

## Review article

# FMR1 triplet arrays: paying the price for perfection

Mark C Hirst

### Abstract

**Our understanding of FMR1 trinucleotide instability has increased dramatically with knowledge of its detailed structures. While most arrays seem to be protected by interspersions, for a few the price of perfection is instability. Although there remain many unanswered questions, diagnosis in the "grey zone" can be greatly improved by studying array content. For the future, as we strive to delineate normal from premutation, we should increasingly be able to estimate rates of instability for future generations and predict the risk of conversion to the full mutation.**

(J Med Genet 1995;32:761-763)

In the normal population, FMR1 trinucleotide array length has a discontinuous bimodal distribution ranging up to 54 repeats.<sup>1</sup> Ninety-five percent of arrays are less than 40 repeats in length and only 1 to 2% are longer than 46. A large number of molecular studies of fragile X families has helped to establish diagnostic boundaries for the various classes of mutation. On premutation chromosomes, the repeat array is greater than 54 in length, is genetically unstable, and expands at a rate proportional to its length until a critical point is reached between 150 and 200 repeats. Above this size, and dependent upon transmission through the female germline, massive expansion to the full mutation can occur. This results in hypermethylation, gene inactivation, late replication and the manifestation of the fragile X site, FRAXA. The detection of expanded arrays is sufficient in most cases for diagnostic analysis of families segregating for the fragile X mutation. Difficulty can occur, however, in the area between 45 and 55 repeats, an area often referred to as the "grey zone", by judging solely upon the array length, whether it is normal or premutant. While the boundary is set at 54 repeats, unstable arrays as small as 46 have been described, whereas others of that length appear to be stable.<sup>2</sup> Rare reports of mutant arrays reverting to within the normal size range have also added to the problem, leaving uncertain the chance of overlooking a normal sized, small, unstable or premutation array. While the only test for instability is, by definition, that of meiotic in-

stability itself, our only assay has been to measure the overall length of the array. With contradictory observations on stability in the range 45 to 55, such a measure can fall short of its objective. New information on FMR1 repeat array structure and the development of techniques for analysing this should aid the identification and study of unstable arrays.<sup>3-6</sup>

### Architecture of the FMR1 array

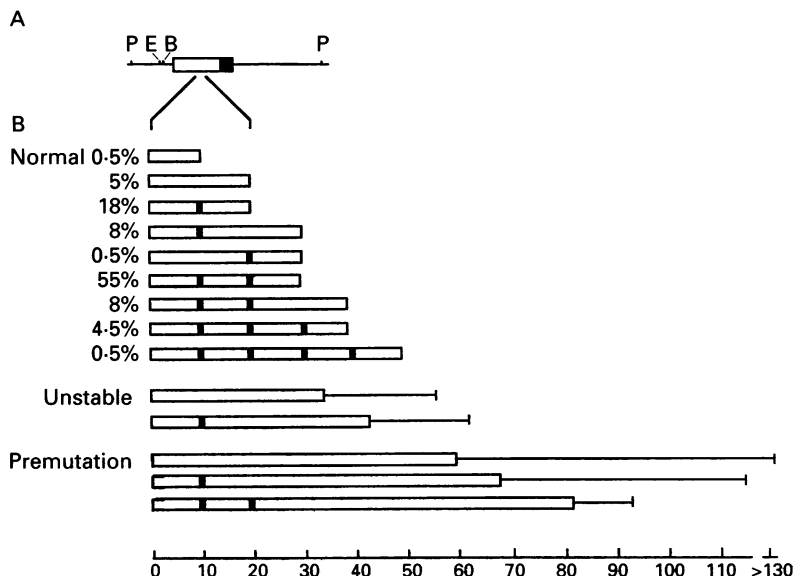
Detailed sequence analysis has shown that the majority of FMR1 triplet arrays are compound repeats consisting of short tracts of  $(CGG)_n$  and single interspersed AGG triplets arranged such that, on average, every tenth triplet is AGG (figure B). Allelic variation is a result of both variation in the  $(CGG)_n$  lengths (from seven to 13, with a modal length of nine triplets) and the number the  $(CGG)_n(AGG)$  blocks present. Thus the modal population lengths of 20 and 30 repeats contain two or three of these interrupted blocks. Ethnic differences between population groups appear to exist, as shown by the presence of a modal group centred at 35 repeats in 10-15% of the Japanese and Chinese populations.<sup>7-9</sup> A sizeable number of normal sized arrays, approximately 10 to 15%, contain portions of  $(CGG)_{n>17}$ , either as uninterrupted arrays or at the 3' end of compound arrays (although several have been identified at the 5' end). Thus, while arrays may be of identical length, they can differ in their composition, a difference which may cause differential rates of instability between arrays.

