33. Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162**, 156-159 (1987).
34. Gubler, V. & Hoffman, B. J. *Gene* **25**, 263-269 (1983).
35. Sive, H. L. & St John, T. *Nucleic Acids Res.* **16**, 10937 (1988).
36. Short, J. M., Fernandez, J. M., Sorge, J. A. & Huse, W. D. *Nucleic Acids Res.* **16**, 7583-7600 (1988).
37. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbour Laboratory Press, N.Y., 1989).
38. Uhl, G. R. *In Situ Hybridization in Brain* (Plenum, New York and London, 1986).

Characterization and localization of the *FMR-1* gene product associated with fragile X syndrome

Coleta Verheij, Cathy E. Bakker, Esther de Graaff,
Joke Keulemans, Rob Willemsen,
Annemieke J. M. H. Verkerk,
Hans Galjaard, Arnold J. J. Reuser,
André T. Hoogeveen & Ben A. Oostra*

MGC Department of Clinical Genetics, Erasmus University, PO Box 1738, 3000 DR Rotterdam, The Netherlands

THE fragile X syndrome is the most frequent form of inherited mental retardation after Down's syndrome, having an incidence of one in 1,250 males[1,2]. The fragile X syndrome results from amplification of the CGG repeat found in the *FMR-1* gene [3–6]. This CGG repeat shows length variation in normal individuals and is increased significantly in both carriers and patients[3–6]; it is located 250 base pairs distal to a CpG island[6] which is hypermethylated in fragile X patients[4–7]. The methylation probably results in downregulation of *FMR-1* gene expression[8]. No information can be deduced about the function of the *FMR-1* protein from its predicted sequence. Here we investigate the nature and function of the protein encoded by the *FMR-1* gene using polyclonal antibodies raised against the predicted amino-acid sequences. Four different protein products, possibly resulting from alternative splicing, have been identified by immunoblotting in lymphoblastoid cell lines of healthy individuals. All these proteins were missing in cell lines from patients not expressing *FMR-1* messenger RNA. The intracellular localization of the *FMR-1* gene products was investigated by transient expression in COS-1 cells and found to be cytoplasmic. Localization was also predominantly cytoplasmic in the epithelium of the oesophagus, but in some cells was obviously nuclear.

As a first step in the identification and characterization of the *FMR-1* gene product, antibodies were raised against different regions of the predicted amino-acid sequence of the *FMR-1* protein[3]. Two different methods were used. A complementary DNA fragment of *FMR-1* containing nucleotides 940–1,325 was cloned in the *Escherichia coli* expression vector pGEX[9] and antibodies were raised in rabbits against the *FMR-1* fusion protein (α765). The second approach was to use a synthetic oligopeptide corresponding to the carboxy-terminal end (position 632 to 656) of the *FMR-1* protein[3] as antigen (α1079)[10].

These antibodies were then used to analyse the *FMR-1* protein (FMRP) in lymphoblastoid cell lines from patients (*n*=5) and controls (*n*=3). FMRP was immunoprecipitated with α765 and analysed by immunoblotting. Four species (M_rs 74K, 72K, 70K and 67K) (Fig. 1*A*) that were present in the controls were absent in four of five patients. The lack of cross-reactive
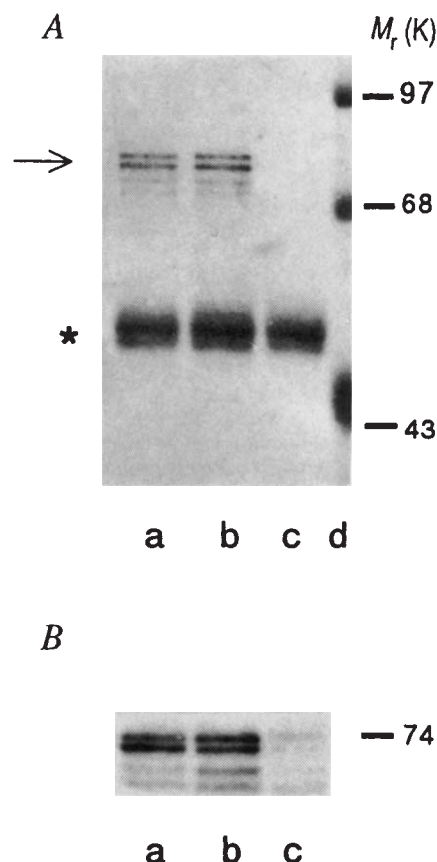
*To whom correspondence should be addressed.

