Frequent small amplifications in the FMR-1 gene in fra(X) families: limits to the diagnosis of 'premutations'

James N Macpherson, David L Nelson, Patricia A Jacobs

Abstract

In five of 40 fra(X) families reinvestigated using the new intragenic probe StB12.3, small amplifications of the DNA fragment appeared unexpectedly in addition to the mutations found in the probands. This suggests that enlargements of the FMR-1 gene detectable by Southern blotting using this probe must be present at an appreciable frequency in the general population. A proportion of these may be classifiable as 'premutations', or precursors of the much amplified, hypermethylated, and somatically unstable fragment associated with the fragile X syndrome, while others will merely represent stable polymorphisms in fragment length. Hence, accurate diagnosis of some fra(X) carriers will depend upon a more precise measurement of insert size than is currently provided by the newly available molecular probes.

(J Med Genet 1992;29:802-6)

Several cloned DNA fragments within and around the fragile X gene have recently been isolated¹⁻³ which, together with the cloning of much of the putative gene sequence designated FMR-1,⁴ have enabled the direct diagnosis of fra(X) carrier status and given us considerable insight into the molecular aspect of the mutation. It appears that the mutation present in non-expressing carriers is a small insertion or amplification of 100 to 500 bp in size, while in affected subjects the insertion is larger (>0.6 kb), invariably accompanied by hypermethylation of a nearby CpG island and often by somatic instability, resulting in a smear of bands rather than a single band upon autoradiography of the Southern blots.^{5,6} The molecular basis for the instability is purported to be a variable number of trinucleotide repeats which also fluctuates stably, within certain limits, in the normal population.^{7,8}

We have used probe StB12.3² to reinvestigate 40 known fra(X) families, classifying each subject (including unrelated spouses) for their insertion mutation status.⁹ In almost all families the probe was found to be a very good diagnostic tool, detecting large inserts in all subjects expressing the fra(X) at greater than background levels, and small inserts in obligate carriers. Many previously unsuspected normal transmitting males and carrier females were identified, and in common with other workers we found no example of a de novo mutation.

In five families, we found six amplifications of the DNA fragment which were not predicted either from phenotype or position in the pedigree, and these are the subject of the present report. Two were in the unrelated spouses of fra(X) carriers, three were on the other X chromosome of known fra(X) female carriers, and one in the daughter of a normal female in a fra(X) family. These small 'inserts', although real and repeatable, tended to be smaller than the average size for a premutation. We have estimated the frequency of such small inserts in the general population based upon their incidence on those X chromosomes in our sample considered to be independent of ascertainment, and conclude that they are sufficiently common to pose a genuine problem for the diagnosis of fra(X) carrier status.

Material and methods

Thirty-three of the 40 families were referred to the Wessex Regional Genetics Laboratory for cytogenetic and DNA investigation, the remaining seven being ascertained during a survey of mentally handicapped schoolchildren in Coventry, England.¹⁰

DNA was extracted from peripheral blood leucocytes by the salt precipitation method¹¹ and stored in TE at 4°C. Samples of genomic DNA from each patient were digested with restriction enzymes *EagI* and *EcoRI* as previously described,⁹ electrophoresed, and blotted according to standard conditions. Membranes were hybridised with radioactively labelled probe DNA¹² and washed in 3 × SSC, 0.1% SDS three times at room temperature then three times at 65°C before drying and exposing to x ray film at -80°C for one to 10 days.

The PCR method and conditions used to synthesise the FMR-1 repeat region have been described previously.⁸

Results

Our protocol for restriction enzyme digestion and Southern blotting, following Rousseau *et al.*,⁶ enables simultaneous detection both of the size of the insert and of the methylation status of the adjacent CpG island. Normal males thus display one band at 2.8 kb (the *EagI/EcoRI* fragment) and normal females a 2.8 kb and a 5.2 kb band, the latter corresponding to the inactive X chromosome on which the *EagI* site is methylated and hence resistant to cleavage by this endonuclease. The double digest allows for the sensitive detection of small inserts, and

Wessex Regional
Genetics Laboratory,
Salisbury District
Hospital, Odstock,
Salisbury SP2 8BJ.
J N Macpherson
P A Jacobs

Baylor College of
Medicine, Institute for
Molecular Genetics,
One Baylor Plaza,
Houston, Texas 77030,
USA.
D L Nelson

Correspondence to
Dr Macpherson.

Received 4 June 1992.
Revised version accepted
1 July 1992.

also makes an absolute distinction between those small inserts corresponding to 'premutations' and the larger inserts which represent full mutations, since all inserts >0.6 kb are methylated even on the active X chromosome.

