A candidate gene for mild mental handicap at the FRAXE fragile site.

Lisa Chakrabarti*, Samantha J. L. Knight*, Angela V. Flannery‡ and Kay E. Davies*

Genetics Laboratory, Department of Biochemistry, University of Oxford, South Parks Road, Headington, Oxford, OX1 3QU, UK

Received November 24, 1995; Revised and Accepted December 5, 1995

The cytogenetic expression of the folate sensitive fragile site, FRAXE, is due to the expansion of a GCC repeat in proximal Xq28 of the human X chromosome and is associated with a mild form of mental handicap. Normal individuals have 6–35 copies of the repeat whereas cytogenetically positive, developmentally delayed males have >200 copies and show methylation of the associated CpG island. Through the use of conserved sequences adjacent to the FRAXE GCC repeat, we have isolated a 1495 bp cDNA which begins 331 bp distal to the FRAXE site and extends to a region >170 kb distal in Xq28. The cDNA sequence possesses both a putative start of translation and a poly-A tail. The predicted protein has amino acid motifs which share significant homologies with the human AF-4 gene which encodes a putative transcription factor. On northern analysis, the cDNA detects a 9.5 kb transcript in human brain, placenta and lung. This transcript is present in multiple human brain tissues, but is more abundant in the hippocampus and the amygdala, thus providing possible functional insights. RT-PCR of normal adult brain RNA provides evidence for the existence of the 1495 bp transcript represented by the isolated cDNA.

INTRODUCTION

The cytogenetic expression of the folate sensitive fragile site, FRAXE, in proximal Xq28 is associated with a mild form of mental handicap (1–4). Using conventional cytogenetic techniques, FRAXE cannot be distinguished from either the FRAXA or FRAXF fragile sites which are also located in the Xq27.3-q28 region of the human X chromosome (4–7). The cytogenetic expression of FRAXA is associated with the fragile X syndrome, and has been shown to be due to the expansion of a CGG trinucleotide repeat (8–10). Normal individuals have between 2 and 50 copies of the CGG repeat whereas affected individuals have amplifications of >200 copies (8–16). The CGG amplification in patient DNA results in the methylation of the CpG residues and transcriptional silencing of the gene (17,18). Gene function may also occasionally be lost by point mutation or deletion events (19–21). Cloning of the FRAXE site has shown that the fragility is also due to the amplification of a GCC repeat and that the methylation of CpG residues is associated with a form of mental handicap which is generally milder than that of the fragile X syndrome (1–4). Affected FRAXE individuals have >200 copies of the GCC repeat compared to the 6–35 copies found in normal individuals (1,22). Phenotypically, FRAXE males show no consistent dysmorphology and the main clinical features appear to be learning difficulties, in particular, speech delay, reading and writing problems. There may also be behavioural deficits (1,2).

The difficulty in defining exact criteria by which the FRAXE phenotype can be defined has caused some debate concerning whether the cytogenetic expression and methylation observed at FRAXE is truly associated with a phenotype or whether the observed mental handicap in FRAXE individuals is a function of ascertainment bias. Although there are a number of studies reported in the literature where clinical similarities between mentally retarded FRAXE individuals support the existence of an associated phenotype, the low frequency of FRAXE families being identified and the existence of apparently normal individuals showing cytogenetic expression of FRAXE have made the ascertainment bias theory difficult to disprove (2–4, S.J.L. Knight et al., in preparation). To overcome this problem and to encourage genetic screening strategies for the diagnosis of FRAXE, it has become increasingly necessary to isolate candidate cDNAs in the FRAXE region and to identify and characterise the gene responsible for FRAXE mental handicap.

