REVIEW ARTICLE

Dynamic mutations on the move

Grant R Sutherland, Robert I Richards

It is only a short time since the isolation and characterisation of the fragile X syndrome mutation¹⁻³ uncovered a new genetic element and mechanism of mutation. The genetic element was an unstable DNA sequence resulting from amplification of a naturally occurring polymorphic trinucleotide repeat, p(CCG)n. The mechanism of mutation, which we have termed dynamic mutation,⁴ is the change (increase or decrease) in copy number of the trinucleotide repeat with the rate of change related to the number of copies present at any time. This process, in which an initial change to a DNA sequence alters the chance of further changes to it, contrasts with classical or static mutation in which the product of a mutation is no more likely to undergo further changes than was the initial DNA sequence.

The instability of the amplified p(CCG)n repeat in fragile X syndrome is such that different cells in different tissues, and even different cells in the same tissue, can have different numbers of copies of the repeat. Furthermore, when the repeat is transmitted from parent to offspring the number of copies can change markedly. Hence different subjects in a single family who carry the mutation can have it in different forms. Normal X chromosomes have six to about 50 copies of the repeat. The mutation appears to begin as a small increase in copy number out of the normal range. This is transmitted unstably over a few generations, gradually increasing when transmitted by women but with little change in size when transmitted by males. Subjects with small size increases (up to about 200 copies) are clinically normal and said to carry premutations at this locus.

The properties of this new genetic element provided possible mechanisms for a variety of phenomena which were poorly understood.⁵ These included incomplete penetrance, variable expression, other fragile sites, and anticipation. The prediction that anticipation, especially as shown by myotonic dystrophy (DM), was the result of an unstable sequence was soon confirmed.⁶⁻⁸ In this disease the trinucleotide was p(AGC)n. This finding of a dynamic mutation in DM allowed the recognition that anticipation was a legitimate clinical observation,^{9,10} now with an acceptable genetic mechanism, and not an artefact resulting from biased ascertainment.¹¹

The number of disorders resulting from unstable amplification of trinucleotide se-

quences is rapidly increasing (table). These include another fragile site (FRAXE) that may be associated with mild mental retardation.¹² This fragile site has turned out to be similar in structure to the fragile X (FRAXA), involving the same trinucleotide p(CCG)n and being close to a CpG island which is hypermethylated when the copy number exceeds approximately 200.*

The dynamic mutations characterised to date fall into two categories probably determined by whether the trinucleotide repeat is in a translated or untranslated region of a gene. The mutations in known or presumed untranslated regions, FRAXA, FRAXE, and DM, appear to have little constraint on the number of repeats, which can range up to several thousand copies. When the mutation is in a presumptively translated region of the gene (SBMA, HD, SCA1) the amplification is restrained and rarely exceeds 100 copies of the repeat. The molecular basis for this limit will hopefully become clearer after more detailed analysis of the biology of the respective encoded proteins.

When compared to other types of single gene defect the properties of dynamic mutation diseases afford distinct advantages and disadvantages for diagnosis. The mutations seen in most single gene defects are characteristically variable and numerous, occasionally a single mutation may predominate (as does $\Delta F508$ in cystic fibrosis), but it still remains one of many at that locus. The dynamic mutation disorders are remarkably homogeneous. Apart from fragile X syndrome where a few deletions and a point mutation in the gene (FMR1) have been described, no other mutations in the genes involved in dynamic mutation disorders have been recorded. This is a great advantage for the diagnostic laboratory where the primary diagnosis of these disorders can now be made with confidence. The major

**Nomenclature.* There are only 10 possible different trinucleotides which can be produced from the four primary bases from which DNA is constructed. Only two of these have been identified in dynamic mutations and these are p(CCG)n and p(AGC)n. Confusion is rampant in published reports because there is no consistent nomenclature for these sequences. The p(CCG)n repeat associated with fragile sites on the X chromosome has been expressed as p(CGG)n, p(CCG)n, and p(GCC)n. The convention proposed by the gene mapping workshops for polymorphic oligonucleotide repeats is that the bases be written in alphabetical order from 5' to 3', thus p(AGC)n and not p(CTG)n for the DM repeat. We are of the view that adhering to this convention will make the body of publications much more intelligible to those readers who wish to follow this area of genetics but who are not intimately involved with it.

Centre for Medical Genetics, Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, Adelaide, SA 5006, Australia.
G R Sutherland
R I Richards

Correspondence to Dr Sutherland.

disadvantages relate to uncertainty over the relationship between genotype and phenotype. Somatic variation can mean that copy number, determined from peripheral blood lymphocytes, is not an accurate determination of the size of the repeat in the affected tissue(s). In addition (perhaps as a consequence) there can be overlap in copy number for the different phenotypic categories (table). For example, in HD there is a very small area of overlap in the copy number distributions of normal and mutant chromosomes. About 2% of mutant chromosomes have copy numbers at the top of the normal range.¹³ In practice, the vast majority of direct diagnoses of dynamic mutation diseases based upon repeat copy number are reliable.