FIG. 1 Immunoprecipitation of lymphoblastoid cell lines with α765 antibodies. *A*, Control cells: lanes a (C49) and b (ROS31) and fragile X patient cells in lane c (RJK1411). Sizes of the protein bands are 74, 72, 70 and 67K. Lane d, molecular weight standards. Nonspecific band (IgG) is indicated by an asterisk, the position of FMRPs is indicated by an arrow. *B*, Control cells: lane a (ROS31) and b (GM7381), cells from a mosaic patient lane c (RJK2368). Molecular weight of upper band indicated on the left.

METHODS. Polyclonal antibodies were raised in rabbits against a GST (glutathione *S*-transferase)–*FMR-1* fusion protein obtained using the pGEX plasmid expression system[9]. A cDNA fragment containing nucleotides 940 to 1,325 of the coding and 3,557–3,765 of the non-coding sequence of the previously published *FMR-1* cDNA[3] was cloned in the *Eco*RI site of pGEX-3X. Expression of the construct in the protease-deficient *E. coli* strain B121 resulted in the synthesis of a *FMR-1* polypetide of 134 amino acids fused with the C terminus of Sj26, a 26K glutathione *S*-transferase. The *FMR-1* polypetide consisted of amino acids 314 to 442 of the *FMR-1* gene plus five additional amino acids (Cys-Thr-His-His-Leu) as a result of the cloning procedure. Expression and purification of the fusion protein were as described[9]. Antibodies were affinity-purified using columns of GST and (GST)-*FMR-1* fusion protein successively. Pellets of lymphoblastoid cell lines were homogenized in a buffer containing 10 mM HEPES, 300 mM KCl, 100 μM $CaCl_2$, 5mM $MgCl_2$, 0.05% Tween and 0.45% Triton X100, pH 7.4. Cell homogenates were spun down for 10 min (10,000*g*). The supernatants were incubated with protein A–Sepharose (Pharmacia) for 2 h at 4°C to remove the IgG present in the lymphoblastoid cells. Protein A–Sepharose was spun down and the supernatant was incubated overnight with α765 and protein A–Sepharose at 4° C. After washing (3×) the immunoprecipitate, sample buffer was added and the precipitates were electrophoresed in SDS–polyacrylamide (8% polyacrylamide gel, 1% crosslinking) and electroblotted. Polypeptides were visualized using 1,000× diluted α765 and a 1,000× diluted alkaline phosphatase-conjugated goat anti-rabbit IgG as a second antibody. Polypeptides were detected using naphthol AS-MX phosphate and 4-aminodiphenylamine diazonium sulphate or AMPPD (TROPIX) as substrates.
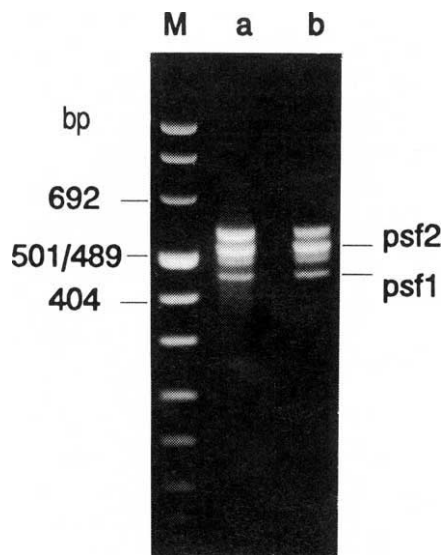
FIG. 2 Different splice products from the 3′ end of the *FMR-1* gene visualized by reverse transcriptase PCR amplification of cDNA. Total RNA was isolated from leukocytes (lane a) and fetal brain (lane b)[14]. Alternative splicing involved exons 15 and 17 of *FMR-1* (D. L. Nelson, personal communication). The lengths of the PCR products were 563, 512, 488 and 437 bp. Length of DNA marker (lane M; factor VIII; Boehringer) bands are indicated in base pairs. The different splice variants shown were found to be present in isolated cDNAs (data not shown).
METHODS. Total RNA isolation was according to ref. 14. The LiCl method was used (procedure C) with several modifications. After overnight incubation in 3 M LiCl/6 M urea, samples were spun down for 20 min at 25,000 r.p.m. at 4° C and treated with proteinase K (10 μg ml⁻¹) for 30 min at 37°C before extraction with phenol/chloroform. Five μg RNA was reverse-transcribed as described in ref. 8, except that instead of precipitating the cDNA, 2 μl was directly used for PCR. PCR was done on 2 μl cDNA solution with the primer set K7 and K8. Primer K7: 5′-GCTAGTTCTAGACCACCACCAAAT-3′ and primer K8: 5′-TTAGGGTACTCCATTCACGAG-3′ were derived from positions 1,462–1,485 and 1,954–1,974 of the published *FMR-1* cDNA sequence[3]. Amplification and analysis of PCR products was done as described[15].
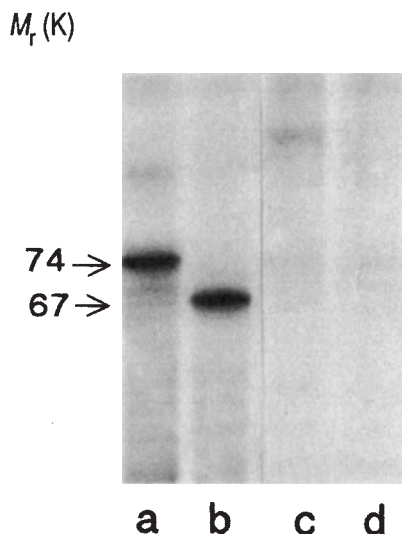

