Cell, Vol. 65, 905-914, May 31, 1991, Copyright © 1991 by Cell Press

Identification of a Gene (*FMR-1*) Containing a CGG Repeat Coincident with a Breakpoint Cluster Region Exhibiting Length Variation in Fragile X Syndrome

Annemieke J. M. H. Verkerk,* Maura Pieretti,† James S. Sutcliffe,‡ Ying-Hui Fu,† Derek P. A. Kuhl,† Antonio Pizzuti,† Orly Reiner,† Stephen Richards,† Maureen F. Victoria,† Fuping Zhang,‡ Bert E. Eussen,* Gert-Jan B. van Ommen,§ Lau A. J. Blonden,§ Gregory J. Riggins,‡ Jane L. Chastain,‡ Catherine B. Kunst,‡ Hans Galjaard,* C. Thomas Caskey,† David L. Nelson,† Ben A. Oostra,‖ and Stephen T. Warren‡
*Department of Clinical Genetics
Erasmus University
Rotterdam
The Netherlands
†Institute for Molecular Genetics
Human Genome Center
Howard Hughes Medical Institute
Baylor College of Medicine
Houston, Texas 77030
‡Departments of Biochemistry and Pediatrics
Emory University School of Medicine
Atlanta, Georgia 30322
§Department of Human Genetics
Sylvius Laboratory
Leiden
The Netherlands
‖Department of Cell Biology
Erasmus University
Rotterdam
The Netherlands

Summary

Fragile X syndrome is the most frequent form of inherited mental retardation and is associated with a fragile site at Xq27.3. We identified human YAC clones that span fragile X site-induced translocation breakpoints coincident with the fragile X site. A gene (*FMR-1*) was identified within a four cosmid contig of YAC DNA that expresses a 4.8 kb message in human brain. Within a 7.4 kb EcoRI genomic fragment, containing *FMR-1* exonic sequences distal to a CpG island previously shown to be hypermethylated in fragile X patients, is a fragile X site-induced breakpoint cluster region that exhibits length variation in fragile X chromosomes. This fragment contains a lengthy CGG repeat that is 250 bp distal of the CpG island and maps within a *FMR-1* exon. Localization of the brain-expressed *FMR-1* gene to this EcoRI fragment suggests the involvement of this gene in the phenotypic expression of the fragile X syndrome.

Introduction

Fragile X syndrome is the most frequently encountered form of inherited mental retardation in humans, with a prevalence estimated to be 1 in 1250 males (Gustavson et al., 1986; Webb et al., 1986). Fragile X syndrome segregates as an X-linked dominant disorder with reduced penetrance, since either sex, when carrying the fragile X mutation, may exhibit mental deficiency. Sherman et al. (1984, 1985) have shown that approximately 30% of carrier females are penetrant and that 20% of males carrying the fragile X chromosome are phenotypically normal but may transmit the disorder and have fully penetrant grandsons. In addition to the mental retardation, which is variable in severity, penetrant males exhibit additional phenotypic involvement including macroorchidism and distinctive facies (Nussbaum and Ledbetter, 1990). Since fully penetrant males rarely reproduce, it has been suggested that the frequency of new fragile X mutations may be as high as 1 in 3000 germ cells to maintain the population frequency (Brown, 1990).

Fragile X syndrome, as implied by the name, is associated with a fragile site, expressed as an isochromatid gap in the metaphase chromosome, at map position Xq27.3 (Krawczun et al., 1985). The fragile X site is induced by cell culture conditions that perturb deoxypyrimidine pools and is rarely observed in greater than 50% of the metaphase spreads (Sutherland and Hecht, 1985). Neither the molecular nature of the fragile X site nor its relationship to the gene(s) responsible for the clinical expression of the syndrome is understood. However, based upon genetic linkage studies as well as in situ hybridizations, the fragile X site and its associated gene(s) are tightly linked, if not coincident (Oostra et al., 1990; Rousseau et al., 1991; Hirst et al., 1991).

To elucidate the fragile X site at the molecular level, somatic cell hybrids were isolated that contained translocations between rodent chromosomes and the human fragile X chromosome, retaining either human Xpter→q27.3 or human Xq27.3→qter, referred to as proximal or distal translocations relative to the fragile X site (Warren et al., 1987, 1988, 1990). Since the high frequency and specificity of the chromosome breakage were not observed in normal X hybrids and since the translocation breakpoints map within the same interval defined by polymorphic loci that flank the fragile X locus (Warren et al., 1990; Rousseau et al., 1991; Hirst et al., 1991), these breakpoints are likely to coincide with the fragile X site.

A yeast artificial chromosome (YAC) has been isolated that spans some of these translocation breakpoints and includes polymorphic loci that flank the fragile X locus (Heitz et al., 1991). Within this region, a fragile X-related CpG island was identified that is aberrantly hypermethylated in patients and most carriers of the fragile X syndrome (Vincent et al., 1991; Bell et al., 1991). Although the significance of this CpG-island hypermethylation remains unclear, these data do imply the presence of a gene, perhaps inactivated by methylation, within a genomic region that includes the fragile X-associated hybrid breakpoints. We report below the isolation and characterization of a brain-expressed gene, designated *FMR-1*, that is associated with this CpG island; we show that the majority of the

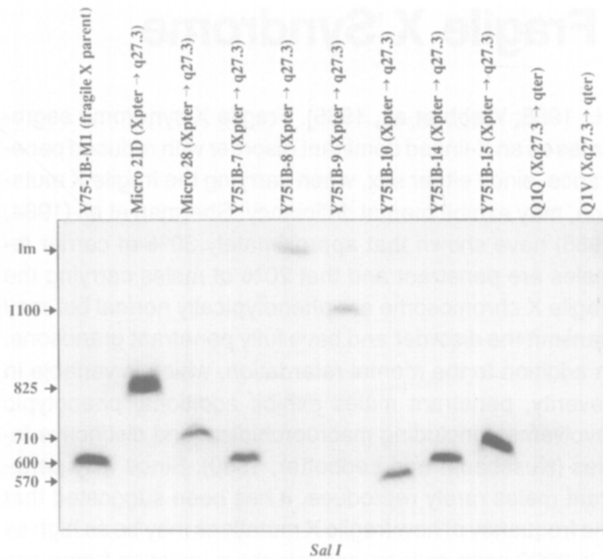

Figure 1. Southern Blot Analysis of PFG-Resolved SalI-Digested DNA of Proximal Translocation Hybrids Probed with p46–1.1

Y75–1B-M1 is a somatic cell hybrid containing the intact fragile X chromosome from which all other hybrids were derived. Lanes 2–9 are proximal translocation hybrids containing centric human Xpter→q27.3 translocated to different rodent chromosome arms. Q1Q and Q1V are distal translocation hybrids containing human Xq27.3→qter translocated to different centric rodent chromosomes. The distal translocation hybrids have lost the human sequence detected by p46–1.1. Hybrids Y751B-7 and Y751B-14 show the same 600 kb SalI fragment as the parental hybrid; however, all other proximal translocation hybrids show variant bands indicating that probe p46–1.1 detects sequence within 600 kb of these translocation breakpoints.

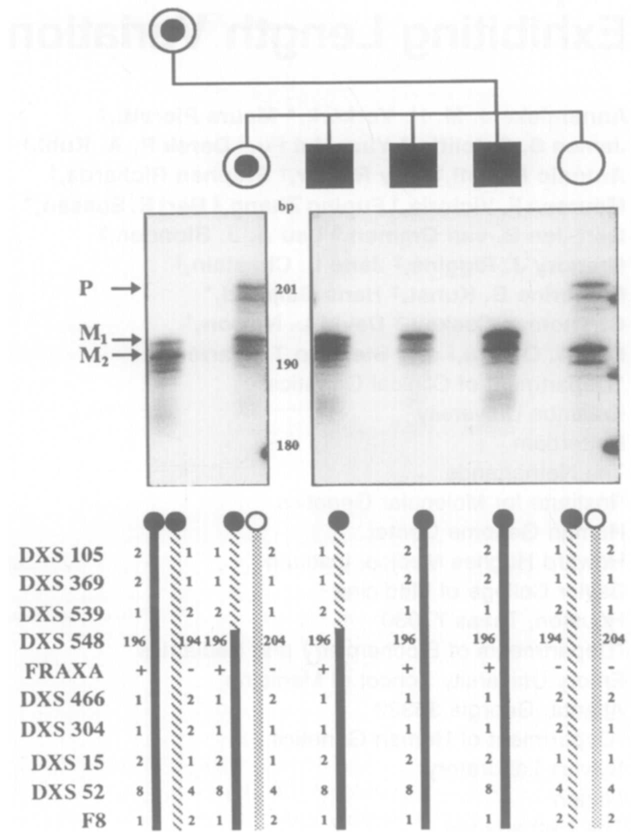

Figure 2. PCR Analysis of DXS548 Alleles in a Fragile X Family with Key Recombinant Individuals

The carrier mother shows two DXS548 alleles at 196 and 194 bp (M_1 and M_2, respectively). The paternal 204 allele of the father (not shown) is seen in the carrier daughter (II-1) who also inherited the maternal 196 bp allele. All three affected males inherited the 196 bp maternal allele (compare with the 194 allele of the normal daughter II-5). The carrier daughter (II-1) and affected son (II-2) are both recombinants between proximal markers DXS105, DXS369, and DXS539. However, these individuals are nonrecombinant with DXS548, placing this locus distal to the crossovers and closer to the fragile X locus.

hybrid translocation breakpoints cluster within a 7.4 kb fragment, containing both the CpG island and sequences at the 5′ end of *FMR-1*, which exhibits length variation in patients, suggesting the association of the fragile X site with *FMR-1*.

