Human Molecular Genetics, 1993, Vol. 2, No. 4 399–404

# Alternative splicing in the fragile X gene *FMR1*

Annemieke J.M.H.Verkerk, Esther de Graaff, Kristel De Boulle[1], Evan E.Eichler[2], David S.Konecki[3], Edwin Reyniers[1], Antonella Manca[3], Annemarie Poustka[3], Patrick J.Willems[1], David L.Nelson[2] and Ben A.Oostra*

Department of Clinical Genetics, Erasmus University, Dr Molewaterplein 50, 3153 DR Rotterdam, The Netherlands, [1]Department of Medical Genetics, University of Antwerp-UIA, 2610 Antwerp, Belgium, [2]Institute for Molecular Genetics, Baylor College of Medicine, Houston, TX 77030, USA and [3]Deutsches Krebsforschungszentrum, Heidelberg, Germany

Received December 10, 1992; Revised and Accepted February 1, 1993 EMBL accession no. X69962

## ABSTRACT

**The *FMR1* gene, associated with fragile X syndrome, has recently been cloned and the sequence of partial cDNA clones is known. We have determined additional cDNA sequences both at the 5′ and 3′ end. We have characterized the expressed gene by means of RT-PCR in various tissues and have found that alternative splicing takes place in the *FMR1* gene, which does not seem to be tissue specific. When the different alternative splicing events are combined, 12 distinct mRNA products could result from *FMR1* expression in each tested tissue. In all these transcripts the open reading frame is maintained until the same stop codon. At the 3′ end alternative use of polyadenylation signals is found. The alternative splicing allows functional diversity of the FMR-1 gene. Whether all the possible proteins will be synthesized and whether they will be functionally active has to be determined.**

## INTRODUCTION

Fragile X syndrome is an X-linked heritable disease and is the most common cause of inherited mental retardation, with an estimated prevalence of 1 in 1250 males (1, 2). Two important clinical features, mental retardation and macroorchidism, indicate that brain and testis are affected tissues. The recent isolation of the fragile X gene *FMR1* (3) revealed a CGG repeat at the 5′ end of the gene. Variation in the length of this repeat comprises the majority of mutations leading to fragile X syndrome. Methylation of a CpG island (4, 5, 6) that is located 250 bp upstream of the CGG repeat correlates with loss of expression of the *FMR1* gene in patients (7). In the normal population the CGG repeat is polymorphic and varies in length from 6–50 repeats (8). In phenotypically normal transmitting males the extended repeat number ranges from 52 up to 200 (premutation). In most fragile X patients the CGG repeat is significantly increased in length to more than 200 repeats (full mutation) (6, 8, 9, 10). In addition to the full mutation, a premutation is found in about 20% of male patients (7). In most of the cells in these mosaic patients a full mutation is found and in these cells there is no mRNA or protein production. The premutation alleles in the other cells are transcribed, resulting in protein. Overall, there is an apparent insufficiency of protein production in the appropriate tissues resulting in the abnormal phenotype.

A few patients lacking expansion of the CGG repeat have been described with the fragile X phenotype. In one patient the entire *FMR1* gene was deleted, as well as approximately 2,5 megabases of flanking sequence (11). In a second patient at least five exons of the *FMR1* gene were missing (12). Recently, transcription of the *FMR1* gene has been described in a patient with the fragile X phenotype without CGG expansion or cytogenetic expression. In this patient the phenotype was the result of a single point mutation (13). An A → T mutation was found in the open reading frame (ORF) resulting in an Ile → Asn substitution in the *FMR1* gene of this patient. In order to facilitate screening for *FMR1* mutations in patients with the fragile X phenotype without CGG expansion, RT-PCR studies of transcription of the normal *FMR1* gene in different tissues were performed. This revealed several alternatively spliced products of the gene that could give rise to as many as 12 different mRNAs and to 12 possible proteins.

## RESULTS

### Transcription of the *FMR1* gene

Transcription of the *FMR1* gene was studied by RT-PCR. Total RNA was isolated from HEPG2 cells and control EBV transformed human lymphocytes. After first strand cDNA synthesis, PCR was performed with 4 partially overlapping primer sets, K1–K2, K3–K4, K5–K6 and K7–K8 (for position of the primers see Material and Methods and fig. 3). PCR reactions with primer sets K1–K2 and K3–K4 each showed one discrete band on ethidium bromide stained agarose gels. The size of the PCR products was in agreement with the size expected from the cDNA sequence (3) (Fig. 1, lanes 1 and 2). However, PCR with primer set K5–K6 showed 2 bands (Fig. 1, lanes 3), whereas 4 bands were visible with primer set K7–K8 (Fig. 1, lanes 4), indicating the possibility of several alternative splice sites in the *FMR1* gene.

