Molecular cloning and analysis of the fragile X region in man

Alexander Dietrich, Petra Kioschis, Anthony P. Monaco^{1,*}, Baerbel Gross, Bernhard Korn, Sarah V. Williams¹, Denise Sheer¹, Dominique Heitz², Isabel Oberle², Daniela Toniolo³, Stephen T. Warren⁴, Hans Lehrach¹ and Annemarie Poustka*

Deutsches Krebsforschungszentrum, Institut fuer Virusforschung, Im Neuenheimer Feld 506, D-6900 Heidelberg, FRG, ¹Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK, ²Laboratoire de Genetique Moleculaire des Eucaryotes, Unite de Biologie Moleculaire et de Genie Genetique U184, 11 rue Humann, 67085 Strasbourg Cedex, France, ³Istituto di Genetica Biochimica Ed Evoluzionistica, Via Abbiategrosso 207, Pavia, 27100, Italy and ⁴Departments of Biochemistry and Pediatrics, Division of Medical Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA

Received April 5, 1991; Accepted April 18, 1991

ABSTRACT

The fragile X syndrome (FraX), the most common inherited form of mental retardation, has been located to Xq27.3. As a step in the molecular analysis of this mutation, we have cloned a contiguous 1.8 Mb region containing the entire fragile X region in YAC and cosmid clones. The cloned area defines a region of 50 kb containing a CpG island, found to be selectively methylated in patients expressing the fragile X phenotype. In this 50kb area we have localised the breakpoints of four somatic cell hybrids selected to break at the position of the fragile site. Fluorescence in-situ hybridisation of cosmids flanking this area shows that the breakpoints, the CpG island and the fragile site coincide.

INTRODUCTION

The fragile X mental retardation syndrome is the most frequent inherited form of mental retardation and the second most common cause of mental retardation after Downs syndrome (1). It is characterised cytogenetically by the presence of a fragile site at the position Xq27.3, which can be induced in-vitro by growth of cells under conditions reducing the availability of precursors for DNA synthesis (2). The relationship between the fragile site and the syndrome is not understood, and it is not clear, if the position defining the region of fragility will coincide with the position of the mutation, which can be localised by the analysis of recombination events. The position of the mutation, in turn, can be different from the position or positions of the genes, which are involved in the expression of the phenotype of this disease (mental retardation, macroorchidism, facial abnormalities).

The FraX syndrome is unusual in showing partial penetrance in both males and females. Especially striking is the observation of phenotypically normal males, who transmit the mutation to daughters, who generally again show no phenotype, while their grandchildren show normal penetrance of the disorder (3). To explain some of the features of the disease and its unusual pattern of inheritance, a number of models have been proposed, which might become testable during the molecular analysis of the region surrounding the mutation. First indications of selective methylation of sequences close to the fragile site (4,5) can be taken as evidence in favor of a model proposed by Charles Laird (6,7), which postulates a local defect in the reactivation of inactivated X chromosomes as the primary cause of mutation, leading to a region of late replication, and therefore to the appearance of a fragile site close to the position of the mutation.

MATERIALS AND METHODS

Derivation of probes

Probes VK21, Do33 and 2.34 have been described before (8,9). All other probes were derived during the course of this work. A12 is a 900bp HindIII fragment isolated from cosmid A12c. 35B-2.35kb EcoRI-SalI, 6.2-2.1kb EcoRI-SalI, BM-3kb EcoRI, BN-5kb EcoRI are fragments cloned from phages, derived by cloning YF and Y16 in Embl3. 45B-400bp, 391-2.7kb, 189-1.25kb and 191-900bp are fragments cloned by vectorette-PCR (10) using vectorette oligonucleotides AM1089 and AM1090, and adaptors for the restriction sites: BamHI, BclI, BglI, NheI, SpeI, StyI and XbaI. Restriction enzymes and ligase were obtained from NEB. Taq DNA polymerase was obtained from Elmer Cetus.