Fig 1 shows the DNA patterns of a typical three generation fra(X) family with transmitting male (I), carrier female (II.1), and affected male (III.2). No new mutation was found: all 'sporadic' affected males had mothers with an insert, and no subject carrying an insert was born to parents carrying no insert.

The correlation between insert size and cytogenetic or clinical expression was very good: all subjects carrying a small (S) insert are unaffected and cytogenetically negative, while all males and 85% of females with a large (L) insert were fra(X) positive.⁹ Conversely all males and females scored as fra(X) positive were found to have a large insert. We found no example of a 'mosaic' as reported by Rousseau *et al*⁶ in our 40 families, although one such case in a family recently referred for prenatal diagnosis is currently being investigated.

Our most important finding was of six people in five families with unexpected small inserts. The first of these families is shown in fig 2. The proband's mother, I.1, clearly has both her 'active' bands raised above the 2.8 kb normal sized fragment, and consequently her daughter, II.1, has an apparently mutated band, although linked probe U6.2 had previously shown II.1 to carry the low risk allele. Fig 3 shows a fra(X) positive female (lane 2) who carries both a large and a small insert, compared with a normal female (lanes 1 and 3);

Figure 1 Autoradiograph of fra(X) family showing inserts of typical size representing the premutation (I and II.1) and full mutation (III.2). A vertical line through a symbol means that subject has been tested clinically and cytogenetically.

this woman has an affected son with a large insert and also a son and a daughter each with a small insert (not shown). The compound heterozygote status of the mother explains the occurrence of both large and small inserts among her children. Fig 4 shows a fragile X family where an unexpected small insert in the unrelated spouse (II.5) of obligate carrier II.4 has been transmitted to daughters III.3 and III.4; hence III.4 has inherited an insert from each parent. Although linked RFLPs were uninformative in this family, it seems likely that the larger of these two inserts has transformed into the full mutation in proband IV.

Fig 5 also shows a family where an obligate carrier woman I.2 has married a man with an unexpected small insert (I.1), and again it seemed probable that the small inserts in II.2 and III were the same as that in I.1. Using flanking probes RN-1A and VK21, we showed that this X chromosome segment in III is indeed grandpaternal in origin.

Finally, fig 6 shows a distant branch of a large fra(X) family where the daughter and grandson of an apparently normal mother (I) each has a small insert, suggesting that it originated from II's father who is not available for testing, although the possibility of a new mutation should not be entirely discounted.

We have referred previously⁹ to all females who carry an insert on each of their X chromosomes as 'compound heterozygotes' which supposes that the unexpected insert in each case is a premutation, making these women effectively fra(X) homozygotes. However, most of the unexpected inserts tend to be smaller than those presumed to have amplified into the full mutation, and thus they may confer much lower risk of such transformation. It seems that the number of CGG repeats, known to vary in the 'normal' population of X chromosomes, is stably transmitted through generations up to a limit of around 50 copies with a modal average of 29⁸; hence it may be that most 'inserts' not ascertained through a proband are simply at the high end of the normal range of alleles. As it is not clear from autoradiographs of Southern blots where the distinction in size between normality and a premutation should be drawn, the only reliable way to determine this is from a precise estimation of the numbers of CGG repeats by PCR and high resolution gel electrophoresis.

Alleles up to a maximum of about 200 repeats in length may be analysed by this procedure, above which limit the reaction does not work reliably and the size estimates are subject to increasing error from about 120 repeats upwards. Alleles between about 52 and 200 repeats are thought to correspond to premutations, but only those over 70 repeats seem predisposed to convert to full mutations at high frequency; hence the PCR analysis is of most use for the normal alleles and those smaller premutations with a relatively low risk of conversion.

In the table, estimates of the size of the amplification (Δ) as numbers of CGG repeats are given for the five unexpected small inserts in our families. It will be seen that two of these

Figure 2 *Fra(X)* family showing 'compound heterozygote' female I.1 with inserts on both her chromosomes relative to the normal band size. The smaller 'insert' appears to have been transmitted unchanged in size to her daughter II.1.

Estimates of ΔS for the five small inserts appearing unexpectedly in our sample.

Carrier of small insert	Insert size by autoradiography (bp)	No of CGG repeats by PCR
Fig 2, I.1/II.1	85	43
Fig 3, 2	100	52
Fig 4 II.5/III.3/III.4	90	41
Fig 5, I.1/II.2/III	70	41
Fig 6 II/III	95	54

Right hand column gives absolute number of repeats; ΔS may be obtained by subtracting 29 from each figure.

(II and III, fig 6 and 2, fig 3) fall on or above the putative borderline and are therefore candidate premutations; the remaining three, although all above the modal average of 29 repeats for a normal allele, are nevertheless of a size that is currently considered stable within families. Thus, of 40 families analysed, a possible two unexpected premutations were detected and three other inserts were found at the top end of the normal range which we had falsely classified as premutations.