To test whether the alternative splicing would be tissue specific, RT-PCR products of various tissues were analyzed, including testis and brain as these two tissues demonstrate the most interesting clinical features (mental retardation and macroorchidism) of the fragile X syndrome. Using primer set K5–K6 2 bands were seen in all tissues analyzed: adult human testis and EBV transformed lymphocytes, fetal pancreas, liver and brain and adult mouse brain (Fig. 2a). Sequencing both bands

* To whom correspondence should be addressed

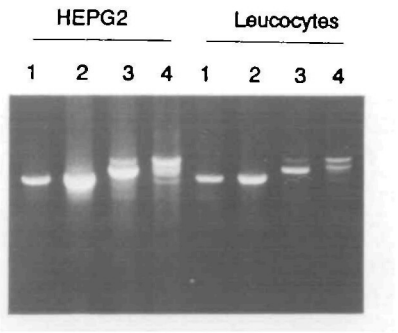


**Figure 1.** RT-PCR products of HEPG2 cells and EBV transformed lymphocytes using 4 overlapping primer sets covering the ORF of the *FMR1* gene, were separated on 3% ethidium bromide stained agarose gels. Lane 1 to 4 show RT-PCR products with primer sets K1−K2, K3−K4, K5−K6 and K7−K8, respectively.

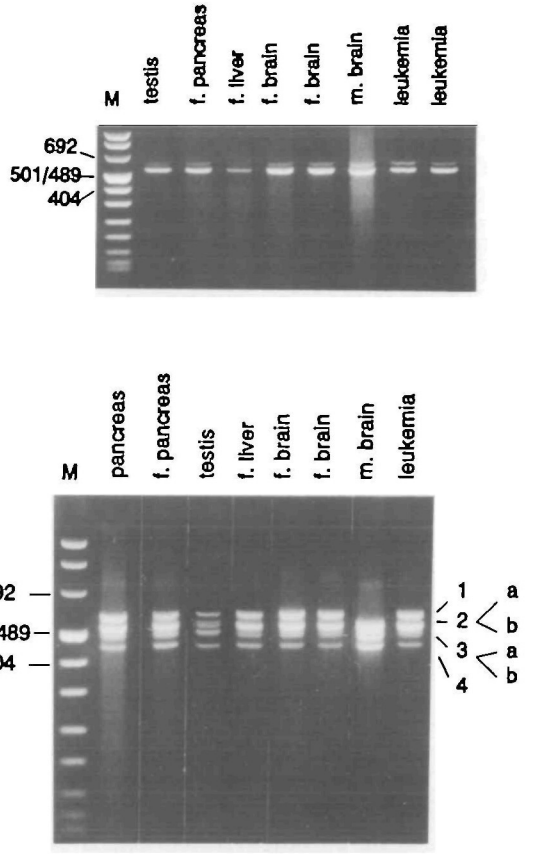


**Figure 2.** RT-PCR products from different tissues, using primer set K5−K6 (**a**) and primerset K7−K8 (**b**); f=fetal tissue, m=mouse tissue. Fetal brain was from a 16 and a 17 week old fetus, respectively. In Figure 2b the bands contain: band 1: +75 bp, +51 bp = C and D in Fig. 3; band 2a: +39 bp, +51 bp = C (last 39 bp) and D in Fig. 3; band 2b: +75 bp = C in fig. 3, −51 bp; band 3a: −75 bp, +51 = D in Fig. 3; band 3b: +39 bp = C (last 39 bp) in Fig. 3, −51 bp; band 4: −75 bp, −51 bp. The original published BC22 cDNA clone corresponds to band 2b.

revealed an additional 63 basepairs between positions 1126−1188 in the upper RT-PCR product (Fig. 2a: upperband and Fig. 3: B). The originally published human cDNA did not contain this 63 bp sequence and corresponds to the lower band. In mouse this 63 bp sequence is found as well (Warren, Nelson, Bakker, unpublished results). The ratios of the upper to the lower band were about 1:4 in all tissues analyzed (Fig. 2a).