* To whom correspondence should be addressed

+ Present address: Institute for Molecular Medicine-ICRF, John Radcliffe Hospital, Oxford OX3 9DU, UK

Pulsed field gel electrophoresis

The PFG map of the genomic DNA was determined as described in Poustka et al. (11). DNA fragments were separated in a Pulsaphor Chef apparatus (Pharmacia-LKB). Exact running conditions are given in the figure legends. Chromosomes from *S. pombe*, *S. cerevisiae* and *Candida albicans* were used as size markers. Restriction enzymes were obtained from Boehringer Mannheim (NotI, MluI, NruI, BssHII, SacII, NaeI) and Amersham (SplI).

Isolation of YAC clones

YAC clones were obtained by filter hybridisation from a YAC library constructed from GM1416B DNA (12). Restriction maps

of YAC clones were determined by digestion of YAC DNA prepared in agarose blocks and detection of fragments by hybridisation with vector DNA, human repeat sequences and the specific probes described above. Filter transfer and hybridisation was performed as in Herrmann et al. (13).

Cosmid isolation and analysis

Cosmids were isolated from a library from flow sorted X chromosomes from the cell line GM1416B (normal X chromosome) (14), as well as a library constructed from the cell line Q1Z (fragile X chromosome) by isolating clones hybridising to Alu repeat sequences (A.P. unpublished) (Fig.3b). For chromosome walking whole cosmids were labelled and hybridised

Figure 1. Rare cutting enzyme restriction map of genomic DNA (cell line GM1416B) and the YAC contig covering the fragile X region. Restriction sites observed both in chromosomal DNA and in cloned DNA are shown only in the map of the chromosome. Only a subset of sites for the enzymes NruI, NaeI and SalI is shown. Positions of probes used are indicated. The crosshatched area between Y47 and Y1 shows the region of YAC clone overlap. The hatched box at the bottom indicates the region covered by cosmid clones.

Figure 2. a) Hybridisation of DNA of YAC1 (size 380 kb) digested with 1(NotI), 2(NruI), 3(SacII), 4(BssHII) and 5(undigested) to probe 2.34 (left panel) and probe 189 (right panel). Fragment sizes: Left panel: NotI-380 kb (undigested), NruI-360 kb, SacII-170 kb, BssHII-300 kb, undigested-380 kb (Yac1 does not contain a NotI restriction site. Therefore lane 1 and lane 5 show bands of identical size.) Right panel: all enzymes except NotI give a similar sized band of app. 20 kb. DNA fragments of up to 500 kb were separated by electrophoresis for 20 hours at 200 Volts with a switching time of 40 seconds in a Pulsaphor Chef apparatus (Pharmacia-LKB). b) Hybridisation of Alu repeat probe and specific probes to Southern blots of EcoRI digested YAC DNA. Lane 1 (YF), lane 2 (Y47), lane 3 (Y1), lane 4 (Y2), lane 5 (Y16). Alu fingerprinting shows that YF and Y47 and Y1, Y2 and Y16 overlap to a large extent. Probe A12 recognises YF and Y47 both coming from the distal side (both EcoRI fragments were not seen in the Alu pattern). Probe 189 detects the common fragment between Y47 and Y1 (also not observed in the Alu pattern), while probe 2.34 (the sequence on the proximal side) hybridises to Y1, Y2 and Y16.

as described in Herrmann et al. (15) to the high density gridded libraries. Filters were prehybridised for four hours in 100 ug sheared human placenta DNA, 60 ug vector DNA and 60 ug t-RNA per 10 ml of hybridisation solution.

Fluorescence in-situ hybridisation

Fragile sites were induced by incubation with excess thymidine for 24 hours before harvest. Fluorescence in-situ hybridisation was carried out as described (16) with the addition of sonicated total human DNA to the reaction as competitor. The signal was detected using FITC conjugated avidin. Chromosomes were counterstained with propidium iodide.