The PCR size estimates did not vary by more than one or two repeats between generations in the families described here, so there is no evidence for meiotic instability in any of these unexpected small inserts. However, the number of meioses and generations involved is too small to make a categorical judgement on this issue.

Discussion

These findings show that small amplifications of DNA within the FMR-1 gene, which can be confused with premutations on autoradiographs of Southern blots, are not uncommon in the general population. We can estimate their frequency using two assumptions: that the spouses analysed in our *fra(X)* families represent an unbiased sample of X chromosomes; and that in females ascertained for a mutation on one of their X chromosomes, their other X remains unbiased by this ascertainment. These assumptions are probably valid provided that there is no consanguinity or identity by descent in the families that is unknown to us.

Each male spouse contributes one unbiased X and each female spouse two, while each unrelated carrier female within the fragile X family contributes one unbiased X. The estimated frequency, q_s , of small inserts in the general population is therefore given by:

$$q_s = \frac{S_s + S_c}{N_{sm} + 2N_{sf} + N_c}$$

where S = number of small inserts in spouses (S_s) and carriers (S_c), N = number of male spouses (N_{sm}), female spouses (N_{sf}), and female carriers (N_c).

Not included in N_c are any daughters of males included in N_{sm} or any two females whose 'sampled' chromosomes are obviously identical.

In our data this represents

$$q_s = \frac{2 + 3}{38 + 16 + 55} = 4.6\%.$$

If, as initially thought, all of the small inserts were considered as premutations, then this would place the *fra(X)* gene frequency on a par with that of the cystic fibrosis mutation in northern Europe. As it is, considering only two as potential premutations, it is difficult to arrive at an unbiased estimate for its frequency but the closest we can get is to say that it may exceed 1% from these data. Such a frequency would still have to be taken into account in family studies, emphasising the need to analyse all spouses in *fra(X)* families and ultimately to screen the general population to determine the true frequency of small inserts and the proportion of premutations.

The definition of a premutation on size alone, even with a precise number of CGG repeats, is difficult given the current uncertainty about the origin and expansive behaviour of the mutation. At present the observed stability through generations of inserts under 52 repeats is used as our main criterion for defining the upper limit of normality, while the lower limit of the premutation range remains unclear owing both to ascertainment bias and to the probability of overlap between the two.

Among the known premutations analysed to date, measurements of DNA fragment size, by Southern blotting or PCR techniques, indicate a strong tendency for the insert to increase in size from generation to generation (at least in families ascertained through a ΔL): carrier daughters tend to have larger insert sizes than their carrier mothers, although occasional reversions in size are observed from transmitting males to their daughters. Although clearly dependent on the initial insert size, the amount by which the insert size increases (if at all) in each generation is unpredictable, as though a chance element is involved in the degree of instability which a given number of CGG repeats confers on the DNA strand. Nevertheless, the number of CGG repeats in a premutation can be shown to correlate strongly with its probability of transformation to a full mutation (table 9 of Fu *et al*⁸).

From our family studies reported previously,⁹ a multistep model was suggested^{13,14} for the mutation process which postulates four alleles; N, S, Z, and L (fig 7). The conversion from N to S has not yet been observed experimentally, but once established we know that S can be transmitted through several generations giving only S progeny, before assuming a state (Z) which converts to L with a very high probability.

Clearly the molecular basis for the switch from S to Z is critical for diagnostic purposes. A qualitative mechanism for this might be provided by Laird's model, whereby the small insert would become 'primed' upon passing through a cycle of Lyonisation and reactivation in a female.¹⁵ Alternatively the switch could have a purely quantitative basis, as a stochastic response to the number of CGG repeats in the segment. Present data with their

Figure 3 'Compound heterozygote' female (lane 2) with both small and large inserts relative to the normal band size.

Figure 4 *Fra(X)* family in which an unexpected small insert in a spouse (II.5) has been passed on to an obligate carrier woman (III.4) whose maternal X chromosome also carries a small insert. A spot in the middle of a symbol denotes an obligate carrier.

Figure 5 *Fra(X)* family showing transmission of an unexpected small insert from a male spouse (I.1) to his otherwise normal daughter (II.2) and thence to her baby son (III). The haplotype for closely linked probes RN-1A and VK21 was conserved through this transmission and distinct from that on the X chromosome of obligate carrier I.2 which gave rise to affected boys II.1 and II.3.

absence of new mutations and generally unidirectional change in size of the insert through generations seem to favour a progressive amplification with certain thresholds of instability as the basis for the shift towards the full mutation.

