lymphoblastoid cells (Sutherland et al., 1985; Nussbaum and Ledbetter, 1986; Richards and Sutherland, 1992) and later by Southern blot analysis of the *FMR1* region (Kremer et al., 1991; Yu et al., 1991; Rousseau et al., 1991). In Chapter 3, the production of antibodies to FMR1 is described. The Western blotting results using the antibodies show that the patients of the fragile X syndrome have absent or markedly reduced levels of FMR1 protein (Devys et al., 1993; Verheij et al., 1993). The antibodies to FMR1 will offer a specific and sensitive diagnosis tool for fragile X syndrome.

To better understand the function and evolution of the *FMR1* gene product, *FMR1* cDNA is cloned from Xenopus laevis. In the course of the experiment, a novel gene, termed *FXR1*, that is highly homologous by amino acid sequence to *FMR1* is discovered. FXR1, like FMR1, contains two KH domains and an RGG box and is a cytoplasmic RNA-binding protein. These findings indicate that *FMR1* is not a one-of-a-kind gene but rather is a member of a gene family. The antibodies to X. laevis FMR1 and FXR1 are raised and show that these proteins are expressed in X. laevis oocytes, indicating that the *FMR1* and *FXR1* mRNAs could be maternal. In Chapter 4, all about Xenopus laevis *FMR1* and *FXR1* are described. *FXR1* cDNA is also isolated from human and completely sequenced, which is described in Chapter 5. Interestingly, cells of a fragile X patient that do not have any detectable FMR1 express normal levels of FXR1. Unlike *FMR1*, *FXR1* is not located on the X chromosome: rather it is an autosomal gene located at 12q13. Moreover, unlike *FMR1*, there is no CGG repeats in the 5' UTR of the *FXR1* mRNA. Instead, *FXR1* mRNA has a most

unusual feature; both X, laevis and human *FXR1* mRNAs contain an approximately 90 nucleotide perfect inverted repeat in the 5' and 3' UTRs. It suggests that *FXR1* expression is subject to particular post-transcriptional regulation or that the *FXR1* mRNA has an unusual function.

Chapter 1

The protein product of the fragile X gene, *FMR1*, has characteristics of an RNA-binding protein

1-1 Introduction

Although the gene, *FMR1*, which causes fragile X syndrome has been recently cloned and sequenced (Figure 1-1), the normal function of FMR1 protein product is not still known. The expression of the protein is apparently ubiquitous in living cells and extraordinary cross-species conserved (Verkerk et al., 1991), which suggest that it may have a 'housekeeping' role. The amino acid sequence of the *FMR1* gene product can be predicted from the DNA sequence of the gene (Verkerk et al., 1991), but its protein product has not yet been identified and there are no biological parameters to assess the activity of FMR1. To give a clue to study the function of FMR1, the homology search for the peptide sequence is carried out. It is found that the predicted amino acid sequence of the *FMR1* gene product contains two motifs that strongly suggest that it is an RNA binding protein, one of which is the K homology (KH) domain (Siomi et al., 1993; Gibson et al., 1993b; Ashley et al., 1993a) and the another one is an RGG box (Kiledjian and Dreyfuss, 1992; Ashley et al., 1993a). In this chapter, it is demonstrated that FMR1 does indeed bind to RNA in vitro. This observation suggests possible functions for FMR1 in RNA metabolism or in RNA-containing cellular structures and focuses on the role of defective FMR1 function in the pathogenesis of this disorder.

1-2 Experimental Procedures

cDNA cloning and construction of expression plasmids for the FMR1 proteins

FMR1 cDNA fragment was isolated by polymerase chain reaction (PCR). Two PCR oligodeoxynucleotide primers were synthesized; one in a 5' to 3' direction within the *FMR1* coding regions (27XM7:1272-1301) and one in a 3' to 5' direction within the *FMR1* 3' noncoding region (27X31:2124-2156; the coordinates are based on the numbering used in Verkerk et al., 1991). 2mg of oligo (dT)-selected HeLa RNA was reverse transcribed using the primer 27X31 according to the manufacturer's suggested conditions (Perkin Elmer Cetus).

```
  1 GACGGAGGCGCCCGTGCCAGGGGGCGTGCGGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAGGC
101 GGCGGCGGCGGCGGCGGCGGCTGGGCCTCGAGCGCCCGCAGCCCACCTCTCGGGGGCGGGCTCCCGGCGCTAGCAGGGCTGAAGAGAAG ATG
                                                                                              M   1

199 GAG GAG CTG GTG GTG GAA GTG CGG GGC TCC AAT GGC GCT TTC TAC AAG GCA TTT GTA AAG GAT GTT CAT GAA GAT
     E   E   L   V   V   E   V   R   G   S   N   G   A   F   Y   K   A   F   V   K   D   V   H   E   D   26

274 TCA ATA ACA GTT GCA TTT GAA AAC AAC TGG CAG CCT GAT AGG CAG ATT CCA TTT CAT GAT GTC AGA TTC CCA CCT
     S   I   T   V   A   F   E   N   N   W   Q   P   D   R   Q   I   P   F   H   D   V   R   F   P   P   51

349 CCT GTA GGT TAT AAT AAA GAT ATA AAT GAA AGT GAT GAA GTT GAG GTG TAT TCC AGA GCA GCA AAT GAA AAG CCT
     P   V   G   Y   N   K   D   I   N   E   S   D   E   V   E   V   Y   S   R   A   A   N   E   K   P   76

424 TGC TGT TGG TGG TTA GCT AAA GTG AGG ATG ATA AAG GGT GAG TTT TAT GTG ATA GAA TAT GCA GCA TGT GAT GCA
     C   C   W   W   L   A   K   V   R   M   I   K   G   E   F   Y   V   I   E   Y   A   A   C   D   A  101

499 ACT TAC AAT GAA ATT GTC ACA ATT GAA CGT CTA AGA TCT GTT AAT CCC AAC AAA CCT GCC ACA AAA GAT ACT TTC
     T   Y   N   E   I   V   T   I   E   R   L   R   S   V   N   P   N   K   P   A   T   K   D   T   F  126

574 CAT AAG ATC AAG CTG GAT GTG CCA GAA GAC TTA CGG CAA ATG TGT GCC AAA GAG GCG GCA CAT AAG GAT TTT AAA
     H   K   I   K   L   D   V   P   E   D   L   R   Q   M   C   A   K   E   A   A   H   K   D   F   K  151

649 AAG GCA GTT GGT GCC TTT TCT GTA ACT TAT GAT CCA GAA AAT TAT CAG CTT GTC ATT TTG TCC ATC AAT GAA GTC
     K   A   V   G   A   F   S   V   T   Y   D   P   E   N   Y   Q   L   V   I   L   S   I   N   E   V  176

724 ACC TCA AAG CGA GCA CAT ATG CTG ATT GAC ATG CAC TTT CGG AGT CTG CGC ACT AAG TTG TCT CTG ATA ATG AGA
     T   S   K   R   A   H   M   L   I   D   M   H   F   R   S   L   R   T   K   L   S   L   I   M   R  201

799 AAT GAA GAA GCT AGT AAG CAG CTG GAG AGT TCA AGG CAG CTT GCC TCG AGA TTT CAT GAA CAG TTT ATC GTA AGA
     N   E   E   A   S   K   Q   L   E   S   S   R   Q   L   A   S   R   F   H   E   Q   F   I   V   R  226

874 GAA GAT CTG ATG GGT CTA GCT ATT GGT ACT CAT GGT GCT AAT ATT CAG CAA GCT AGA AAA GTA CCT GGG GTC ACT
     E   D   L   M   G   L   A   I   G   T   H   G   A   N   I   Q   Q   A   R   K   V   P   G   V   T  251

949 GCT ATT GAT CTA GAT GAA GAT ACC TGC ACA TTT CAT ATT TAT GGA GAT GAT GCA GTG AAA AAA GCT AGA
     A   I   D   L   D   E   D   T   C   T   F   H   I   Y   G   E   D   Q   D   A   V   K   K   A   R  276

1024 AGC TTT CTC GAA TTT GCT GAA GAT GTA ATA GAA GTT CCA AGG AAC TTA GTA GTA ATA GGA AAA AAT GGA AAG CTG
      S   F   L   E   F   A   E   D   V   I   Q   V   P   R   N   L   V   V   I   G   K   N   G   K   L  301

1099 ATT CAG GAG ATT GTG GAC AAG TCA GGA GTT GTG AGG GTG AGG ATT GAG GCT GAA AAT GAC AAA AAT GTT CCA CAA
      I   Q   E   I   V   D   K   S   G   V   V   R   V   R   I   E   A   E   N   D   K   N   V   P   Q  326

1174 GAA GAG GAA ATT ATG CCA CCA AAT TCC CTT CCT TCC AAT AAT TCA AGG GGT GGA CCT AAT GCC CCA GAA GAA AAA
      E   E   E   I   M   P   P   N   S   L   P   S   N   N   S   R   G   G   P   N   A   P   E   E   K  351

1249 AAA CAT TTA GAT ATA AAG GAA AAC AGC ACC TAT TTT TCA CAA CCT ATG ACT ACA AAA GTC CAG AGG GGT ATG GTA
      K   H   L   D   I   K   E   N   S   T   Y   F   S   Q   P   M   T   T   K   V   Q   R   G   M   V  376

1324 CCA TTT GTT TTT GTG GGA ACA AAG GAC AGC ATC GAT CAT GCC ACT GTT TTG GAT TAT TAC CAC CTG AAC TAT TTA
      P   F   V   F   V   G   T   K   D   S   I   A   N   A   T   V   L   D   Y   H   L   N   Y   L  401

1399 AAG GAA GTA GAC CAG TTG CGT TTG GAG AGA TTA CAA ATT GAT GAG CAG TTG CGA CAG ATT GGA GCT AGT TCT AGA
      K   E   V   D   Q   L   R   L   E   R   L   Q   I   D   E   Q   L   R   Q   I   G   A   S   S   R  426

1474 CCA CCA CCA AAT CGT ACA GAT AAG GAA AAA AGC TAT GTG ACT GAT GAT GGT CAA GGA ATG GGT CGA GGT AGT AGA
      P   P   P   N   R   T   D   K   E   K   S   Y   V   T   D   D   G   Q   G   M   G   R   G   S   R  451

1549 CCT TAC AGA AAT AGG GGG CAC GGC AGA CGC GGT CCT GGA TAT ACT TCA GGA ACT AAT TCT GAA GCA TCA AAT GCT
      P   Y   R   N   R   G   H   G   R   R   G   P   G   Y   T   S   G   T   N   S   E   A   S   N   A  476

1624 TCT GAA ACA GAA TCT GAC CAC AGA GAC GAA CTC AGT GAT TGG TCA TTA GCT CCA ACA GAG GAA GAG AGG GAG AGC
      S   E   T   E   S   D   H   R   D   E   L   S   D   W   S   L   A   P   T   E   E   E   R   E   S  501

1699 TTC CTG CGC AGA GGA GAC GGA CGG CGG CGT GGA GGG GGA GGA AGA GGA CAA GGA GGA AGA CGT GGA GGA GGC
      F   L   R   R   G   D   G   R   R   R   G   G   G   G   R   G   Q   G   G   R   G   R   G   G   G  526

1774 TTC AAA GGA AAC GAC GAT CAC TCC CGA ACA GAT AAT CGT CCA CGT AAT CCA AGA GAG GCT AAA GGA AGA ACA ACA
      F   K   G   N   D   D   H   S   R   T   D   N   R   P   R   N   P   R   E   A   K   G   R   T   T  551

1849 GAT GGA TCC CTT CAG AAT ACC TCC AGT GAA GGT AGT CGG CTG CGC ACG GGT AAA GAT CGT AAC CAG AAG AAA GAG
      D   G   S   L   Q   N   T   S   S   E   G   S   R   L   R   T   G   K   D   R   N   Q   K   K   E  576

1924 AAG CCA GAC AGC GTG GAT GGT CAG CAA CCA CTC GTG AAT GGA GTA CCC TAA ACTGCATAATTCTGAAGTTATATTTCCTATAC
      K   P   D   S   V   D   G   Q   Q   P   L   V   N   G   V   P   *

2007 CATTTCCGTAATTCTTATTCCATATTAGAAAACTTTGTTAGGCCAAAGACAAAATAGTAGGCAAGATGGGCACAGGGCATGAAATGAACACAAATTATGCTA
2107 AGAATTTTTTATTTTTTTTTTTTTTTTT... (sequence continues)
```

Figure 1-1. The DNA sequence and the predicted amino acid sequence
of human FMR1.   (Verkerk et al., 1991)

PCR reactions were done on 50ul of cDNA solution with the primers 27XM7 and 27X31 according to the manufacturer's suggested conditions (Perkin Elmer Cetus). The amplified fragment was digested with KpnI and cloned into pGEM7Z (Promega) cut with KpnI-SmaI to create pFX1. Human *FMR1* full length cDNA was obtained by screening of a λgt11 HeLa cell cDNA library with the RT-PCR fragment (pFX1) as a probe. A cDNA was digested with EcoRI-KpnI and cloned into pFX1 cut with EcoRI-KpnI to create pF27X. The nucleotide sequence of the insert was verified by DNA sequencing. The insert contains a fragment of *FMR1* cDNA corresponding to the coding region of *FMR1* (nucleotides 197-2132, based on the numbering used in Verkerk et al., 1991). Sequencing the pF27X revealed an insertion of 6 bases, GCAAAG, positioned between 881 and 882. This change added Gly-Lys codons. The correct amino acid sequence including these two amino acid residues is listed in Figure 1-2B. The fragment EcoRI-NsiI of pF27X was then inserted into the EcoRI and NsiI cut pHHSI, under transcriptional control of the T7 promoter, to create plasmid pHHSI-F27X. The pHHSI plasmid was derived from pcDNAI (Invitrogen) by insertion of a 72-base pair BstXI fragment encoding a nine amino acid epitope from the influenza virus hemaglutinin protein HA1 (Wilson et al., 1984) with an initiator methionine followed by SmaI, SalI and EcoRI restriction sites.

In vitro transcription and translation

pHHSI F27X was linearized with NsiI to generate template for in vitro RNA synthesis of the full length of *FMR1* with T7 RNA polymerase. To generate C-terminal truncated mutants of FMR1, the plasmid was linealized with either AccI or SspI, followed by in vitro RNA systhesis with T7 RNA polymerase. The resultant RNAs were translated in rabbit reticulocyte lysate in the presence of [^{35}S] methionine (Amersham) according to the manufacturer's suggested conditions (Promega Biotech).

RNA binding assays

Binding of in vitro produced proteins to ribohomopolymers was carried out essentially as described in Swanson and Dreyfuss (1988) with minor modifications. Briefly, ribonucleotide homopolymer (Pharmacia or Sigma)

binding reactions were carried out with an equivalent of 10^5 counts per min (c.p.m.) of trichloroacetic acid (TCA)-precipitable protein in a total of 0.5ml of binding buffer [10mM Tris-HCl(pH7.4), 2.5mM $MgCl_2$, 0.5% Triton X-100, 2mg/ml Pepstatin, 2mg/ml Leupeptin, 0.5% Aprotinin] with the stated NaCl concentration for 10 min on a rocking platform at 4°C. The beads were pelleted with a brief spin in a microfuge and washed five times with binding buffer prior to resuspension in 50 ul of SDS-PAGE loading buffer. Bound protein was eluted from the nucleic acid by boiling, resolved on a 12.5% SDS-PAGE gel and visualized by fluorography.

Expression and purification of fusion protein

The expression vector pET15 F2 was constructed by cleaving pHHSI F27X with XhoI and BamHI and inserting a XhoI-BamHI fragment [nucleotide 845-1855, based on the numbering used in (Verkerk et al., 1991)] of FMR1 into the XhoI and BamHI sites of the plasmid pET15b-HA. The pET15b-HA was derived by insertion of a duplex DNA linker containing the codons for the nine amino acid HA epitope with 5' NdeI and 3' XhoI overhangs into the NdeI and XhoI sites of pET15b vector (Novagen). For production of the His-HA-FMR1 fusion protein, the plasmid pET15 F2 was introduced into BL21(DE3)pLysE bacteria and induced with isopropyl-β-D-thiogalactopyranoside(IPTG) as described (Studier et al., 1990; Rosenberg et al., 1987). Cells expressing reasonable amounts of the fusion protein were screened by using anti-HA mouse monoclonal 12CA5. For purification of the fusion protein, bacterial sonicate was incubated at 4°C with 1x packed volume of DEAE-Sephacel (Pharmacia). The resulting supernatant solution was applied to a 2.5-ml His·Bind^{TM}Resin (Novagen) column, washed and eluted as described by the manufacturer.

Northwestern blotting assay

Poly(G) was 5' end labeled with T4 polynucleotide kinase. Blots immobilizing the overexpressed and purified His-HA-FMR1 fusion peptide were treated for 1h at room temperature in binding buffer [10mM Tris-HCl(pH7.4), 50mM NaCl, 1mM EDTA, 1x Denhardt's solution). The blots were then probed at room temperature for 1h with ^{32}P-labeled poly(G)RNA (100,000 cpm per lane) in binding buffer containing 20mg of E. coli tRNA (Boehringer Mannheim

Biochemicals) and 1mg of heparin (porcine intestinal mucosa) per ml. Blots were washed three times for 15 minutes each with binding buffer, air dried, and exposed to X-ray film for autoradiography.