RESULTS AND DISCUSSION

Isolation of Yac contig and verification by PFG analysis

We have recently established a 12 Mb PFG map over the entire Xq28 region (17) using somatic cell hybrids generated by X chromosome breakage following induction of the fragile site (18,19). The breakpoints of these hybrids were localised on this map in a region of 700 kb between the probes VK21 (DXS296) (8) and 2.34 (DXS477) (9). To investigate if this region coincides with the fragile site we have used these probes as well as Do33 (DXS465) (9), A12 and 391 (B.G. and B.K. unpublished)(Fig.1) located close to the fragile X locus to isolate clones from a YAC

Figure 3. a) Localisation of the fragile X associated cell line breakpoints between probe 189 and probe 6.2. GM1416B: lymphoblastoid (48,XXXX) cell line obtained from Human Genetic Mutant Cell Repository, Camden, New Jersey, 578: human X chromosome on hamster (24), Y2I: hamster cell line (18), micro 21D: translocation chromosome Xpter-q27.3, QIZ, QIV, QIAD and h-g+: the reciprocal translocations Xqter-q27.3. Left panel: Genomic DNA of different cell lines digested with PstI hybridised to probe 189 shows that 189 is contained in all Xqter-q27.3 hybrids except in QIAD which breaks between probe 35B and 189 (Fig. 3b), but is not contained in micro 21D, the Xpter-q27.3 cell line. Right panel: EcoRI digested DNA of the same cell lines hybridised to probe 6.2 shows hybridisation to micro21D and not to any of the distal cell lines. b) Schematic drawing of the area surrounding the CpG island found to be methylated in patients expressing the FraX phenotype. Localisation of breakpoints with respect to the newly isolated probes (Fig. 1, 3a) and cosmids isolated in this region in two different libraries. For chromosome walking cosmids were isolated with the two outmost markers (6.2, A12) and then whole cosmids were hybridised back to the two libraries. Orientation was determined by rescreening with the existing cloned probes (top line) and with whole cosmids. Maps were derived by the information from single and double digest experiments. A gap was found between cosmid B2 and A12 closed by cosmid IV from the GM1416B library. No connection to A6 was found in the QIZ library, most likely due to lack of Alu repeats in that region.

Figure 4. The position of cosmids IV, A6 and 7172 (Fig. 3b) relative to the fragile site are shown by fluorescence in situ hybridisation to X chromosomes from fragile X patient cell lines induced to express the fragile site. The centromere probe pSV2X5 (25) allows identification of the X chromosome. The positions of the centromeres are marked by arrowheads, the unique probes are marked by arrows. Top: X chromosomes from different metaphases hybridised to cosmid 7172 : (left and middle: proximal, right: distal hybridisation pattern) Bottom: Hybridisation with 7172 (left), A6 (middle), and IV (right) (distal patterns). Cosmid 7172 therefore hybridises in different metaphases either at a proximal, or at a distal location, with a majority of the metaphases (app. 60%) showing a proximal hybridisation pattern. Cosmids A6 and IV never showed proximal hybridisation. Identical results were obtained in cell lines from two different patients (GM7730A and GM5025B, Camden Mutant Cell Repository).

library constructed from DNA from GM1416B, a lymphoblastoid cell line containing four X chromosomes (12). To verify the fidelity of the clone coverage, and to rule out the products of possible rearrangements or coligation events from further analysis, restriction maps of the isolated clones were determined using enzymes cutting rarely in mammalian DNA. To be able to compare the structure of the clones with that of the genomic DNA, a precise PFG map was also established in GM1416B, the cell line used to construct the YAC library (Fig.1). This comparison is especially useful, since the DNA of GM1416B is relatively undermethylated, in contrast to both leukocyte DNA as well as the DNA of the cell hybrids used by us in establishing the map of the entire region (17), allowing an easier comparison to the map constructed on (inherently non-methylated) YAC clone DNA. Four overlapping YAC clones (Y4, Y3, YF, Y47) were identified by probes located distal to the FraX region. Three YAC clones (Y2, Y1, Y16) were found by the proximal probes. Most isolated YAC clones ended close to the position of a CpG island (Fig.1), which contains sites for the restriction enzymes *NarI*, *NaeI*, *NruI*, *SacII* and *BssHII*, and has been found (4,5) to be selectively methylated in mentally retarded FraX males.