Sequencing the K5−K6 PCR products also revealed 6 basepairs that were not present in the published *FMR1* cDNA (3). These 6 bases, GCAAAG, are positioned between base 881−886 (Fig. 3: A) and were consistently present in all mRNA products and cDNA clones analyzed. The fact that all sequences analyzed contain the 6 bases makes it likely that the absence of this sequence in the original BC22 cDNA clone could have been due to a splicing artefact. Analysis of these tissues with primer set K7−K8 showed 4 bands in all tissues tested, except in testis where a fifth band was found (Fig. 2b). The upper band (band 1) was missing in mouse brain. Sequencing band 1 disclosed 51 additional bases compared to the published cDNA sequence, positioned between bases 1738−1788 (Fig. 3: D). These 51 basepairs were also found in band 3. Sequence data also showed that band 3 and 4 (Fig. 2b) were missing 75 bases positioned from base 1472−1546 (Fig. 3: C), as compared to the published cDNA sequence. Direct sequencing of the RT-PCR product in band 2 from fetal brain and EBV transformed lymphocytes showed overlapping sequences, indicating more than one fragment in this band. We presumed that as in testis material, this band consisted of 2 different PCR products (band 2 → a and b) that were not separated due to relatively high amounts of these 2 products with a minor size difference. To investigate this further primer A3 (see Fig. 3 and Material and Methods) was used in combination with K7. RT-PCR with this primer set excluded the 51 bases. Only the area around the 75 basepairs was amplified and this gave 3 well separated bands in testis, fetal brain (Fig. 4), EBV transformed lymphocytes and mouse (data not shown). Sequencing the upper product identified a sequence of 242 bases in complete alignment with the published cDNA sequence. The lower band of 167 bp was missing the above mentioned 75 bp. This is consistent with data found with primer set K7−K8. However, the middle band contained only part (39 bases present) of these 75 bases, with 36 bp missing from the 5′ end, indicating another splice site situated within the 75 basepairs (Fig. 3: C). Using primers K7 and K8, a sixth band could be visualised after extended running time (Fig. 2b: band 3b). According to the size of this product, this band contains the 39 bases without the 51 bases. When using primers that select for the presence or absence of the 51 basepairs, in mouse we only found bands indicating the absence of this sequence. In human tissues there is a 1:1 ratio for the presence or absence of splice site D (data not shown). In table I the sequences that are present in the alternatively spliced regions C and D found with primers K7 and K8 as well as their relative abundance are summerized.

In figure 3 the above mentioned differential sequences are visualised as grey areas in the original cDNA. The individual differential sequences B,C and D were found to be present in different cDNA clones. In all cases the open reading frame is maintained until the original stop codon.

## Additional sequence from fetal brain and liver FMR1 cDNA clones

New *FMR1* clones were isolated from human fetal brain and liver cDNA libraries to extend the sequence of the original published *FMR1* cDNA at the 5′ end. Several clones that were longer at the 5′ end, contained the same 24 basepairs at the 5′ end preceding the original sequence (underlined in Figure 3 at the