1-3 Results

FMR1 has sequence motifs characteristic of RNA binding proteins

In examining the predicted sequence of FMR1 (Verkerk et al., 1991), it has been noticed that it contains sequence motifs characteristic of RNA-binding proteins. First, several RNA binding proteins have recently been found to contain an arginine and glycine-rich domain that contains a cluster of the tripeptide repeat (Arg-Gly-Gly) called the RGG box (Kiledjian and Dreyfuss, 1992). This motif has been found in a considerable number of nuclear and nucleolar RNA-binding proteins and has been demonstrated to have RNA-binding activity (Kiledjian and Dreyfuss, 1992; Dreyfuss et al., 1993). Figure 1-2A shows the sequence of an RGG box near the carboxyl end of FMR1. This RGG box bears particularly striking similarity to those found in two other proteins, fibrillarin (Aris and Blobel, 1991) and pre-mRNA binding (hnRNP) A1 (Buvoli et al., 1988), both of which are RNA-binding proteins (Tyc and Steitz, 1989; Cobianchi et al., 1988; Nadler et al., 1991). Second, amino acids 286-321 and 347-382 (Figure 1-2B) are two internal repeats. These repeats have significant similarity to each other (39% similarity) but they bear even stronger similarity to a sequence motif, the K homology (KH) domain, that has been recently described in the hnRNP K protein (Siomi et al., 1993) (Figure 1-2B). It consists of the highly conserved I/L/V-I-G-X_2-G-X_2-I sequence and regularly spaced hydrophobic residues. This sequence motif, which extends over ~40 amino acids, was originally found in the hnRNP K protein (Siomi et al., 1993; Matunis et al., 1992; Dreyfuss et al., 1993) and in several other proteins including the archeabacterial ribosomal protein S3 (Spiridonova et al., 1989) and the yeast meiosis-specific splicing regulator MER-1 (Engebrecht and Roeder, 1990; Engebrecht et al., 1991). We found that KH domains are also present in several additional proteins, including the GAP-associated tyrosine phosphoprotein p62 (Wong et al., 1992), a differentiation-associated protein, vigilin (Schmidt et al., 1992) and a glycine-rich putative hnRNP protein, GRP33 (Cruz-Alvarez and
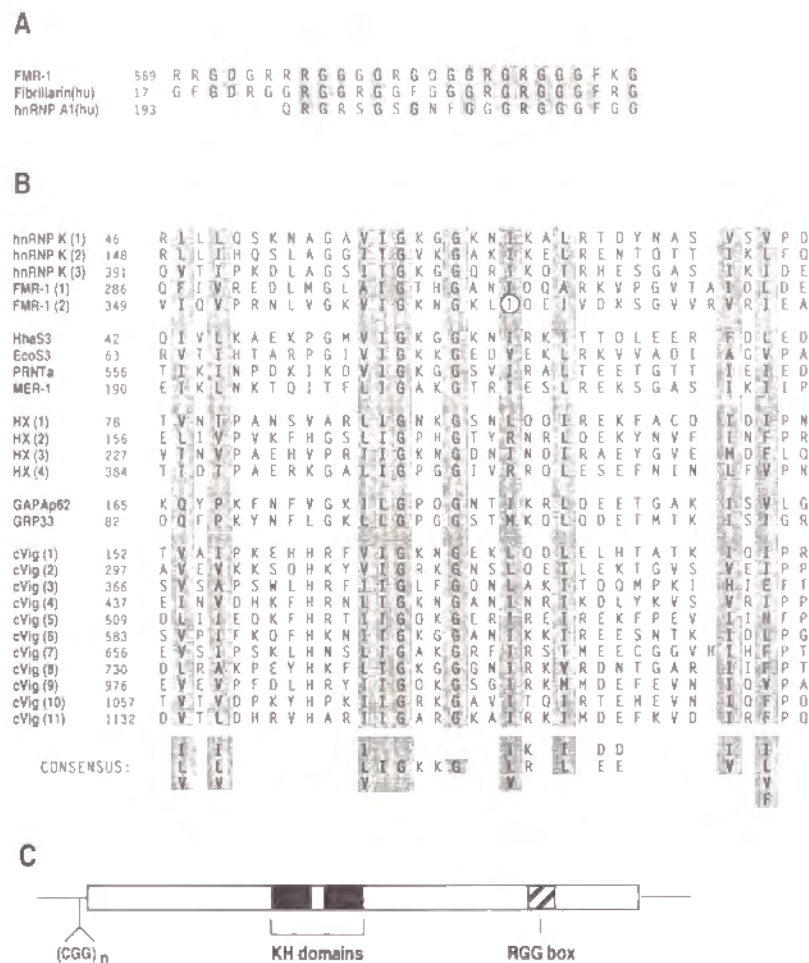
A

```
FMR-1          569   R R G D G R R R G G G Q R G Q G G R G R G G G F K G
Fibrillarin(hu) 17   G F G D R G G R G G G R G G F G G G R G R G G G F R G
hnRNP A1(hu)   193                 Q R G R S G S G N F G G G R G G G F G G
```

B

```
hnRNP K (1)    46    R L L L Q S K N A G A V I G K G G K N I K A L R T D Y N A S   V S V P D
hnRNP K (2)    148   R L L I H Q S L A G G I T G V K G A K I K E L R E N T Q T T   I K L F Q
hnRNP K (3)    391   Q V T I P K D L A G S I T G K G G Q R I K Q I R H E S G A S   I K I D E
FMR-1 (1)      286   Q F I V R E D L M G L A I G T H G A N I Q Q A R K V P G V T A I Q L D E
FMR-1 (2)      349   V I Q V P R N L V G K V I G K N G K L Q E A V D K S G V V R V R V R I E A

HhaS3          42    Q I V L K A E K P G M V I G K G G K N I R K Y T T Q L E E R   F D L E D
EcoS3          63    R V T I H T A R P G I V I G K K G E D V E K L R K V V A D I   A G V P A
PRNTa          556   T I K I N P D K I K Q V I G K G G S V I R A L T E E T G T T   I E E E D
MER-1          190   E T K L N K T Q I T F L I G A K G T R I E S L R E K S G A S   I K I I P

HX (1)         78    T V N T P A N S V A R L I G N K G S N L Q Q I R E K F A C Q   L D I P N
HX (2)         156   E L I V P V K F H G S L I G P H G T Y R N R L Q E K Y N V F   I N F P R
HX (3)         227   V T N V P A E H V P R I I G K N G D N I N D I R A E Y G V E   M D F L Q
HX (4)         384   T I Q T P A E R K G A L I G P G G I V R R Q L E S E F N I N   L F V P N

GAPAp62        165   K Q Y P K F N F V G K I L G P Q G N T I K R L Q E E T G A K   I S V L G
GRP33          82    Q Q F P K Y N F L G K L L G P G G S T M K Q L Q D E T M T K   I S I G R

cVig (1)       152   T V A I P K E H H R F V I G K N G E K L Q D L E L H T A T K   I Q I P R
cVig (2)       297   A V E V K K S Q H K Y V I G R K G N S L Q E I L E K T G V S   V E I P P
cVig (3)       366   S V S A P S W L H R F I I G L F G Q N L A K T T Q Q M P K I   H I E F T
cVig (4)       437   E I N V D H K F H R N L I G K N G A N I N R I K D L Y K V S   V R I P P
cVig (5)       509   D L I I E D K F H R T I I G Q K G E R I R E I R E K F P E V   L I N F P
cVig (6)       583   S V P I F K Q F H K N I I G K G G A N I K K I R E E S N T K   I Q L P G
cVig (7)       656   E Y S I P S K L H N S L I G A K G R F I R S I M E E C G G V H I H F P T
cVig (8)       730   Q L R A K P E Y H K F L I G K G G G N I R K Y R D N T G A R   L I I F P T
cVig (9)       976   E V E V P F D L H R Y I I G K G G S G I R K H M D E F E V N   I Q V P A
cVig (10)      1057  T V T V D P K Y H P K I I G R K G A V I T Q I R T E H E V N   I Q F P Q
cVig (11)      1132  Q V T L D H R V H A R I I G A R G K A I R K I M D E F K V D   I R F P Q
```

```
CONSENSUS:           I        I                      I I G K K G        I K    I      D D                          I   I
                     L        L                                            R    L      E E                          V   L
                     V        V                                                                                         V
                                                                                                                        F
```

C

(CGG)n   KH domains   RGG box

Figure 1-2.   Primary structure of the FMR1 protein and its sequence similarity with RNA-binding proteins.

(A) Comparison of the RGG box in the FMR1 protein to similar domains in other RNA-binding proteins.
(B) The KH domains in the protein show similarity to a number of known and putative RNA-proteins.
(C) Schematic representation of structural motifd identified in the FMR1 protein. The open bar indicates the coding region.

Pellicer, 1987) (Figure 1-2B). Because several of the proteins that contain KH domains bind RNA, it appears very likely that this domain is involved in RNA binding. It is of interest that a single mutation in FMR1, I367N [based on the numbering used in (Verkerk et al., 1991)], which changes one of the most highly conserved residues of the KH motif to asparagine, can cause fragile X syndrome (De Boulle et al., 1993), (Figure 1-2B). The finding that a highly conserved residue within the KH domain is required for normal function of FMR1 supports the view that the KH domain is of functional significance. The FMR1 KH1 domain is most similar to KH3 of the human and Xenopus laevis hnRNP K protein (14 identities and 18 similarities over 36 amino acids), and FMR1 KH2 is most similar to KH4 of chicken vigilin (14 identities and 17 similarities over 36 amino acids) and to KH domains in hnRNP K, archaebacterial ribosomal protein S3 and polynucleotide phosphorylase (11 identities each). Taken together, these sequence features strongly suggest that FMR1 is an RNA binding protein. The position of these motifs in FMR1 is indicated in Figure 1-2C.

FMR1 binds RNA in vitro

To test the possibility that FMR1 is an RNA-binding protein directly, we cloned *FMR1* [predicted open reading frame beginning with methionine 66 (Caskey et al., 1992): based on the numbering used in (Verkerk et al., 1991)] into an expression vector that contains a T7 RNA polymerase promoter for in vitro transcription as a fusion protein with the influenza hemaglutinin peptide (HA) as an epitope tag starting at the putative initiator methionine. The transcription product was translated in vitro and the protein product was assayed for RNA binding activity using RNA homopolymers immobilized on agarose beads, an assay that has been useful to assess RNA binding for many other RNA-binding proteins (Kiledjian and Dreyfuss, 1992; Swanson and Dreyfuss, 1988). The translation product of FMR1 had an apparent molecular weight by SDS-PAGE of 85K daltons, larger than expected from the predicted amino acid sequence (Figure 1-3A). This discrepancy between molecular mass and mobility in SDS-PAGE is frequently observed for proteins that contain stretches rich in acidic amino acids such as hnRNP C1/C2 (Swanson et al., 1987) and the yeast transcriptional activator GCN4 (Hope and Struhl, 1986). FMR1 contains several clusters of acidic amino acids (e.g. immediately amino terminal to the RGG box)
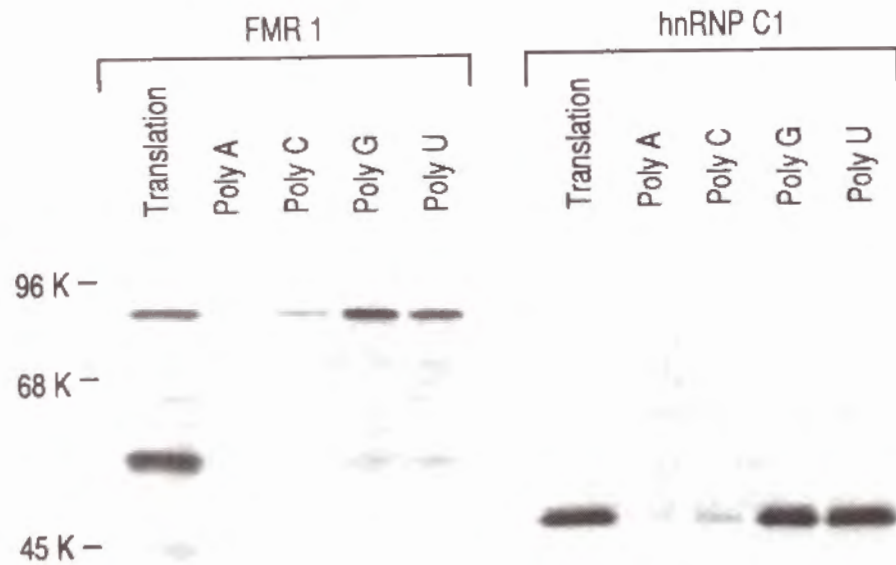
Figure 1-3. RNA binding properties of FMR1 protein.

(A) Binding of the FMR1 protein to ribinucleotide homopolymers. An amount equivarent to 20% of the materiual used for each binding reaction is shown in the lanes marked "Transltion". In vitroproduced proteins were bound to the indicated ribonucleotide homopolymers at 100mM NaCl. The position of molecular weight markers are indicated on the left.

and this may explain its slower mobility in SDS-PAGE. The FMR1 polypeptide showed strong binding to poly(G), a little weaker but significant binding to poly(U), and very little binding to poly(A) and poly(C). The failure of the major 50kD translation by-product to bind to the RNA homopolymers further verifies the RNA-binding activity of the FMR1 protein. The hnRNP C1 protein, a tenacious RNA binding protein used as a control in the same experiment, showed binding to poly(U), less to poly(G) and very little binding to poly(A) and poly(C), as previously demonstrated (Swanson and Dreyfuss, 1988). The binding of FMR1 to poly(G) and poly(U) was stable in NaCl concentrations up to 0.25M (Figure 1-3B). Both FMR1 and C1 showed stronger binding to ribohomopolymers than to ssDNA (data not shown), and preferential binding to some ribohomopolymers. This binding profile is characteristic of RNA binding proteins (Piñol-Roma et al., 1988).

To assess the role of the conserved sequence motifs described above in RNA binding, we produced translation products of truncated transcripts generated by digestion of the *FMR1* cDNA with AccI and SspI (Figure 1-3C) and compared the RNA binding ability with the full length FMR1. Whereas the full length protein bound well to poly(G) and poly(U) at 0.1M NaCl, the truncated proteins showed little or no binding to either. The polypeptide produced from the AccI-truncated construct is of particular interest as it is missing the RGG box but still contains the KH domains and it does not bind RNA, supporting the possibility that the RGG box is essential for RNA binding.

The RNA binding activity of FMR1 and its fragments was also confirmed by Northwestern blotting. With this approach, the proteins, immobilized on nitrocellulose membrane after SDS-PAGE, are probed with radioactively-labeled RNA and visualized by autoradiography. For this experiment, a fragment of FMR1, which contains all the KH motifs and the RGG box, was fused at the carboxyl terminus to His6-HA. This permitted rapid purification from E. coli extracts on a nickel column (Hochuli et al., 1987; Smith et al., 1988) and afforded detection with anti-HA antibodies. Repeated attempts to produce the full length protein in E. coli were unsuccessful. The purified material contained three bands, the largest of which (ca. 50K), corresponded to the desired polypeptide, and the two smaller ones corresponded to carboxyl terminal deletions as determined by their reactivity with anti-HA antibody (data not shown). Figure 1-4 shows that
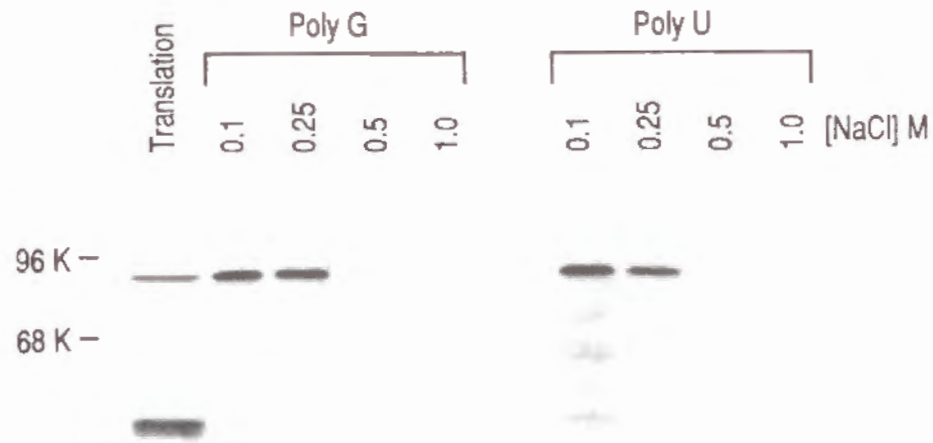
Figure 1-3. RNA binding properties of FMR1 protein.

(B) FMR1 binds poly(G) and poly (U) in a salt resistant manner. In vitro translated FMR1 protein was bound to the indicated ribonucleotide homopolymers at the indicated salt concentrations.
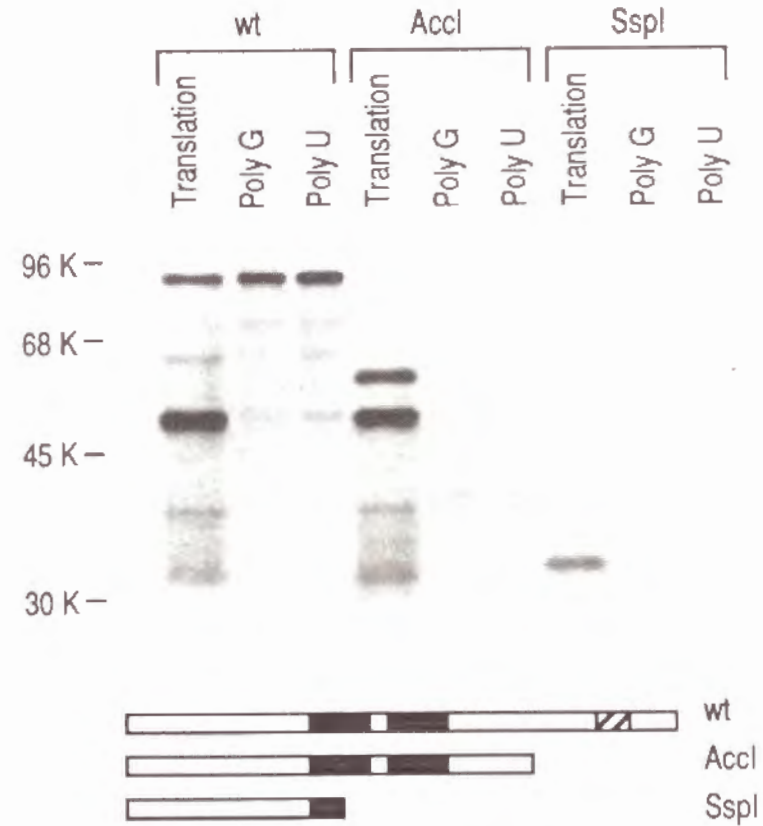
Figure 1-3. RNA binding properties of FMR1 protein.

(C) The ribonucleotide homopolymer binding domain is located at the C-terminal end of the FMR1 protein. wt, AccI and SspI represent for the full length FMR1, the truncated FMR1 by AccI and SspI site respectively. The structure of the C-terminal deletion mutants is shown below, the stippled boxed representing the KH domains, and the cross-hatched box, the RGG box of the FMR1. In vitro produced proteins were bound to the indicated ribonucleotide homopolymers at 200mM NaCl concentration.
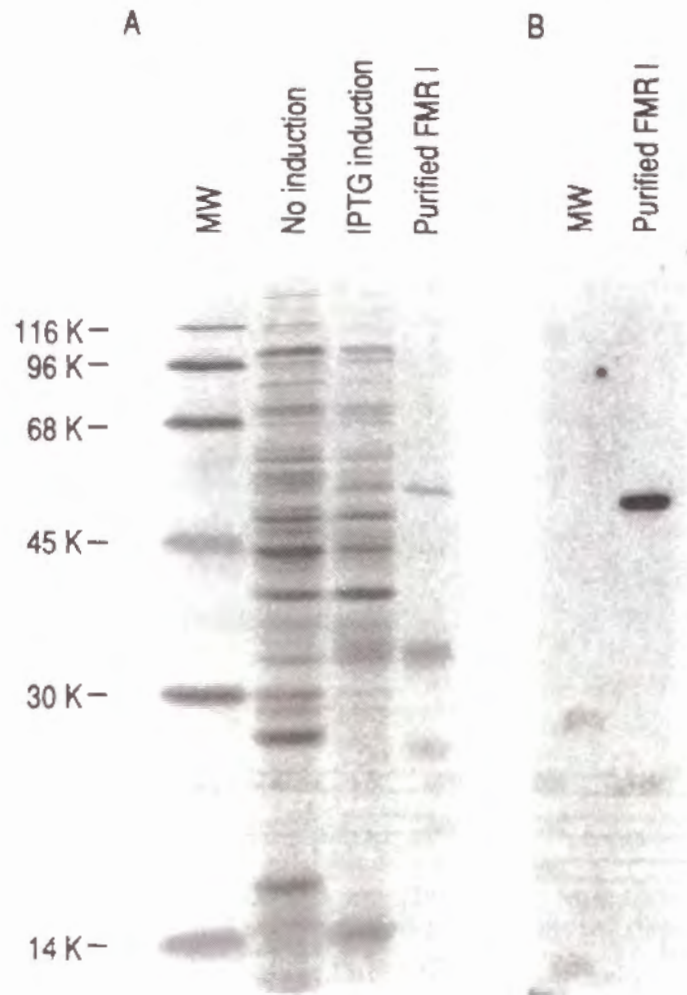
Figure 1-4. RNA binds to immobilized FMR1 fusion protein in a heparine-resistant manner.

(A) Production of FMR1 fusion protein. E. coli strain BL21(DE3)pLysE (lane; no induction) harboring the plasmid pET15 F2 containing a portion of the FMR1 coding sequence was induced with IPTG (lane; IPTG induction) and lysed. The IPTG-induced cell lysate was then purified by metal chelation chromatography (lane; "purified FMR1"). Protein size markers are shown (lane; "MW").

(B) Northwestern blot analysis. 2mg of purified recombinant FMR1 protein and 2mg of each of protein size markers were subjected to SDS gel electrophoresis and transfered to nitrocellulose membrane. The filter was incubated with ^{32}P-labeled poly(G) in 50mM NaCl plus 1mg of heparin (porcine intestinal muscosa) per ml and autoradiographed.

the full length product, but not the smaller fragments, bound [^{32}P]-poly(G) at 50mM NaCl in the presence of 1mg/ml heparin. The heparin-resistant binding (Swanson and Dreyfuss, 1988; Piñol-Roma et al., 1988) and the lack of binding to the proteins in the molecular weight marker lane (lysozyme, carbonic anhydrase, ovalbumin, bovine serum albumin, phosphorylase b and β-galactosidase) demonstrate the stringency and specificity of the binding of FMR1 to RNA. Based on these experiments, it is concluded that FMR1 is an RNA binding protein in vitro and thus has the capacity to bind RNA also in vivo.