PFG analysis of Y1 showed that this clone extended beyond the position of the CpG island by app. 23 kb (Fig. 2a). To find out if the clones hybridising to the probes distal of the island and clones coming from the proximal side overlap, Southern blots of YAC DNA digested with *EcoRI* were hybridised to an Alu repeat probe (20) as well as to probes from both sides of the FraX region (Fig. 2b). This experiment shows that groups of YAC clones overlap, but did not yet detect the overlap between the proximal and the distal cluster of YAC clones. Such an

overlap could however be identified by a hybridisation with the probe 189 (B.K. unpublished), the end of Y1, a 1.25 kb fragment amplified by the PCR-Vectorette system (10), which detects a common band between the clones Y47 and Y1.

Isolation of cosmids

To improve the level of resolution of the clones we covered the more immediate area of the CpG island by a cosmid walk of 250 kilobases (Fig. 3b). Cosmids were isolated from an X chromosome specific library constructed from the cell line GM1416B (14) as well as from an Xq28 specific cosmid library constructed from the cell line Q1Z (A.P., unpublished), one of the cell hybrids selected to break at the position of the fragile site (19), and translocated onto a rodent chromosome.

Localisation of cell line breakpoints

From the PFG analysis and from the maps of the YAC and cosmid clones it became clear, that the probes Do33 and 2.34 previously shown to map closest to the fragile X mutation by genetic and physical mapping techniques (9) were separated by at most 300 kb and that they were likely to flank the breakpoints of the cell hybrids selected to break at or close to the fragile site. We therefore used probe 189 (the end of Y1) and probe 6.2 (end of Y16) flanking the CpG island to hybridise Southern blots of restriction digests of the FraX hybrids (18,19). The results (Fig. 3a, 3b) demonstrate the very close clustering of the breakpoints. While four out of five hybrids (Q1Z, Q1V, h-g+, micro-21D), which were all derived in independent experiments break between probes 189 and 6.2, the breakpoint of hybrid Q1AD maps more distally, between 35B and 189 (data not shown).

Figure 5. Selective methylation of rare cutter restriction sites in the CpG island within the breakpoint region in a FraX patient. a) Analysis of DNA from normal male (lane 1) and FraX patient (patient 15, lane 2) digested with enzymes NaeI (A), BssHII (B) and SacII (C) hybridised with probe 189 located distal to the CpG island. The normal male shows a 600 kb band, which is not observed in the patient DNA. PFG runs were performed in $0.25 \times$ TBE, 1% agarose, at 150 V with a pulse time of 90 seconds. Gels were run for 40h at 12 deg. C. b) Analysis of BssHII digested DNA from three normal males (lanes 1,3,4) and the same patient as above (lane 2) using PFG conditions resolving larger fragments ($0.5 \times$ TBE, 0.8% agarose, pulse time 10–20 min. for 3 days at 70V and 12 deg. C). Both the proximal probe (6.2) and the distal probe (189) identify the same large band of app. 2.7 Mb in the patient DNA, corresponding in size to the sum of the bands observed in the DNA of normal males.

Since both the majority of the cell lines containing Xq27.3-qter (Q1V, Q1Z and h-g+, distal hybrids) and the cell line extending from Xq27.3 to pter (micro-21D, proximal hybrid) are found to break within a distance of app. 50 kb it is likely, that this interval defines the region of chromosomal fragility associated with the fragile site. This assumption is exciting, since the close proximity of the breakpoints of distal hybrids with the breakpoint in the proximal hybrid seems to rule out the possibility of exonucleolytic degradation of the chromosome fragments before their ligation to hamster chromosomes.