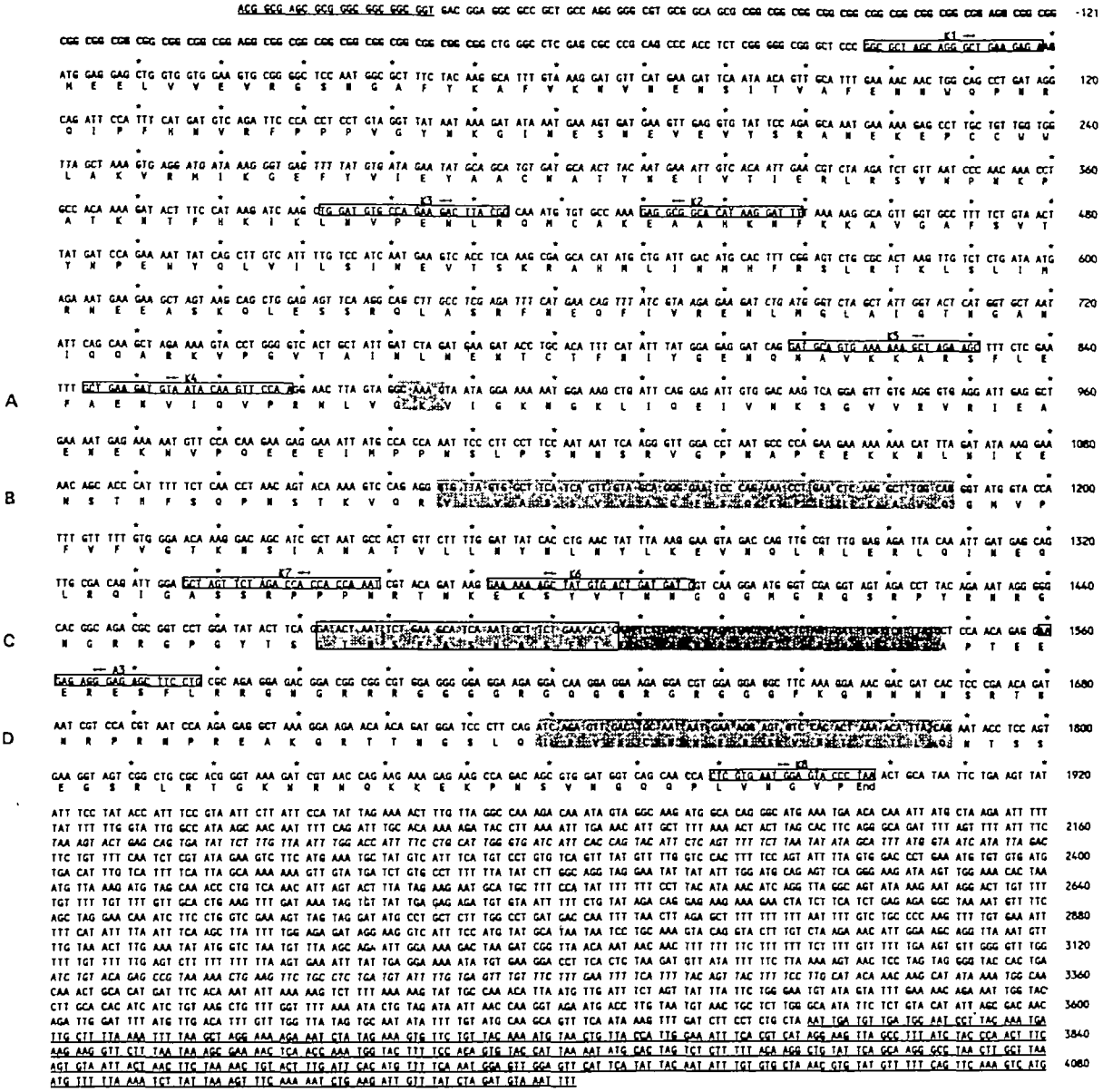

**Figure 3.** *FMR1* cDNA sequence. Additional (exon) sequences are integrated in the sequence of the original BC22/BC72 cDNA clones (3). Sequences that undergo alternative splicing are indicated in grey and are indicated A, B, C and D in the left margin. Additional sequences (underlined) have been added at the 5′ and 3′ end of the *FMR1* cDNA. The location of the primers is given and the sequences are contained in boxes. The sequence has been renumbered as compared to the prior published sequence (3) and the numbering starts at the first ATG 3′ from the CGG repeat.

5′ end) which is in agreement with the genomic sequence published by Fu et al (8). In front of this (underlined) shared sequence, all the clones show a different sequences.

The original BC72/22 clone was isolated from fetal brain and contains a polyadenylation signal and the first A of a poly A tail. Several clones isolated from a liver cDNA library show the same 3′ end, but a number extended beyond the published sequence at the 3′ end (underlined in Figure 3 at the 3′ end). The original 3′ sequence and the extended sequence are adjacent in the genomic sequence.

## Intron−exon boundaries

To determine the nature of the alternative splicing, the relevant splice boundaries were analyzed. Sequences of the intron−exon

**Table 1.** Sequences present in the alternatively spliced regions C and D found with primers K7/K8 and the relative abundance of the different transcripts

| Product | Bases present at C | Bases present at D | Total bases present | Relative abundance | | |
|---|---|---|---|---|---|---|
| | | | | Human somatic | Human testis | Mouse brain |
| 1 | +75 | +51 | 126 | +++ | + | 0 |
| 2a | +39 | +51 | 90 | ++ | + | 0 |
| 2b | +75 | 0 | 75 | +++ | + | ++++ |
| 3a | 0 | +51 | 51 | ++ | + | 0 |
| 3b | +39 | 0 | 39 | + | +/− | ++ |
| 4 | 0 | 0 | 0 | + | + | ++ |

The relative abundance of the different transcripts was determined by comparing the ratios of the different PCR products obtained with different primersets in the alternatively spliced regions C and D.

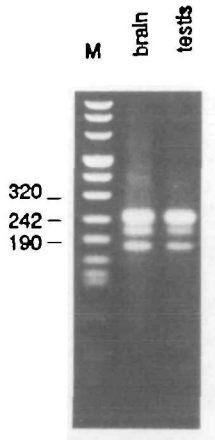


**Figure 4.** RT-PCR products from fetal brain and testis, using primers K7 and A3. For EBV transformed lymphocytes and mouse brain the same bands were found in the same ratios (data not shown).

| | 5′ splice site | 3′ splice site |
|---|---|---|
| A | TTAGTAGgtaagtgcaga | atttcttagGCAAAG |
| | | ttagcaaagTAATAG |
| B | CCAGAGGgtaagaattac | actgcttagGTGTTA |
| | TTGGCAGgtaggaaaaca | ctttttcagGGTATG |
| C | ACTTCAGgtacanctaag | ccncgatagGAACTA |
| | | ctgaaacagAATCTG |
| | | ggtcattagCTCCAA |
| D | CCTTCAG..... | ttgttttagATCAGA |
| | | acattacagAATACC |
| consensus | AGgt | agG |

**Figure 5.** Intron/exon boundaries of the alternative splice sites. A, B, C and D give the intron/exon sequences for the 6, 63, 75, and 51 basepairs, respectively. Capitals indicate exon sequences, lower case letters (not underlined) indicate intron sequences. Underlined lower case letters indicate exon sequences that are (alternatively) spliced out of the mRNA.