Results

Isolation of YACs Spanning the Fragile X Translocation Breakpoints

Through regional mapping of YAC clones containing DNA inserts derived from the distal human Xq (Nelson et al., 1991), an 80 kb YAC (RS46) was found to map within Xq27.3, proximal to the fragile X–associated hybrid breakpoints. A 4.0 kb subclone (p46–1.1) of RS46 identified a normal 600 kb SalI fragment on pulsed-field gel (PFG) electrophoresis that was altered in size in 6 of 8 proximal translocation hybrids (Figure 1). Previous PFG electrophoresis analyses of these hybrids, with more distant X-linked probes, had shown identical band sizes and therefore similar methylation patterns, as might be expected since the hybrids were all derived from the same parental fragile X somatic cell hybrid (Y75–1B-M1). These data therefore suggest that in 75% of the proximal translocation hybrids, the human breakpoint is within the 600 kb SalI fragment observed in the parental, intact fragile X hybrid. In the translocation hybrids, the distal human SalI site is lost and replaced by heterologous translocations containing different rodent SalI sites.

Since YAC RS46 does not hybridize to the DNA of the distal translocation hybrids and therefore does not cross these translocation breakpoints, additional YACs were sought of this region. A YAC library developed at the Human Polymorphism Study Center (CEPH; Albertsen et al., 1990) was screened using RS46-specific oligonucleotide primers 1625 and 1626 (Nelson et al., 1991). A YAC of 475 kb (209G4) was identified that completely overlaps YAC RS46 and includes sequences distal to the proximal translocation breakpoints that are present in 13 of 14 distal translocation breakpoints (data not shown). YAC 209G4 encompasses 86% (19 of 22) of both the proximal and distal translocation breakpoints and thus identifies a fragile X–associated breakpoint cluster region. In situ hybridization using YAC 209G4 showed localization to the expressed fragile X site with signal on both flanking boundaries of the isochromatid gap of the fragile site as well as within the gap itself, suggesting the presence of uncon-

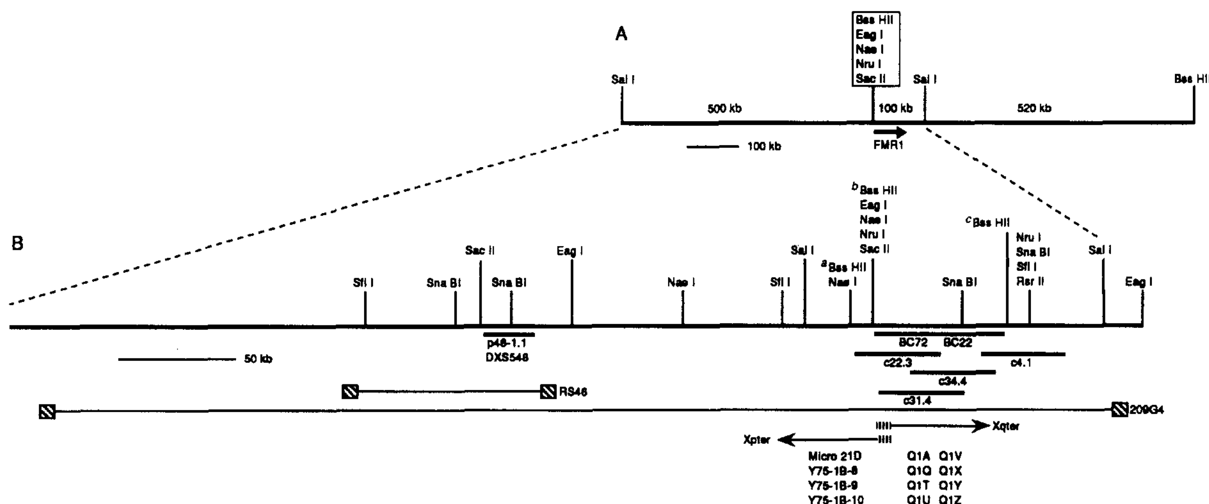

Figure 3. Physical Map of the Fragile X Region of Genomic and YAC DNA

(A) Physical map of the fragile X chromosome in the vicinity of the fragile X locus. Shown are the SalI sites, which give rise to the 600 kb fragment seen in hybrid Y75–1B-M1 probed with p46–1.1 and the normal 620 kb BssHII fragment observed in normal X chromosomes. The sites within the box are those previously shown to be methylated on the fragile X chromosome. The position and orientation of *FMR-1* are shown.

(B) Higher resolution physical map derived from both YAC inserts and genomic DNA. Probe p46–1.1 and the DXS548 loci are shown, as are the positions of cDNAs and cosmids. YACs RS46 and 209G4 are shown below in alignment with the map (cross-hatched boxes indicate YAC vector sequences). The positions of the translocation breakpoints are shown as well as the orientation of the map relative to the X chromosome telomeres.

densed DNA within the fragile site and indicating that YAC 209G4 includes this region (data not shown).

The close proximity of these YACs to the fragile X locus was independently supported by genetic linkage studies between a polymorphism identified in YAC RS46 and the fragile X locus. DXS548 is a dinucleotide repeat that reveals nine alleles of variable length that are informative in >80% of fragile X families (G. J. R., unpublished data). In highly selected families previously shown to have crossovers with tightly linked flanking markers, DXS548 cosegregated, without recombination, with the fragile X locus (LOD score 6.95 at $\theta = 0$). As shown in Figure 2, a carrier daughter and affected son are recombinant between the fragile X locus (*FRAXA*) and proximal markers DXS539 (probe JH89) and DXS369 (probe RN1), which map approximately 5 cM proximal to *FRAXA* with LOD scores greater than 40 (Oostra et al., 1990; B. Oostra, unpublished data). Therefore, DXS548 positions YACs RS46 and 209G4 near the mutation responsible for the clinical phenotype of the fragile X syndrome.

A physical map of YAC 209G4 and of the corresponding genomic region was developed and is shown in Figure 3. A CpG island containing five infrequently cleaving restriction endonuclease sites was identified 150 kb distal to DXS548. This CpG island is that identified by a previous independent isolation of YAC 209G4 from the CEPH library (Heitz et al., 1991). This CpG island appears hypermethylated on the fragile X chromosome. Vincent et al. (1991) and Bell et al. (1991) have both reported the absence of a normal 620 kb BssHII fragment (Figure 3A) in patients and most carriers of the fragile X syndrome. The absence of the fragment appears to be due to the methylation (and therefore resistance to cleavage) of the BssHII site (b in Figure 3B), leading to a very large band that fails to resolve on PFG electrophoresis. Since CpG islands often are found 5′ to mammalian genes (Lindsay and Bird, 1987) and since methylation of such islands may influence expression of associated genes (Wolf et al., 1984; Yen et al., 1984), it is possible a gene may reside near this fragile X-related CpG island and its expression (or lack of) may be responsible for at least a portion of the fragile X phenotype.

Cosmid Contig Surrounding the Fragile X-Related CpG Island and Breakpoint Cluster Region

To characterize the region surrounding this CpG island, a cosmid library was constructed from the yeast clone harboring YAC 209G4, and cosmids containing human DNA were identified by hybridization to human-specific repetitive elements. A four cosmid contig was identified that spans the fragile X-related CpG island (Figure 3B) from BssHII site a (cosmid 22.3) through BssHII site c (cosmid 4.1).