1-4 Discussion

FMR1 plays a central role in the pathogenesis of fragile X syndrome. Sequence similarity between FMR1 and other RNA binding proteins was described and it was demonstrated that the FMR1 protein can bind RNA directly in vitro. This focuses the investigation of the function of this protein on its RNA binding activity and, ultimately, on the role of defective FMR1 function in the pathogenesis of this disorder. The RNA-binding activity of FMR1 was also demonstrated by Ashlay et al. (Ashlay et al., 1993a), in which it was shown that FMR1 selectively bind not only to 4% by mass of the human fetal brain message, but also to sense (Kd=5.7nM in this case) or antisense *FMR1* own RNA. These data confirmed my results.

RNA binding proteins can be involved in a wide range of cellular processes in the nucleus and the cytoplasm (Dreyfuss et al., 1988; Frankel et al., 1991) and they can regulate gene expression post-transcriptionally, including regulation of pre-mRNA splicing, mRNA stability, translation efficiency, and possibly the transport of RNAs between the nucleus and the cytoplasm. Several RNA binding proteins can also function as DNA binding proteins (e.g. TFIIIA) and regulate transcription of specific genes (Honda and Roeder, 1980; Pelham and Brown, 1980; Murray et al., 1992). A role for RNA binding proteins as developmental regulators has also been previously noted (Bandziulis et al., 1989), and several RNA binding proteins that affect development of the nervous system in Drosophila, particularly *elav* (Robinow et al., 1988), have been described. To determine the function of FMR1 it will be necessary to determine which cellular RNA(s) it interacts with and its subcellular localization. Additional information

obtained from the study of proteins that contain similar structural features such as hnRNP K (containing KH domains) and fibrillarin (containing RGG box) should also yield valuable information about the structure and mechanism of action of FMR1. The arginine residues in the RGG box of several proteins such as fibrillarin and hnRNP A1 have been found to be methylated to dimethylarginines (Christensen and Fuxa, 1988; Lischwe et al., 1985; Riva et al., 1986). It will be of interest to determine if FMR1 is also methylated and if so, to determine the effect of this posttranslational modification on the activity of this protein.

Both the KH domain and the RGG box are strong predictors of RNA binding activity. It is of interest that fragile X syndrome can result not only from lack of expression of the protein but also from expression of a mutant FMR1 protein with point mutation (Ile 367 to Asn) in the KH domain (De Boulle et al., 1993). Ile 367 is one of the most highly conserved residues of the KH domain (see Figure 1-2B), and this suggests that the highly conserved residues of the KH domain have important function. It will be important to determine if the point mutation Ile 367 to Asn in FMR1 which causes fragile X syndrome has a reduced RNA binding activity. Although it is premature to speculate on the specific function of FMR1, the fact that the protein is not essential for viability raises various interesting possibilities such as that it may somehow regulate the expression (e.g.: at the level of splicing or mRNA stability) of specific mRNAs in the nervous system.

1-5 Summary

Fragile X syndrome is one of the most common human genetic diseases and the most common cause of hereditary mental retardation. The gene that causes fragile X syndrome, *FMR1*, was recently identified and sequenced and found to encode a putative protein of unknown function. In this chapter it has been reported that FMR1 contains two types of sequence motifs recently found in RNA binding proteins: an RGG box and two hnRNP K homology (KH) domains and demonstrated that FMR1 binds to RNA in vitro. The RNA-binding activity of FMR1 opens the way to understanding the function of FMR1.

Chapter 2

Essential role for the KH domains of FMR1 in RNA binding

2-1 Introduction

The vast majority of patients with fragile X syndrome show a folate-sensitive fragile site at Xq27.3 (FRAXA) at the cytogenetic level, and both amplification of the (CGG)n repeat and hypermethylation of the CpG island in the *FMR1* at the molecular level. The *FMR1* gene of a patient with the fragile X phenotype but without cytogenetic expression of FRAXA, a (CGG)n repeat of normal length and an unmethylated CpG island has been studied and it was found that a T to A single point mutation in *FMR1* resulted in an Ile 304 [based on the numbering used in (Caskey et al., 1992)] Asn substitution (De Boulle et al. 1993). This *de novo* mutation was absent in the patients' family and in 130 control X chromosomes, suggesting that the mutation causes the clinical abnormalities. In other words, the mutations in *FMR1* are directly responsible for fragile X syndrome, irrespective of possible secondary effects caused by FRAXA.

Ile304 is at one of the most highly conserved residues of the KH domain (Figure 2-1), which was originally described in the pre-mRNA-binding (hnRNP) K protein, contains approximately 50 amino acids and was found in a diverse group of proteins many, if not all of which, are RNA-binding proteins. These include the archaebacterial ribosomal protein S3, the yeast meiosis-specific splicing regulator MER1, the *E.coli* antiterminator NusA and the human GAP-associated p62 phosphoprotein (Siomi et al., 1993; Gibson et al., 1993a). A common feature to these proteins with KH domains is physical or functional association with RNA molecules, implying that the KH domain is both involved in RNA binding and participates in regulating RNA metabolism (Siomi et al., 1993; Gibson et al., 1993a). However experimental evidence for a function of the KH domain is still lacking.

It is expected that Ile304, or the KH domain which contains the residue is required for the normal function of FMR1, possibly for RNA binding. Since the residue was changed, the Ile304 to Asn, FMR1 mutant would not bind to RNA. To confirm that, the Ile304 to Asn mutant and also the analogous mutant in the
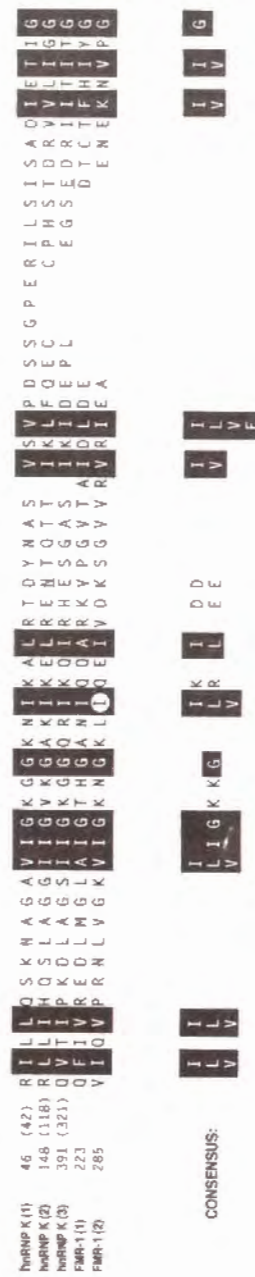
Figure 2-1. Amino acid alignment of KH domains od hnRNP K and FMR1.

KH domain sequence of hnRNP K (Matunis et al., 1992) are aligned with the sequence of FMR1 (Verkerk et al., 1991). Numbers on the left indicate positions of the initial amino acid of the KH domain, and parentheses show the position of X. laevis hnRNP K (Siomi et al., 1993). The position of FMR1 are based on the numbering used in Verkerk et al. (1993) and Ashley et al. (1993b). Highly conserved positins are highlighted and gaps are introduced for optimal alignment. The consensus sequence of the KH domains was derived from the amino acid sequences od the KH domains contained in the proteins listed in Siomi et al. (1993) and Gibson et al. (1993a). Underlined amino acids in the hnRNP K sequences mark the positions that are not identical between human and X. laevis hnRNP K proteins (Siomi et al., 1993). The position of a single mutation in FMR1 resulting in an Ile304 to Asn substitution found in a fragile X patient (De Boulle et al., 1993) is circled.

first KH domain of FMR1 (Ile241 to Asn) are constructed and tested to see whether those mutants bind to RNA in vitro. It is known that the FMR1 protein shows strong binding to poly(G) and poly(U) in an RNA homopolymer binding assay (Chapter 1). Here it is demonstrated that KH domains play a role in RNA binding and are showed that the RNA binding of the FMR1 Ile304 to Asn mutant is strongly impaired.

2-2 Experimental Procedures

Construction of FMR1 mutants

The template DNA for mutagenesis was prepared by inserting the cDNA encoding human FMR1 from pHHSI-27X (Chapter 1) with SalI and SacI to pALTER1, which is a vector for the Altered Sites in vitro Mutagenesis System (Promega). The mutants of FMR1 were generated by deoxyoligonucleotide-directed mutagenesis by following the manuscript of the system. Two deoxyoligonucleotides (m1 primer and m2 primer) that introduced the desired amino acid changes, which were made by DNA synthesizer, were as follows.

m1 primer (for Ile241 to Asn; nucleic acid # 908-926, antisense)

5'-GCTTGCTGATTATTAGCAC-3'

m2 primer (for Ile304 to Asn; nucleic acid #1086-1103, antisense)

5'-ACAATCTCCTGATTCAGCTTTCCATTT-3'

After the mutagenesis all mutants were confirmed by DNA sequencing (Sanger et al., 1977). When the cDNA encoding human FMR1 was inserted in pALTER1 with SalI and SacI, the DNA fragment encoding a nine amino acid epitope HA was left in pHHSI-27X. Thus, an alternative short DNA fragment with XhoI and SalI encoding the nine amino acid epitope HA with an additional methionine, was reintroduced in the constructs.

In vitro transcription and translation

Wild-type and its mutants DNAs were linearized with ClaI and transcribed with T3 RNA polymerae (Promega), followed by translation of the resultant RNA in rabbit reticulocyte lysate (Promega) in the presence of [^{35}S]methionine (Amersham).

RNA Binding assay

Binding assay of in vitro produced proteins was carried out as described in Chapter 1. Briefly, ribonucleotide homopolymer (Pharmacia) binding reactions were carried out with an equivalent of 100,000 cpm of trichloroacetic acid-precipitable protein in a total of 0.5ml of binding buffer (10mM Tris-HCl [pH7.4], 2.5mM MgCl$_2$, 0.5% Triton X-100, 0.2% Pepstatin, 0.2% Leupeptin, 0.5% Aprotinin) with a salt concentration of either 100mM or 250mM for 10 min on a rocking platform at 4°C. The beads were pelleted with a brief spin in a micro centrifuge and washed five times with binding buffer prior to resuspension in 50ul of SDS-PAGE loading buffer. Bound proteins were eluted from the nucleic acid by boiling, resolved on a SDS-polyacrylamide (12.5%) gel, and visualized by fluorograpy. An amount equivalent to 20% of the material used for each binding reaction was used to show in the lanes marked "T" standing for Translation (Figure 2-2).

2-3 Results

It is known that one male patient who shows a severe phenotype of fragile X syndrome but expressing *FMR1* gene has a missense mutation (Ile304 to Asn) in it. The discovery raised a question to ask if the corresponding Ile to Asn missense mutation in *FMR1* has an effect on its binding to RNA. To address to it, the Ile304 to Asn mutant (m2) and also the equivalent mutant in the first KH domain of FMR1 (Ile241 to Asn: m1) were constructed (Figure 2-2). The natural RNA substrate of FMR1 is not yet known, but wild-type FMR1 binds well to poly(G) and poly (U) but not to poly(C) and poly(A) (Chapter 1). To asses the effect of these mutations on RNA binding, poly(G) and poly(U) binding assays were performed using both wild-type and mutants of FMR1 products synthesized in reticulocyte lysates. The binding to poly(G) of both mutants (m1 and m2), was only slightly reduced at 250mM compared with the wild-type FMR1. However, in contrast with wild-type FMR1, neither of the two mutants bound to poly (U) at a salt concentration of 250 mM. These data demonstrate that the point mutation Ile304 to Asn in FMR1 that causes fragile X syndrome (De Boulle et al., 1993) results in a protein with impaired RNA binding, establishing a connection between the RNA binding-activity of the
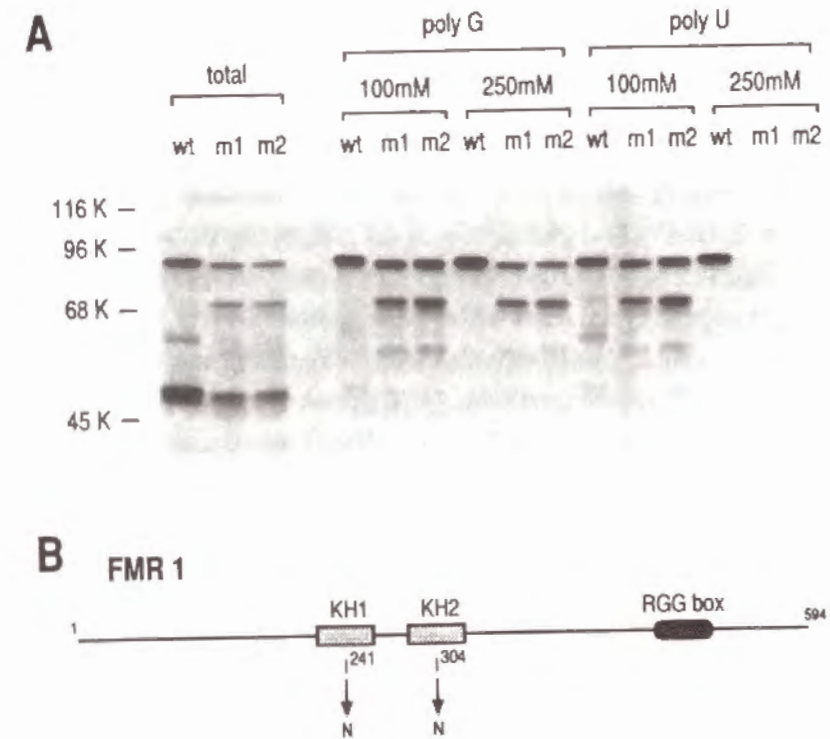
Figure 2-2. Mutation in each of the KH domains of FMR1 abolish poly(U) binding in vitro.

(A) In vitro translated wild-type and mutant FMR1 proteins (m1, Ile 241 to Asn; m2, Ile304 to Asn) were bound to the indicated ribonucleotide homopolymers at the indicated salt concentrations. The largest molecular mass band corresponds to the full-length FMR1 (indicated by arrow) and the smaller ones probably result from internal initiations, premature terminations of translation in the in vitro system, or both.
(B) Schematic representation of the FMR1 protein and structure of mutants in the KH domains. Two KH domains (stippled) and an RGG box (closed) are shown. The Ile304 to Asn point mutation was found in a patient with very severe fragile X syndrome(De Boulle et al., 1993).

2-4 Discussion

From this experiment, it was shown that the mutation Ile304 to Asn affects the RNA-binding properties of FMR1. Changing this Ile residue in either the first or second KH domains of FMR1 to Asn abolished poly(U) binding, suggesting that the two KH domains constitute a bipartite RNA binding surface. Alternatively, since Ile304 residue is in a region of the protein that is predicted to have a strong propensity to form an α-helix (Gibson et al., 1993b), an Asn substitution (hydrophobic to hydrophilic) could impair the stability of the folded domain. In addition to the two KH domains, FMR1 contains an RGG box near the C-terminus. The preliminary domain mapping experiments of FMR1, which has been demonstrated in Chapter 1, showed that one C-terminal deletion mutant which is missing the RGG box but still contains the KH domains, does not bind either poly(G) and poly(U). It appears that poly(U) binding is a more sensitive indicator of defective FMR1 binding to RNA and that it requires both KH domains and the RGG box.

The same experiment has been done with hnRNP K protein, in which the KH domain had been originally described. The Ile to Asn mutation was introduced in each of three KH domains of human hnRNP K and RNA binding assay was done with those mutants. The experimental data also showed that the RNA-binding activities of the mutants were abolished, namely the conserved amino acid was essential in RNA binding (Siomi et al., 1994).

KH domains can be located anywhere in proteins and occur singly or in several copies (Siomi et al., 1993; Gibson et al., 1993a). The findings here suggest that KH domains function collectively rather than independently. Therefore, it will be interesting to determine how proteins containing a single KH domain, such as MER1 and S3, interact with RNA.

The Ile 304 to Asn mutation in the KH domain of FMR1 in a patient with severe fragile X syndrome (De Boulle et al., 1993) provided a framework for a better understanding of the role of the *FMR1* gene. The tissue specificity of *FMR1* gene expression is consist with involvement in the fragile X phenotype; in situ hybridization and immunostaining reveal widespread but not ubiquitous expression of *FMR1*, with particularly high level in neuronal cells in the brain and in the testis (Hinds et al., 1993; Devys et al., 1993; Abitbol et al., 1993; Bachner et al., 1993). In the vast majority of fragile X patients, the syndrome results from loss of expression of the *FMR1* gene. Therefore, it is likely that lack of *FMR1* or disturbance of FMR1 function in these tissues is the cause of their clinical manifestation. It appears that the normal role of FMR1 is to regulate gene expression posttranscriptionally in these developing tissues. The impaired RNA binding that results from the Ile 304 to Asn change in FMR1 (Figure 2-2) may explain how this can lead to fragile X syndrome, namely, by interfering with the proper interaction of FMR1 with its cognate RNA(s). It would be of particular inerest to identify the cellular RNA(s) with which FMR1 interacts. This would permit the identification of relevant RNA-binding sites for FMR1 and facilitate a more direct analysis of the function of FMR1. The observation that another mutant, Ile 241 to Asn, is also impaired in RNA binding predicts that such a mutation and many others that may affect the RNA binding activity of FMR1 would also result in fragile X syndrome. Together, these findings reduce at least one aspect of intelligence to a molecular issue of protein-RNA interaction.

2-5 Summary

While fragile X syndrome is usually caused by lack of expression of *FMR1*, a severely retarded fragile X patient has been reported who expresses *FMR1* that has a mutation in a highly conserved residue of one of it's two KH domains (Ile304 to Asn). The hnRNP K homology (KH) domain is a highly conserved ca. 50 amino acid sequence motif present in many RNA-associated proteins from widely divergent organisms including eukaryotes, eubacteria and archaebacteria. Several proteins have been described which contain from one to fourteen KH domains. Although FMR1 and several other KH domain proteins have been shown to bind RNA, the function of KH domains is unknown. To assess the role of KH domains in RNA-binding of FMR1, mutagenesis of the KH domains was carried out and examined their effects on RNA-binding. It was found that the RNA-binding of this mutant is severely impaired. These results demonstrate that KH domains have an essential role for FMR1 in RNA-binding.

Furthermore, they strengthen the connection between fragile X mental retardation and loss of the RNA-binding activity of FMR1.

Chapter 3

Production of antibodies for FMR1 protein and the application to the diagnosis for the fragile X syndrome

3-1 Introduction

In Chapter 1 and 2, it was shown that FMR1 protein contains two RNA-binding motifs, one of which is the KH domain (Siomi et al., 1993; Gibson et al., 1993b) and the another is an RGG box (Kiledjian and Dreyfuss, 1992), and demonstrated that it indeed binds to RNA in vitro. When the point mutation that had been found in FMR1 protein of one fragile X male patient with a severe phenotype (De Boulle et al., 1993) was introduced into FMR1 protein by site-directed mutagenesis, the RNA-binding was impaired in a specific manner. This was the first biological parameter to assess the activity of *FMR1* gene product since the gene of *FMR1* had been identified and sequenced in 1991. However, its protein product has not been identified yet. Moreover, the real substrates to be bound and the physiological function of FMR1 in vivo are still not known.

In this chapter, the production of antibody to FMR1 protein is described. It would not be only the best way to address the questions mentioned above, but also very useful to study the expression of FMR1 in cells at the protein level. The observation that the anti-FMR1 antibodies can detect the majority of the fragile X cases has considerable potential for improved diagnosis of the fragile X syndrome.

3-2 Experimental procedures

Expression and purification of fusion protein

The expression and purification of FMR1 partial peptide fused with His-HA has been described in Chapter 1.