boundaries of the different splice sites were determined by comparing genomic and cDNA sequences (Fig. 5). The sequence of 63 bases (B in Fig. 5) is a complete exon that has intron sequences on either side in genomic DNA. Intron sequences are also found adjacent to the other 5′ splice sites C and D and to site A (Fig. 5). However, at the 3′ splice boundaries of these sites, splicing occurs inside exon sequences (Fig. 5, C, D and A). Splice junction sites in higher eukaryotes are characterized by a short but well conserved consensus sequence. By comparing the actual splice sites in a number of known gene sequences, Shapiro and Senapathy (14) have found that splice sites can be identified with a great degree of accuracy in different species. The most frequently occurring consensus sequence found in primate mRNAs at the 5′ junction site (CAG/gta) is present in all above described alternative splice sites in *FMR1*, except in the first 5′ junction site of splice site B (Fig. 5) where AGG/gta is found. All introns preceding A, B, C and D start with gt, which is found at 5′ junctions of all introns studied in different

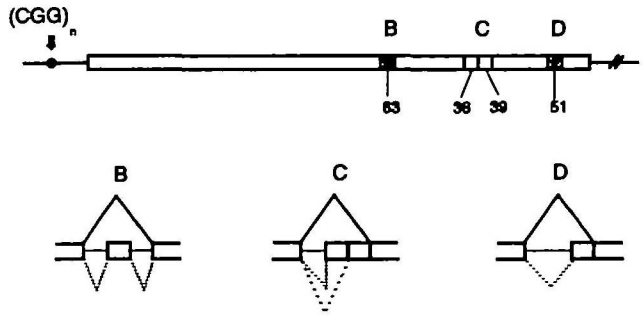


**Figure 6.** Schematic representation of the different splice sites B, C and D (not drawn to scale). In the upper part of the figure the *FMR1* cDNA is represented containing the different regions B, C and D that undergo splicing. In the lower part of the figure the different splice sites are depicted.

organisms, including primates, plants and invertebrae. For splice site D the 5′ intron sequence was not determined. All 3′ ends of the introns of the *FMR1* splice sites of A, B, C and D conform to the ag consensus sequence (tccag/G) (14).

It is striking that 3 of the 3′ acceptor sites (Fig. 5 and 6, C and D) are in the middle of exon sequences. The AG consensus sequence is found to directly precede the 3′ acceptor site, however here it is found in exon sequences. The 6 bases at site A in Figure 6 are present in all tested cDNA clones, with the exception of the originally described BC22 (3). Therefore the absence of this short sequence in BC22 is probably a splicing artefact. The splicing in BC22 has occurred at an AG in exon sequence that is downstream from the ag of the intron. In this case the original splice site has been bypassed, which is very unusual: when 2 ag sequences are within 10 bases of each other, splicing normally occurs at the upstream ag (15).

Splicing at the alternative sites (Fig. 5, B, C and D) occurs at different ratios, but no differences were detected between the various tissues analyzed.

## DISCUSSION

Our goal was to pursue qualitative transcription information in various normal tissues and to examine if *FMR1* would be differentially expressed in various tissues, especially with regard to those tissues, brain and testis, that are affected in fragile X patients. Northern blot analysis indicated high expression of *FMR1* in brain, testis, placenta and lymphocytes, whereas in other organs a lower level of expression is seen (3). As RT-PCR is a very sensitive and rapid method to study mRNA transcripts and is also the method of choice to detect rare mRNAs or mRNAs in small amount of tissues or cells, we used RT-PCR to study normal transcription of *FMR1* in different tissues.

We focused on the open reading frame of *FMR1* as alternative transcripts might encode isoproteins responsible for functional diversity of *FMR1*. No alternative splicing was found in the first half of the *FMR1* gene. The CGG repeat is on the same (first) exon as the start ATG and cannot be spliced out without removing the start ATG (8). Alternative splicing occurs in the *FMR1* gene but does not appear to be tissue specific as similar ratios of transcripts were found in all tissues analyzed. In these tissues, using primers K7 and K8, PCR products 1 and 2 are considerably more abundant than 3 and 4. In testis we find slightly different ratios of transcripts using primers K7 and K8. Although RT-PCR

is not a quantitative method, in one reaction the same primers are used resulting in different alternative spliced mRNA products in different ratios.