### Unstable and premutation arrays and progression to the full mutation

Premutation and other unstable arrays contain long portions of  $(CGG)_n$  and have been found as completely uninterrupted arrays or with one or two 5' AGG interruptions (figure B). The length of uninterrupted  $(CGG)_n$  repeat appears to be similar for each array class, although the  $(CGG)_n$  tract lengths appear to be smaller on arrays with 5' interrupting AGG triplets. For example, premutations with a single 5' AGG have a smaller overall range than those with no 5' AGG. This suggests that the 5' interruption has some stabilising influence upon these arrays. In fragile X families, premutation array length changes occur exclusively because of

**Molecular Genetics Group,  
Institute of Molecular Medicine,  
John Radcliffe Hospital,  
Headington,  
Oxford OX3 9DB, UK**  
M C Hirst

Present address:  
Program in Molecular Medicine,  
Fred Hutchinson Cancer Research Center,  
1124 Columbia Street,  
Seattle, WA 98104,  
USA.



*FMR1* array structure from sequenced alleles. (A) *FMR1* exon I is shown relative to the important diagnostic restriction sites *Pst*I (P), *Bss*HII (B), and *Eag*I (E). Hatching indicates the translated portion of exon I. (B) Representations of the *FMR1* array structure in various groups. Nine normal *FMR1* array structures were found on 184 non-selected chromosomes in the normal size range.<sup>3,5</sup> Open boxes indicate (CGG)<sub>n</sub> and solid boxes indicate AGG trinucleotides. Representation of unrelated premutation (n=37) and genetically unstable non-premutation arrays (n=9) (for distinction between these two groups, see text) which have been examined by sequence analysis<sup>4,5</sup> or by internal restriction site mapping.<sup>6</sup> The range of the array lengths is indicated by scale bars. The arrays are shown relative to a scale measuring the total triplet number.

alterations in the length of the uninterrupted portion of the repeat. In the one fragile X family in which progression of the premutation to full mutation was followed, the uninterrupted portion of the array expanded without loss of a 5' AGG.<sup>4</sup>

Several arrays have been identified as being genetically unstable, but are not known to have been associated with fragile X syndrome (figure B). In this class of arrays the shortest array exhibiting instability was (CGG)<sub>34</sub>,<sup>6</sup> although an array of (CGG)<sub>11</sub>(AGG)(CGG)<sub>30</sub> has also been shown as being unstable.<sup>4</sup> Thus uninterrupted tracts as short as (CGG)<sub>30</sub> may exhibit instability. When an overall comparison is made between (CGG)<sub>n</sub> length and genetic instability, the degree of instability correlates well with its increasing length.<sup>6</sup> Therefore the longer the uninterrupted (CGG)<sub>n</sub> portion the more unstable the array appears to be. One would expect that these longer tracts of perfect repeat would have a higher probability of slip strand mispairing during replication which would lead to their further expansion. This appears to be the case for the SCAI associated (CAG)<sub>n</sub> repeat array where 98% of arrays are interrupted, whereas the disease associated arrays are perfect arrays.<sup>10</sup> Similarly for DM, a rare stable array within the unstable size range was shown to carry an interruption.<sup>11</sup>