Cosmid 22.3 was found to include the breakpoints of 11 of 16 tested translocation hybrids (4 of 5 proximal translocations and 7 of 11 distal translocations; all 16 breakpoints map within YAC 209G4). As shown in Figure 4A, nine bands (including doublet bands at 5.6 and 5.5 kb), surveying approximately 44 kb of genomic DNA, are observed on Southern analysis of EcoRI-digested DNA of the intact fragile X hybrid (Y75–1B-M1) following hybridization with radiolabeled and preannealed cosmid 22.3. Of these nine bands, three are present in the proximal hybrid micro21D (with a novel 3.8 kb junctional fragment) and five are present in the distal hybrid Q1X (with a novel 4.8 kb junctional fragment). The 7.4 kb band of the intact X hybrid Y75–1B-M1 is absent in both translocation hybrids, indicating that both breakpoints fall within this interval. The other nine hybrids all exhibited patterns similar to either micro21D or

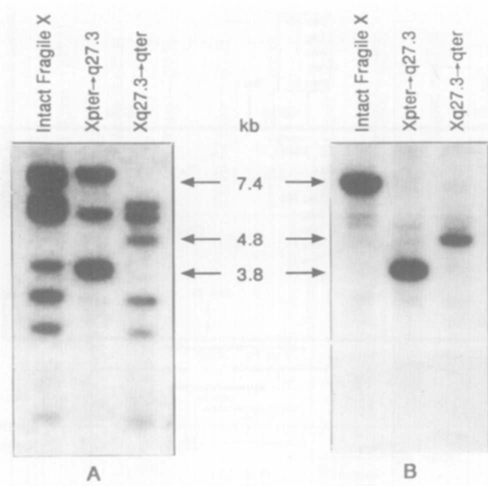

Figure 4. Southern Blot Analysis of Fragile X-Associated Translocation Breakpoints

DNA of hybrid Y75-1B-M1, containing the intact fragile X chromosome, proximal hybrid micro21D, and distal hybrid Q1X was cleaved with EcoRI and 10 μg of each sample loaded onto the gel. (A) Southern blot hybridized with cosmid 22.3 and (B) the same filter hybridized with pE5.1.

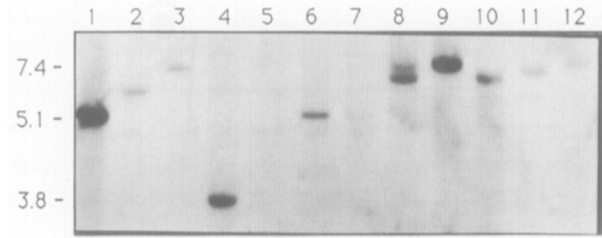

Figure 6. Length Variation of EcoRI Fragments from Normal and Fragile X Human Chromosomes Detected with Probe pE5.1

Lanes 1, 6, and 7 show hybridization of the normal 5.1 kb EcoRI fragment in placental DNA (lane 1) and cloned into a cosmid (22.3) or YAC vector (209G4) and seeded into hamster DNA at single-copy level. Somatic cell hybrids containing portions of fragile X chromosomes in hamster backgrounds show bands of altered size from the normal 5.1 kb fragment. Lane 2 contains the hybrid X3000-11.1 (Nussbaum et al., 1986a). Lane 3 contains DNA from micro28D, a proximal hybrid with a breakpoint distal to the fragile site; lane 4 contains DNA from micro21D, a proximal hybrid with the same chromosome as micro28D, however with a breakpoint detected by pE5.1 (see Figure 5). Lane 5 contains hamster DNA. Lanes 8–12 contain DNA from five unrelated fragile X patients' lymphoblastoid lines. These are GM03200, GM04025c, GM06897, GM06912, and GM07294. Bands altered from the normal 5.1 kb are seen in each fragile X sample.

Q1X, with distinct junctional fragments allowing identification of a fragile X-associated breakpoint cluster region (FXBCR) with this 7.4 kb fragment.

The 7.4 kb EcoRI fragment observed above on the fragile X chromosome was not observed in restriction digests of the overlapping cosmids 22.3 and 31.4. However, comparison of the cosmid restriction maps with the genomic EcoRI fragments detected by cosmid 22.3 shows the replacement of the 7.4 kb fragment of the fragile X chromosome with a 5.1 kb fragment of normal DNA, which, similar to the 7.4 kb fragment, includes the BssHII site of the fragile X-related CpG island (Figure 5A). The normal 5.1 kb fragment was subcloned to further analyze this difference. As shown in Figure 4B, the 5.1 kb fragment (pE5.1; Figure 5B) hybridizes specifically to the 7.4 kb EcoRI fragment of the fragile X chromosome and clearly shows the junctional fragments in micro21D and Q1X. Thus, a fragment length difference exists between the normal DNA used to construct YAC 209G4 and the fragile X chromosome of hybrid Y75-1B-M1, and this fragment identifies the FXBCR.

Fragile X Breakpoint Cluster Region Rearranged in Fragile X Patients

Figure 6 shows the results of Southern hybridization of EcoRI-digested DNA from two normal and seven unrelated

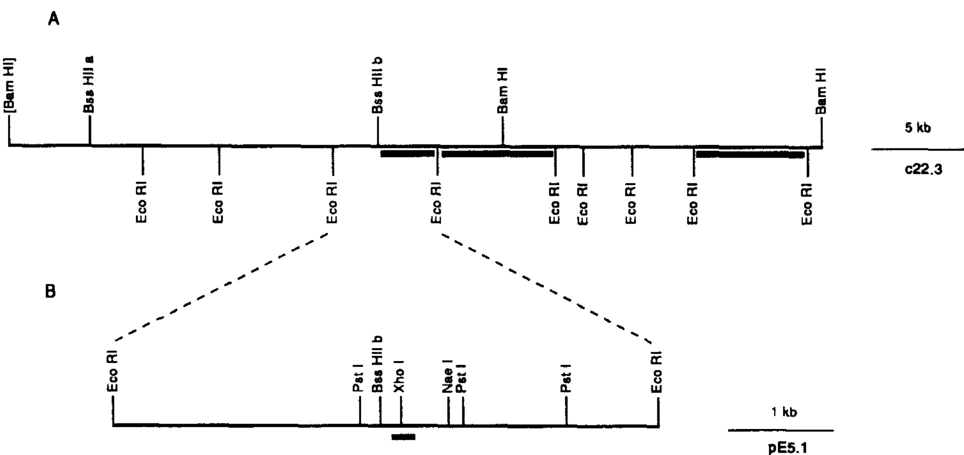

Figure 5. Restriction Map of Cosmid 22.3 and pE5.1

(A) Cosmid 22.3 showing BssHII sites a and b as well as EcoRI and BamHI sites. BamHI site in brackets was destroyed during cloning. Solid lines below the map show fragments that hybridize to cDNAs BC72 and BC22.

(B) Map of the cloned 5.1 kb EcoRI fragment of cosmid 22.3 (pE5.1). Solid line below map shows the position of the *FMR-1* exonic sequence, which contains the XhoI site.

A

5' open reading frame 3' TAA
BC72
BC22
Xho I Xho I Hind III Hind III
1000 bp

Figure 7. Map and Sequence of *FMR-1* cDNA Clones

Map (A) and DNA sequence (B) of *FMR-1* cDNA clones. The 3765 nucleotides of DNA sequence generated from cDNA clones BC72 and BC22 defining the *FMR-1* gene are shown with translation of the 657 amino acid open reading frame. Nucleotides 1–1027 derive from BC72, and nucleotides 934–3765 are from BC22. The CGG repeat encoding 30 contiguous Arg residues begins with base 37 and extends to base 127.