From data accumulated by PCR these thresholds can be identified as approximately 50 repeats for the N/S transition, and perhaps around 70 repeats for the S/Z transition, although these boundaries should certainly not be regarded as hard and fast. They may, however, come to provide arbitrary guidelines for diagnostic purposes. Methylation, rather than size, may be the more significant consequence of the final Z→L mutation and is associated with absence of FMR-1 gene expression.¹⁶ Meanwhile, the relationship of the mutation to the Xq27.3 fragile site still remains obscure.

The PCR method has proved very valuable to date in resolving important sequence length differences at the boundary between normal (N) and premutation (S) alleles. Unfortunately, the technical difficulties inherent in PCR analysis of long GC rich regions have tended to preclude a similar assessment of the transition from S to Z, as a proportion of the mutated fragments giving rise to L alleles in the next generation have proved refractory to PCR synthesis. We have so far failed to observe from Southern blots any consistent dif-

ference in size between those S fragments which generate only S, and those which give rise to only L sib pairs in the next generation, and it may thus await further refinements in the PCR technique before this issue is resolved.

In the meantime, the high frequency of small inserts in the general population poses a dilemma for diagnostic laboratories and genetic counsellors in terms of the risk to be assigned. It is of equal importance to avoid creating false positives by labelling all such inserts as premutations as to omit a true premutation by lumping it as a high end normal allele. Both such errors are likely to be encountered by use of the Southern transfer method

Figure 6 Unexpected small insert appearing in the daughter of a normal female (I) which has probably been contributed by I's absent spouse.

Figure 7 Model for *fra(X)* mutation. M= male. F= female. N= normal. S= small insert at low risk of converting to L. Z= small insert at high risk of converting to L. L= large insert.

and probes currently available: the *EcoRI*/StB12.3 combination is relatively insensitive to the detection of small inserts, while the *PstI*/pfxa3 combination of Yu *et al.*¹ tends to be oversensitive. In addition the precise relationship between size and stability of the repeat sequence remains to be firmly established. By rigorous molecular screening, using methods of equivalent sensitivity to those used here, of all known fra(X) families including spouses, in addition to embarking on widespread population screening programmes, it should eventually be possible to determine the true incidence of both the premutated and the fully mutated forms of this most common inherited cause of mental retardation.

We are grateful to M F Victoria for technical assistance.

- 1 Yu S, Pritchard M, Kremer E, *et al.* Fragile X genotype characterized by an unstable region of DNA. *Science* 1991;252:1179-81.
- 2 Oberlé I, Rousseau F, Heitz D, *et al.* Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* 1991;252:1097-102.
- 3 Hirst MC, Nakahori Y, Knight SJL, *et al.* Genotype prediction in the fragile X syndrome. *J Med Genet* 1991;28:824-9.
- 4 Verkerk AJMH, Pieretti M, Sutcliffe JS, *et al.* Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile-X syndrome. *Cell* 1991;65:905-14.
- 5 Heitz D, Rousseau F, Devys D, *et al.* Isolation of sequences that span the fragile-X and identification of a fragile-X related CpG island. *Science* 1991;251:1236-9.
- 6 Rousseau F, Heitz D, Biancalana V, *et al.* Direct diagnosis by DNA analysis of the fragile X syndrome of mental retardation. *N Engl J Med* 1991;325:1673-81.
- 7 Kremer EJ, Pritchard M, Lynch M, *et al.* Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)_n. *Science* 1991;252:1711-14.
- 8 Fu YH, Kuhl DPA, Pizzuti A, *et al.* Variation of the CGG repeat at the fragile-X site results in genetic instability - resolution of the Sherman paradox. *Cell* 1991;67:1047-58.
- 9 Macpherson J, Harvey J, Curtis G, *et al.* A reinvestigation of thirty three fragile(X) families using probe StB12.3. *Am J Med Genet* 1992;43:905-12.
- 10 Webb TP, Bunday S, Thake A, Todd J. The frequency of the fragile X chromosome among school children in Coventry. *J Med Genet* 1986;23:396-9.
- 11 Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988;16:1215.
- 12 Feinberg AP, Vogelstein B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 1983;132:6-13.
- 13 Mandel JL, Hagerman R, Froster U, *et al.* Conference Report. Fifth International Workshop on Fragile X and X-linked Mental Retardation. *Am J Med Genet* 1992;43:5-19.
- 14 Morton NE, Macpherson JN. Population genetics of the fra(X) syndrome: a multiallelic model for the FMR1 locus. *Proc Natl Acad Sci USA* 1992;89:4215-7.
- 15 Laird CD. Proposed mechanism of inheritance and expression of the human fragile-X syndrome of mental retardation. *Genetics* 1987;117:587-99.
- 16 Pieretti M, Zhang F, Fu YH, *et al.* Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* 1991;66:817-22.