Raising Antibodies to His-HA-FMR1 fusion protein

Antisera to FMR1 protein were raised in BALB/c mice injected with the purified recombinant His-HA-FMR1 fusion protein produced in E. coli. About 100ug of the fusion protein in phosphate-buffered saline (PBS) was used to

immunize a mouse per injection. The interval between injections was 2 weeks. Totally 4 times injections were done for a mouse.

Cell Lines

Blood samples were obtained from Family FX06 with informed consent. The lymphoblastoid cell lines were established by Epstein-Barr virus transformation of peripheral B lymphocytes by previously published techniques (Wilson et al., 1983).

Western Blot Analysis

Cells were grown to subconfluence, lysed in SDS-PAGE sample buffer, sonicated, and then heated at 95°C for 5 min. Proteins were resolved on a 12.5% SDS-PAGE gel and transferred to nitrocellulose using a Bio Trans Model B transblot apparatus (Gelman Sciences) according to the manufacturer's instructions. Filters were incubated in blotting solution (phosphate-buffered saline, 5% non-fat dry milk) for at least 30 min at room temperature and then incubated with primary antibody for 1 hr at room temperature. Filters were washed three times in phosphate-buffered saline/ 0.05% Tween 20, and bound antibody was detected using the iodinated polyclonal anti-mouse IgG. Western blots were incubated with polysera at a 1:200 dilution.

Immunofluorescence on HeLa cells

HeLa cells were grown on cover glasses to subconfluence, fixed with 2% formaldehyde in PBS and permialized with cold acetone. After washing with cold PBS, cells were incubated with polysera diluted at 1:200 by 3% BSA in PBS for 1 hr at room temperature, followed by washing with PBS extensively. FITC-conjugated anti-mouse IgG secondary antibody was diluted with 3% BSA in PBS, applied on cells and incubated for 1 hr at room temperature. The localization of *FMR1* gene product in HeLa cells were detected and their pictures were taken under microscope.

3-3 Results

Production of antibodies to FMR1: Expression in divergent organisms

To examine the expression of the predicted FMR1 protein, antibodies were produced by immunizing mice with the His-HA-FMR1 fusion protein. Immunoblotting of the fusion protein produced in E. coli (data not shown) and of HeLa total cell material (Figure 3-1) demonstrated that the polyclonal antibodies produced against the FMR1 fusion protein also recognized a protein of the expected molecular weight (based on the in vitro transcription/translation described in Chapter 1) in HeLa cells (Figure 3-1). This serum was then used to detect the protein by probing immunoblots of cell lysates from evolutionarily divergent organisms (Figure 3-1). Southern blotting has previously shown cross-hybridizing DNA fragments in several organisms from human to S. cerevisiae, but no hybridization was detected in Drosophila melanogaster (Verkerk et al., 1991). In agreement with these findings, the anti-FMR1 antibodies detected cross-reactive proteins in monkey, mice, bovine, X. laevis and S. cerevisiae but not in D. melanogaster (Figure 3-1). The size of the proteins in other mammals appears similar to that in humans, but the S. cerevisiae protein is considerably smaller (ca. 55K).

Expression of FMR1 in patients with fragile X syndrome

It was of particular interest to probe cell lysates from fragile X patients with the anti-FMR1 antibodies. Lymphoblastoid (LBL) cell lines were established from all members of family FX06 by standard procedures (Wilson et al., 1983). A partial pedigree of family FX06 is shown on Figure 3-2A. All members of the family were examined by Dr. Robert L. Nussbaum. FX06-06, FX06-07, FX06-25, FX06-21 are all affected males; they are moderately to severely retarded, are in special education classes, and live at home. The diagnosis of fragile X syndrome was confirmed initially by finding 50% expression of the fragile site in transformed LBL cells (Sutherland et al., 1985; Nussbaum and Ledbetter, 1986; Richards and Sutherland, 1992) and later by Southern blot analysis of the FMR1 region (Kremer et al., 1991; Yu et al., 1991; Rousseau et al., 1991). All show absent or markedly reduced levels of FMR1 protein (Figure 3-2). Of interest, FX06-07, who shows the most protein of all
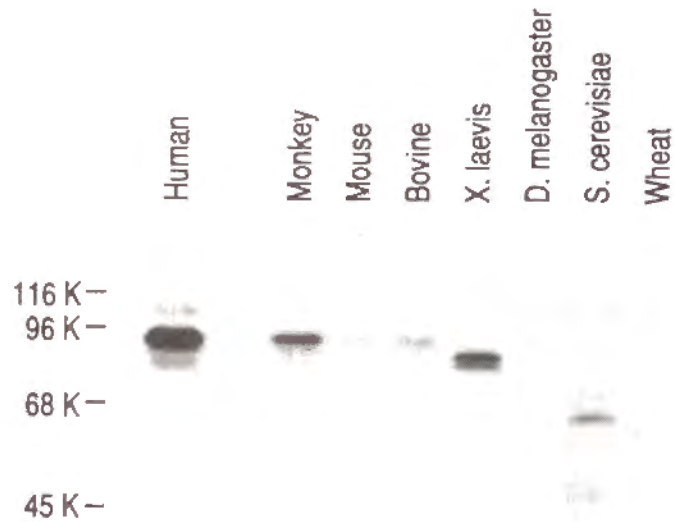
Figure 3-1. Identification of the protein product of the *FMR1* gene in divergent organisms.

Total cellular proteins from the indicated organisms were resolved by SDS-PAGE, transferred to nitrocellulose membrane, and probed with the anti-serum. Total cellular proteins were obtained from the following: human HeLa cell, monkey COS1 cell, mouse NIH3T3 cell, Madin-Darby bovine kidney cell, Xenopus laevis kidney epithelial cell, Drosophila melanogaster Schneider's cell, Saccharomyces cerevisiae and wheat germ. Molecular weight markers are indicated on the left.
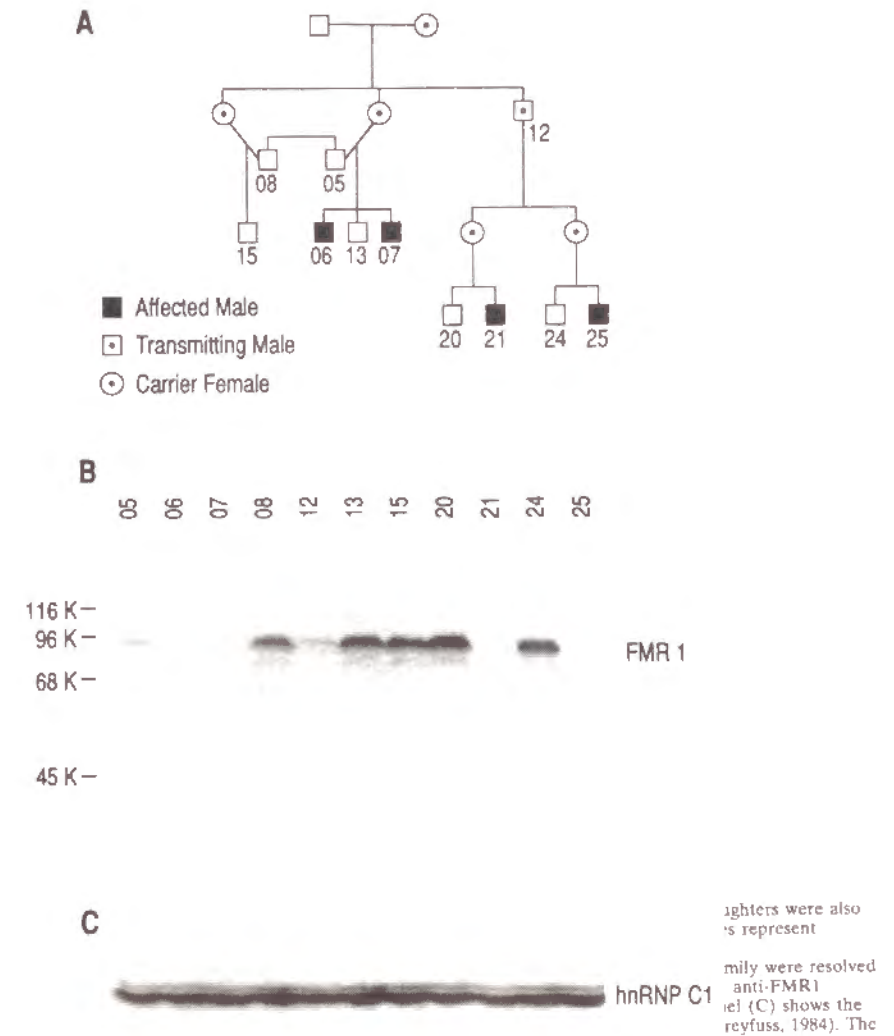
Figure 3-2. Analysis of fragile X family.

(A) In this family (FX06) the grandmother was a carrier female, her two daughters were also carriers and her son was a transmitting male. Squares represent males; circles represent females.
(B) Total cellular proteins from lymphoblastoid cell lines of the fragile X family were resolved by SDS-PAGE, trasferred to nitrocellulose membrane, and probed with the anti-FMR1 serum. Lane numbers correspond to individuals in pedigree. The lower panel (C) shows the same membrane probed with an anti-hnRNP C1 antibody 4F4 (Choi and Dreyfuss, 1984). The position of molecular weight markers are indicated.

the affected males, has a 47, XXY karyotype due to paternal nondisjunction as determined by linkage analysis. Males FX06-13, FX06-15, FX06-24, and FX06-20 were all clinically unaffected and had <30 copies of the CGG repeat (data not shown) as determined by polymerase chain reaction amplification across the fragile site region (Verkerk et al., 1991; Fu et al., 1991). Individual FX06-12 is a transmitting male who expressed the fragile site in 5% of his lymphoblasts but did not exhibit signs of the fragile X syndrome. He shows decreased but readily detectable expression of FMR1 protein. FX06-05 and FX06-08 are spouse controls. The western blotting results (Figure 3-2) show excellent correspondence to the clinical picture. They also further confirm that the antibodies are specific for FMR1. Although the available anti-FMR1 antibodies may not be able to detect these rare point mutations, they can readily detect the majority of the fragile X cases. It can thus be anticipated that antibodies to FMR1 will offer a specific and sensitive diagnostic tool for fragile X syndrome. This should augment and complement cytologic and DNA-based methods for the detection and study of this disorder.

Cellular Localization of FMR1 protein in living cells

The cellular localization of the endogenous gene product was studied by immunofluorescence on HeLa cells and showed the cytoplasmic localization with no significant staining over the nucleus (Figure 3-3). No significant staining was observed when cells were treated with sera from mice without immunization (data not shown). The cellular localization of FMR1 protein was confirmed in COS1 cells transfected with FMR1 expression vector pHHSI-F27X (Chapter 1). In cells transfected with the complete protein coding sequence, the anti HA monoclonal antibody detected the overexpressed FMR1 protein with a nine amino acid epitope (HA) as cytoplasmic localization (data not shown).

3-4 Discussion

The production in E.coli of His-tagged fusion proteins containing the C terminal half of the FMR1 protein coding sequence allowed to obtain polyclonal antibodies, which could detect the endogenous protein in human cells by either

Figure 3-3.   Cellular localization of FMR1 protein in HeLa.

Immunofluorescence were performed using anti-FMR1 antibodies on HeLa. Cells were incubated at a 1:400 dilutions.

western blot or immunofluorescence. In western blot analysis of cell lysate from evolutionarily divergent organisms, the anti-FMR1 antibodies detected cross-reactive proteins in monkey, mice, bovine, X. laevis and S. cerevisiae but not in D. melanogaster (Chapter 1). This result confirmed the data of the southern blotting analysis showing cross-hybridizing DNA fragments in many organisms from human to S. cerevisiae with the exception of D. melanogaster (Verkerk et al., 1991). The *FMR1* gene and the protein product expression are highly conserved through evolution. The isolation and sequencing of the homologous clones from other organisms may reveal conserved protein domains within FMR1, indicating regions of functional importance. Also such homologies may allow experimental approaches in other organisms to unravel the function of FMR1.

Immunoblotting experiments on the lymphoblastoid cell lines presented here using antibodies to FMR1 protein extend the conclusions made at the mRNA level (Pieretti et al., 1991) to the actual expression of the FMR1 protein and demonstrate that the defect in fragile X syndrome results from the lack of expression of FMR1 (Figure 3-2). Until recently, laboratory diagnosis of fragile X syndrome was carried out by cytogenetic analysis utilizing specialized growth media conditions (Jacky et al., 1991, Dewald et al., 1992). When the guidelines for fragile site induction are carefully followed, the sensitivity for detection of FRAXA in affected males is quite high. Cytogenetic studies, however, are generally insensitive for detection of premutation carriers. In addition, the presence of three other fragile sites in distal Xq (FRAXD, FRAXE, and FRAXF) (Flynn et al, 1993; Hirst et al., 1993) which might not be distinguishable cytogenetically from FRAXA, indicates a positive cytogenetic finding may not be specific for fragile X syndrome and requires confirmation by direct molecular testing. For these reasons, cytogenetic diagnosis is rapidly becoming obsolete and is being replaced by molecular diagnosis of FRAXA. The fundamental molecular assay for the fragile X syndrome is the measurement of the length of the (CGG)n repeat in the *FMR1* gene by either Southern blots or PCR (reviewed in Nussbaum and Ledbetter, 1986). Because of its ease and speed, the PCR method has obvious advantage over Southern blotting for assessing the size of the trinucleotide repeat in the *FMR1* gene. Unfortunately, the PCR method has been very difficult to use to detect the full mutation because of technical difficulties in performing a PCR across hundreds of tandemly repeated CGG triplets (Fu et al., 1991), though progress adapting PCR for detecting the full fragile X mutation is continuing to be made (Pergolizzi et al., 1992). The experimental data shown in this chapter indicate that antibodies to FMR1 would be particularly valuable for the diagnosis of fragile X syndrome, might be a better way than both cytogenetic method and DNA-based method.

It had been suggested previously that the FMR1 could be a nuclear protein, as it contains a short lysine rich sequence (KKEK) in the C-terminal region which could constitute a nuclear localization signal (Verkerk et al., 1991). However, it was shown that FMR1 protein locates predominantly in the cytoplasm as determined by the immunocytofluorescence and immunohistochemical studies with anti-FMR1 antibodies in HeLa cells and in transfected COS1 cells (Devys et al., 1993; Verheij et al., 1993). The findings from the experiment described in this chapter confirmed those observations. A surprising observation, however, has been reported that the N-terminal half of FMR1, when overexpressed in COS1 cells, had a predominantly nuclear localization, which may suggest that under some physiological conditions, FMR1 or derivatives of it may have a function in the nucleus (Devys et al., 1993).

With RNA binding assay results shown in Chapter 1 and 2, FMR1 could have a function such as translational regulation. The availability of antibodies should allow the investigation of its precise intracellular localization, of possible specific interactions with other cellular materials, or post translational modifications under physiological and pathological conditions.

3-5 Summary

Using antibodies to FMR1, its expression is detected in divergent organisms and in cells of unaffected humans, but little or no in fragile X-affected patients. These findings demonstrate that FMR1 expression is directly correlated with the fragile X syndrome and suggest that anti-FMR1 antibodies will be important for diagnosis of fragile X syndrome. The immunocytofluorescence study show clearly that FMR1 protein has cytoplasmic localization on HeLa cells. Though its function still remains unclear, it is suggested that FMR1 protein could have a function such as regulation of gene expression posttranscriptinally.

Chapter 4

cDNA cloning and characterization of FMR1
and its homolog, FXR1 from Xenopus laevis

4-1 Introduction

mRNA in situ hybridization studies and northern blot analysis were performed in mouse and human, respectively to demonstrate the normal gene expression patterns of *FMR1* (Hinds et al., 1993). In mice strong expression of *FMR1* was located in several regions of the brain and tubules of the testes, which are the major organs affected in fragile X syndrome. Universal and very strong expression was observed in early mouse embryos, with differentially decreasing expression during subsequent stages of embryonic development.

It is of interest to follow the expression of FMR1 throughout oogenesis in order to understand its function. When Western blotting using antisera raised against the human FMR1 protein was carried out, the homologous protein was detected in Xenopus laevis cells as well as mammalian organisms (Chapter 3). Because Xenopus laevis is a useful system for studying oogenesis, a cDNA encoding FMR1 was obtained from Xenopus laevis.

A novel cDNA, designated *FXR1* (for *FMR1* crossreacting relative), was obtained when the X. laevis cDNA was screened with the human *FMR1* cDNA. In the case of human *FMR1*, extensive alternative splicing has been demonstrated at the mRNA level (Verkerk et al., 1993; Eichler et al., 1993). Nonetheless a novel clone like *FXR1* that is highly homologous to *FMR1* has not been reported yet. These findings indicate that *FMR1* is a member of a gene family. Studying FXR1 in addition to FMR1 may yield some clues to better understand the function of FMR1 and the correlation between FMR1 and fragile X syndrome.

4-2 Experimental procedures

Isolation of cDNA clones, in vivo excision and sequencing

The full length human *FMR1* cDNA was used as a probe to screen a λZAPII Xenopus ovary cDNA library. Six clones reacting with the probe were obtained from 10^6 plaques, and all of them were purified by further screening. In vivo excision was done for three of those positive clones according to the manufacturer's instruction (Stratagen), creating pXF1-61, pXF1-43 and pXF1-45. Nucleotide sequencing was performed according to the dideoxy chain termination method of Sanger et al. (1977). DNA sequence was determined on both strands of those clones.

RNA Binding Assay

Binding of in vitro-produced proteins to ribonucleotide homopolymers was carried out as described above (Chapter 1). The NaCl concentration of the binding buffer was 250mM.

Expression and purification of fusion peptide and production of antisera against X. laevis FMR1 and FXR1

PCR was done to create a NdeI site in the middle and a BamHI site at the 3' end of the X. laevis *FMR1* open reading frame. The PCR product was digested with NdeI and BamHI and inserted into pET15b (Novagene) to construct the expression vector pEXFMR1. To raise antisera specifically to X. laevis FXR1, pXF1-45 was digested with BamHI and the 600bp fragment encoding just the C terminal region was inserted into pET15b to create the expression vector pEXFXR1. For production of the His-FMR1 peptide and His-FXR1 peptide, the plasmids pEXFMR1 and pEXFXR1 were introduced into BL21(DE3) bacteria and induced with isopropyl-β-D-thiogalactopyranoside as described (Studier et al., 1990; Rosenberg et al., 1987). For purification of the fusion peptides, bacterial sonicates were applied to 2ml His-Bind resin (Novagen) columns, washed and eluted as described by the manufacturer.

Antisera against X. laevis FMR1 and FXR1 were raised in BALB/c mice injected with the purified recombinant His-FMR1 and His-FXR1 fusion peptides respectively. Western blots were incubated at a 1:400 dilution.

Western Blot analysis

SDS-PAGE sample preparation and Western blotting procedures were described in Chapter 1.

Immunoflorescence on Xenopus laevis cells

Xenopus laevis cells were grown on cover glasses to subconfluence, fixed with 2% folmaldehyde in PBS and permeabilized with cold acetone. After washing with cold PBS, cells were incubated with polysera either for Xenopus FMR1 or FXR1 diluted at 1:400 with 3% BSA in PBS for 1hr at room temparture, followed by washing with PBS extensively. FITC-conjugated anti-mouse IgG secondary antibody was diluted with 3% BSA in PBS, applied to the cells and incubated for 1hr at room temparature. The localization of *FMR1* and *FXR1* gene products in Xenopus laevis cells were detected and pictures were taken under the microscope.

Biological materials

Mature female Xenopus laevis were purchased from ISCO. Oocytes were separated manually into six stages according to Dumont (1972). Forty oocytes from each stage (stage I and II were not separated) were ground in 50mM Tris-HCl pH7.5 with 0.05% PMFS, followed by spinning down. About 80ul of protein solution were obtained from each stage and mixed with 1 volume of SDS-PAGE loading buffer.