B

```
   1: GAC GGA GGC GCC CGT GCC AGG GGG CGT GCG GCA GCG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG AGG CGG CGG CGG CGG CGG CGG CGG
   1: asp gly gly ala arg ala arg gly arg ala ala ala arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg

  91: CGG CGG AGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CTG GGC CTC GAG CGC CCG CAG CCC ACC TCT CGG GGG CGG GCT CCC GGC GCT AGC
  31: arg arg arg arg arg arg arg arg arg arg arg arg leu gly leu glu arg pro gln pro thr ser arg gly arg ala pro gly ala ser

 181: AGG GCT GAA GAG AAG ATG GAG GAG CTG GTG GTG GAA GTG CGG GGC TCC AAT GGC GCT TTC TAC AAG GCA TTT GTA AAG GAT GTT CAT GAA
  61: arg ala glu glu lys met glu glu leu val val glu val arg gly ser asn gly ala phe tyr lys ala phe val lys asp val his glu

 271: GAT 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 CCT GTA GGT TAT
  91: asp ser ile thr val ala phe glu asn asn trp gln pro asp arg gln ile pro phe his asp val arg phe pro pro pro val gly tyr

 361: AAT AAA GAT ATA AAT GAA AGT GAT GAA GTT GAG GTG TAT TCC AGA GCA AAT GAA AAA GAG CCT TGC TGT TGG TGG TTA GCT AAA GTG AGG
 121: asn lys asp ile asn glu ser asp glu val glu val tyr ser arg ala asn glu lys glu pro cys cys trp trp leu ala lys val arg

 451: ATG ATA AAG GGT GAG TTT TAT GTG ATA GAA TAT GCA GCA TGT GAT GCA ACT TAC AAT GAA ATT GTC ACA ATT GAA CGT CTA AGA TCT GTT
 151: met ile lys gly glu phe tyr val ile glu tyr ala ala cys asp ala thr tyr asn glu ile val thr ile glu arg leu arg ser val

 541: AAT CCC AAC AAA CCT GCC ACA AAA GAT ACT TTC CAT AAG ATC AAG CTG GAT GTG CCA GAA GAC TTA CGG CAA ATG TGT GCC AAA GAG GCG
 181: asn pro asn lys pro ala thr lys asp thr phe his lys ile lys leu asp val pro glu asp leu arg gln met cys ala lys glu ala

 631: GCA CAT AAG GAT TTT AAA AAG GCA GTT GGT GCC TTT TCT GTA ACT TAT GAT CCA GAA AAT TAT CAG CTT GTC ATT TTG TCC ATC AAT GAA
 211: ala his lys asp phe lys lys ala val gly ala phe ser val thr tyr asp pro glu asn tyr gln leu val ile leu ser ile asn glu

 721: GTC 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 AAT GAA GAA GCT
 241: val thr ser lys arg ala his met leu ile asp met his phe arg ser leu arg thr lys leu ser leu ile met arg asn glu glu ala

 811: AGT AAG CAG CTG GAG AGT TCA AGG CAG CTT GCC TCG AGA TTT CAT GAA CAG TTT ATC GTA AGA GAA GAT CTG ATG GGT CTA GCT ATT GGT
 271: ser lys gln leu glu ser ser arg gln leu ala ser arg phe his glu gln phe ile val arg glu asp leu met gly leu ala ile gly

 901: ACT CAT GGT GCT AAT ATT CAG CAA GCT AGA AAA GTA CCT GGG GTC ACT GCT ATT GAT CTA GAT GAA GAT ACC TGC ACA TTT CAT ATT TAT
 301: thr his gly ala asn ile gln gln ala arg lys val pro gly val thr ala ile asp leu asp glu asp thr cys thr phe his ile tyr

 991: GGA GAG GAT CAG GAT GCA GTG AAA AAA GCT AGA AGC TTT CTC GAA TTT GCT GAA GAT GTA ATA CAA GTT CCA AGG AAC TTA GTA GTA ATA
 331: gly glu asp gln asp ala val lys lys ala arg ser phe leu glu phe ala glu asp val ile gln val pro arg asn leu val val ile

1081: GGA AAA AAT GGA AAG CTG ATT CAG GAG ATT GTG GAC AAG TCA GGA GTT GTG AGG GTG AGG ATT GAG GCT GAA AAT GAG AAA AAT GTT CCA
 361: gly lys asn gly lys leu ile gln glu ile val asp lys ser gly val val arg val arg ile glu ala glu asn glu lys asn val pro

1171: CAA GAA GAG GAA ATT ATG CCA CCA AAT TCC CTT CCT TCC AAT AAT TCA AGG GTT GGA CCT AAT GCC CCA GAA GAA AAA AAA CAT TTA GAT
 391: gln glu glu glu ile met pro pro asn ser leu pro ser asn asn ser arg val gly pro asn ala pro glu glu lys lys his leu asp

1261: ATA AAG GAA AAC AGC ACC CAT TTT TCT CAA CCT AAC AGT ACA AAA GTC CAG AGG GGT ATG GTA CCA TTT GTT TTT GTG GGA ACA AAG GAC
 421: ile lys glu asn ser thr his phe ser gln pro asn ser thr lys val gln arg gly met val pro phe val phe val gly thr lys asp

1351: AGC ATC GCT AAT GCC ACT GTT CTT TTG GAT TAT CAC CTG AAC TAT TTA AAG GAA GTA GAC CAG TTG CGT TTG GAG AGA TTA CAA ATT GAT
 451: ser ile ala asn ala thr val leu leu asp tyr his leu asn tyr leu lys glu val asp gln leu arg leu glu arg leu gln ile asp

1441: GAG CAG TTG CGA CAG ATT GGA GCT AGT TCT AGA CCA CCA CCA AAT CGT ACA GAT AAG GAA AAA AGC TAT GTG ACT GAT GAT GGT CAA GGA
 481: glu gln leu arg gln ile gly ala ser ser arg pro pro pro asn arg thr asp lys glu lys ser tyr val thr asp asp gly gln gly

1531: ATG GGT CGA GGT AGT AGA CCT TAC AGA AAT AGG GGG CAC GGC AGA CGC GGT CCT GGA TAT ACT TCA GGA ACT AAT TCT GAA GCA TCA AAT
 511: met gly arg gly ser arg pro tyr arg asn arg gly his gly arg arg gly pro gly tyr thr ser gly thr asn ser glu ala ser asn

1621: GCT 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 TTC CTG CGC AGA
 541: ala ser glu thr glu ser asp his arg asp glu leu ser asp trp ser leu ala pro thr glu glu glu arg glu ser phe leu arg arg

1711: GGA GAC GGA CGG CGG CGT GGA GGG GGA GGA AGA GGA CAA GGA GGA AGA GGA CGT GGA GGA GGC TTC AAA GGA AAC GAC GAT CAC TCC CGA
 571: gly asp gly arg arg arg gly gly gly gly arg gly gln gly gly arg gly arg gly gly gly phe lys gly asn asp asp his ser arg

1801: ACA GAT AAT CGT CCA CGT AAT CCA AGA GAG GCT AAA GGA AGA ACA ACA GAT GGA TCC CTT CAG AAT ACC TCC AGT GAA GGT AGT CGG CTG
 601: thr asp asn arg pro arg asn pro arg glu ala lys gly arg thr thr asp gly ser leu gln asn thr ser ser glu gly ser arg leu

1891: CGC ACG GGT AAA GAT CGT AAC CAG AAG AAA GAG AAG CCA GAC AGC GTG GAT GGT CAG CAA CCA CTC GTG AAT GGA GTA CCC TAA
 631: arg thr gly lys asp arg asn gln lys lys glu lys pro asp ser val asp gly gln gln pro leu val asn gly val pro ***
                                 ---------------

1975: ACT GCA TAA TTC TGA AGT TAT ATT TCC TAT ACC ATT TCC GTA ATT CTT ATT CCA TAT TAG AAA ACT TTG TTA GGC CAA AGA CAA ATA GTA
2065: GGC AAG ATG GCA CAG GGC ATG AAA TGA ACA CAA ATT ATG CTA AGA ATT TTT TAT TTT TTG GTA TTG GCC ATA AGC AAC AAT TTT CAG ATT
2155: TGC ACA AAA AGA TAC CTT AAA ATT TGA AAC ATT GCT TTT AAA ACT ACT TAG CAC TTC AGG GCA GAT TTT AGT TTT ATT TTC TAA AGT ACT
2245: GAG CAG TGA TAT TCT TTG TTA ATT TGG ACC ATT TTC CTG CAT TGG GTG ATC ATT CAC CAG TAC ATT CTC AGT TTT TCT TAA TAT ATA GCA
2335: TTT ATG GTA ATC ATA TTA GAC TTC TGT TTT CAA TCT CGT ATA GAA GTC TTC ATG AAA TGC TAT GTC ATT TCA TGT CCT GTG TCA GTT TAT
2425: GTT TTG GTC CAC TTT TCC AGT ATT TTA GTG GAC CCT GAA ATG TGT GTG ATG TGA CAT TTG TCA TTT TCA TTA GCA AAA AAA GTT GTA TGA
2515: TCT GTG CCT TTT TTA TAT CTT GGC AGG TAG GAA TAT TAT ATT TGG ATG CAG AGT TCA GGG AAG ATA AGT TGG AAA CAC TAA ATG TTA AAG
2605: ATG TAG CAA ACC CTG TCA AAC ATT AGT ACT TTA TAG AAG AAT GCA TGC TTT CCA TAT TTT TTT CCT TAC ATA AAC ATC AGG TTA GGC AGT
2695: ATA AAG AAT AGG ACT TGT TTT TGT TTT TGT TTT GTT GCA CTG AAG TTT GAT AAA TAG TGT TAT TGA GAG AGA TGT GTA ATT TTT CTG TAT
2785: AGA CAG GAG AAG AAA GAA CTA TCT TCA TCT GAG AGA GGC TAA AAT GTT TTC AGC TAG GAA CAA ATC TTC CTG GTC GAA AGT TAG TAG GAT
2875: ATG CCT GCT CTT TGG CCT GAT GAC CAA TTT TAA CTT AGA GCT TTT TTT TTT AAT TTT GTC TGC CCC AAG TTT TGT GAA ATT TTT CAT ATT
2965: TTA ATT TCA AGC TTA TTT TGG AGA GAT AGG AAG GTC ATT TCC ATG TAT GCA TAA TAA TCC TGC AAA GTA CAG GTA CTT TGT CTA AGA AAC
3055: ATT GGA AGC AGG TTA AAT GTT TTG TAA ACT TTG AAA TAT ATG GTC TAA TGT TTA AGC AGA ATT GGA AAA GAC TAA GAT CGG TTA ACA AAT
3145: AAC AAC TTT TTT TTC TTT TTT TCT TTT GTT TTT TGA AGT GTT GGG GTT TGG TTT TGT TTT TTG AGT CTT TTT TTT TTA AGT GAA ATT TAT
3235: TGA GGA AAA ATA TGT GAA GGA CCT TCA CTC TAA GAT GTT ATA TTT TTC TTA AAA AGT AAC TCC TAG TAG GGG TAC CAC TGA ATC TGT ACA
3325: GAG CCG TAA AAA CTG AAG TTC TGC CTC TGA TGT ATT TTG TGA GTT TGT TTC TTT GAA TTT TCA TTT TAC AGT TAC TTT TCC TTG CAT ACA
3415: AAC AAG CAT ATA AAA TGG CAA CAA ACT GCA CAT GAT TTC ACA AAT ATT AAA AAG TCT TTT AAA AAG TAT TGC CAA ACA TTA ATG TTG ATT
3505: TCT AGT TAT TTA TTC TGG GAA TGT ATA GTA TTT GAA AAC AGA AAT TGG TAC CTT GCA CAC ATC ATC TGT AAG CTG TTT GGT TTT AAA ATA
3595: CTG TAG ATA ATT AAC CAA GGT AGA ATG ACC TTG TAA TGT AAC TGC TCT TGG GCA ATA TTC TCT GTA CAT ATT AGC GAC AAC AGA TTG GAT
3685: TTT ATG TTG ACA TTT GTT TGG TTA TAG TGC AAT ATA TTT TGT ATG CAA GCA GTT TCA ATA AAG TTT GAT CTT CCT CTG CTA CCG
```