Recently, Gedeon et al. (1995) described overlapping submicroscopic deletions in Xq28 in two unrelated boys with developmental disorders and demonstrated that the probe VK21A (DXS296) was deleted in one of the boys (23). Northern analysis using the same probe revealed transcripts of 9.5 kb in placenta in brain and a smaller ~2.5 kb transcript in other tissues. Although VK21A is located ~175 kb distal to FRAXE, the transcripts identified nevertheless reflect the existence of one or more genes in the FRAXE candidate region. Gedeon et al. (23) briefly discussed preliminary results from the analysis of exons from a corresponding cDNA, but whether this cDNA represents the gene responsible for the developmental delay observed in FRAXE individuals remains unknown. Until this can be elucidated, the continued search for cDNAs within the FRAXE candidate region remains necessary.
The aim of the work presented here was to identify cDNAs in the FRAXE region and to isolate the gene responsible for FRAXE mental handicap. As sequence conservation analyses have contributed to the isolation of several important disease genes including those responsible for Duchenne muscular dystrophy, retinoblastoma, cystic fibrosis and the testis determining factor (24–27), we decided to adopt a similar strategy. Initial efforts focussed on the genomic region containing the FRAXE CpG island because it is known that CpG islands are often found at the 5′ ends of genes (28). We describe the isolation of a 1495 bp cDNA which commences 331 bp distal to the FRAXE GCC repeat and intriguingly, possesses a poly-A tail >170 kb further distal in Xq28. This cDNA is currently being investigated as a candidate gene for FRAXE mental handicap.

RESULTS

Identification of conserved sequences in the FRAXE region

Single copy probes from cosmids spanning the FRAXE region were used as hybridisation probes against Southern blots containing EcoRI digested DNA from a number of animal species (zooblots). Three probes adjacent to the FRAXE CpG island, OxE18, OxE19 and OxE20 were shown to be conserved in human, cow, pig and mouse. The relative map positions of these probes and an example of the zooblot hybridisation with OxE19 are shown in Figure 1a and b respectively. The isolation of OxE20 has been described previously (1). OxE18 is a subclone of the ∼2 kb EcoRI/NotI fragment immediately proximal to the FRAXE CpG island whereas OxE19 is a 435 bp PCR product selectively cloned because of high sequence conservation with a homologous mouse bacteriophage clone isolated previously (unpublished observations).

Isolation and mapping of cDNAs from a human fetal brain cDNA library

OxE18, OxE19 and OxE20 were hybridised independently to duplicate membrane lifts from a human fetal brain cDNA library (Clontech). In a screen of 1 × 10⁶ bacteriophage plaques, no positive clones were detected using OxE18 or OxE20. However, OxE19 detected six positive clones which remained positive through subsequent rounds of plaque purification. Restriction enzyme analysis revealed all of these to be identical. Each cDNA insert was ∼1.5 kb in size and possessed four EcoRI fragments (unpublished data). These fragments were subcloned into pBluescript KS(+) and the insert fragments (designated Ox19.1, Ox19.2, Ox19.3 and Ox19.4) used as hybridisation probes against blots with the human X chromosome on mouse background, somatic cell hybrid, ThyBX (29). They were also used against panels of DNA from cosmids and YACs which formed part of a contig spanning and extending distal to the FRAXE site (see Fig. 2a). The results showed that the Ox19 cDNA mapped to the X chromosome and was present in cosmids 102A which spans the FRAXE site, but was absent in the cosmids extending ∼100 kb distal (unpublished data). Independent hybridisation of the EcoRI subclones revealed that Ox19.2 mapped to the 2.7 kb EcoRI fragment of cos102A which contains the FRAXE CpG island. Interestingly, the remaining cDNA subclones (Ox19.1, Ox19.3 and Ox19.4) did not map to this cosmid, but gave positive hybridisation signals >100 kb distal in an overlapping region of ∼140 kb between YACs 2A3, H7 and D10 (Fig. 2a). This was deduced by comparing the hybridisation pattern of the subclones with those obtained by previous hybridisations of pYAC4 vector arm probes to the same Southern blot containing BssHII digested DNA from the YACs. An example of the mapping of Ox19.1 is shown in Figure 2b. Ox19.1 hybridises to the same BssHII fragments of YACs D10 and 2A3 as pYAC4-R and to the same BssHII fragment of YAC H7 as pYAC4-L.