The mouse *FMR1* sequence has a high homology to the human sequence. Using the primers K5/K6 the same alternative splicing is found in the mouse and human *FMR1* gene with the upperband containing the 63 basepairs in both species. But, when primers K7/K8 are used, a difference is observed. In adult mouse brain using these primers, the largest transcript is missing. In mouse brain no transcripts are found containing the 51 basepairs. In human tissues there is a 1:1 ratio for the presence or absence of splice site D (data not shown). Taking into account the relative abuncance of the different transcripts in human somatic tissues (Table 1), it seems that the majority of *FMR1* RNA products use the upstream 3′ splice site at C (no bases excluded). The upstream and downstream 3′ splice sites at D are used equally efficient in human. In mouse only the downstram 3′ splice site at D is used. The product ratios at splice site B in favour of the absence of B is not influenced by the size of the preceding intron (which is larger than 500 bp), as there is a minimum size of about 80 base pairs for efficient splicing (15).

The likelihood that all the transcripts found are exon sequences and do not represent incompletely spliced mRNA products, is strengthened by the maintenance of an open reading frame and the apparent prevalence of different splice products in cDNA libraries. Two types of alternative splicing are found in *FMR1*. The 63 basepair sequence is a complete exon located between two introns and spliced into or out of the mRNA in a conventional way. Alternative splicing at sites A, C and D involves use of alternative acceptors. 3′ splicing at sites A, C and D occurs inside exon sequences with AG conserved and preceding all 3′ acceptor sites. As the absence of the 75 bases and the absence of 36 bases are not independent events, 3 combinations of different transcripts are possible for C (Fig. 6). Combining these 3 possible transcripts with the splice sites B and D, 12 distinct mRNA products could result from *FMR1* expression, each of which could be translated into protein. Western blot analysis using antibodies directed to a portion of the *FMR1* protein has shown the presence of 4 discrete proteins of different mobilities in lymphoblastoid cell lines. These proteins are absent in cell lines of fragile X patients (C. Verhey et al, submitted). These 4 products could represent more than 4 proteins with overlapping lengths (the largest possible product being 631 amino acids and the smallest being 568 amino acids). Expression of different cDNA constructs in COS cells showed distinct single proteins with different lengths excluding extensive post-translational modification events. Whether all these proteins will have a different function has to be determined.

The finding of 12 possible different mRNAs in *FMR1* does not necessarily indicate that 12 isoforms of the protein are produced. However, the presence of at least 4 proteins has been confirmed by using *FMR1* specific antibodies.

New *FMR1* cDNA clones have been isolated from human fetal brain and liver cDNA libraries. The cDNA clones that were extended at the 5′ end as compared to the sequence of the original published *FMR1* cDNA showed a different sequence in front of the underlined shared sequence (Figure 3). A possible explanation for this phenomenon could be the presence of unstable secondary structures through which the reverse transcriptase is not able to proceed correctly. At the 3′ end the original BC72/22 cDNA clone contains a polyadenylation signal and the first A of a poly A tail (followed by vector sequence CGG (3)). This was confirmed by sequencing cDNA clones that did have longer poly A tails. Several liver cDNA clones are extended at the 3′ end beyond the originally published sequence. This suggests alternative use of different polyadenylation signals, although no new polyadenylation signal was found at the 3′ end of these liver clones, which are probably incomplete. Mouse cDNA clones that extend even further than the human clones at the 3′ end are found as well (Warren, pers. comm.). These mouse clones do contain a polyadenylation signal and are at the 3′ end very homologous to the human sequence. Combining all the sequence data from the different cDNA clones, the longest *FMR1* cDNA sequence consists of 4362 basepairs.

Alternative splicing is found in many genes and can introduce functional diversity into the products of a single gene. In most cases this gives rise to protein isoforms sharing extensive regions of identity and varying only in specific domains, thus allowing for the fine regulation of protein function. The functional significance of alternative splicing is described for several genes. 1) The function of differential splice products can be tissue specific, e.g.: transcripts derived from the *CALC1* gene are differentially processed in a tissue specific manner to include exon 4 in the thyroidal C-cells (encoding the precursor to the hormone calcitonin) and to exclude this exon in neuronal cells (encoding the precursor of the neuropeptide CGRP) (16,17). 2) Differential splicing products can have different functions in separate developmental stages, e.g.: mRNA transcribed from the transformer (*tra*) gene of Drosophila melanogaster is subjected to sex-specific alternative splicing during embryogenesis (18,19). 3) Alternative splicing can create an altered coding frame, for instance in myotonic dystrophy. The 3′ ends of the ORFs of brain and heart transcripts of the DM-kinase gene differ in length and codon usage (20). Consequently, different hydrophobic C termini are predicted for the putative kinases. Differential splicing could have a regulatory function by creating differential anchoring of the kinase to different cellular structures. 4) By means of alternative splicing domains with suggested functions (glycosylation or phosphorylation sites etc.) can be added to transcripts. Also in Wilms' tumor alternative splicing in the *WT1* gene results in protein isoforms with differing binding affinities to the EGR-1 consensus sequence (21,22). For the *FMR1* protein no homology is found to any other known protein, so no functions can be assigned to different domains. Which function can be attributed to the alternative splicing in the *FMR1* gene will be subject to further research. At least splicing in this gene does not seem to be tissue specific. However, *FMR1* like DM shows alternative splicing confined to the 3′ half of the gene.