FIG. 3 Immunoprecipitation of COS-1 cells transfected with *FMR-1* (psf-2, lane a), or an alternative splice variant of *FMR-1* (psf-1, lane b). Immunoprecipitation of mock-transfected COS-1 cells or COS-1 cells transfected with *FMR-1* cloned in pSG5 (ref. 12) in the antisense direction are shown in lanes c and d, respectively. The molecular weight of the immunoprecipitated proteins is indicated on the left. COS-1 cells were labelled for 2 h with ³⁵S-methionine and ³H-leucine, 72 h after transfection. Cells were homogenized and used for immunoprecipitation with polyclonal antibody α765 raised against a fusion protein of glutathione S-transferase.
METHODS. *FMR-1* cDNA was cloned into the *Eco*RI site of the eukaryotic expression vector pSG5 (ref. 12) either as a 3,765-bp cDNA[3] (psf-2) or as a 3,690-bp *FMR-1* splice variant missing in the 3′ end of the 75-nucleotide coding sequence (1,598–1,672) but leaving the reading frame intact (psf-1). Transfections were achieved using the DEAE–dextran method[16].

material in the lymphoblastoid cell lines from these patients is in agreement with the absence of *FMR-1* mRNA in these cells (ref.8, and data not shown).

In the cell line of one patient, however, we found the same molecular species as in the controls (Fig. 1*B*). The patient has a mosaic DNA pattern, with one-third of the cells carrying a premutation and the rest a full mutation (data not shown). The CpG island preceding the CGG repeat is unmethylated in the premutation, in contrast to the full mutation. The premutation allele of the mosaic patients is expressed into *FMR-1* mRNA[8] (data not shown), and subsequently translated into FMRPs (Fig. 1*B*, lane c).