Interestingly, the probe VK21C (DXS296) has been observed previously to map to the same overlapping region of the YACs (1). In the current studies, the Ox19.1 and Ox19.3 subclones were found to be positive for the probe VK21A which, like VK21C, forms part of the DXS296 locus. A comparison of the physical mapping data published by Gedeon et al. (23) and Knight et al. (1) indicated that VK21A and, by extrapolation, the distal portion of the Ox19 cDNA are located ∼175 kb distal to the FRAXE CpG island.
Sequence analysis of the Ox19 cDNA

The EcoRI subclones, Ox19.1, Ox19.2, Ox19.3 and Ox19.4, were sequenced and the internal fragments oriented by comparing the sequences with those obtained by sequencing across the EcoRI sites of the cloned PCR insert of the Ox19 cDNA. The order of the subclones plus an additional EcoRI fragment (Ox19.5), discovered through sequencing the PCR insert, was found to be 5'-Ox19.2, Ox19.5, Ox19.4, Ox19.3, Ox19.1'-3'. The Ox19 cDNA sequence begins 331 bp from the FRAXE GCC repeat and continues for only 135 bp of genomic DNA before being spliced to sequences mapping >170 kb distal. The splice site occurs 49 bp from the 5' end of the 435 bp OxE19 genomic sequence. The complete 1495 bp sequence of the Ox19 cDNA clone aligned with the OxE19 sequence is shown in Figure 3. The predicted amino acid translation is also shown. The cDNA possesses a methionine at amino acid position 30, a TAA stop codon at amino acid position 457 and a short poly-A tail. The 5' end of the Ox19 cDNA is situated ~480 bp distal to the island of methylation sensitive restriction enzyme sites (NotI, BssHII, SacII and EagI).

The predicted open reading frame (ORF) has amino acid motifs and sequences which have homology with the human AF-4 gene (Fig. 3), a ubiquitously expressed gene thought to encode a transcriptional factor (30). There are sequences which correspond very closely to a Kozak consensus sequence (31) but no obvious TATA box or TTGACA sequences.

At the nucleic acid level there is sequence identity of 71% over a stretch of 78 bp and at the level of translation, there are 40 identical matches in a stretch of 77 amino acids and the similarity score is 66% between the Ox19 cDNA and the gene AF-4.

Ox19 cDNA northern blot analysis and RT-PCR

The Ox19.3 subclone was used as a hybridisation probe against Clontech northern blots containing human poly-A+ mRNAs from adult and fetal tissues and adult brain tissues. The hybridisation pattern of the Ox19.3 subclone versus the Clontech adult multiple tissue northern (MTN) blot is shown in Figure 4a. This revealed the existence of a 9.5 kb transcript which gave a strong signal in placenta (lane 3) and comparatively weaker signals in brain and lung (lanes 2 and 4). On a human fetal MTN a 9.5 kb transcript was observed in kidney and in brain (unpublished data). On the MTN containing various tissues from an adult brain, again only the 9.5 kb transcript was observed (Fig. 4b). This was present in...
Figure 3. 5′–3′ sequence and predicted amino acid translation of the Ox19 cDNA. The alignment with the genomic sequence is as shown. The OxE19 49 bp overlap is indicated by a line above the genomic sequence. The methionine at the putative start of translation is enclosed within a box and the main amino acid motifs homologous to the AF-4 gene are underlined. *putative stop codon.

most of the tissues tested, but the signal was notably more intense in the hippocampus and the amygdala. Despite longer exposures and less stringent washing, no other transcripts were observed on any of the northern blots.
The 1.5 kb cDNA may either represent part of the 9.5 kb transcript or a smaller alternatively spliced transcript. In the fetal brain library, six positives out of 1 x 10^6 recombinants corresponded to the 1.5 kb cDNA. We have been unable to extend the clone in this library. RT-PCR was used in order to rule out the possibility of the Ox19 clone being a product of some form of rearrangement that may have occurred during library construction. Primers 10245 and 8752, derived from the 5′ ends of the cDNA respectively were used to amplify adult human brain mRNA. Primers 10245 and 8752, derived from the 5′ ends of the cDNA corresponded to the 1.5 kb cDNA. We have been unable to extend the sequence mapping >170 kb distal in Xq28. This implies the existence of an unusually large genomic intron (at least 170 kb) and helps explain previous difficulties in isolating cDNA sequences from the FRAXE region. The Ox19 cDNA sequence does not show any homology to FMR1, but does show a high level of identity with the chromosome 4 gene, AF-4 or FEL. At the nucleic acid level there is sequence identity of 71% over a stretch of 78 bp. At the level of translation, there are 40 identical matches in a stretch of 77 amino acids and the similarity score is 66% between the Ox19 cDNA and the gene AF-4.