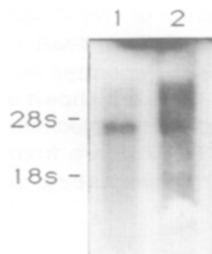

Figure 8. Northern Blot Analysis of Poly(A) RNA Hybridized with cDNA BC22

Five micrograms of poly(A)-selected RNA from human brain (lane 1) and normal placenta (lane 2) was electrophoresed, blotted onto a GeneScreen Plus filter, and hybridized with radiolabeled BC22 insert. A single hybridizing species of approximately 4.8 kb is observed in each lane.

fragile X individuals using pE5.1 as probe. The normal samples (2 of 5 normal samples are shown) exhibit the expected 5.1 kb fragment, while all seven fragile X patient DNAs exhibited larger EcoRI fragments with variable increases in size, including the 7.4 kb fragment observed above from hybrid Y75-1B-M1. These data suggest an insertion or amplification event within the normal 5.1 kb fragment that is specific for the fragile X chromosome and is coincident with the fragile X-associated breakpoint cluster region and the fragile X-related CpG island.

Identification and Characterization of *FMR-1*

To search for transcripts associated with the fragile X region, the cosmid subclones of YAC 209G4 were used as hybridization probes to screen a cDNA library derived from normal human fetal brain RNA. Cosmid 4.1, containing BssHII site c (see Figure 3B), identified cDNA clone BC22. Restriction digestion and sequence analysis revealed an insert in BC22 of 2832 bp, with an open reading frame at one end extending 1033 bp to a stop codon (Figure 7). Since the reading frame remains open at the 5′ end of the clone, BC22 was used to identify related cDNAs from the same library. Several overlapping clones were isolated, one of which, BC72, was characterized in greater detail. This clone extended the cDNA sequence another 933 bp in the 5′ direction and overlapped BC22 for approximately 2000 bp toward the 3′ end. Sequence analysis demonstrated that the same reading frame remained open through the 5′ end of BC72, indicating that the 5′ end of the mRNA has not yet been reached and allowing prediction of a portion (657 amino acids) of the encoded protein. It remains unclear whether the entire 3′ portion was also isolated since no poly(A) tract was found at the end of BC22; however, a putative polyadenylation addition signal is observed in position 3741 following numerous in-frame stop codons (Figure 7). A consensus sequence for nuclear translocation signal KKXK, often followed by a proline, has been described (Roberts, 1989) and is underlined in Figure 7B.

Northern hybridization using the BC22 insert as probe detects an mRNA of approximately 4.8 kb in human brain and placenta (Figure 8), indicating that the 3.8 kb of cDNA obtained does not contain the entire mRNA of this gene.

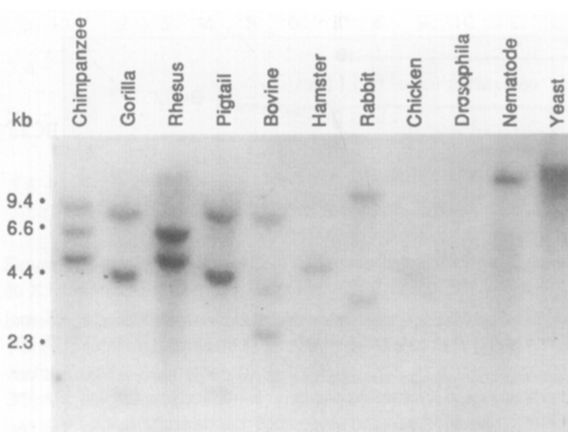

Figure 9. Zoo Blot Analysis of DNA Isolated from Several Species Hybridized with cDNA BC22

Ten micrograms of DNA from each species was cleaved with EcoRI and electrophoresed and blotted onto nylon membrane. Hybridization was carried out with labeled cDNA overnight using standard conditions and washed to a final stringency of 0.2× SSC for 5 min at 65°C.

The probe failed to detect signal in human liver, fetal lung, and fetal kidney, but did detect message in lymphocytes (data not shown).

Figure 9 shows hybridization of BC22 to DNA samples from a number of different organisms. Hybridization signals were observed with all organisms with the exception of Drosophila melanogaster. Since this blot was washed under very stringent conditions (final wash in 0.2× SSC at 65° C for 5 min), cross-hybridization may be observed in Drosophila under less stringent conditions. However, the high stringency of the final wash does indicate the highly conserved nature of this sequence, particularly in Caenorhabditis elegans.

A repeated DNA sequence is found close to the 5′ end of BC72 with 28 CGG triplets interspersed with two AGG triplets. In the predicted open reading frame, this repeat would generate a protein domain composed of 30 contiguous arginine residues. Homology searches with the predicted protein sequence identify significant overlaps with a number of arginine-rich proteins, although none contain a polyarginine stretch of equivalent length. The remainder of the protein shows no significant homology in protein data base searches. However, searches against DNA sequence data bases identify several related sequences, the strongest of which is with the human androgen receptor (AR). This is an X-linked gene (mapping to Xq12) with an identical, though smaller, CGG repeat in the first exon that encodes a polyglycine stretch (Tilley et al., 1989).

Location of the *FMR-1* Gene Relative to the Fragile X-Related CpG Island and FXBCR

BC22 demonstrates hybridization to the 70 kb fragment of YAC 209G4 between BssHII sites b and c as well as to cosmids 4.1, 34.4, 31.4, and 22.3 (see Figure 3), indicating exons spanning over 80 kb of DNA. The proximal/distal orientation of the transcript was determined by hybridizing end fragments of BC22 to the cosmid contig. Since the 3′ end of BC22 detected cosmid 4.1 and the 5′ end detected

cosmid 22.3, the transcriptional orientation was distal from BssHII site b toward the Xq telomere. This suggests the potential involvement of the fragile X–related CpG island in the regulation of this gene. A 1 kb 5′ fragment of BC72 (to the HindIII site at position 1026 of Figure 7) was used to study the location of the exons encoding this portion of the mRNA in the cosmid and YAC clones. In cosmid 22.3, this probe identifies three EcoRI fragments (see Figure 5A) distal to the BssHII site b. One of the fragments contains the BssHII site (b) as well as the breakpoint cluster region and exhibits length variation in fragile X patients. Restriction mapping and direct sequencing of the 5.1 kb EcoRI fragment using a primer derived from BC72 sequence (position 223 to 246) demonstrated the presence of an exon immediately distal to the BssHII site b. This exon contains an XhoI site (position 137 in the *FMR-1* cDNA sequence) that is found 310 nucleotides from the BssHII site in genomic DNA (see Figure 5B), allowing the 5′ end of BC72 to be positioned 220 bp from the BssHII site. This exon, within a 1 kb PstI fragment (Figure 5B), also contains the block of CGG repeats that are seen in the sequence analysis of the genomic DNA as well. This PstI fragment has been found to contain two translocation breakpoints (micro21D and Q1X) distal to the BssHII site (b) as well as to be the region of length variation observed in fragile X chromosomes (Yu et al., 1991). Therefore, the CGG repeat is found within the fragile X–related CpG island, constituting a portion of this CpG-rich region, and is in the immediate proximity of the breakpoint cluster region and the region of length variation.