In conclusion, we have detected 12 possible *FMR1* transcripts that could result in 12 possible isoproteins. The next challenge will be to decipher the functions of these proteins and to determine whether the different isoforms will have different functions in different tissues.

## MATERIAL AND METHODS

### FMR1 liver cDNA isolation and characterization

A human liver cDNA library utilizing the vector lambda gt11 was constructed as described by Konecki et al. (23). Approximately 3 million recombinants were screened by filter hybridization using $^{32}$P-labelled Bgl I/BamH I fragment (base 430–1730) from the FMR-1 isolated from a fetal human brain library (Stratagene). Hybridization-positive recombinants were plaque-purified by three rounds of purification. Phage DNA inserts were isolated, purified and subcloned into pBlueScript II-SK (−) vector for DNA sequence analysis. Sequencing was performed by the double-stranded dideoxy chain termination technique using the Pharmacia T7 sequencing kit.

### RNA isolation

RNA was isolated according to Auffray and Rougeon (24). The LiCl method was used (procedure C) with several modifications. After overnight incubation at 0°C in 3M LiCl/6M urea the samples were spun down at 25K at 4°C for 20 min. Proteinase K treatment (10 μg/ml) was performed (30 min at 37°C) prior to phenol/chloroform extraction.

### First strand cDNA synthesis

5 μg of total RNA was reversed transcribed as described by Pieretti et al (7) with minor modifications. Instead of precipitating the cDNA, 2 μl of the reverse transcribed reaction was directly used for PCR.

### PCR

PCR reactions were done on 2 μl of cDNA solution with different primer sets. Amplification was carried out in a total volume of 50 μl containing 10 mM Tris-HCl (pH 8.3), 50 mM KCL, 1.5 mM $MgCl_2$, 0,25 mM of each dNTP and 0,5 units of Taq polymerase. The reaction was heated to 94°C for 5 min, followed by 28 cycles of DNA denaturation (1 min at 94°), annealing (1 min at 55°C) and extension (1 min and 45 sec. at 72°C). Final extension was at 72°C for 10 min. 5−10 μl of PCR product was analyzed on a 1 or 3% agarose gel (1.5% regular and 1.5% nusieve agarose) depending on band sizes, stained with ethidium bromide.
The following primers were used:
Primer K1: 5′ GGCGCTAGCAGGGCTGAAGAGA 3′
Primer K2: 5′ AAAATCCTTATGTGCCGCCTCTTT 3′
K1 and K2 were derived from positions −24 to −3 and 430−449 (Fig. 3)
Primer K3: 5′ TTGAACTTGTATTACATCTTCAGC 3′
Primer K4: 5′ TTGGAACTTGTATTACATCTTCAGC 3′
K3 and K4 were derived from positions 392−414 and 844−868.
Primer K5: 5′ GATGCAGTCAAAAAAGCTAGAAGC 3′
Primer K6: 5′ CATCATCAGTCACATAGCTTTTTTC 3′
K5 and K6 were derived from positions 808−831 and 1372−1396.
Primer K7: 5′ GCTAGTTCTAGACCACCACCAAAT 3′
Primer K8: 5′ CTTAGGGTACTCCATTCAGGAG 3′
K7 and K8 were derived from positions 1336−1359 and 1878−1898.
Primer A3: 5′ CAGGAAGCTCTCCCTCTCTT 3′ is derived from position 1559−1578.

### Sequencing

To sequence the PCR products bands were excised from the agarose gel and equilibrated in 1 ml of 300 mM NaAc pH 6.5 + 1 mM EDTA for 15 min at room temperature. The agarose containing the DNA was then cut in little pieces and transferred to 0.5 ml eppendorf tube that was punctured at the bottom and on top of some glasswool. The tube with agarose was kept at −80°C for 30 min. and spun down in a 1.5 ml tube for 10 min at room temperature at 13000 rpm. To the supernatant 1/100 vol 1M $MgCl_2$ + 10% HAc was added and the DNA was precipitated with 2 volumes of ethanol at −80°C for 30 min. The DNA was spun down, washed with 70% ethanol and dissolved in 50μl of Tris-HCl (10 mM)/EDTA (0,1mM). Direct sequencing of approximately 30 ng of RT-PCR-product was performed using the BRL cycle sequencing kit.