AF-4 was originally cloned because of its involvement in the t(4;11)(q21;q23) acute lymphoblastic leukaemia (ALL). Non-random aberrations involving chromosome 11 band q23 are frequently seen in human acute lymphoblastic and myeloid leukemias. The t(4;11) breakpoints fuse an 11q23 gene (known as MLL, HRX, Htx-1 or ALL-1) to the 4q21 AF-4 (or FEL) gene (32–36). The 11q23 gene is widely expressed in a variety of haematopoietic cell lines as well as epithelial and glial cell-originated tumour cell lines.

It is thought to encode a transcriptional factor (30,37). The function of the AF-4 gene is still uncertain, but its ubiquitous expression indicates a significant role in the cellular function of many tissues (38). The predicted protein possesses a nuclear localisation consensus sequence as well as serine/proline-rich and hydrophilic regions, suggesting that it may be a transcription factor. These domains are highly homologous to those of two other genes, AF-9 and ENL also known to be involved in translocation fusions to ALL-1 (39). However, the motifs shared between AF-4, AF-9 and ENL are not the same as those showing high homology between AF-4 and the Ox19 cDNA. Whether alternative transcripts related to the Ox19 cDNA possess such motifs remains to be discovered. In addition, the functional significance of the amino acid motifs known to be shared between AF-4 and the Ox19 cDNA will be extremely interesting to elucidate.

When hybridised against an adult multiple tissue northern, the Ox19 cDNA detected a 9.5 kb transcript which was relatively strong in placenta, but weaker in brain and lung. Interestingly, the adult brain multiple tissue northern revealed the same 9.5 kb transcript in several brain tissues, but the level of expression appeared to be higher in the amygdala and the hippocampus. This may have functional significance as the amygdala and the hippocampus are intimately related at the neuroanatomical level (40–46). The precise functions of these structures are still being elucidated, but it is widely accepted that the hippocampus is involved in attention, learning and memory whereas the amygdala is thought to play a role in emotional and social behaviour (41–46). Thus, the expression of a FRAXE gene product in both the amygdala and the hippocampus would be extremely provocative, particularly in view of the developmental delay, the speech, learning and behavioural problems, which are often the presenting features of individuals with FRAXE mental handicap (1–4, S.J.L. Knight et al., in preparation). However, more detailed studies of these and other brain tissues are required before it will be possible to provide more informed insights regarding the possible role of a FRAXE gene product.

It seems likely that the 9.5 kb transcript detected by the Ox19.3 subclone of the Ox19 cDNA is the same as that reported by Gedeon et al. (23). This is because the Ox19 cDNA is positive for VK21A (DXS296), the probe which was used by these authors for northern analysis. However, the Ox19.3 subclone did not detect the 2.5 kb transcript reported by Gedeon et al. (23). This transcript may be a splice variant or, alternatively, may represent an adjacent, independent gene in Xq28.