4-3 Results

Isolation and characterization of a cDNA specific for Xenopus FMR1 and a related cDNA, FXR1

The human *FMR1* cDNA was used as a probe to screen a Xenopus ovary cDNA library constructed in λZAPII. Six clones were isolated and DNA sequences of three of them were determined from both strands using T7 and T3 primers (Sanger et al., 1977) after in vivo excision. pXF1-61 was 1.2kbp long but incomplete, lacking the 5' terminus. The DNA sequence of pXF1-43 (2.1kbp) was identical to pXF1-61 but was complete. The deduced amino acid sequence of pXF1-43 showed that it encodes Xenopus FMR1 (564 amino acids with a predicted molecular weight of 64kD)(Figure 4-1A). It was found that Xenopus FMR1 lacks 45 amino acids just after the two KH domains as compared with human FMR1 (amino acids 331-375)(Figure 4-2). The segment containing amino acids 331-375 of human FMR1 corresponds exactly to exon 11 of the
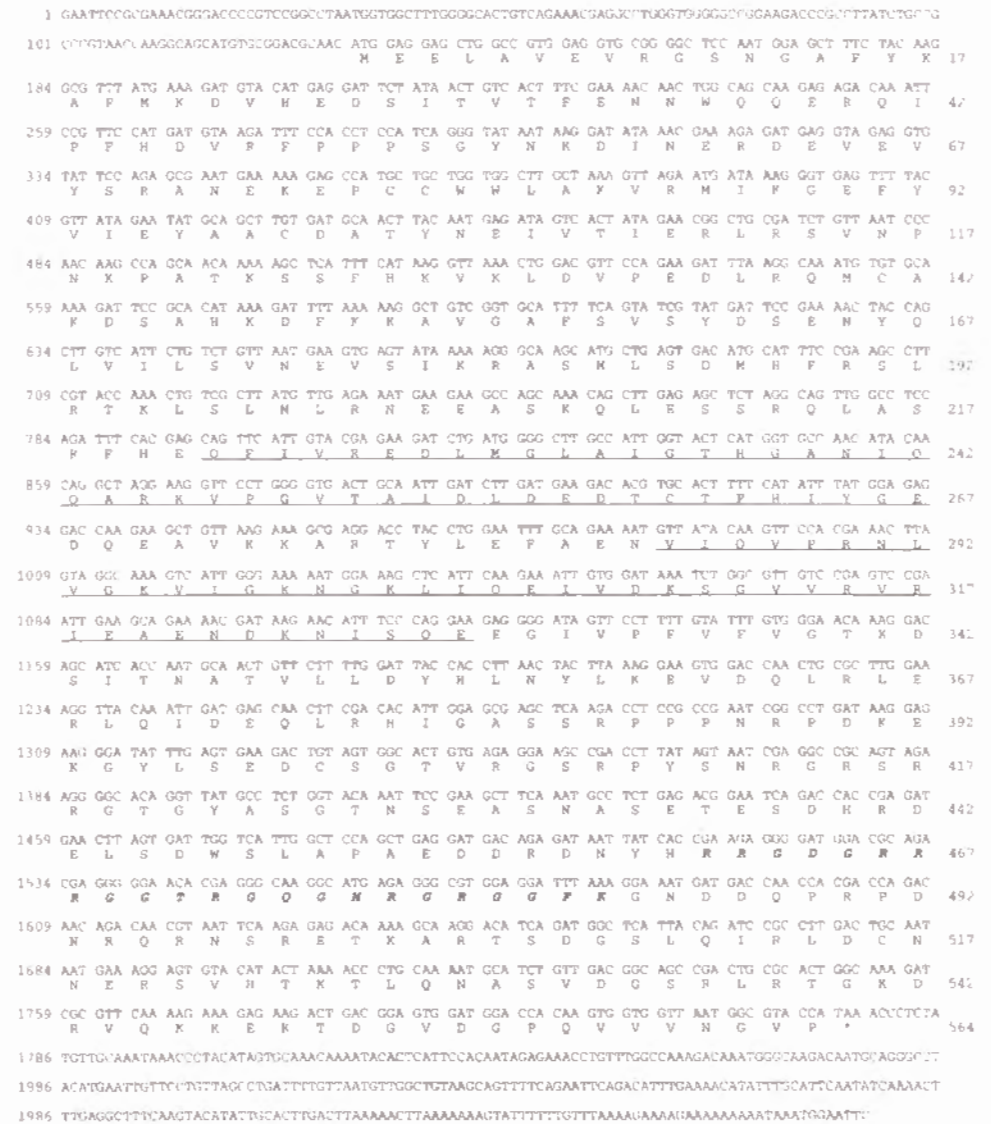
Figure 4-1. (A)The DNA sequence and the predicted amino acid sequence of X. laevis FMR1. The KH domains are underlined and the amino acids in the RGG box are shown by italic letters.

Figure 4-1. (B) The DNA sequence and the predicted amino acid sequence of X. laevis FXR1. The KH domains are underlined and the amino acids in the RGG box are indicated by italic letters.

40

(Figure 4-1 B: DNA sequence with predicted amino acid translation; sequence numbered from 1 to 2351, with KH domains underlined and RGG box amino acids in italics.)

41

```
XFMR1   1 MEELAVEVRG SNGAFYKAFM KDVHEDSITV TFENNWQQER QIPFHDVRFP PPSGYNKDIN ERDEVEVYSR ANEKEPCCWW LAKVRMIKGE FYVIEYAACD
            ::::: :::: :::::::::  :::::::::: :::::: :::  :::::::::: :: :::::::: : :::::::: :::::::::::: :::::::::: ::::::::::
HFMR1   1 MEELVVEVRG SNGAFYKAFV KDVHEDSITV AFENNWQPDR QIPFHDVRFP PPVGYNKDIN ESDEVEVYSR ANEKEPCCWW LAKVRMIKGE FYVIEYAACD
            :::: :::::  :::::::  :  :::::: ::  :::::::::   :::: :::  :: :::  ::   :::::::: :::: : :::: :::: ::::  :::::::::::
XFXR1   1 MEDMTVEVRG SNGAFYKGFV KDVHEDSLTV VFENNWQPER QVPFDEVRMP PLPDIKKEIT EGDEVEVYSR ANDDQEPCGWW LAKARMQMKGE FYVIEYAACD

XFMR1 101 ATYNEIVTIE RLRSVNPNKP ATKSSFHKVK LDVPEDLRQM CAKDSAHKDF KKAVGAFSVS YDSENYQLVI LSVNEVSIKR ASMLSDMHFR SLRTKLSLML
            ::::::::::  ::::::::::  :::::  :::  :::::::::: :::: ::::: :::::::::  ::: :::::: :::: :: ::   ::::: ::: :::::::: ::
HFMR1 101 ATYNEIVTIE RLRSVNPNKP ATKDTFHKIK LDVPEDLRQM CAKEAAHKDF KKAVGAFSVT YDPENYQLVI LSINEVTSKR AHMLIDMHFR SLRTKLSLIM
            :::::::::: :::   :::::  :: ::: ::   ::::::::   :: :  :::  :::: :::::  ::::  ::  ::::  : :::   ::  ::::  :: :: ::
XFXR1 101 ATYNEIVTFE RLRPVNQNKT VTKNSPFKCT VDDVPEDLRES CSNENVHKEF KKAVGACRVY FHAETNQLII LSACESTVKR VNILSDMHLR SIRTKLMLMS

XFMR1 201 RNEFASKQLE SSRQLASRFH EQFIVREDLM GLAIGTHGAN IQQARKVPGV TAIELDEDTC TFHIYGEDQE AVKKARTYLE FAENVIQVPR NLVGKVICKN
            :::::::::: :::::::::: :::::::::: :::::::::: :::::::::: ::::::: ::: :::::::   :::::: :::  :: :   :::: :::::::::::
HFMR1 201 RNEEASKQLE SSRQLASRFH EQFIVREDLM GLAIGTHGAN IQQARKVPGV TAIDLDEDTC TFHIYGEDQD AVKKARSFLE FAEDVIQVPR NLVGKVIGKN
            ::::: ::   :: ::::::: ::  :::::: ::: ::: :  :::::::: : ::: ::::: :: ::  :  ::  :::::: ::: :: ::  ::  :::::::::
XFXR1 201 RNEEATKHLE CTKQLASRFH EEFIVREDLM GLAIGTHGSN IQQARKVPGI TAIELDEDSG TERIYGESAE AVKKARSYLE FVEDRIQVPR NLVGKVIGKN

XFMR1 301 GKLIQEIVDK SGVVRVRIEA ENDKNISQEE ---------- ---------- ---------- ---------- --------GIVPF VFVGTKDSIT NATVLLDYHL
            ::::::::::  :::::::: :  :::   :                                                          :::: :::::: :: ::::::::::
HFMR1 301 GKLIQEIVDK SGVVRVRIEA ENEKNVPQEE EIMPPNSLPS NNSRGGPENAP EEKYHLDIKE NSTHFSQPNS TKVQRQMVPF VFVGTKDSIA NATVLLDYHL
            :: ::::::: :  :::::: : :::: :  :                                                       :::::: ::: :  ::: :::: ::
XFXR1 301 GKVIQEIVDK SRVVRVRIEG DNETKLPRED ---------- ---------- ---------- ---------- --------QMVPF VFVGTKENIG NVQVLLEYHI

XFMR1 356 NYLKEVDQLR LERLQIDEQL RHIGASSRPF PNR-PDKEKG YLSEDCSGTV -RGSRPYSNR GRSRRGTGYA S--GTNSEAS NASETESDHR DELSDWSLAP
            :::::::::: :::::::::: ::  :::::  :::  :: :: ::  :::: :   ::::::  :: :  :::: ::  ::  ::::::: :::::::::: ::::::::::
HFMR1 401 NYLKEVDQLR LERLQIDEQL RQIGASSRPP PNR-TDKEKS YVTDDGQGMG -RGSRPYRNR GHGRRGPGYT S--GTNSEAS NASETESDHR DELSDWSLAP
            ::::: ::: :::::::: ::  ::::  ::   : ::::::   :::  ::::     :: ::  :: :: :: :    ::  ::: ::: : ::::   ::::::: :
XFXR1 356 AYLKEVEQLR MERLQIDEQL RQIGMGFRPS SSRGTEKEKG YATDESTASS GRGRRGPNYT SGYGTNSELS NPSFTESERK EELSDWSLA-

XFMR1 452 AEDDRDNYHK RGDGRRR-GG TRGQGMHGH- GGFKGNDDQP RPDNRQRNSR ETKARTSDGS LQIRLDCNNE RSVHTKTLQN ASVDGSRLRT GKDRVQKKEK
            :::::::::  ::::::::  ::::: :::   ::::::::::  :: : : ::  ::::: :::: :: :: ::  : :::::::::: :      :::: ::::::::::
HFMR1 497 TEEERESFLR RGDGRRRGG- CRGQGGRGRG GGFKGNDDHS RTDNRPRNPR EAKGRITDGS LQIRVNCNNE RSVHTKTLQN TSSEGSRLRT GKDRNQKKEK
            :  ::::    :::: ::::   ::: ::: :  :: ::  :   ::::  :::: : :  :  :   ::   ::    :: ::     :   ::::  :::: ::: :::
XFXR1 455 GEDERESRQO R-USRRRPGG -RGRSGSA-G RG-RG-G--S R-GGKSSTSS VLKDPDSN-P YSLLDNTESD QTADTDASES HHNTNRR RR SR RRRTDED

XFMR1 550 TDGVDGPQVV VNGVP*
            : ::: ::   :::::
HFMR1 597 PDSVDGQQPL VNGVP*

XFXR1 544 SSLMDGNTEL DNASVNENGL VTTADYISRA ESCSRQRNLP KETLAKGKKE KVXDVIEEHG PSEKVINGPR ASSDKASNHR PHPQKETKQA VKMEVNKKP*
```

Figure 4-2. Amino acid sequence alignment of Xenopus laevis FMR1 and FXR1 with human FMR1.
The amino acid sequences of X. laevis FMR1 and FXR1 were aligned with human FMR1 (Verkerk et al., 1991). The numbering of human FMR1 are based on the numbering used in Verkerk et al. (1993). Gaps are introduced for optimal alignment. The KH domains and the RGG box are boxed with black line and gray line respectively. The RGG box of X. laevis FXR1 is underlined.

FMR1 gene (Verkerk et al., 1993; Eichler et al., 1993). pXF1-45 was 2.4kbp long and did contained an entire open reading frame encoding a protein of 649 amino acids with a predicted molecular weight of 73kD. Interestingly, the DNA sequences of pXF1-43 and pXF1-45 were not identical and the similarity was less than 70%. pXF1-45 is similar to *FMR1* but is a novel clone. It was therefore designated *FXR1*, for *FMR1* crossreacting relative clone 1. The predicted amino acid sequence of FXR1 (Figure 4-1B) showed that it also has the 45 amino acid deletion as seen in Xenopus FMR1. The carboxy terminus of the protein was unique compared to that of Xenopus FMR1, which was shown in Figure 4-2. The amino acid sequence identity of the region between Xenopus FMR1 and FXR1 was only 8%, although the identity of the other regions was 87%. Both Xenopus FMR1 and FXR1 have two KH domains, which are highly conserved with human FMR1. An RGG box was also found in both Xenopus FMR1 and FXR1. However the RGG box between human FMR1 and Xenopus FMR1 is more similar than between human FMR1 and Xenopus FXR1 (Figure 4-2). The CGG repeats in the 5' untranslated region of the mRNA, a characteristic of human *FMR1*, were not seen in either Xenopus *FMR1* or *FXR1*. It is known that mice contain a much smaller number of triplets in the same region of the 5'UTR as do humans (Ashley et al., 1993). It thus appears that nucleotide triplet repeats may be a mammalian phenomenon.

Both Xenopus FMR1 and FXR1 have similar RNA-binding profile to human FMR1

To examine whether Xenopus FMR1 and FXR1 also bind to RNA in vitro as human FMR1 does, these cDNAs were transcribed with T7 RNA polymerase and the transcripts were translated in vitro. The protein products were assayed for RNA binding activity using RNA homopolymers immobilized on agarose beads, an assay that has been used for human FMR1 protein product and other RNA-binding proteins (Kiledjian and Dreyfuss, 1992; Swanson and Dreyfuss, 1988). As human FMR1 protein showed strong binding to poly(G), weaker but significant binding to poly(U) and very little binding to poly(A) and poly(C) (Chapter 1), Xenopus FMR1 and FXR1 also showed a similar RNA-binding profile to human FMR1, binding at 250mM NaCl to poly(G) and poly(U) (Figure 4-3).
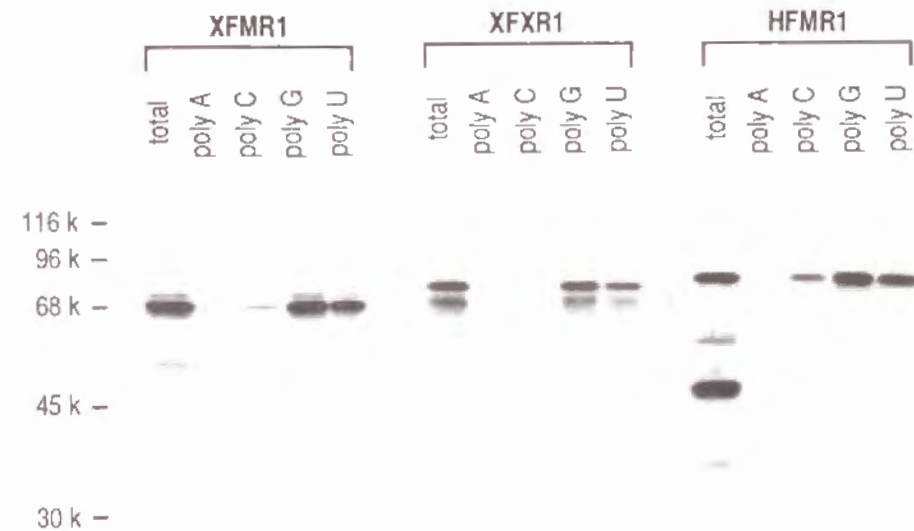

Figure 4-3. Xenopus FMR1 and FXR1 show similar RNA-binding profile to human FMR1.

Xenopus FMR1 and FXR1 were produced by in vitro transcription-translation of pXF1-43 and pXF1-45 truncated outside the coding region (ClaI). HA-tagged human FMR1 was produced by in vitro transcription-translation of pHHS1-F27X truncated outoside the coding region (NsiI). The in vitro transcribed RNAs were translated in reticulocyte lysate in the presence of [35S]methinine. An amount equivalent to 20% of the materials used for each binding reaction is shown in the lane marked "T". In vitro produced proteins were bound to 30ul of the indicated ribinucleotide homopolymers at 250mM NaCl and analyzed by SDS-PAGE as described (Kiledjian anf Dreyfuss, 1992; Swanson and Dreyfuss, 1988). The positions of molecular mass markers are indicated on the left.

Production of antibodies to Xenopus FMR1 and FXR1: Expression in divergent organisms

The carboxy half of Xenopus FMR1 and the carboxy terminus of Xenopus FXR1, which is a unique region, were overexpressed in E coli as fusion proteins with 6 His amino acids and purified on nickel columns (Hochuli et al., 1987; Smith et al., 1988). To examine the expression of Xenopus FMR1 and FXR1 proteins, specific antibodies to those proteins were produced by immunizing mice with the fusion proteins. Immunoblotting of the fusion proteins produced in E.coli (data not shown) and Xenopus total cell materials demonstrated that the polyclonal antibodies produced against the FMR1 and FXR1 fusion proteins also recognized proteins of the expected molecular mass (based on the in vitro transcription and translation described above) in Xenopus cells (Figure 4-4 for FXR1; for FMR1 data not shown). The antisera against Xenopus FXR1 was then used to detect the protein by probing immunoblots of cell lysates from evolutionarily divergent organisms (Figure 4-4). The anti-FXR1 antibodies detected cross-reactive proteins in human, monkey, chicken, and D. melanogaster. Both Southern blotting and Western blotting have shown that no cross-reactive genes and proteins of FMR1 were detected in D. melanogaster (Chapter 3; Verkerk et al., 1991). Therefore, it is very interesting that D. melanogaster does contain some cross-reactive proteins to FXR1, although the size of the protein is smaller (~50kD) than of other organisms.

Cellular localization of FMR1 and FXR1 proteins in living Xenopus laevis cells

The cellular localization of the endogenous gene products of *FMR1* and *FXR1* were studied by immunofluorescence on Xenopus laevis cells using antibodies specific for them and showed the cytoplasmic localization for both with no significant staining in the nucleus (Figure 4-5). It has already been known that human FMR1 has cytoplasmic localization (Chapter 3; Devys et al., 1993). The data obtained from the experiment with X. laevis were consist with these results.

Figure 4-4. Identification of the protein product of the *FXR1* gene in divergent organisms.

Total cellular proteins from the indicated organisms were resolved by SDS-PAGE, tranferred to nitrocellulose membrane, and probed with the anti-Xenopus FXR1 antibodies. Total cellular proteins were obtained from the following: human HeLa cell, monkey COS1 cell, Xenopus laevis kidney epithelial cell, chicken MSB cell, and D. melanogaster Schneider's cell. Molecular mass is indicated on the left.
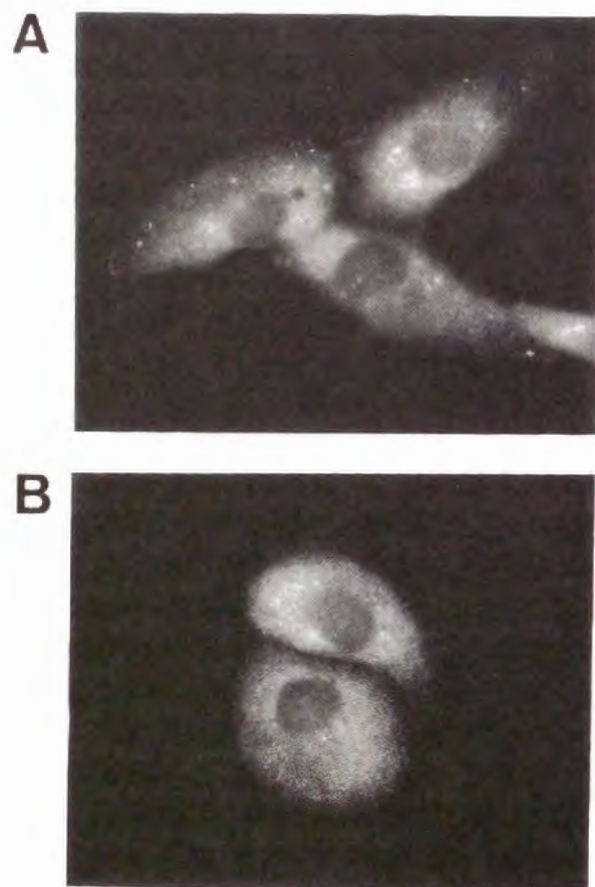
Figure 4-5. Cellular localization of FMR1 and FXR1 proteins in X. laevis cells.