The finding of four protein products rather than one was unexpected, and the possibility was envisaged that this could be due to alternative splicing. Figure 2 shows that there were several *FMR-1* mRNA species in lymphoblastoid cells and fetal brain. The set of primers used in the polymerase chain reaction (PCR) were chosen from near the 3′ end of the open reading frame and amplified at least four splice products, and we have evidence for an even higher number of splice variants[11]. Several of these splice variants have been found in isolated cDNAs.

Two of the *FMR-1* cDNA clones representing the alternative splice variants psf-2 and psf-1 (Fig. 2) were cloned in expression vector pSG5 (ref. 12) and these constructs were transiently expressed in COS-1 cells. Both the larger and the smaller

construct seemed to encode stable proteins that were recognized by antibodies α765 (Fig. 3, lanes a and b) and α1079 (data not shown). The estimated M_rs were 74K and 67K for the proteins encoded by psf-2 and psf-1, respectively. No protein product was detected in mock-transfected COS-1 cells or in COS-1 cells transfected with *FMR-1* cDNA cloned in PSG5 in the antisense orientation (Fig. 3, lanes c and d).

The intracellular localization of the protein products of these different splice variants was investigated in COS-1 cells. It was shown by immunofluorescence microscopy using α765 antibodies that the protein encoded by psf-2 was to be found in the cytoplasm, despite the presence of a putative nuclear location signal (Fig. 4). The results were the same with the protein encoded by psf-1 and antibody α1079 raised against a synthetic peptide spanning the nuclear location signal. This observation was confirmed in mouse oesophagus, which is known to strongly express FMR-1 (ref. 13). FMRP was seen in the cytoplasm of the keratinocytes throughout the thick stratified squamous epithelial lining, with no staining of the keratin layer (Fig. 4*c*). Some cells showed an intense nuclear labelling.

Until now there has been uncertainty as to whether the CGG repeat is translated into protein. Our experiments (Fig. 1) on the expression of the *FMR-1* gene in lymphoblastoid cell lines show that the protein products of the controls and the mosaic patient are of the same size. Mosaic patients have a full mutation (>200

FIG. 4 Cytoplasmic localization of FMRP in transfected COS-1 cells using an indirect immunocytochemical method[17]. The α765 antibodies were visualized using goat anti-rabbit immunoglobulins conjugated with fluorescein. a, COS-1 cells transfected with psf-1. b, COS-1 cells transfected with psf-2. As a control, non-transfected COS-1 cells were used. c, Cryostat section (6 μm) of the oesophagus from an adult mouse incubated with polyclonal antibody α765, raised against fusion protein. Reaction product (dark grey staining) is seen in the cytoplasm of the cell, forming a thick stratified squamous epithelium, and not in the keratin layer. No product was seen using preimmune sera. Some cells show an intense labelling of the nucleus. E, stratified squamous epithelium; S, skeletal muscle; arrows indicate nuclei labelled for FMRP.