Mapping of the fragile site by in-situ hybridisation experiments

To show that this interval is very close to the position of the fragile site, we carried out fluorescence in-situ hybridisation using cosmids from the region (Fig. 3b) to metaphase chromosomes from FraX patient cell lines induced to express the fragile site. This cytogenetic analysis positions cosmid A6 distal (Fig. 4), while cosmid 7172 is found either proximal or distal to the

cytogenetic fragile site in different metaphases of two different patients analysed (Fig. 4). This localisation of the cytogenetically observed fragile site to this region has also been verified in parallel experiments by Heitz et al. (21), in which a YAC clone spanning part of the region described here was found to hybridise to both sides and throughout the fragile site.

Selective methylation of sites in the breakpoint region

As shown in figure 5, this short region defined by the cell hybrid breakpoints might be directly implicated in the expression of the FraX phenotype, since a number of sites in the CpG island (BssHII, SacII, NaeI) are highly methylated in fragile X patients (15 out of 16). While control males show shorter restriction fragments with the either distal (189) or proximal probes (6.2) (Fig. 5a,5b), a common large fragment corresponding to the sum of the two shorter fragments can be observed in FraX patients. The observation of selective methylation of sites in the CpG island is in good agreement with the hypothesis of Laird (6), proposed to explain the unusual pattern of inheritance of this mutation. Since methylation of sites in this CpG island has been found to be absent in males which carry the mutation, but do not express the FraX phenotype, it is tempting to speculate, that some or all phenotypic consequences of the disease could be due to the influence of this methylation on the expression of genes located in the area.

High resolution comparison of the FraX and wild-type form of the region

Since the high rate of mutation postulated for FraX (22,23) seems to make its formation by point mutations unlikely, we carried out a detailed comparison of sequences of the FraX and wild type forms of a large part of the region. A 170 kb region distal to the CpG island was analysed by comparing restriction digests of clones isolated from libraries GM1416B (normal chromosome) and QIZ respectively (FraX chromosome) (Fig. 3b). In addition the sequences spanned by the cosmids XII, 494 and 7172 were analysed. These experiments were carried out by hybridising radiolabelled cosmids prehybridised to total human DNA to suppress hybridisation of repeat sequences to Southern blots of normal and patient DNA restricted with EcoRI or PstI (data not shown). No obvious differences were found, though differences in the low molecular weight bands, as well as in the remaining gap in the cosmid clone coverage can not be ruled out. Results up to now indicate the absence of larger insertions or deletions in this area.

Though this analysis has therefore been able to localise the area involved in the formation of the fragile site, and has identified a selectively methylated region likely to be involved either directly or indirectly in the fragile site formation, we have not yet localised the position of the fragile X mutation itself. Analysis of the region in FraX and wild-type chromosomes does however indicate the absence of major rearrangements (insertions or deletions larger than a few kb) within a large fraction of the region of 250 kilobases surrounding the methylated CpG island. If verified over the entire region, this would either indicate a smaller mutation, or could be compatible with a mutation caused by a larger rearrangement located outside the region analysed.

ACKNOWLEDGEMENTS

We thank Grant Sutherland for providing VK21, Gaby Langenstein for excellent technical help, Mark Ross and Sebastian Meier-Ewert for high density filters of the YAC library, Dean

Nizetic, Guenther Zehetner and Christal Douglas for filters of the X specific cosmid library. We are very grateful to Prof. Bullerdiek and Dr. Albrecht for bloodsamples of FraX patients. A.P.M was supported by a fellowship of the muscular dystrophy association of America. This work was partly supported by a NIH grant to S.T.W., a grant of the Hereditary Disease foundation to H.L. and a grant from the Deutsche Forschungsgemeinschaft to A.P.