#### Precursor arrays in the populations

The realisation that unstable arrays have longer 3' CGG tracts raises a question as to whether normal sized arrays which carry these may be precursors that will eventually progress into the premutation range. They appear to occur more

frequently upon chromosomes carrying flanking marker haplotypes which are associated with fragile X chromosomes, suggesting that this is indeed the case.<sup>4,5</sup> Such arrays fulfil the criteria for a group of alleles predisposed to transition to premutation (S alleles) as proposed by Morton and MacPherson.<sup>12</sup> Their presence in the current population indicates their relatively stability and longevity, while the longer tract of (CGG)<sub>n</sub> would be expected to give them an increased propensity for further mutation. Their structure also suggests that they have arisen by the loss of the interspersed AGG triplet, either by deletion or conversion to CGG, although by what mechanism such a loss might occur and apparently be directed toward the 3' end of the array is unknown. Thus, such rare events could give rise to founder chromosomes from which most fragile X arrays are descendent.

If (CGG)<sub>n</sub> length is the only critical factor in array instability, then at what point do we consider arrays as potentially capable of progression to premutation? By comparison with the lengths of triplet arrays associated with other triplet repeat expansion disorders, it has been suggested that a critical threshold is reached at 24 uninterrupted triplets.<sup>3</sup> Others have suggested that all long uninterrupted tracts should be considered candidates for further expansion events.<sup>4,5</sup> If instability ranges from 100% as seen with premutation arrays through the intermediate level observed in (CGG)<sub>30-50</sub> tracts, smaller arrays ((CGG)<sub>>17</sub>) may well have a low level of instability, its measurement hampered by the limited number of observable meiotic events in family studies. It may be possible to obtain such information, at least for the male germline, by sperm typing, as has been performed for the Huntington and Kennedy disease genes.<sup>13,14</sup> Additionally, there may be factors acting in *cis* or in *trans* which might influence array stability in some as yet undetermined manner.

#### Diagnostic issues and strategies: interpreting array instability

For diagnostic purposes, several difficulties can arise with the interpretation of *FMR1* array lengths. Firstly, within the grey zone of 45 to 55 repeats, can premutations be distinguished from long normal arrays? At what length should an array be considered a premutation? Current diagnostic thresholds have been determined through extensive family studies and have served well. At the moment, only a small number of sequenced premutations have been reported (37 independent chromosomes). In order to redefine these thresholds in terms of (CGG)<sub>n</sub> content many more premutation arrays will have to be analysed. Additionally, it is not known whether any other feature other than length of uninterrupted (CGG)<sub>n</sub> plays a role in instability. Rare cases of reversion into the normal size range could provide valuable insights into this.<sup>15</sup>

How should we consider an array with a long (CGG)<sub>n</sub> portion in terms of its instability? Arrays in this size range should be analysed for

(CGG)<sub>n</sub> content, either with direct sequencing or the coupled PCR restriction digest analysis. The predicted level of instability will depend upon the profile of AGG interruptions; thus for two arrays of 52 repeats, one with 10A 10A 9A 9A 10 structure would be considered as normal with a low risk, whereas an array of 9A 42, should be considered as high risk. In the latter case, the only currently available option for further investigation is to determine the degree of instability by extended family analysis. This will inevitably lead to more arrays being considered as potential pre-mutations. As no data are available to provide a quantitative estimate of instability in the intermediate range, such arrays will prove difficult for counselling purposes as no clear cut answer is available. Several points should be borne in mind with these cases. Firstly, the risk of conversion to a full mutation is still dependent upon passage through the female germline; thus for these alleles transmitted from a male the risk will be zero. Secondly, arrays within this intermediate zone of instability are unlikely to carry a significant risk of conversion to the full mutation in a single generation, as with all the fragile X families which have been analysed to date no pre-mutation has been reported smaller than 54 repeats in length. In the limited sequence analyses to date, the smallest array which has given rise to a fragile X chromosome contained 56 uninterrupted repeats and expanded over two generations.<sup>6</sup> In contrast, an array of 55 repeats progressed through six unstable transmissions without converting to a full mutation.<sup>16</sup> This latter example should be treated with a degree of caution as this array is not known to have ever progressed to a fragile X array.