An important observation is that the Ox19 cDNA is only 1495 bp in size compared with the 9.5 kb transcript which is detected by Ox19.3 on northern analysis. The RT-PCR results presented here suggest that the Ox19 cDNA sequence is expressed in human adult brain and therefore it is unlikely that this cDNA is the result of either a deletion or a rearrangement which might have occurred during the cDNA library construction. The two most feasible explanations for the observed size discrepancy are that either the Ox19 cDNA represents only a portion of the FRAXE gene and the remainder has yet to be isolated, or that the Ox19 cDNA contains the complete coding region of a gene which is alternatively spliced. If the Ox19 cDNA represents only a partial transcript, the 5′ coding region of the FRAXE gene may begin at an unknown location proximal to the FRAXE CpG island and both the CpG island and the FRAXE GCC repeat would be predicted to lie in the 5′ untranslated region of the gene (UTR). Analyses of the genomic sequence in this region revealed no consensus acceptor splice sites between these and the Ox19
CpG islands are generally found at the 5′ ends of genes (unpublished data). Although the gene may extend further because the analogous CGG repeat expansions associated with the fragile X syndrome (FMR-1) and Jacobsen disease (CBL2) are restricted to the 5′ UTRs of the respective genes (9, 28, 47). The cDNA presented here has several stop codons in every frame at its 3′ end and a short poly A tail. The absence of a full polyadenylation consensus sequence allows room to speculate that this may not be the true poly A tail and 3′ end at its 3′ end because the analogous CGG repeat expansions associated with the fragile X syndrome where the expansion of the CGG repeat occurs in the 5′ untranslated region of the gene for FRAXE mental handicap. The isolated cDNA is unusual because the coding region is 1284 bp in length and spans at least one intron encompassing 170 kb of genomic sequence. Further analyses are required in order to fully characterise the 5′ region of the gene and to define the promoter activity more accurately. In order to gain possible functional insights, studies of the protein structure and tissue localisation will also be necessary. Importantly, the expression profile must be investigated in both normal and FRAXE patient samples in order to determine whether FRAXE developmental delay can be attributed to the disruption of this gene and to try to establish the importance of the expression of the 9.5 kb transcript, detected on northern blots in the hippocampus and amygdala of the human brain. We conclude that the close proximity of the Ox19 cDNA relative to the FRAXE CpG island makes it an extremely credible candidate gene for FRAXE mental handicap.

MATERIALS AND METHODS

Identification of conserved sequences in the FRAXE region

Human, cow, pig, mouse, chicken and Drosophila DNAs (Promega) were digested with EcoRI (Boehringer Mannheim) and the samples allowed to migrate through a 20 cm × 25 cm 0.8% agarose gel at 65 V for 10 h. Electrophoresed samples were transferred onto Hybond-N membrane (Amersham) and hybridised with the appropriate [32P]-dCTP radiolabelled DNA probe (48). Hybridisation was carried out at 65°C overnight in a solution containing 0.5 M sodium phosphate pH 7.4, 7% SDS and 5% dextran sulphate. Membranes were washed in 3x SSC, 0.1% SDS at 65°C and exposed to Kodak XAR-5 film at ~70°C overnight.

Isolation of cDNAs from a human fetal brain cDNA library (Clontech)

PFUs (250 000) from a human fetal brain cDNA library (Clontech) were plated onto each of 4243 × 243 × 18 mm Bioassay plates (Nunc) containing NZY agar using Y1090 host cells. Duplicate filter lifts were hybridised with radiolabelled OxE18, OxE19 and OxE20 as described for the hybridisation of zooblots. Following autoradiography, only OxE19 produced positive hybridisation signals. Nine positively hybridising plaque regions were identified and six positive bacteriophages, from independent physical regions, plaque purified. Bacteriophage DNAs were isolated using the 'Wizard λ-preps DNA purification system' (Promega) according to manufacturer’s instructions. To release the cDNA inserts from the bacteriophage arms, DNAs from the six positive bacteriophages were digested with EcoRI (Boehringer Mannheim) and subjected to agarose gel electrophoresis. The resulting restriction patterns revealed that each of the cDNAs consisted of four identical EcoRI fragments, indicating that the isolated cDNA clones represented the same clone. One representative cDNA, designated ‘Ox19 cDNA’ was studied further.

Subcloning and mapping the Ox19 cDNA

The four EcoRI fragments of the Ox19 cDNA were subcloned into pBluescript KS(+) (Stratagene). The resulting subclones were designated Ox19.1, Ox19.2, Ox19.3 and Ox19.4 and had insert sizes of 692 bp, 363 bp, 295 bp and 135 bp respectively. The entire Ox19 cDNA was also subcloned: the PCR insert was obtained by standard PCR using 3 μl of the Ox19 cDNA bacteriophage lysate as template and 2gt11-F (5′-GGTGGCGACGACTCCTGGAGCCCG-3′) and 2gt11-R (5′-TTGACACCCAGACCAACTTGTAATG-3′) as primers. The product was electroelphoresed through a 0.8% LMP agarose, excised and purified through a Wizard PCR column (Promega). The purified insert was subcloned using the ‘pGem T-vector System I’ (Promega) and maximum efficiency DH5-α competent cells (Stratagene) according to manufacturer’s instructions.