Immunofluorescence was performed using (A) anti-X. laevis FMR1 antibodies and (B) anti-X. laevis FXR1 antibodies on Xenopus cells. Cells were incubated at 1:400 dilutions.

Expression of FMR1 and FXR1 during oogenesis

The expression of FMR1 and FXR1 was followed by probing immunoblots of Xenopus oocytes from various stages with anti-FMR1 and anti-FXR1 antibodies (Figure 4-6). It was shown that both FMR1 and FXR1 are expressed at very early stages, demonstrating that mRNAs of *FMR1* and *FXR1* are maternal, and suggesting that those proteins may have important roles in oogenesis.

4-4 Discussion

In situ studies of the mouse homolog of *FMR1* showed strong, if not universal, expression of the gene during early stages of development, suggesting an important functinal role in embryonic development (Hinds et al., 1993). However, the expression of *FMR1* and/or the protein product through oogenesis had not been studied. The result of Western blot analysis which has been shown in Chapter 3 demonstrated that a cross-reactive protein to human FMR1 is present in Xenopus laevis, which is a useful system to study oogenesis. Therefore, a cDNA for Xenopus laevis *FMR1* was isolated. Mice have a much smaller number of CGG triplets in the same region of the 5'UTR than humans do (Asbley et al., 1993). The Xenopus *FMR1* mRNA did not have any repeats in the same region of 5'UTR. It seems likely that the exsistence and the length of the CGG tripets in this region are related to evolution. The overall organization of the FMR1 protein is essentially the same in human and X. laevis including two KH domains and an RGG box. A 45 amino acid sequence immediately after the second KH domain is deleted in Xenopus FMR1. In the case of human *FMR1*, extensive alternative splicing has been demonstrated at the mRNA level (Verkerk et al, 1993; Eichler et al., 1993) and it is known that the segment encoding the 45 amino acid region corresponds to exon 11 in human *FMR1* gene. It is likely that this 45 amino acid region is not essential in the function of FMR1, and may have developed into the human exon.

FXR1, a novel gene that is highly homologous to *FMR1*, was obtained from Xenopus. It also contains two conserved KH domains and an RGG box. The RNA binding assay showed that Xenopus FXR1 binds to RNA in vitro in the same pattern as human FMR1, which is a reasonable observation in terms of
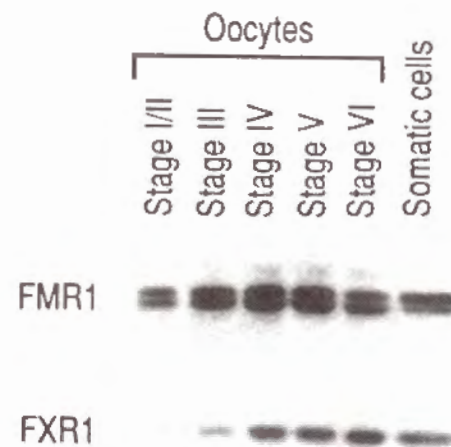
Figure 4-6.   Analysis of Xenopus FMR1 and FXR1 expression in oogenesis.

Total oocyte proteins from the indicated stages were resolved by SDS-PAGE, transferred to nitrocellulose membrane, and probed with the anti-Xenopus FXR1 or the anti-Xenopus FXR1 antibodies.

the high conservation of the KH domains between FMR1 and FXR1. Immunofluorescence microscopy demonstrated that Xenopus FMR1 and FXR1 are localized in the cytoplasm, like human FMR1. Taken together, FMR1 and FXR1 may have very similar functions in cells, although the cognate RNA targets of FMR1 and FXR1 and their precise functions have not been elucidated yet. In contrast to the conserved N-terminal regions, the C-terminal regions of FXR1, beginning with the region containing the RGG box, differ markedly both in size and sequence from FMR1. This suggests that they may have specific regulatory functions such as protein-protein interactions. The Western blot analysis showed that humans also express a cross-reactive protein to FXR1. It would be very interesting to isolate and characterize the human cDNA.

The Western blot analysis using Xenopus laevis oocytes demonstrated that both FMR1 and FXR1 are expressed throughout oogenesis even at very early stages, revealing that those mRNAs are maternal and suggesting that these proteins potentially have important functions in oogenesis.

4-5 Summary

To better understand the function and evolution of the *FMR1* gene product, a cDNA encoding Xenopus *FMR1* has been isolated and sequenced. In the course of the experiment, in addition to the X. laevis *FMR1*, a novel gene, termed *FXR1*, that is highly homologous by amino acid sequence to *FMR1*, was also isolated and sequenced. Xenopus FMR1 and FXR1, like human FMR1, are cytoplasmic RNA-binding proteins. FXR1, like FMR1, is highly conserved through evolution. Unlike FMR1, though, FXR1 is present in Drosophila melanogaster. mRNAs of both *FMR1* and *FXR1* are maternal, implying that these proteins may have important functions in oogenesis.

Chapter 5

Human FXR1: cDNA cloning and characterization

5-1 Introduction

As discussed in Chapter 4, FXR1 is a protein highly homologous to FMR1 first found in Xenopus laevis. Using an antiserum raised against Xenopus FXR1, a cross-reactive protein was detected in HeLa cells (Figure 4-4 in Chapter 4). A cDNA encoding human FXR1 was isolated from HeLa cDNA library and sequenced. The deduced amino acid sequence revealed that it also contains two KH domains and an RGG box like human FMR1, Xenopus FMR1 and FXR1. The amino acid sequence identity of FXR1 between X. laevis and human is about 80%. An alternatively spliced isoform of *FXR1*, which has a different, shorter carboxy terminus, was also obtained. To determine the expression of *FXR1* mRNA in different tissues, RT-PCR was performed. *FXR1* mRNAs were detected in all tissues tested and different size bands were observed in various tissues as expected. Interestingly, cells of a patient with fragile X syndrome that do not have any detectable FMR1 express normal levels of FXR1. *FXR1* is located on human autosomal chromosome 12 at 12q13. The mRNA of *FXR1* does not contain CGG repeats, a characteristic of *FMR1*, but it does have an unusual and striking sequence. There is an approximately 90 nucleotide perfect inverted repeat sequence in the 5' and 3' untranslated region of the mRNA. The inverted repeat sequence of human and Xenopus are 90% identical, more highly conserved than the protein coding sequence. This is the first such mRNA inverted repeat found in animal cells, and it is likely to have an important regulatory function.

5-2 Experimental procedures

Isolation of cDNA clones and sequencing

The partial Xenopus *FXR1* cDNA was used as a probe to screen a λgt11 HeLa cDNA library. The probe was made using the *FXR1*-specific region to avoid isolating other FMR1-like clones. Ten clones reacting with the probe were obtained from 10^6 plaques, and all of them were purified by further screening. PCR was done on the positive phage DNAs using primers hybridizing to λ arms according to conditions suggested by the manufacturer (Perkin-Elmer Cetus). The amplified fragments were cloned into the pCRII vector (Invitrogen). The amino terminal half of human *FXR1* cDNA was obtained by rescreening the same library using the longest cDNA from the first screening as a probe, which contained the carboxy-terminal half of human *FXR1*. RT-PCR was done on HeLa poly $(A)^+$RNA to obtain the full length cDNA as one fragment using primers hybridizing to the 5' and 3' UTRs. The RT-PCR fragment was cloned into the pCRII vector. Nucleotide sequencing was performed by the dideoxy chain termination method of Sanger et al. (1977). DNA sequence was determined on both strands of those clones.

Chromosome mapping of FXR1

Somatic cell hybrid panel #2 was purchased from the Coriell Institute Cell Repository. This panel consists of DNA isolated from 24 human/rodent somatic cell hybrids. All but two of the hybrids retain a single intact chromosome. Primers were designed to generate a PCR product of 145 bp from a portion of the carboxy terminal end of the *FXR1* open reading frame derived from a cDNA clone isolated from a HeLa cDNA library. The primer sequences are

Forward-5': GATGACATTTCTAAGCTACAGC-3' (1870-1892)

Reverse-5': TGTACAAGCACTATTGTAAATG-3' (1993-2015)

The numbers in the parenthesis of the primers above were based on the numbering in Figure 5-1. PCR reactions were performed according to conditions suggested by the manufacturer (Perkin-Elmer Cetus).

RT-PCR on poly(A)⁺RNAs from HeLa, human organs and the lymphoblastoid cell lines

Poly(A)$^+$RNA of human heart, brain, kidney, and testis were purchased from Clontech. Poly(A)$^+$RNA of HeLa and the lymphoblastoid cell lines were manually prepared using DYNABEADS mRNA DIRECT KIT (DYNAL). The RNAs (100ng) were reverse transcribed using the oligo(dT) primer according to conditions suggested by the manufacturer (Stratagen). PCR reactions were done on 5ul of cDNA solution with the primers specifically bound to *FMR1* or *FXR1*,

namely 27XM7 and 27X31 for *FMR1* (27XM7, 1272-1301; 27X31, 2124-2156, the coordinates are based on the numbering used in Verkerk et al., 1991) or XF-E and XF-B1 for *FXR1* (XF-E, 1328-1348; XF-B1, 2009-2033, the coordinates are based on the numbering used in Figure 5-1) according to conditions suggested by the manufacturer (Perkin-Elmer Cetus). In order to confirm the existence of the inverted repeats in the 5' and 3' UTRs of *FXR1* cDNA, poly(A)$^+$RNA (100ng) from HeLa cells was reverse transcribed using the cDNAprimer1 for 5'UTR repeat and the oligo(dT) primer (=cDNAprimer2) for 3'UTR repeat. PCR reactions were done on 5ul of cDNA solution with the 5'PCRprimer1 and 3'PCRprimer1 for 5'UTR repeat and 5'PCRprimer2 and 3'PCRprimer2 for 3'UTR repeat (5'PCRprimer1, 1-25; 3'PCRprimer1, 612-627; 5'PCRprimer2, 1328-1248; 3'PCRprimer2, 2168-2192, the coordinates are based on the numbering used in Figure 5-1). The DNA sequence of 3'PCRprimer2 is as same as 5'PCRprimer1. The amplified fragments were resolved on a 1.2% agarose gel and visualized with EtBr.

Cell lines

The preparation of the lymphoblastoid cell lines was described in Chapter 1.

Western blot analysis

Western blot analysis was carried out using the cell lines established from FX06-24 (as a normal) and FX06-25 (as a fragile X patient) (Chapter 3) and HeLa cells. The analysis was done essentially as described in Chapter 1 with minor modifications using either anti-human FMR1 antibodies (Chapter 3) or anti-Xenopus FXR1 antibodies (Chapter 4). Bound primary antibodies were detected using the peroxidase-conjugated goat anti-mouse immunogloblin G + M and ECL reagent (Amersham). Western blots were incubated at 1:400 dilutions.

5-3 Results

Isolation and characterization of human FXR1 cDNA

The X. laevis *FXR1* cDNA was used in hybridization screening to isolate the human *FXR1* cDNA from a HeLa cDNA library. The clones were isolated and DNA sequences of all clones were determined from both strands after cloning into pCRII vector. None of the clones contained the entire open reading frame encoding FXR1, therefore the same library was rescreened using the cDNA encoding more amino terminal region of FXR1 as a probe. One cDNA encoding the amino terminal end of FXR1 was obtained, however was missing the carboxy terminal region. In order to obtain the full length cDNA of *FXR1* as one fragment, RT-PCR was done on HeLa poly(A)$^+$RNA using two primers for 5' and 3' untranslated regions of *FXR1*. The nucleotide sequence and the predicted amino acid sequence of the human FXR1 are shown in Figure 5-1. FXR1 has 86% amino acid sequence identity to human FMR1 in the region containing the KH domains (Siomi et al., 1993; Burd and Dreyfuss, 1994) and is very similar to FMR1 over the amino terminal domain (70% identity), but human FMR1 and FXR1 have entirely different carboxy domains (8% identity)(Figure 5-2A and B). The same phenomenon was seen between X. laevis FMR1 and FXR1. Sequencing of several cDNA revealed that there are alternative spliced forms of FXR1 that differ in a miniexon in the carboxyl portion of the protein. The shorter form diverges from the longer form (shown in Figure 5-2A) beginning with amino acid 535, and contains instead the sequence GKRCD as its carboxy terminus.

The FXR1 gene is located on human chromosome locus 12q13

Mapping of *FXR1* was carried out to determine the chromosomal location of *FXR1*. Reaction conditions allowed specific amplification of the human gene in a background of rodent DNAs. In the mapping panel (Figure 5-3), the cell line containing chromosome 12 contained an amplified fragment of the correct size, 145bp, whereas none of the other samples contained the amplified fragment of interest. Therefore, *FXR1* was tentatively assigned to human chromosome 12. The cell line containing chromosome 21 also showed a very faint amplified fragment at the same size, which may be due to another member of *FMR1* gene family.

Recently, further chromosome mapping of *FXR1* was carried out by fluorescence in situ hybridization (FISH). First, the same set of primers was used to screen pools of yeast artificial chromosome (YACs) from the Washington University CGM with the same conditions described for the mapping panel. Two
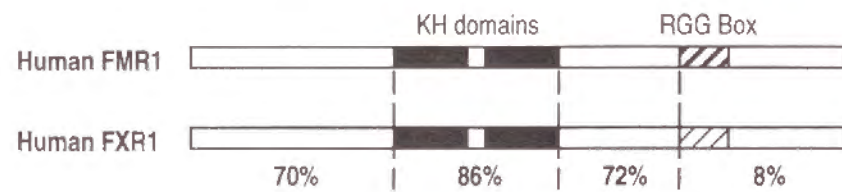
Figure 5-1.

The DNA sequence and the predicted amino acid sequence of human FXR1.
The KH domains are indicated with closed boxes and the amino acid sequence
of the RGG box is underlined.

```
HFMR1    1  MEELVVEVRG SNGAFYKAFV KDVHEDSITV AFENNWQPDR QIPFHDVRFP PPVGYNKDIN ESDEVEVYSR ANEKEPCCWW LAKVRMIKGE
XFMR1    1  MEELAVEVRG SNGAFYKAFM KDVHEQSITV TFENNWQQER QIPFHDVRFP PPSGYNKDIN ERDEVEVYSR ANEKEPCCWW LAKVRMIKGE
HFXR1    1  MADVTVEVRG SNGAFYKGFI KDVHEDSLTV VFENNWQPER QVPFNEVRLP PPPDIKKEIS EGDEVEVYSR ANDQEPCGWW LAKVRMMKGE
XFXR1    1  MEDMTVEVRG SNGAFYKGFV KDVHEDSLTV VFENNWQPER QVPFDEVRMP PLPDIKKEIT EGDEVEVYSR ANDQEPCGWW LAKARMMKGE

HFMR1   91  FYVIEYAACD ATYNEIVTIE RLRSVNPNKP ATKDTFHKIK LDVPEDLRQM CAKEAAHKDF KKAVGAFSVT YDPENYQLVI LSINEVTSKR
XFMR1   91  FYVIEYAACD ATYNEIVTIE RLRSVNPNKP ATKSSFHKVK LDVPEDLRQM CAKDSAHKDF KKAVGAFSVS YDSENYQLVI LSVNEVSIKR
HFXR1   91  FYVIEYAACD ATYNEIVTFE RLRPVNQNKT VKKNTFFKCT VDVPEDLREA CANENAHKDF KKAVGACRIF YHPETTQLMI LSASEATVKR
XFXR1   91  FYVIEYAACD ATYNEIVTFE RLRPVNQNKT VTKNSEFKCT VDVPEDLRES CSNENVHKEF KKAVGACRVY FHAETNQLII LSACESTVKR

HFMR1  181  AHMLIDMHFR SLRTKLSLIM RNEEASKQLE SSRQLASRFH EQFIVREDLM GLAIGTHGAN IQQARKVPGV TAIDLDEDTC TFHIYGEDQD
XFMR1  181  ASMLSDMHFR SLRTKLSLML RNEEASKQLE SSRQLASRFH EQFIVREDLM GLAIGTHGAN IQQARKVPGV TAIDLDEDTC TFHIYGEDQE
HFXR1  181  VNILSDMNLR SIRTKLMLMS RNEEATKHLE CTKQLAAAFH EEFVVREDLM GLAIGTHGSN IQQARKVPGV TAIELDEDTG IFRIYGESAD
XFXR1  181  VNILSDMHLR SIRTKLMLMS RNEEATKHLE CTKQLASRFH EEFIVREDLN GLAIGTHGSN IQQARKVPGI TAIELDEDSG TFRIYGESAE

HFMR1  271  AVKKARSFLE FAEDVIQVPR NLVGKVIGKN GKLIQEIVDK SGVVRVRIEA ENEKNVPQEE EIMPPNSLPS NNSRVGPNAP EEKKHLDIKE
XFMR1  271  AVKKARTYLE FAENVIQVPR NLVGKVIGKN GKLIQEIVDK SGVVRVRIEA ENDKNISQEE ---------- ---------- ----------
HFXR1  271  AVKKARGFLE FVEDFIQVPR NLVGKVIGKN GKVIQEIVDK SGVVRVRIEG DNENKLPRED ---------- ---------- ----------
XFXR1  271  AVKKARSYLE FVEDFIQVPR NLVGKVIGKN GKVIQEIVDK SRVVRVRIEG DNETKLPRED ---------- ---------- ----------

HFMR1  361  NSTHFSQPNS TKVQRGMVPF VFVGTKDSIA NATVLLDYHL NYLKEVDQLR LERLQIDEQL RQIGASSRPP PNR-TDKEKS YVTDDGQGMG
XFMR1  331  ---------- -----GIVPF VFVGTKDSIT NATVLLDYHL NYLKEVDQLR LERLQIDEQL RHIGASSRPP PNR-PDKEKG YLSEDCSGTV
HFXR1  331  ---------- -----GMVPF VFVGTKESIG NVQVLLEYHI AYLKEVEQLR MERLQIDEQL RQIG------ ---------- ----------
XFXR1  331  ---------- -----GMVPF VFVGTKENIG NVQVLLEYHI AYLKEVEQLR MERLQIQEQL RQIGMGFRPS SSRGTEKEKG YATDESTASS

HFMR1  450  -RGSRPYRNR GHGRRGPGYT S--GTNSEAS NASETESDHR DELSDWSLAP TEEERESFLR NGDGRRRGGG -GRGQ-GGRG RGGGFKGNDD
XFMR1  405  -RGSRPYSNR GRSRRGTGYA S--GTNSEAS NASETESDHR DELSDWSLAP AEDDRDNYHR RGDGRRR-GG -TRGQ-GMRG RGG-FKGNDD
HFXR1  380  ---SRSYSGR GRGRRGPNYT SGYGTNSELS NPSETESERK DELSDWSLA- GEDNRDSRHQ R-DSRRRPGG RGRSVSGGRG RGGPRGGKSS
XFXR1  406  VRGSRSYSGR GRGRRGPNYT SGYGTNSELS NPSETESERK EELSDWSLA- GEDERESRQQ R-DSRRRPGG RGRSGSAGRG RGGSRGGKSS

HFMR1  535  HSRTDNRPR- NPREAKGRTT DGSLQIRVNC NNERSVH-TK TLQNTSSEGS RLRTGKDRNQ KKEKPDSVDG QQPLVNGVP*
XFMR1  488  QPRPDNRQR- NSRETKARTS DGSLQIRLDC NNERSVH-TK TLQNASVDGS RLRTGKDRVQ KKEKTDGYDG PQVVVNGVP*
HFXR1  464  ISSVLKDPQS NPYSLLDNT- E-SDQTADTD ASE-SHHSTN RRRRSRRRRT DEDAVLMDGM TESDTASVNE NGLVTVADYI SRAESQSRQR
XFXR1  493  ISSVLKDPQS KPYSLLDNT- E-SDQTADTD ASE-SHHNTN RRRRSRRRRT DEDSSLMDGM TELDNASVNE NGLVTVADYI SRAESQSRQR

HFXR1  551  NLPRETLAKN KKEMAKDVIE EHGPSEKAIN GPTSASGDDI SKLQRTPGEE KINTLKEENT- QEAAVLNGVS*
XFXR1  580  NLPKETLAKG KKEKVKDVIE EHGPSEKVIN GPRASSADKA SNPQ-TASTE RNQASCQDGSK QEA-IMNGVS*
```

Fifure 5-2. (A)Amino acid alignment of human FMR1, human FXR1, X. laevis FMR1 and X. laevis FXR1.