## ACKNOWLEDGEMENTS

We thank Jo Höppener, Annelies de Klein, Marianne Kroos and Marcel Koken for providing mRNA of different tissues, and Richard A.Gibbs and the DNA sequence core of the BCM Human Genome Center for sequence data, and M.F.Victoria and A.McAughan for technical assistance. This research was supported by the Foundation of Clinical Genetics to A.J.M.H. Verkerk en E.de Graaff, a grant of the MGCZWN to B.A.Oostra, and NIH grant NICHHDIROIHD29256 and NIH Human Genome Center grant P30-HG00210 to D.L.Nelson.

## REFERENCES

1. Gustavson, K.H., Blomquist, H. and Holmgren G. (1988) Am. J. Med. Genet., 23, 58−588.
2. Webb, T.P., Bundey, SE., Thake, A.I. and Todd, J. (1986) Am. J. Med. Genet. 23, 573−580.
3. Verkerk, A.J.M.H., Pieretti, M., Sutcliffe, J.S., Fu, Y-H., Kuhl, D.P.A., Pizzuti, A., Reiner, O., Richards, S., Fuping Zhang, M.F.V., Eussen, B.E., van Ommen, G.J.B., Blonden, L.A.J., Riggins, G.J., Chastain, J.L., Kunst, C.B., Galjaard, H., Caskey, C.T., Nelson, D.L., Oostra, B.A. and Waren, S.T. (1991) Cell, 65, 905−914.
4. Vincent, A., Heitz, D., Petit, C., Kretz, C., Oberle, I. and Mandel, J.L. (1991) Nature 349, 624−626.
5. 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., Kerr, B., Turner, G., Lindenbaum, R.H., Winter, R., Pembrey, M., Thibodeau, S. and Davies, K.E. (1991) Cell 64, 861−866.
6. Oberle, I., Rousseau, F., Heitz, C., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M.F. and Mandel J.L. (1991) Science, 252, 1097−1102.
7. Pieretti, M., Zhang, F., Fu, Y-H., Warren, S.T., Oostra, B.A., Caskey, C.T. and Nelson, D.L. (1991) Cell, 66, 1−20.
8. Fu, Y-H., Kuhl, D.P.A., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J.M.H., Holden, J.J.A., Fenwick, Jr., R.G., Warren, S.T., Oostra, B.A., Nelson, D.L. and Caskey, C.T. (1991) Cell, 67, 1047−1058.
9. Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, I.C., Warren, S.T., Schlessinger, D., Sutherland, G.R. and Richards, R.I. (1991) Science, 252, 1179−1181.
10. Kremer, E., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S.T., Schlessinger, D., Sutherland, G.R., Richards, R.I. (1991) Science 252, 1711−1714.
11. 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. (1991) Nature Genetics 1, 341−344.
12. Wöhrle, D., Kotzot, D., Hirst, M.C., Manca, A., Korn, B., Schmidt, A., Barbi, G., Rott, H.D., Poustka, A., Davies, K.E. and Steinbach, P. (1992) Am. J. Hum. Genet. 51, 299−306.
13. De Boulle, K., Verkerk, A.J.M.H., Reyniers, E., Vits, L., Hendrickx, J., Van Roy, B., van den Bos, F., de Graaff, E., Oostra, B.A. and Willems, P.J. (1993) Nature Genetics, in press.
14. Shapiro, M.B. and Senapathy, P. (1987) Nucleic Acids Res. 15, 7155−7174.
15. Wieringa, B., Hofer, E. and Weissman, C. (1984) Cell 37, 915−925.
16. Amara, S.G., Jonas, V., Rosenfeld, M.G., Ong, E.S. and Evans R.M. (1982) Nature 298, 240−244.
17. Cote, G.J., Stolow, D.T., Peleg, S., Berget, S.M. and Gagel, R.F. (1992) Nucleic Acids Res. 20, 2361−2366.
18. Baker, B.S. (1989) Nature 340, 521−524.
19. O'Neil, M.T. and Belote, J.M. (1992) Genetics 131, 113−128.
20. Jansen, G., Mahadevan, M., Amemiya, C., Wormskamp, N., Segers, B., Hendriks, W., O'Hoy, K., Baird, S., Sabourin, L., Lennon, G., Jap, P.L., Iles, D., Coerwinkel, M., Hofker, M., Carrano, A.V., de Jong, P.J., Korneluk, R.G. and Wieringa, B. (1992) Nature Genetics 1, 261−266.
21. Brenner, B., Wildhardt, G., Schneider, S. and Royer-Pokora, B. (1992) Oncogene, 7, 1431−1433.
22. Bickmore, W.A., Oghene, K., Little, M.H., Seawright, A., van Heyningen, V. and Hastie, N.D. (1992) Science 257, 235−237.
23. Konecki, D.S., Wang, Y., Trefz, F.K., Lichter−Konecki, U. and Woo, S.L.C. (1992) Biochemistry 31, 8363−8368.
24. Auffray, C. and Rougeon, F. (1980) Eur. J. Biochem. 107, 303−314.