Small alterations in array length can also result from the interchange of repeats caused by normal recombinational activity at the FMR1 array which is unlikely to be associated with the mutational process of the fragile X syndrome. The interrupted nature of the FMR1 array bestows a larger repeat feature upon each array, such that each block of (CGG)<sub>n</sub>AGG can function as a repeat unit in itself much like

a minisatellite. Thus, sudden jumps in size, increasing or decreasing the array length, can occur. One such exchange event has been described which resulted in a new array of normal size which had increased in size by 10 triplets through an addition of a (CGG)<sub>9</sub>(AGG) unit during male meiosis.<sup>16</sup> One other published case appears to be a similar type of event.<sup>7</sup> The sequencing of such arrays should confirm a normal interrupted pattern.

- 1 Fu Y, Kuhl D, 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.
- 2 Reiss A, Kazazian H, Krebs C, *et al.* Frequency and stability of the fragile X pre-mutation. *Hum Molec Genet* 1994;3:393-8.
- 3 Kunst C, Warren S. Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell* 1994;77:853-61.
- 4 Snow K, Tester D, Kruckerberg K, Schaid D, Thibodeau S. Sequence analysis of the fragile X trinucleotide repeat: implications for the origin of the fragile X mutation. *Hum Molec Genet* 1994;3:1543-51.
- 5 Hirst M, Grewal P, Davies K. Precursor arrays for triplet repeat expansion at the fragile X locus. *Hum Molec Genet* 1994;3:1553-60.
- 6 Eichler E, Holden J, Popovich B, *et al.* Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nature Genet* 1994;8:88-94.
- 7 Aranami T, Asano M, Kobayashi K, Yanagi H, Hamaguchi H. Data on the CGG repeat at the fragile X site in the non-retarded Japanese population and family suggest the presence of a subgroup of normal alleles predisposed to mutate. *Hum Genet* 1993;92:431-6.
- 8 Zhong N, Liu X, Gou S, *et al.* Distribution of FRM1 and associated microsatellite alleles in the normal Chinese population. *Am J Med Genet* 1994;51:417-22.
- 9 Richards R, Kono I, Holman K. Haplotype analysis at the FRAXA locus in the Japanese population. *Am Med Genet* 1994;51:412-16.
- 10 Chung M, Ranum L, Duvik L, Servadio A, Zoghbi H, Orr H. Evidence for a mechanism predisposing to intergenerational CAG instability in spinocerebellar ataxia type I. *Nature Genet* 1993;5:254-8.
- 11 Leeflang E, Arnheim N. A novel repeat structure at the myotonic dystrophy locus in a 37 repeat allele with unexpected stability. *Hum Molec Genet* 1995;4:135-6.
- 12 Morton N, MacPherson J. Population genetics of the fragile X syndrome: multi-allelic model for the FMR1 locus. *Proc Natl Acad Sci USA* 1992;89:4215-17.
- 13 Zhang L, Fischbeck K, Arnheim N. CAG repeat length variation in sperm from a patient with Kennedy's disease. *Hum Molec Genet* 1995;4:303-5.
- 14 Telenius H, Almqvist E, Kremer B, *et al.* Somatic mosaicism in sperm is associated with intergenerational (CAG)<sub>n</sub> changes in Huntington disease. *Hum Molec Genet* 1995;4:189-95.
- 15 Vits L, Debouille K, Reyniers E, *et al.* Apparent regression of the CGG repeat in FMR1 to an allele of normal size. *Hum Genet* 1994;94:523-6.
- 16 MacPherson J, Curtis G, Crolla J, *et al.* Unusual (CGG)<sub>n</sub> expansion and recombination in a family with fragile X and Di George syndrome. *J Med Genet* 1995;32:236-9.