Discussion

We report the YAC cloning of 475 kb of human DNA corresponding to the region involved in the fragile X syndrome and show the presence of a gene, designated *FMR-1*, adjacent to the fragile X–related CpG island. It is further demonstrated that within a 7.4 kb EcoRI fragile X genomic fragment, which contains the CpG island and *FMR-1* exonic sequences, is a breakpoint cluster region where the majority of fragile X–associated translocation breakpoints position. This same EcoRI fragment, which is 5.1 kb in normal X chromosomes, undergoes a variable increase in size in the fragile X chromosome. The fragile X variation and breakpoint clustering may involve a highly unusual CGG repeat found within the coding region of *FMR-1* and also present within this same EcoRI fragment, 220 bp distal to the BssHII site, demonstrating fragile X–related methylation.

The data presented above suggest the identification of the region containing the fragile X site. A breakpoint cluster region is demonstrated where the majority of fragile X-associated translocation breakpoints fall. These translocations were identified by the segregation of the marker loci *HPRT* and *G6PD*, which flank the fragile X site (Warren et al., 1987, 1990). Since the translocation breakpoints are flanked by the same polymorphic loci that genetically flank the fragile X syndrome locus (Rousseau et al., 1991; Hirst et al., 1991), the breakpoints were believed to cluster at the fragile X site. We show above that these breakpoints indeed cluster within a 7.4 kb EcoRI fragment. Since almost 70% (11 of 16) of tested translocation breakpoints fall within 7.4 kb, an interval of less than 0.03% of the estimated 30 Mb that separate the marker loci used for selection, the clustering is quite remarkable and is compelling evidence for the presence of the fragile X site within this fragment. This is further supported by the observation that this EcoRI fragment is markedly smaller (5.1 kb) in normal chromosomes, and when the normal fragment (pE5.1) is used as probe on Southern blot analysis of fragile X DNA, significantly increased size variation is observed. This suggests that an insertion and/or amplification event within the normal 5.1 kb fragment occurred in the fragile X chromosome. Finally, in situ hybridization of cloned material from this region maps precisely within the isochromatid gap characteristic of the fragile X site.

Within this 5.1 kb fragment, the fragile X-related CpG island is found (Heitz et al., 1991), which contains a BssHII site hypermethylated in fragile X chromosomes (Vincent et al., 1991; Bell et al., 1991). The correlation of this CpG island and its methylation status also supports the involvement of the 5.1 kb fragment, reported above, in the fragile X syndrome. Indeed, linkage analysis using a highly polymorphic dinucleotide repeat (DXS548) located within 150 kb (Figure 3) of the CpG island demonstrates tight linkage with the fragile X locus (*FRAXA*) without recombination. Such linkage is of significance, as it represents evidence independent of the physical mapping studies and is correlated to the clinical expression of the fragile X syndrome, indicating the close proximity of the fragile X site and a gene involved in phenotype.

Such a gene is described above, designated *FMR-1* (fragile X mental retardation-1). There is compelling evidence that *FMR-1* is important in the phenotypic consequences of inheriting the fragile X locus: first, the 5′ association of *FMR-1* with the CpG island shown to be selectively associated with hypermethylation in the fragile X chromosome; second, the presence of a *FMR-1* exon within the 5.1 kb fragment, which is increased in size in the fragile X chromosome and contains the fragile X-associated breakpoint cluster region; third, the presence of a highly unusual trinucleotide repeat CGG, which is reiterated 30 times in the normal mRNA of *FMR-1*, and the placement of this repeat 250 bp distal to the BssHII site of the CpG island within the 5.1 kb fragment; and, fourth, the expression of *FMR-1* in human brain, as would be expected for an inherited disorder whose major phenotype is mental retardation. Although the finding of *FMR-1* does not preclude the additional involvement of other genes in the fragile X phenotype, *FMR-1* is a likely candidate, particularly since there are no other readily apparent CpG islands in the immediate vicinity (Figure 3 and Heitz et al., 1991).

Further study of *FMR-1* is required to show a definitive involvement with the fragile X syndrome, including currently ongoing expression studies in normal and affected individuals. However, the data presented above clearly indicate this gene to be of considerable interest independent of this involvement. The CGG repeat within the coding region is quite striking since the reading frame of the message would suggest a protein containing 30 contiguous

arginine residues. If this repeat is not processed off, *FMR-1* would encode an exceptionally basic protein. No significant protein matches were found on data base searches, except for the polyarginine stretch found in histones and protamines of different organisms, although the arginine stretches in these proteins are shorter and interspersed with other amino acids (Lee et al., 1987; Martin-Ponthieu et al., 1991). Considering that DNA-binding proteins such as histones and protamines contain arginine stretches of variable length and that a nuclear translocation signal is found, it might well be considered that *FMR-1* has a DNA-binding function in the cell nucleus. *FMR-1* appears to have been strongly conserved through evolution with stringent hybridization through yeast. Hybridization to such organisms should allow the isolation and sequencing of the homologous sequences, which may reveal conserved protein domains within *FMR-1*, indicating regions of functional importance. Also, such homologies, particularly with C. elegans, may allow experimental approaches in other organisms to unravel the function of *FMR-1*.

It is tempting to speculate that the unusual *FMR-1* CGG repeat, found within the same 5.1 kb fragment associated with the breakpoint cluster region and fragile X patient length variation, is a component of the fragile site itself. It has long been speculated that the fragile X site is a repeat of variable length (Nussbaum et al., 1986b; Ledbetter et al., 1986; Warren et al., 1987). This possibility awaits further analysis of the repeat region in normal and fragile X chromosomes as well as the junctional sequences of the translocation chromosomes. If the fragile site is not the CGG repeat, the data presented above strongly suggest that the responsible sequence is adjacent to the repeat within the implicated EcoRI fragment. However, the high CpG density of this repeat may influence local methylation status and if expanded may result in the abnormal methylation levels observed in fragile X chromosomes.

The finding of larger fragment lengths in fragile X chromosomes relative to normal chromosomes does indicate a molecular diagnostic approach that should be superior to either linkage analysis or cytogenetic examination for the fragile X site. In particular, Southern blot hybridization with pE5.1 of DNA digested both with EcoRI and BssHII should document both the methylation status of the BssHII site and the size of the resulting fragment(s).

In summary, we report the presence of a brain-expressed gene (*FMR-1*) containing a CGG repeat that is coincident with a fragile X breakpoint cluster region, which itself exhibits length variation in fragile X chromosomes. The breakpoint cluster region and the interval of length variation are likely to contain the sequence responsible for the cytologic expression of the fragile X site. *FMR-1* exonic sequences, particularly the CGG repeat, also map to this same interval and therefore are likely related in some manner. Although the exact nature of the fragile X site, the mechanisms of the mutation, and phenotypic consequences remain to be elucidated, these data provide the elements needed to now explore the molecular biology, biochemistry, and cell biology of this unique and puzzling genetic disease.

Experimental Procedures

PFG Electrophoresis

Southern blot analysis of genomic DNA or YAC DNA resolved by PFG electrophoresis was performed essentially as described (Smith et al., 1988). Trypsinized and washed mammalian cells were suspended in molten agarose (final concentration 0.5% w/v; Baker) prepared in SE buffer (75 mM NaCl, 25 mM EDTA [pH 8.0]) at a final concentration of 1.5×10^7 cells/ml. Chromosomal DNAs were isolated from YAC clones as described (Anand et al, 1989). Yeast cells from a 10 ml saturated culture were harvested, rinsed once in 50 mM EDTA (pH 8.0), and recovered in 0.5 ml of SBE-zymolase (1 M sorbitol, 25 mM EDTA [pH 8.0], 14 mM 2-mercaptoethanol, 1 mg/ml zymolase [ICN]); 0.5 ml of 1% seaplaque agarose (FMC Corp.) in SBE (without zymolase) was added, and the suspension was transferred to plug molds. Spheroplast generation (for yeast cells) was for 5 hr to overnight in SBE-zymolase. Cell lysis (mammalian or yeast cells) was for 2 days in ESP (0.5 M EDTA [pH 9.5], 1% N-lauroylsarcosine, 1 mg/ml proteinase K) at 50°C. Restriction endonuclease digestion was performed using the manufacturer's recommended buffers and conditions with a 50 µl plug slice in 250 µl of buffer containing 50 U of enzyme. For double digests, the plugs were rinsed and equilibrated, following digestion with the first enzyme, with the second buffer several times prior to digestion with the second enzyme. PFG electrophoresis was carried out on a Bio-Rad contour-clamped homogeneous electric field DRII apparatus through 1% agarose (Bethesda Research Laboratories) at 200 V and 14°C in 0.5× TBE buffer (45 mM Tris–borate, 1 mM EDTA). For resolution of fragments of ~200–1200 kb, switch time was 60 s for 17 hr followed by 90 s for 10 hr; for resolution of fragments of ~10–500 kb, the switch times were ramped from 5 s to 50 s over 27 hr. Southern blotting and hybridization were carried out as described (Southern, 1975; Feinberg and Vogelstein, 1983) with the exception that acid depurination in 0.25 M HCl was allowed to proceed 20 min for PFGs. Radiolabeled probes were synthesized by random priming from 50 ng of gel-purified fragments, except when intact cosmids were used that were nick translated (Boehringer Mannheim kit; following manufacturer's recommendations). For genomic probes containing repetitive elements, repeat suppression was accomplished by preassociation with 1–3 mg of sonicated human placental DNA in 100–300 µl of 5× SSC (1× SSC is 150 mM NaCl, 15 mM sodium citrate [pH 7.0]) for 3–10 min at 65°C prior to addition to the filter. Washing was carried out to a final stringency wash of 0.2× SSC for 15 min at 65°C prior to autoradiography. Saccharomyces cerevisiae strain YNN295 chromosomes (Bio-Rad), concatamers of phage λ (Bio-Rad), or high molecular weight markers (BRL) were used as size standards.