REFERENCES

1. Nussbaum, R. L. & Ledbetter, D. H. (1986) *Ann. Rev. Genet.* **20**, 109–145
2. Sutherland, G. R. (1979) *Am. J. Med. Genet.* **31**, 125–135
3. Sherman, S. L., Jacobs, P. A., Morton, N. E., Froster-Iskenius, U. & Howard-Peebles, P. N. (1985) *Hum. Genet.* **69**, 289–299
4. Vincent, A., Heitz, D., Petit, C., Kretz, C., Oberle, I., Mandel, J. L. (1991) *Nature* **349**, 624–626
5. Bell, M. V., Hirst, M. C., Nakahori, Y., MacKinnon, R. N., Roche, A., Flint, T. J., Jacobs, P. A., Tommerup, N., Tranebjaerg, L., Froster-Iskenius, U., Kerr, B., Turner, G., Lindenbaum, R. H., Winter, R., Penbrey, M., Thibodeau, S., Davies, K. E. (1991) *Cell*, **64**, 861–866
6. Laird, C. D. (1987) *Genetics* **117**, 587–599
7. Laird, C. D., Lamb, M. M., Thorne, J. L. (1990) *Am. J. Hum. Genet.* **46**, 696
8. Suthers, G. K., Callen, D. F., Hyland, V. J., Kozman, H. M., Baker, E., Eyre, H., Harper, P. S., Roberts, S. H., Hors-Cayla, M. C., Davies, K. E. (1989) *Science* **246**, 1298–1300
9. 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. (1990) *Am. J. Human. Genet.* in press
10. Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J. C., Markham, A. F. (1990) *Nucleic Acids Res.* **18**, 2887–2890
11. Poustka, A., Lehrach, H., Williamson, R. & Bates, G. (1988) *Genomics* **2**, 337–345
12. Larin, Z., Monaco, A. P. & Lehrach, H. (1991) *Proc. Natl. Acad. Sci. USA* in press
13. Herrmann, B. G., Barlow, D. P. and Lehrach, H. (1987) *Cell* **48**, 813–825
14. Nizetic, D., Zehetner, G., Monaco, A. P., Young, B. D., Lehrach, H. (1991) *Proc. Natl. Acad. Sci. USA* in press
15. Herrmann, B. G., Labeit, S., Poustka, A., King, T., R. & Lehrach, H. (1990) *Nature* **343**, 617–622
16. Edwards, J., Williams, S., West, L., Lipowicz, S., Sheer, D., Attwood, J., Spurr, N., Sarkar, R., Saha, N. and Povey, S. (1990). *Ann. Hum. Genet.* **54**, 131–139
17. Poustka, A., Dietrich, A., Langenstein, G., Toniolo, D., Warren, S. T., Lehrach, H. (1990) *Proc. Natl. Acad. Sci. USA* in press
18. Warren, S. T., Zhang, F., Licameli, G. R. and Peters, J. F. (1987) *Science* **237**, 420–423
19. Warren, S. T., Knight, S. L., Peters, J. F., Stayton, C. L., Consalez, G. G., Zhang, F. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 3856–3860
20. Perlino, E., Paonessa, G. & Ciliberto, G. (1985) *Nucleic Acids Res.* **13**, 8369–8377
21. Heitz, D., Roesseau F., Devys, D., Saccone S., Abderrahim H., Le Paslier D., Cohen D., Vincent A., Toniolo D., Della Valle G., Johnson S., Schlessinger D., Oberlé I., Mandel J.L. (1991) *Science*, **251**, 1237–1239
22. Sherman, S. L., Morton, N. E., Jacobs, P. A. & Turner, G. (1984) *Ann. hum. Genet.* **48**, 21–37
23. Sved, J. A. and Laird, C. D. (1990) *Am. J. Hum. Genet.* **46**, 443–451
24. Wieacker, P., Davies, K. E., Cooke, H. J., Pearson, P. L., Williamson, R., Bhattacharya, S., Zimmer, J. and Ropers, H. H. (1984) *Am. J. Hum. Genet.* **36**, 265–268
25. Willard, H., F., Smith, K., D. & Sutherland, J., (1983) *Nucleic Acids Res.* **11**, 2017–2033