Identical positons are highlighted and gaps are introduced for optimal alignment. The KH domains and
the RGG box are boxed with black lines and gray lines respectively.

Human FMR1

Human FXR1

70% | 86% | 72% | 8%

KH domains   RGG Box

Figure 5-2. (B) The structure of human FMR1 and FXR1 are schematically shown, the stippled boxes representing the KH domains and the cross-hatched box representing the RGG box. The identity of amino acid sequence of human FXR1 is expressed as percent relative to human FMR1 amino acid sequence.

A 1 2 3 4 5 6 7 8 C 9 10 11 12 13 14 15 16 G 17 18 19 20 21 22 X Y A C G T - H M - Ha - T

Figure 5-3. Chromosome mapping of human *FXR1*.

Hybrids 1 to 22, X and Y were screened for the presence of genomic fragment of human *FXR1* by PCR. PCR was done using two ^{32}P-labeled primers specific for human *FXR1*. The 145bp PCR products are seen in the lane 12 and H (human), indicating that *FXR1* locates on human chromosome 12. Lane Ha (hamster) and M (mouse) do not show any PCR products, suggesting that the primers are specific for human FXR1 and are not cross-reacting to rodent genomic DNA. Lane A, C, G, and T are molecular markers. "-" indicates lanes unprogrammed.

YAC clones containing *FXR1* were identified. Fluorescent in situ hybridization using these two YACs revealed the localization of *FXR1* to chromosome 12q13 (data not shown).

Tissue distribution of FXR1 expression

To determine the expression of *FXR1* mRNA in different human tissues, RT-PCR was performed using specific primers to *FXR1* on poly(A)$^+$RNA from heart, kidney, brain, testis, and HeLa cells. As control, an equivalent experiment was done using two primers specific to *FMR1* cDNA. *FXR1* mRNA was detected in all tissues tested, but different size bands were observed in various tissues (Figure 5-4). For example, while HeLa cells contain only one variant *FXR1* mRNA, at least two forms can be detected in brain and testis, and in heart there is an additional large form. The major smaller HeLa band and the longer testis band were cloned and sequenced and their sequences corresponded to the *FXR1* shorter cDNA form and the *FXR1* longer cDNA form respectively described above. These findings suggest that there is considerable tissue-specific alternative splicing of *FXR1* pre-mRNA at least for the carboxy part and immediate 3'UTR of the mRNA. A similar complex tissue-specific pattern of expression has been reported for *FMR1* (Eichler et al., 1993), although multiple forms of *FMR1* were not observed by RT-PCR with the primers used in this experiment.

Expression of FXR1 in fragile X syndrome patients

Most fragile X syndrome patients do not express *FMR1* mRNA or the protein product. It was, therefore, of particular interest to determine if the expression of the related protein, FXR1, is also affected in these patients. To do so, RT-PCR were (Figure 5-5A) and immunoblotting (Figure 5-5B) carried out on lymphoblastoid cells of a fragile X patient and his normal sibling (Chapter 3). By RT-PCR, both the normal sibling and the patient express *FXR1* mRNA, while the patient, as expected, does not express *FMR1* mRNA. The same is seen for the protein products of *FXR1* and *FMR1*, respectively. Because of inherent limitations of RT-PCR it is not possible to draw quantitative conclusions from this experiment. It does, however, appear from the immunoblotting experiments that the amount of FXR1 produced in the patient cells is not reduced compared
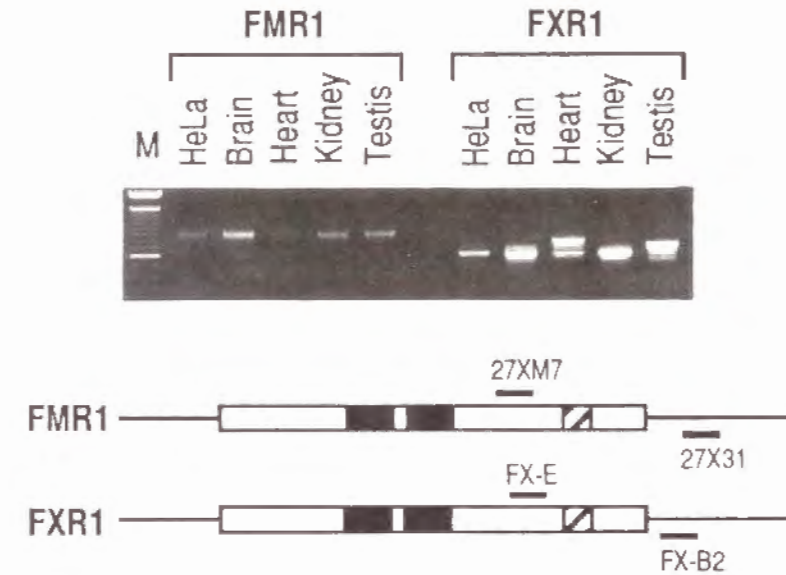

Figure 5-4. *FXR1* gene expression in various human tissues.

RT-PCR was performed on poly(A)$^+$RNAs from human brain, testis, kidney, heart and HeLa cells using the primers specific to human *FMR1* or *FXR1*. The schematic drawing of FMR1 and FXR1 mRNAs are shown below the agarose gel. The closed boxes and the stippled box indicate the **KH** domains and the RGG box respectively. The bars indicate the primers used for PCR. M, 100bp ladder molecular markers.

A

FMR1   FXR1

HeLa Normal Patient   HeLa Normal Patient

B

FMR1   FXR1

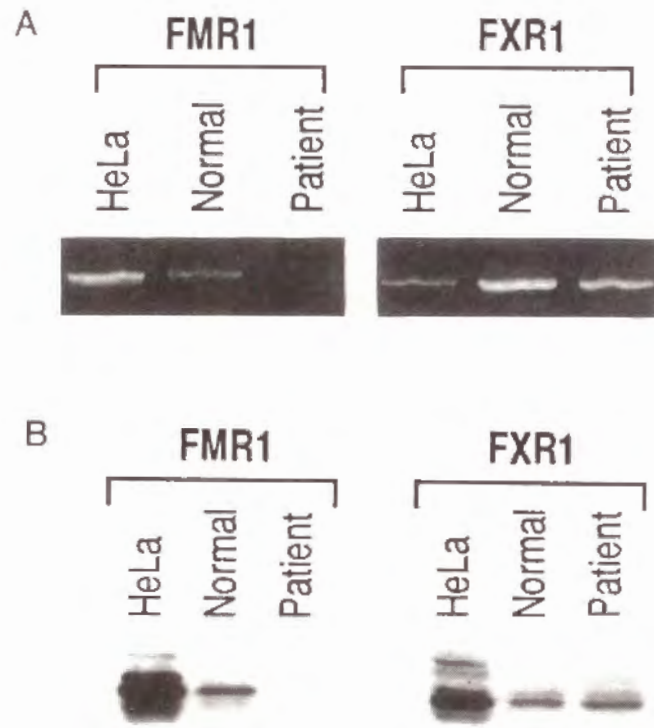HeLa Normal Patient   HeLa Normal Patient

Figure 5-5.   Expression of *FMR1* and *FXR1* mRNAs and the protein products in a fragile X syndrome patient and his normal sibling.

(A) RT-PCR was performed on poly(A)$^+$RNAs from HeLa cells and lymphoblastoid cell lines of a fragile X patient and his normal sibling.
(B) Western blot. Total cellular proteins from HeLa cells and lymphoblastoid cell lines of the fragile X patient and his normal sibling were resolved by SDS-PAGE, transferred to nitrocellulose membrane, and probed with either the anti-FMR1 or anti-FXR1 antibodies.

to normal. Thus, FXR1 expression is not drastically affected by the lack of expression of FMR1, and therefore, *FXR1* gene expression does not appear to be linked to that of the *FMR1* gene.

Inverted repeat in the 5'UTR and 3'UTR of FXR1 mRNA

Examination of the nucleotide sequence of the X. laevis *FXR1* cDNA and the human *FXR1* cDNA revealed a striking and unusual mRNA structure. *FXR1* mRNA from both organisms contains a sequence of approximately 90 nucleotides that is present as a perfect inverted repeat in the 5'UTR and 3'UTR (Figure 5-6). The orientation of the repeats and their distance from the open reading frame are shown in Figure 5-6. The repeats in human and X. laevis are not identical but they are much more highly conserved than in the nucleotide sequence of the coding regions. This suggested that the repeats have an important and specific function. To confirm that the repeats were not the result of some artifact introduced in the course of the construction, propagation or handling of the cDNA libraries, an RT-PCR experiment was carried out. First of all, poly(A)$^+$RNA from HeLa was reverse transcribed using the oligo(dT) primer (=cDNAprimer2), followed by PCR with either 5'PCRprimer1 and 3'PCRprimer1, or 5'PCRprimer2 and 3'PCRprimer2. With the latter set of primers, one fragment of the correct size, 864bp, was amplified (lane3, Figure 5-6C). With the first set of primers, a 627bp fragment was amplified as expected (data not shown). The identical fragment was obtained by another RT-PCR using the same set of PCR primers after RNAs were reverse transcribed using cDNAprimer1 instead of the oligo(dT) primer (lane1, Figure 5-6C). In this case, the cDNAs contain only the amino terminal region of *FXR1* gene, so that the RT-PCR product was not from the region containing the 3'UTR repeat, confirming that 5'UTR does contains its own repeat. All PCR products were cloned into pCRII vector and the DNA sequences were verified.

5-4 Discussion

A novel gene, *FXR1*, that is highly homologous to the fragile X mental retardation gene, *FMR1*, found in Xenopus laevis originally (Chapter 4), was isolated from human and sequenced. Like Xenopus FXR1, the amino acid
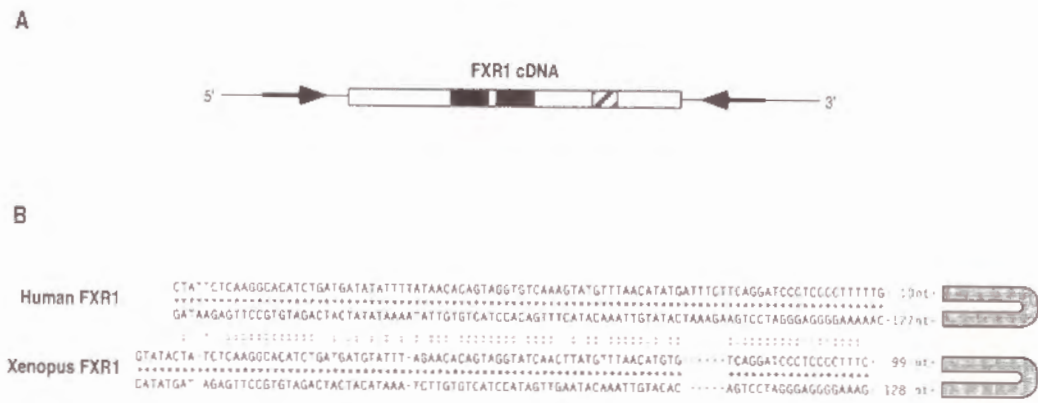
A

B

Figure 5-6. Inverted repeat sequence in the 5' and 3' untranslated regions of the transcript of *FXR1*.

(A) Schematic drawing of the transcript of *FXR1*. The open reading frame and the inverted repeat sequence are indicated by a open box and arrows respectively. The black and the stippled boxes represent the KH domains and the RGG box respectively.
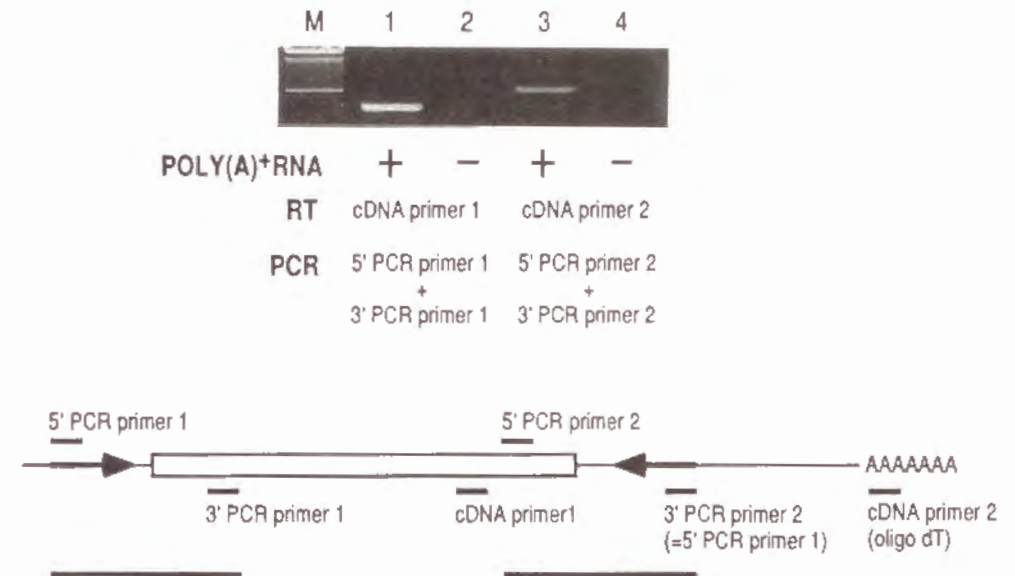(B) The inverted repeat sequence of *FXR1* are aligned between 5' and 3' UTRs and between human and X. laevis.

Figure 5-6. Inverted repeat sequence in the 5' and 3' untranslated regions of the transcript of *FXR1*.

(C) RT-PCR was performed on poly(A)⁺RNA from HeLa cells with the primers indicated below the agarose gel. Lane2 and 4: no poly(A)⁺RNA as negative controls. The human *FXR1* mRNA is shown schematically. cDNA primers, 5'PCR primers, and 3'PCR primers are indicated by bars. The inverted repeat is indicated by soild arrows on the drawing. The bars below the schematic drawing of *FXR1* mRNA indicates the RT-PCR products.

sequence of human FXR1 is very similar to FMR1 over the amino terminal region containing the KH domains. The carboxy portion of human FXR1 is quite different from that of human FMR1, though the region has high similarity between human and X. laevis (78% identity). Like FMR1, FXR1 is expressed in many human tissues. While *FMR1* mRNA and the protein product are not expressed in cells of a patient with fragile X syndrome, FXR1 is found at apparently normal levels in these cells. In Chapter 4, it was speculated that FMR1 and FXR1 may have very similar functions because of the experimental data from RNA-binding assays and immunofluorescence microscopy. However, it seems that FXR1 protein can not compliment the lack of FMR1 protein function in fragile X patients. This may be because FMR1 and FXR1 have specific regulatory functions such as protein-protein interactions, since these proteins contain quite different carboxy terminal regions. Alternatively, they may have to interact with each other or at least exist in the same complex to perform the correct functions. It would be of interest to see what kind of proteins and/or other factors are associated with FMR1 and/or FXR1 proteins.

The gene encoding FXR1 is located on autosomal chromosome 12 at 12q13, while the gene encoding FMR1 is located on X chromosome at Xq27.3. It is not yet known if there are mental retardation genes in this vicinity. *FXR1* gene knockout mice should help to elucidate the physiological function of FXR1.

The mRNA of *FXR1* does not contain the CGG repeats in the 5' UTR that is a characteristic of *FMR1*. Instead, *FXR1* has an unusual structure, approximately 90 nucleotide inverted repeats in the 5' and the 3'UTRs in its mRNA. The inverted repeats are capable of forming a perfect duplex. The inverted repeats of *FXR1* are the first such structures found in animal cells. There are some examples in plants, such as zein mRNAs from Zea mays (Larkins et al., 1984), A-gliadin mRNA(Anderson et al., 1984), phaseolin mRNA(Slightom et al., 1983), patatin mRNA(Mignery et al., 1984), soybean actin mRNA(Shah et al., 1982), and wheat histon H4 mRNA(Tabata et al., 1983). In the case of zein mRNAs, the inverted repeats are also found in the 5' and the 3'UTRs (Spena et al., 1982). However, the size of the inverted repeats are much smaller than *FXR1* and contain some gaps. The secondary structure of the zein mRNAs affect its translational potential and the translational block is released after deletion of the 3' inverted repeat (Spena et al., 1985). It was concluded that

the tlanslational block is caused by hybrid formation between the two inverted repeats. Moreover, the translational efficiency of zein mRNAs is affected by varying the length or the primary structure of the 5' untranslated region (Spena et al., 1985). It must now be determined if the predicted secondary structure actually forms in the *FXR1* mRNA. If so, it is most likely that the secondary structure of *FXR1* mRNA also affect its translational potential. It will be of great interest to determine the function of the inverted repeats, although an obvious thought is that the mRNA serves some structural role and regulates the translation of the *FXR1* mRNA.

5-5 Summary

A cDNA encoding FXR1, that is a highly homologous protein to FMR1, found in Xenopus laevis originally, has been isolated from HeLa cells and sequenced. Like *FMR1*, *FXR1* mRNAs are expressed in many human tissues and tissue-specific alternative splicing of *FXR1* pre-mRNA was found at least for the carboxy part and immediate 3'UTR of the mRNA. Cells of a patient with fragile X syndrome that do not have any detectable FMR1 express normal levels of FXR1, suggesting that FMR1 and FXR1 may have specific regulatory functions such as protein-protein interactions. *FXR1* is located on human autosomal chromosome 12 at 12q13. The mRNA of *FXR1* has an approximately 90 nucleotide perfect inverted repeat sequence in the 5' and 3' UTRs. It is likely to have an important regulatory function such as a translational regulation.

References

Abitbol, M., Menini, C., Delezoide, A.-L., Rhyner, T., Vekemans, M., and Mallet, J. (1993). Nucleus basalis magnocellularis and hippocampus are the major sites of FMR1 expression in the fetal brain. Nature Genet. 4, 147-152.

Anderson, O. D., James, C. L., Gautier, M.-F., and Greene, C. (1984). Nucleic acid sequence and chromosome assignment of a wheat strage protein gene. Nucl. Acids Res. 12, 8129-8144.

Aris, J. P., and Blobel, G. (1991). cDNA cloning and sequencing of human fibrillarin, a conserved nucleolar protein recognized by autoimmune antisera. Proc. Natl. Acad. Sci. USA 88, 931-935.

Ashley, C. T., Jr., Wilkinson, K. D., Reines, D., and Warren, S. T. (1993a). FMR1 protein: Conserved RNP family domains and selective RNA binding. Science 262, 563-566.

Ashley, C. T., Sutcliffe, J. S., Junst, C. B., Leiner, H. A., Eichler, E. E., Nelson, D. L., and Warren, S. T. (1993b). Human and murine FMR1: alternative splicing and translational initiation downstream of the CGG-repeat. Nature Genet. 4, 244-251.