Cosmid Library Construction of YAC 209G4

Agarose plugs (0.5%, seaplaque; FMC) containing 5–10 µg of yeast DNA were prepared as described (van Ommen and Verkerk, 1986). Blocks (100 µl) of DNA were equilibrated on ice in 0.5 ml of MboI digestion buffer, containing 0.1 mg/ml bovine serum albumin (MB grade; Boehringer Mannheim). After 2–3 hr, the buffer was replaced by 150 µl of fresh buffer, to which MboI was added (0.0001–0.0007 U). Following overnight incubation on ice, digestion was carried out for 40 min at 37°C. The agarose blocks were melted, and the DNA was dephosphorylated with 1 U of calf intestinal alkaline phosphatase (Boehringer Mannheim) and treated with 2.5 U of agarase (Calbiochem). The solution was extracted twice with phenol–chloroform, once with chloroform, and the DNA was precipitated with ethanol and dissolved in 10 mM Tris, 0.1 mM EDTA (pH 7.4) at a concentration of 500 ng/µl. DNA (250 ng) was ligated to 500 ng of BstBI (dephosphorylated) and BamHI-digested vector (p2CpG; Dauwerse et al., 1989). Ligation and packaging were carried out according to standard procedures. Cosmids containing human inserts were selected by hybridizing with human-specific Alu repeat probe.

YAC and Cosmid Subcloning

YACs were subcloned following isolation of the intact chromosome by preparative PFG electrophoresis and EcoRI digestion of the DNA in molten agarose (seaplaque; FMC). Fragments were phenol–chloroform extracted, ethanol precipitated, recovered, and ligated into EcoRI cut, dephosphorylated λ ZAP II arms according to manufacturer's rec-

ommendations (Stratagene). Cosmids were subcloned following an alkaline lysis isolation and EcoRI digestion. Fragments were phenol–chloroform extracted and ethanol precipitated prior to ligation into λ ZAP II arms as with YAC fragments. In the case of both cosmids and YACs, 75 ng EcoRI fragments were ligated to 1 μg vector arms. Selected phage were converted into pBluescript II SK⁻ clones following in vivo excision of plasmid with insert, according to manufacturer's guidelines.

Polymerase Chain Reaction Analysis of DXS548 Alleles

Analysis of DXS548 was carried out as described (G. J. R. et al., unpublished data). Amplification was carried out on 0.2–0.5 μg of genomic DNA in a 10 μl total reaction containing 0.25 mM dNTPs, 40 ng of primers 5'-AGAGCTTCACTATGCAATGGAATC and 5'-GTACATTAGAGTCACCTGTGGTGC, and 0.25 U of Taq polymerase in a buffer of 10 mM Tris–HCl, 50 mM KCl, 12 mM MgCl, and 0.01% gelatin. Twenty-three cycles of polymerase chain reaction (PCR) were carried out in the following fashion: three cycles of 1 min each at 97°C, 62°C annealing, and 72°C extension followed by 20 additional cycles with the annealing temperature lowered to 55°C. The reaction volume was then increased to 50 μl with the same reaction components and concentrations except that one primer was 5' end labeled with [γ-^{32}P]ATP as described (Sambrook et al., 1989). PCR was continued for 10 cycles of 1 min each at 95°C denaturation, 62°C annealing, and 72°C extension. PCR products were analyzed by electrophoresis of 2 μl of reaction through a 40 cm 6% polyacrylamide denaturing sequencing gel for approximately 2.25 hr. The gel was dried without fixing and exposed to X-ray film overnight at room temperature.

cDNA Library Screening

A human fetal brain λgt11 cDNA library (Clonetech, Palo Alto, CA) of 1.3 × 10^6 independent clones with insert lengths of 0.7–4.0 kb was used. The library was plated on 15 cm plates at a density of 50,000 pfu per dish using strain LE392. Filter lifts were prepared according to standard techniques, and the library was screened with cosmid DNA hexanucleotide labeled with [^{32}P]dATP and [^{32}P]dCTP. The labeled DNA was first prehybridized with 100 μg of total sheared human genomic DNA and 100 μg of cosmid vector DNA in 5× SSC at 65°C for 2 hr. Following hybridization for 16 hr, the filters were washed to a stringency of 0.1× SSC. The filters were exposed to Fuji film with intensifying screens for 2 days at −80°C.

Northern Blot Analysis

Total RNA was extracted using guanidinium isothiocyanate followed by centrifugation through cesium chloride as described (Sambrook et al., 1989). Poly(A)⁺ RNA was selected by passage through oligo(dT)–cellulose (Aviv and Leder, 1972). Human brain, liver, and fetal poly(A) RNA was purchased from Clontech Laboratories (Palo Alto, CA).

Five micrograms of poly(A)-containing RNA or 25 μg of total RNA was precipitated and dissolved in 20 μl of 50% (v/v) formaldehyde and 1× MEN (20 mM MOPS [pH 6.8], 5 mM sodium acetate, 1 mM EDTA) and incubated for 10 min at 60°C; 5 μl of dye marker (50% sucrose, 0.5% bromophenol blue) was added, and the samples were loaded on a formaldehyde–agarose gel. Electrophoresis was carried out for 3 hr at 100 V, and the gel was then soaked for 30 min in 20× SSC and blotted onto a nitrocellulose or nylon membrane (GeneScreen Plus, Dupont) overnight in 10× SSC (Thomas, 1980). The RNA was fixed to the membranes by baking under vacuum for 2 hr at 80°C. The membranes were prehybridized in 50% formamide, 5× Denhart's solution, 50 mM sodium phosphate (pH 6.8), 10% dextran sulfate, and 100 μg of denatured salmon sperm DNA at 42°C for 2–4 hr. Hybridization with the probe was for 16–20 hr at 42°C in the above buffer. Filters were washed with 3× SSC, 0.1% SDS at 50°C, and then the SSC concentration was lowered according to the level of background with a final wash in 0.1× SSC, 0.1% SDS.

DNA Sequencing

Sequence of BC22 was obtained via a shotgun strategy as described in Bankier et al. (1987) using dideoxynucleotide termination. The sequencing reactions were analyzed on an automated DNA sequencer (ABI 373) using fluorescently labeled oligonucleotide primers. Sequence information derived from 25 random M13 clones was assembled using the sequence assembly manager software developed by the Molecular Biology Information Resource of the Department of Cell Biology, Baylor College of Medicine.

An EcoRI to HindIII fragment at the 5' end of BC72 was subcloned into Bluescript II KS⁻ and sequenced with vector and internal primers. Sequence at the extreme 5' end is incomplete due to the difficulty of extension by polymerases through the repeated CGG sequence.

Acknowledgments

We thank K. E. Bakker, R. M. v. d. Helm, J. F. Peters, and C. L. Stayton for assistance and helpful discussion; S. Ledbetter and D. Ledbetter for their continuous support; and D. Le Paslier of CEPH (Paris) for the generous gift of the YAC library/clone. G. J. R. is a predoctoral fellow of the March of Dimes Birth Defects Foundation (18–90–57). O. R. is a postdoctoral fellow of the Human Frontier Science Program. This research was supported by the Foundation of Clinical Genetics to A. J. M. H. V.; a grant of the MGCZWN to B. A. O.; a DOE grant (FG05–88ER60692) to D. L. N. and C. T. C.; the Howard Hughes Medical Institute and a National Institutes of Health Genome Center Grant to C. T. C.; and NIH grants (HD20521 and HG00038) to S. T. W.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 USC Section 1734 solely to indicate this fact.