Inserts from the subclones were mapped back to the FRAXE region by hybridisation against panels of EcoRI digested DNA from (i) ThyBX, mouse and human and (ii) cosmids cos102A, F9/8, D7/13 and H7R. The inserts were also hybridised against pulsed field blots of BssHII digested DNA from YACs 2A3, D10 and H7. The isolation of cosmids F9/8, D7/13 and H7R and YACs 2A3, D10 and H7 has been described previously (1). Cos102A is an ~40 kb cosmiz which spans the FRAXE CpG island and was obtained by subcloning YAC D10 into SuperCos-1 (Stratagene) and hybridising with radiolabelled OxE20. Cos102A overlaps with cosmid F9/8 resulting in a cosmiz contig of ~120–140 kb across the FRAXE region.
Sequence analysis of the Ox19 cDNA

Subclones Ox19.1, Ox19.2, Ox19.3 and Ox19.4 were sequenced with the US Biochemical Sequenase kit (Amersham) using both vector and insert specific oligonucleotides. The position and orientation of the EcoRI inserts within the Ox19 cDNA were determined by sequencing across the EcoRI sites contained within the Ox19 cDNA T-Vector clone. During this study an additional 10 bp EcoRI fragment was identified (Ox19.5). The sequence of Ox19.2 revealed 100% homology over the 49 bp overlap with the OxEI9 subclone used in the isolation of the Ox19 cDNA. Database searches (Genbank, Swissprot) of the complete Ox19 cDNA sequence were carried out using the ‘BlastN’, ‘BlastX’ and ‘Fasta’ programs of the GCG package (Genetics Computer Group, Madison). Amino acid translations and putative ORFs were identified using the ‘Map’ program of the GCG package.

Northern analysis

Ox19.3 was hybridised against an adult multiple tissue northern (MTN), a fetal MTN and an adult brain MTN (Clontech). The hybridisation conditions were as recommended by the manufacturers with the exception that the hybridisation solution was passed through a 0.45 µm filter prior to use. Washes were performed according to manufacturer’s instructions and the membranes were subjected to autoradiography at –70°C for 1–14 days.

RT-PCR studies

Total RNA was isolated from adult human brain tissue using a total RNA isolation kit (Promega) according to manufacturer’s instructions. Poly-A+ RNA was selected from the total RNA using Dynabeads (Dynal) according to manufacturer’s instructions. In each RT-PCR reaction 5 µg adult total brain RNA was used as template. First strand synthesis, with and without reverse transcriptase, was performed using random poly d(N)6 hexamer primers and poly d(T)12–18 primers. PCR amplification was carried out using primers 10245 (5′-CAATGACATCTATTCAGCTTTTTCAG-3′) and 8752 (5′-CAAGCAGCTTTGTGAGGCTTCG-3′) each at 2 µM in a final reaction volume of 25 µl containing 1 X PCR Buffer (Perkin-Emerly), 1.5 mM MgCl₂, 4% DMSO and 200 µM each of dGTP, dCTP, dATP and dTTP (Boehringer Mannheim). As a control, the primers U13 (5′-GCAACAGAGAAGCTGACTA-3′) and U4 (5′-CACACGCGTTTGCAGCGGCACT-3′) from the Urophrin gene were also used to PCR from the first strand template. The PCR conditions were 94°C for 5 min followed by 30 cycles of 94°C for 1 min, 64°C for 1 min, 72°C for 3 min and a final extension of 72°C for 10 min. The products were electrophoresed through a 0.8% agarose gel, transferred to Hybond-N and hybridised using the radiolabelled Ox19.3 subclone insert as probe.

ACKNOWLEDGEMENTS

We would like to thank Dr Andrew Nesbit for plating and lifting the cDNA library. We would also like to thank Dr Andrea Németh, Dr Dahlia Zaidel and Dr Jon Tinsley for helpful discussion. We are grateful to the Medical Research Council for financial support.

REFERENCES