Bachner, D., Steinbach, P., Wohrle, D., Just, W., Vogal, W., Hameister, H., Manca, A., and Poustka, A. (1993). Enhanced FMR1 expression in testis. Nature Genet. 4, 115-116.

Bandziulis, R. J., Swanson, M. S., Dreyfuss, G. (1989). RNA-binding proteins as developmental regulators. Genes Dev. 3, 431-437.

Bell, M. V., Hirst, M. C., Nakahori, Y., MacKinnon, R. N., Roche, A., Flint, T. J., Jacobs, P. A., Tommerup, N., Tranebjaerg, L., Froster-Iskenius, U. (1991). Physical mapping across the fragile X: hypermethylation and clinical expression of the fragile X syndrome. Cell 64, 861-866.

Burd, C. G., and Dreyfuss, G. (1994). Coserved structures and deversity of functions of RNA-binding proteins. Science 265, 615-621.

Buvoli, M., Biamonti, G., Tsoulfas, P., Bassi, M. T., Ghetti, A., Riva, S. and Morandi, C. (1988). cDNA cloning of human hnRNP protein A1 reveals the existence of multiple RNA isoforms. Nucl. Acids Res. 16, 3751-3770.

Caskey, C. T., Pizzuti, A., Fu, Y.-H., Fenwick, Jr, R. G., Nelson, D. L. (1992). Triplet repeat mutations in human disease. Science 256, 784-789.

Choi, Y. D. and Dreyfuss, G. (1984). Isolation of the heterogeneous nuclear RNA-ribonucleoprotein complex (hnRNP): A unique supramolecular assembly. Proc. Natl. Acad. Sci. USA 81, 7471-7475.

Christensen, M. E. and Fuxa, K. P. (1988). The nucleolar protein, B-36, contains a glycine and dimethylarginine-rich sequence conserved in several other nuclear RNA-binding proteins. Biochem. Biophys. Research Comm. 155,1278-1283.

Cobianchi, F., Karpel, R. L., Williams, K. R., Notario, V., and Wilson, S. H. (1988). Mammalian heterogeneous nuclear ribonucleoprotein complex protein A1. J. Biol. Chem. 263, 1063-1071.

Cruz-Alvarez, M., and Pellicer, A. (1987). Cloning of full-length complementary DNA for an Artemia salina glycine protein. J. Biol. Chem. 262, 13377-13380.

De Boulle, K., Verkerk, A. J. M. H., Reyniers, E., Vits, L., Hendrickx, J., VanRoy, B., VanDenBos, F., deGraaff, E., Oostra, B. A. and Willems, P.J. (1993). A point mutation in the FMR1 gene associated with fragile X mental retardation. Nature Genet. 3, 31-35.

Devys, D., Lutz, Y., Rouyer, N., Bellocq, J.-P., and Mandel, J.-L. (1993). The FMR1 protein is cytoplasmic, most abundand in neuron and appears normal in carriers of a fragile X premutation. Nature Genet. 4, 335-340.

Deward, G. W., Bnckley, D. D., Spurbeck, J. L., and Jalal, S. M. (1992). Cytogenetic Guidelines for fragile X studies tested in routine practice. Am. J. Med. Genet. 44, 816-.

Dreyfuss, G., Philipson, L. and Mattaj, I. W. (1988). Ribonucleoprotein particles in cellular processes. J. Cell Biol. 106, 1419-1425.

Dreyfuss, G., Matunis, M. J., Piñol-Roma, S. and Burd, C. G. (1993). hnRNP proteins and the biogenesis of mRNA. Annu. Rev. Biochem. 62, 289-321.

Dumont, J. N. (1972). Oogenesis in Xenopus laevis. J. Morphol. 136, 153-179.

The Dutch-Belgian Fragile X Consortium (1994). Fmr1 knockout mice: a model to study fragile X mental retardation. Cell 78, 23-33.

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.

Engebrecht, J. and Roeder, G. S. (1990). MER1, a yeast gene required for chromosome pairing and genetic recombination, is induced in meiosis. Mol. Cell. Biol. 10, 2379-2389.

Engebrecht, J. A., Voelkel-Meiman, K., Roeder, G. S. (1991). Meiosis-specific RNA splicing in yeast. Cell 66, 1257-1268.

Flynn, G. A., Hirst, M. C., Knight, S. J. L., MacPherson, J. N., Barber, J. C. K., Flannery, A. V., Davies, K. E., and Buckle, V. J. (1993). Identification of the FRAXE fragile site in two families ascertained for X linked mental retardation. J. Med. Genet. 30, 97-103.

Frankel, A. D., Mattaj, I. W. and Rio, D. C. (1991). RNA-protein interactions. Cell 67, 1041-1046.

Fu, Y.-H., Kuhl, D. P. A., Pizzuti, A., Peiretti, M. Sutcliffe, J. S., Richards, S., Verkerk, A. J. M. H., Holden, J. J. A., Fenwick, R. G., Jr., Warren, S. T., Oostra, B. A., Nelson, D. L., and Caskey, C. T. (1991). Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell 67, 1047-1058.

Gedeon, A. K., Baker, E., Robinson, H., Partington, M. W., Gross, B., Manca, A., Korn, B., Poustka, A., Yu, S., Sutherland, G. R., and Mulley, J. C. (1992). Fragile X syndrome without CCG amplification has an FMR1 deletion. Nature Genet. 1, 341-344.

Gibson, T. J., Thompson, J. D., and Heringa, J. (1993a). The KH domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NusA and is probabry involved in binding to nucleic acid. FEBS Lett. 324, 361-366.

Gibson, T. J., Rice, P. M., Thompson, J. D., and Heringa, J. (1993b). KH domains within the FMR1 sequence suggest that fragile X syndrome stems from a defect in RNA metabolism. Trends Biochem. Sci. 18, 331-333.

Hansen, R. S., Gartler, S. M., Scott, C. R., Chen, S., 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.

Heitz, D., Rousseau, F., Devys, D., Saccone, S., Abderrahim, H., Le Paslier, D., Cohen, D., Vincent, A., Toniolo, D., Della Valle, G. (1991). Isolation of sequences that span the fragile X and identification of a fragile X-related CpG island. Science 251, 1236-1239.

Heitz, D., Devys, D., Imbert, G., Kretz, C., and Mandel, J. L. (1992). Inheritance of the fragile X syndrome: size of the fragile X mutaion is a major determinant of the transition to full mutation. J. Med. Genet. 29, 794-801.

Hinds, H. L., Ashley, C. T., Sutcliffe, J. S., Nelson, D. L., Warren, S. T., Housman, D. E., and Shhalling, M. (1993). Tissue specific expression of FMR1 provides evidence for a functional role in fragile X syndrome. Nature Genet. 3, 36-43.

Hirst, M. C., Barnicoat, A., Flynn, G., Wang, Q., Daker, M., Buckle, V. J., Davies, K. E., and Bobrow, M. (1993). The identification of a third fragile site, FRAXF in Xq27-28 distal to both FRAXA and FRAXE. Hum. Mol. Genet. 2, 197-200.

Hochuli, E., Dobeli, H., Schacher, A. (1987). New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. J. Chromat. 411, 177-184.

Honda, B. M. and Roeder, R. G. (1980). Association of a 5S gene transcription factor with 5S RNA and altered levels of the factor during cell differentiation. Cell 22, 119-126.

Hope, I. A. and Struhl, K. (1986) Functional dissection of a eukaryotic transcriptional activator proteins, GCN4 of yeast. Cell 46, 885-894.

Jacky, P. B., Ahuja, Y. R., Anyane-Yeboa, K., Breg, W. R., Carpenter, N. J., Froster-Iskenius, U. G., Fryns, J. P., Glover, T. W., Gustavson, K.-H., Hoegerman, S. F., Holgren, G., Howard-Peebles, P. N., Jenkins, E. C., Krawczun, M. S., Neri, G., Pettigrew, A., Schapp, T., Schonberg, S. A., Shapiro, L. R., Sppiner, N., Steinbach, P., Vianna-Morgante, A. M., Watson, M. S., and Wilmot, P. L. (1991). Giudelines for the preparation and analysis of the fragile X chromosome in lymphocytes. Am. J. Med. Genet. 38, 400-.

Kiledjian, M. and Dreyfuss, G. (1992). Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. EMBO J. 11, 2655-2664.

Kremer, E. J., Pritchard, M., Lunch, 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(CCG)n. Science 252, 1711-1714.

Larkins, B. A., Pederson, K., Handa, A. K., Hurkman, W. J., and Smith, L. D. (1979). Synthesis and processing of maize strage proteins in Xenopus laevis oocytes. Proc. Natl. Acad. Sci. USA 76, 6448-6452.

Lischwe, M. A., Ochs, R. L., Reddy, R., Cook, R. G., Yeoman, L. C., Tan, E. M., Reichlin, M., and Busch, H. (1985). Purification and partial characterization of a nucleolar scleroderma antigen (Mr=34,000; pI, 8.5) rich in NG, NG-Dimethylarginine. J. Biol. Chem. 260, 14304-14310.

Matunis, M. J., Michael, W. M., Dreyfuss, G. (1992). Characteriation and primary structure of the Poly(C)-binding heterogeneous nuclear ribonucleoprotein complex K protein. Mol. Cell. Biol. 12, 164-171.

Meijer, H., de Graaff, E., Merckx, D. M. L., Jongbloed, R. J. E., de Die-Smulders, C. E. M., Engelen, J. J. M., Fryns, J.-P., Curfs, P. M. G., and Oostra, B. A. (1994). A deletion of 1.6kb proximal to the CGG repeat of the FMR1 gene causes the clinical phentype of the fragile X syndrome. Hum. Mol. Genet. 3, 615-620.

Mignery, G. A., Pikaard, C. S., Hannapel, D. J., and Park, W. D. (1984). Isolation and sequence analysis of cDNAs for the major potato tuber protein, patatin. Nucl. Acids Res. 12, 7987-8000.

Murray, M. T., Schiller, D. L., and Franke, W. W. (1992). Sequence analysis of cytoplasmic mRNA-binding proteins of Xenopus oocytes identifies a family of RNA-binding proteins. Proc. Natl. Acad. Sci. USA 89, 11-15.

Nadler, S. G., Merrill, B. M., Roberts, W. J., Keating, K. M., Lisbin, M. J., Barnett, S. F., Wilson, S. H., and Williams, K. R. (1991). Interactions of the A1 heterogeneous nuclear ribonucleoprotein and its proteloytic derivative, UP1, with RNA and DNA: Evidence for multiple RNA binding domains and salt-dependent binding mode transitions. Biochemistry 30, 2968-2975.

Nussbaum, R. L., Ledbetter, D. H. (1986). Fragile-X syndrome: a unique mutation in man. Annu. Rev. Genet. 20, 109.

Oberle, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boule, J., Bertheas, M. F., Mandel, J. L. (1991). Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science 252, 1097-1102.

Oostra, B. A., and Verkerk, A. J. M. H. (1992). The fragile X syndrome: isolation of the FMR1 gene and characterization of the fragile X mutation. Chromosoma. 101, 381-387.

Pelham, H. R. and Brown, D. D. (1980). A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. Proc. Natl. Acad. Sci. USA 77, 4170-4174.

Pergolizzi, R. G., Erster, S. h., Goonewardena, P., and Brown, W. T. (1992). Detection of full fragile X mutation. Lancet 339, 271-.

Picretti, 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 FMR1 gene in fragile-X syndrome. Cell 66, 817-822.

Piñol-Roma, S., Choi, Y. D., Matunis, M. J., Dreyfuss., G. (1988). Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. Genes Dev. 2, 215-227.

Régnier, P., Grunberg-Manago, M., Portier, C. (1987). Nucleotide sequence of the pnp gene of Escherichia coli encoding polynucleotide phosphorylase. Homology of the primary structure of the protein with the RNA-binding domain of ribosomal protein S1. J. Biol. Chem. 262, 63-68.

Richards, R. I., Sutherland, G. R. (1992). Fragile X Syndrome: The molecular picture comes into focus. Trends in Genetics 8, 249-254.

Riva, S., Morandi, C., Tsoulfas, P., Pandolfo, M., Biamonti, G., Merrill, B., Williams, K. R., Multhaup, G., Beyreuther, K., Werr, H., Henrich, B., and Schafer, K. P. (1986). Mammalian single-stranded DNA binding protein UP1 is derived from the hnRNP core protein A1. EMBO J. 5, 2267-2273.

Robinow, S., Campos, A. R., Yao, K.-M., White, K. (1988). The elav gene product of Drosophila, required in neurons, has three RNP consensus motifs. Science 242, 1570-1572.

Rosenberg, A. H., Lade, B. N., Chui, D. S., Lin, S. W., Dunn, J. J., Studier, F. W. (1987). Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene 56, 125-135.

Rousseau, F., Heitz, D., Biancalana, V., Blumen feld, S., Kretz, C., Boue, J., Tommerup, N., Van Der Hage, C., DeLozier-Blanchet, C., Croquette, M. F., et al. (1991). Direct diagnosis by DNA analysis of the fragile X syndrome of mental retardation. N. Engl. J. Med. 325, 1673-1681.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequence with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Schmidt, C., Henkel, B., Poschl E., Zorbas, H., Purschke, W. G., Gloe, T. R., Muller, P. K. (1992). Complete cDNA sequence of chicken vigilin, a novel protein with amplified and evolutionary conserved domains. Eur. J. Biochem. 206, 625-634.

Shah, D. M., Hightower, R. C., and Mearher, R. B. (1982). Complete nucleotide sequence of a soybean actin gene. Proc. Natl. Acad. Sci. USA 79, 1022-1026.

Sherman, S. L., Morton, N. E., Jacobs, P. A., and Turner, G. (1984). The merker (X) syndrome: a cytogenetic and genetic analusis. Ann. Hum. Genet. 48, 21-.

Sherman, S. L., Jacob, P. A., Morton, N. E., Froster-Iskenius, U., Howard-peebles, p. N., Nielson, K. B., Partington, M. W., Surtherland, G. R., Turner, G., and Watson, M. (1985). Further segregation analysis of the fragile X syndrome with special reference to transmitting males. Human Genet. 69, 289-.

Siomi, H., Matunis, M. J., Michael, W. M., Dreyfuss, G. (1993). The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. Nucl. Acids Res. 21, 1193-1198.

Siomi, H., Choi, M., Siomi, C. M., Nussbaum, R. L., and Dreyfuss, G. (1994). Essential role for KH domains in RNA binding: Impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. Cell 77, 33-39.

Slightom, J. L., Sun, S. M., and Hall, T. C. (1983). Complete nucleotide sequence of a French bean strage protein gene: Phaseolin. Proc. Natl. Acad. Sci. USA 80, 1897-1901.

Smith, M. C., Furman, T. C., Ingolia, T. D., Pidgeon, C. J. (1988). Chelating peptide-immobilized metal ion affinity chromatography. A new concept in affinity chromatography for recombinant proteins. J. Biol. Chem. 263, 7211-7215.

Spena, A., Viotti, A., and Pirrota, V. (1982). A homologous repetitive block structure underlies the heterogeneity of heavy and light chain zein genes. EMBO J. 1, 1589-1594.

Spena, A., Krause, E., and Dobberstein, B. (1985). Translaration efficiency of zein mRNA is reduced by hybrid formation between the 5'- and 3'-untranslated region. EMBO J. 4, 2153-2158.

Spiridonova, V. A., Akhomanova, A. S., Kagramanova, V. K., Köpke, A. K., Mankin. A. S. (1989). Ribosomal protein gene cluster of Halobacterium halobium: nucleotide sequence of the genes coding for S3 and L29 equivalent ribosomal proteins. Can. J. Microbiol. 35, 153-159.

Studier, F. W., Rosenberg, A. H., Dunn, J. J., Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. 185, 60-89.

Sutherland, G. R. (1977). Fragile sites on human chromosomes: demonstration of their dependence on the type of tissue culture medium. Science 197, 256-266.

Sutherland, G. R., Hecht, F., Mulley, J. C., Glover, T. W., Hecht, B. S. (1985). *Fragile Sites on Human Chromosomes*. (Oxford Univ. Press, New York), 304.

Swanson, M. S. and Dreyfuss, G. (1988). Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA binding proteins. Mol. Cell. Biol. 3,2237-2241.

Swanson, M. S., Nakagawa, T. Y., Levan, K., Dreyfuss, G. (1987). Primary structure of the human ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA and pre-rRNA-binding proteins. Mol. Cell. Biol. 7, 1731-1739.

Tabata, T., Sasaki, K., and Iwabuchi, M. (1983). The structural organization and DNA sequence of a wheat histon H4 gene. Nucl. acids res. 11, 5865-5875.

Tommerup, N. (1989). Cytogenetics of the fragile site at Xq27. In the fragile X syndrome. K. E. Davies, ed. (Oxford Unuversity Press), pp. 102-135.

Tyc, K. and Steitz, J. A. (1989). U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. EMBO J. 8, 3113-3119.

Verheij, C., Bakker, C. E., de Graaff, E., Keulemans, J., Willemsen, R., Verkerk, A. J. M. H., Galjaard, H., Ruser, A. J. J., Hoogeveen, A. T., and Oostra, B. A. (1993). Characterization and localization of the FMR1 gene product. Nature 363, 722-724.

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, F., Eussen, B. E., van Ommen, G.-J. B., Blonden, L. A. J., Riggins, G. J., Chastain, J. L., Kunst, C. B., Galijaard, H., Caskey, C. T., Nelson, D. L., Oostra, B. A., and Warren, S. T. (1991). Identification of a gene (FMR1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile-X syndrome. Cell 65, 905-914.

Verkerk, A. J. M. H, de Graaff, E., De boulle, K., Eichler, E. E., Konecki, D. S., Reyniers, E., Mance, A., Poustka, A., Willems, P. J., Nelson, D. L., and Oostra, B. A. (1993). Alternative splicing in the fragile X gene FMR1. Human Mol. Genet. 2, 399-404.

Vincent, A., Heitz, D., Petit, C., Kretz, C., Oberle, I., Mandel, J. L. (1991). Abnormal pattern detected in fragile-X patients by pulsed-field gel electrophoresis. Nature 349, 624-626.

Wilson, J. M., Frossard, P., Nussbaum, R. L., Caskey, C. T., Kelley, W. N. (1983). Human hypoxanthine-guanine phosphoribosyltransferase. J. Clin. Invest. 72, 767-772.

Wilson, I. A., Niman, H. L., Houghten, R. A., Cherenson, A. R., Connolly M. L. and Lerner, R. A. (1984). The structure of an antigenic deteminant in a protein. Cell 37, 767-778.

Wohrle, D., Kotzot, D., Hirst, M.C., Manca, A., Korn, B., Schmidt, A., Barbi, G., Rott, H. D., Poustka, A., Davies, K. E. and et al. (1992). A microdeletion of less than 250 kb, including the proximal part of the FMR1 gene and the fragile-X site, in a male with the clinical phenotype of fragile-X syndrome. Am. J. Hum. Genet. 51, 299-306.

Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M. F., Polakis, P., McCormick, F. (1992). Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. Cell 69, 551-558.

Yu, S., Pritchard, M., Kremer, E., Lunch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, J. C., Warren, S. T., Schlessinger, D., Sutherland, G. R. and Richards, R. I. (1991). Fragile-X genotype characterized by an unstable region of DNA. Science 252, 1179-1181.

List of publications

Chapter 1 and 3

Siomi, H., Siomi, C. M., 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.

Chapter 2

Siomi, H., Choi, M., Siomi, C. M., Nussbaum, R. L., and Dreyfuss, G. (1994). Essential Role for KH Domains in RNA Binding: Impaired RNA Binding by a Mutation in the KH Domains of FMR1 that Causes Fragile X Syndrome. Cell 77, 33-39.

Chapter 4 and 5

Siomi, C. M., Siomi, H., Sauer, H. W., Srinivasan, S., Nussbaum, R. L., and Dreyfuss, G. FXR1, An Autosomal Homolog of the Fragile X Mental Retardation Gene with an Unusual mRNA Structure. manuscript in preparation.