METHODS. For immunohistology cryostat sections were labelled with primary antibody and visualized with peroxidase-conjugated, swine–anti-rabbit immunoglobulins (DAKO, Denmark). Peroxidase was detected with 0.1% 3,3′-diaminobenzidine-HCl (Sigma) With 0.1% hydrogen peroxide. Sections were not counterstained.
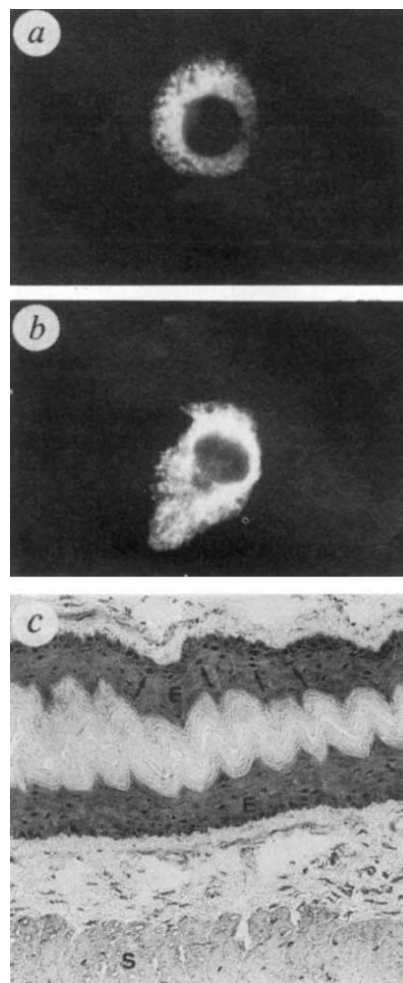
are of the same size. Mosaic patients have a full mutation (>200 repeats) in a proportion of their cells and a premutation (50–200 repeats) in others. The premutation allele is transcribed, unlike in the full mutation[8]. In this patient the premutation allele contains 65 CGG copies more than the average control. If the CGG repeat were translated into a stretch of arginines, then protein products with an increased molecular mass would have been expected in the mosaic patient. As this was not found (Fig. 1B), it can be concluded that the translation of FMR-1 starts distal to the CGG repeat. Expression of psf-1 and psf-2 in COS-1 cells showed that translation can start at an ATG initiation codon distal to the CGG repeat, as no ATG codon is present proximal to the CGG repeat in the open reading frame of the FMR-1 gene (S. T. Warren, personal communication). Considering these data together, it is unlikely that the CGG repeat is part of the coding sequence.

In fragile X syndrome the amplification of the CGG repeat blocks transcription of the FMR-1 gene[8] and this results in the absence of FMRP (Fig. 1) and causes mental retardation. The mosaic patient is mentally retarded despite the expression of FMR-1 proteins. It is therefore assumed that the proportion of cells expressing FMRP is insufficient to maintain normal function. Alternatively, it could be that the mosaic pattern found in lymphoblastoid cell lines is not representative of the situation either in other tissues (like the brain) or during the developmental stages in which the FMRP expression is essential.

With respect to the further elucidation of the function of the FMRPs, it is important that the protein is predominantly present in the cytoplasm. Our results will make it possible to isolate the FMRPs, characterize them further, and study their putative differential function. This could provide insight into the mechanism leading to the manifestations of the fragile X syndrome. □

1. Gustavson, H. K., Blomquist, H. & Holgrem, G. Am. J. med. Genet. 23, 581–588 (1986).
2. Webb, T. P., Bundey, S. E., Thake, A. I. & Todd, J. Am. J. med. Genet. 23, 573–580 (1986).
3. Verkerk, A. J. M. H. et al. Cell 65, 905–914 (1991).
4. Oberlé, I. et al. Science 252, 1179–1181 (1991).
5. Yu, S. et al. Science 252, 1179–1181 (1991).
6. Fu, Y.-H. et al. Cell 67, 1047–1058 (1991).
7. Heitz, D. et al. Science 251, 1236–1239 (1991).
8. Pieretti, M. et al. Cell 66, 817–822 (1991).
9. Smith, D. B. & Johnson, K. S. Gene 67, 31 40 (1988).
10. Merrifield, R. B. J. Am. chem. Soc. 85, 2149 2154 (1963).
11. Verkerk, A.J.M.H. et al. Hum. molec. Genet. 2, 399–404 (1993).
12. Green. S., Isseman, I. & Sheer. E. Nucleic Acids Res. 16. 369 (1988).
13. Hinds. H. L. et al. Nature Genet. 3, 36–43 (1993).
14. Auffray. C. & Rougeon, F. Eur. J. Biochem. 107, 303–314 (1980).
15. DeBoulle, K. et al. Nature Genet. 3, 31–35 (1993).
16. Oshima, A. et al. J. biol. Chem. 263, 2553 -2562 (1988).
17. Van Dongen, J. M., Barneveld, R. A., Geuze. H. J. & Galjaard. H. Histochem. J. 16, 941–954 (1984).