Received May 9, 1991; revised May 16, 1991.

References

Albertsen, H. M., Abderrahim, H., Cann, H. M., Dausset, J., Le Paslier, D., and Cohen, D. (1990). Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. Proc. Natl. Acad. Sci. USA *87*, 4256–4260.

Anand, R., Villasante, A., and Tyler-Smith, C. (1989). Construction of yeast artificial chromosome libraries with large inserts using fractionation by pulsed-field gel electrophoresis. Nucl. Acids Res. *17*, 3425–3433.

Aviv, H., and Leder, P. (1972). Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA *69*, 1408–1412.

Bankier, A. T., Weston, K. M., and Barrel, B. G. (1987). Random cloning and sequencing by the M13 dideoxynucleotide termination method. Meth. Enzymol. *155*, 51–93.

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). Physical mapping across the fragile X: hypermethylation and clinical expression of the fragile X syndrome. Cell *64*, 861–866.

Brown, W. T. (1990). The fragile X: progress towards solving the puzzle. Am. J. Hum. Genet. *47*, 175–180.

Dauwerse, H. G., van Ommen, G. B., Breuning, M. H., and Pearson, P. L. (1989). Cosmid vector pCpG and plasmid vector pKNUNI, and their uses for cloning DNA sequences adjacent to sites for rare cutting restriction endonucleases. Nucl. Acids Res. *17*, 3603.

Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction fragments to high specific activity. Anal. Biochem. *132*, 6–13.

Gustavson, K.-H., Blomquist, H., and Holmgren, G. (1986). Prevalence of fragile-X syndrome in mentally retarded children in a Swedish county. Am. J. Med. Genet. *23*, 581–588.

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

Hirst, M. C., Roche, A., Flint, T. J., MacKinnon, R. N., Bassett, J. H. D., Nakahori, Y., Watson, J. E. V., Bell, M. V., Patterson, M. N., Boyd, Y., Thomas, N. S. T., Knight, S. J. L., Warren, S. T., Hors-Cayla,

M., Schmidt, M., and Davies, K. E. (1991). Linear order of new and established markers around the fragile site at Xq27.3. Genomics *10*, 243–249.

Krawczun, M. S., Jenkins, E. C., and Brown, W. T. (1985). Analysis of the fragile-X chromosome: localization and detection of the fragile site in high resolution preparations. Hum. Genet. *69*, 209–211.

Ledbetter, D. H., Ledbetter, S. A., and Nussbaum, R. L. (1986). Implications of fragile X expression in normal males for the nature of the mutation. Nature *324*, 161–163.

Lee, C.-H., Hoyer-Fender, S., and Engel, W. (1987). The nucleotide sequence of a human protamine 1 cDNA. Nucl. Acids Res. *15*, 7639.

Lindsay, S., and Bird, A. P. (1987). Use of restriction enzymes to detect potential gene sequences in mammalian DNA. Nature *327*, 336–338.

Martin-Ponthieu, A., Wouters-Tyrou, D., Belaiche, D., Sautiere, P., Schindler, P., and Van Dorsselaer, A. (1991). Cuttlefish sperm protamines. Eur. J. Biochem. *195*, 611–619.

Nelson, D. L., Ballabio, A., Victoria, M. F., Pieretti, M., Bies, R. D., Gibbs, R. A., Maley, J. A., Chinault, A. C., Webster, T. D., and Caskey, C. T. (1991). Alu PCR for regional assignment of 110 yeast artificial chromosome clones from the human X chromosome: identification of clones associated with a disease locus. Proc. Natl. Acad. Sci. USA, in press.

Nussbaum, R. L., and Ledbetter, D. H. (1990). The fragile X syndrome. In The Metabolic Basis of Inherited Disease, C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, eds. (New York: McGraw Hill), pp. 327–341.

Nussbaum, R. L., Airhart, S. D., and Ledbetter, D. H. (1986a). A rodent-human hybrid containing Xq24-qter translocated to a hamster chromosome expresses the Xq27 folate-sensitive fragile site. Am. J. Med. Genet. *23*, 457–466.

Nussbaum, R. L., Airhart, S. D., and Ledbetter, D. H. (1986b). Recombination and amplification of pyrimidine-rich sequences may be responsible for initiation and progression of the Xq27 fragile site. Am. J. Med. Genet. *23*, 715–721.

Oostra, B. A., Hupkes, P. E., Perdon, L. F., van Bennekom, C. A., Baker, E., Halley, D. J. J., Schmidt, M., Du Sart, D., Smits, A., Wieringa, B., and van Oost, B. A. (1990). New polymorphic DNA marker close to the fragile site FRAXA. Genomics *6*, 129–132.

Roberts, B. (1989). Nuclear location signal-mediated protein transport. Biochim. Biophys. Acta *1008*, 263–280.

Rousseau, F., Vincent, A., Rivella, S., Heitz, D., Tribioli, C., Maestrini, E., Warren, S. T., Suthers, G. K., Goodfellow, P., Mandel, J.-L., Toniolo, D., and Oberle, I. (1991). Four chromosomal breakpoints and four new probes mark out a 10 cM region encompassing the FRAXA locus. Am. J. Hum. Genet. *48*, 108–116.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 7.19–7.22 and 10.59–10.66.

Sherman, S. L., Morton, N. E., Jacobs, P. A., and Turner, G. (1984). The marker (X) syndrome: a cytogenetic and genetic analysis. Ann. Hum. Genet. *48*, 21–37.

Sherman, S. L., Jacobs, P. A., Morton, N. E., Froster-Iskenius, U., Howard-Peebles, P. N., Nielsen, K. B., Partington, N. W., Sutherland, G. R., Turner, G., and Watson, M. (1985). Further segregation analysis of the fragile X syndrome with special reference to transmitting males. Hum. Genet. *69*, 3289–3299.

Smith, C. L., Klco, S. R., and Cantor, C. R. (1988). Pulsed-field gel electrophoresis and the technology of large DNA molecules. In Genome Analysis: A Practical Approach, K. E. Davies, ed. (Oxford: IRL Press), pp. 41–72.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol. Biol. *98*, 503–517.

Sutherland, G. R., and Hecht, F. (1985). Fragile Sites on Human Chromosomes (New York: Oxford University Press), p. 53.

Thomas, P. S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA *77*, 5201–5205.

Tilley, W. D., Marcelli, M., Wilson, J. D., and McPhaul, M. J. (1989). Characterization and expression of a cDNA encoding the human androgen receptor. Proc. Natl. Acad. Sci. USA *86*, 327–331.

van Ommen, G. J. B., and Verkerk, J. M. H. (1986). Restriction analysis of chromosomal DNA in a size range up to two million base pairs by pulsed-field gradient electrophoresis. In Human Genetic Diseases: A Practical Approach, K. E. Davies, ed. (Oxford: IRL Press), pp. 113–133.

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

Warren, S. T., Zhang, F., Licamelli, G. R., and Peters, J. F. (1987). The fragile X site in somatic cell hybrids: an approach for molecular cloning of fragile sites. Science *237*, 420–423.

Warren, S. T., Zhang, F., Sutcliffe, J. S., and Peters, J. F. (1988). Strategy for the molecular cloning of the fragile X site DNA. Am. J. Med. Genet. *30*, 613–623.

Warren, S. T., Knight, S. J. L., Peters, J. F., Stayton, C. L., Consalez, G. G., and Zhang, F. (1990). Isolation of the human chromosomal band Xq28 within somatic cell hybrids by fragile X site breakage. Proc. Natl. Acad. Sci. USA *87*, 3856–3860.

Webb, T. P., Bundey, S. E., Thake, A. I., and Todd, J. (1986). Population incidence and segregation ratios in the Martin-Bell syndrome. Am. J. Med. Genet. *23*, 573–580.

Wolf, S. F., Jolly, D. J., Lunnen, K. D., Axelman, J., and Migeon, B. R. (1984). Methylation of the hypoxanthine phosphoribosyltransferase locus on the human inactive X chromosome: implications for X chromosome inactivation. Proc. Natl. Acad. Sci. USA *81*, 2806–2810.

Yen, P. H., Patel, P., Chinault, A. C., Mohandas, T., and Shapiro, L. J. (1984). Differential methylation of hypoxanthine phosphoribosyltransferase genes on active and inactive human X chromosomes. Proc. Natl. Acad. Sci. USA *81*, 1759–1763.

Yu, S., Kremer, E., Pritchard, M., Lynch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, J. C., Warren, S. T., Schlessinger, D., Sutherland, G. R., and Richards, R. I. (1991). The fragile X genotype is characterized by an unstable region of DNA. Science, in press.