Parent of origin effects are marked in transmission of dynamic mutations. In fragile X syndrome the premutation only progresses to a full mutation and becomes clinically significant when it is transmitted by a female. In DM, the congenital form of the disease only occurs when a child receives the mutation from its mother. In HD, and possibly spinocerebellar ataxia, juvenile onset of the disease is paternally determined. The instability of the repeats would appear to be different in germline and somatic tissues. In fragile X only small changes, within the premutation range, are apparently transmitted through sperm and possibly also through ova.¹⁴ The large expansions seen in somatic tissues occur postzygotically but, by some imprinting mechanism, these expansions only occur on the maternally transmitted chromosome. The p(AGC)n repeat in HD sperm is particularly unstable and may be the reason why there are greater differences in copy number between fathers and their children than mothers and their children.¹⁵

The parental bias of congenital DM transmission appears to be because of constraints on the size of the repeat which can be transmitted by males; these constraints are not apparent in maternal transmission.^{16,17} In particular, the copy number of the p(AGC)n repeat may

increase when it is transmitted by either parent, but once males have about 600 to 700 copies little further increase occurs and decreases in size are seen. Females can, however, have children with up to approximately 4000 copies of the repeat. Studies on sperm from DM males are yet to be reported.

The basis for the sex differences, including the imprinting associated with the fragile X, are unknown. Whether they reflect the differences in the mechanisms of gamete formation or some other property of the sequences themselves remains unclear. There is evidence in fragile X syndrome that the increases in repeat copy number occur early in development after which time the sequence is stabilised and copy number changes no longer occur in somatic tissues. Jansen *et al*¹⁸ reported findings that suggest imprinting does not contribute to the manifestation of DM.

Effect of repeat copy number

In some of the dynamic mutation disorders there is a relationship between copy number and age of onset of disease. This is most clearly seen in spinocerebellar ataxia where there is a clear curvilinear relationship with only little scatter.¹⁹ In DM there is a general relationship with the congenital form of the disease having the largest number of repeats, but for any given copy number there is a wide range of ages of onset.^{20,21} A similar situation exists for Huntington's disease; high copy numbers are associated with juvenile onset but there appears to be little relation between age of onset and copy number where there are fewer than 50 repeats. Andrew *et al*²² report that 50% of the variation in age of onset in Huntington's disease can be attributed to repeat copy number. In Kennedy disease there is general decreasing age of onset with increasing copy number.²³

In fragile X, generally, when n exceeds about 200, transcription of the FMR1 gene stops (in lymphocytes), the p(CCG)n repeat and adjacent CpG island become hypermethy-

Features of dynamic mutations.

Disease	Fragile X syndrome	Spinobulbar muscular atrophy	Myotonic dystrophy	Huntington's disease	Spinocerebellar ataxia type 1	FRAXE ? Mild mental retardation
Inheritance	X linked dominant with incomplete penetrance (anticipation) Maternal (full mutation)	X linked recessive Variable severity	Autosomal dominant Anticipation	Autosomal dominant Anticipation	Autosomal dominant Anticipation	X linked
Sex bias for transmission of severe form	Maternal (full mutation)	?	Maternal (congenital DM)	Paternal (early onset)	Paternal (possibly)	?
Protein expression	Putative RNA binding protein mRNA widely expressed	Androgen receptor	Putative protein kinase	Unknown function, mRNA widely expressed	ORF (unknown) mRNA widely expressed	?
Disease causing mechanism	Transcription shut down; abnormal DNA methylation	Abnormal protein gain of function? (decrease/increase?)	Altered level of mRNA/protein?	Abnormal protein gain of function?	Abnormal protein gain of function?	?
Repeat location	CCG (interrupted) 5'UTR (CGG)	AGC (perfect) Protein coding (Gln)	AGC (perfect) 3'UTR (CUG)	AGC (perfect) Protein coding (Gln)?	AGC (interrupted) Protein coding (Gln)?	CCG (interrupted) ?UTR
Size of repeat (normal)	10–50	11–31	5–35	9–34	25–36	6–25
Linkage disequilibrium	Yes	Not known	Yes	Yes	?	?
Disease alleles	52–200 (premutation) 200–2000 (full mutation)	40–62	50–80 (protomutation) 80–2000 affected	30–100 ?	43–81	? > 25 < 200 premutation ? > 200 full mutation

lated, and the syndrome is apparent. Below 200 copies (the premutation) there is little clinical effect. The general situation is, however, occasionally disturbed by two forms of tissue mosaicism, repeat copy number and degree of methylation of the CpG island adjacent to the repeat (and possibly of the repeat itself). This situation will probably not be clarified until it is known in which tissue(s) abnormal expression of the FMR1 gene is critical for the various components of the syndrome and whether the inactivation of FMR1 expression in lymphocytes in which the CpG island is methylated also applies to other tissues.

Founder chromosomes

The conventional interpretation of the finding of linkage disequilibrium between a mutation and polymorphic DNA sequences adjacent to it (haplotype) is that the mutation arose only once on each haplotype on which it is present. Hence such disequilibrium is evidence of founder mutations and a very low mutation rate. The fragile X shows marked linkage disequilibrium²⁴ and DM shows complete disequilibrium with a diallelic RFLP.²⁵ Huntington's disease also shows a similar effect.²⁶

How can these data be interpreted for dynamic mutations? Some of the fragile X mutant haplotypes which do not carry the mutation have repeat copy numbers at the top of the normal range.²⁴ Since the rate of mutation of simple repeat sequences is proportional to copy number, there may be a group of normal alleles with high copy number which are predisposed to mutate. Alternatively, since there can be imperfections in the repeat on normal alleles, and since mutation may be proportional to the length of perfect repeat, some normal alleles without imperfections may also be predisposed to mutation. The linkage disequilibrium studies may therefore not be evidence of rare ancestral mutation but of a small number of normal alleles which have much higher rates of copy number increase, and thus mutation, than other normal alleles.

Genotype to phenotype

The mechanisms by which dynamic mutations produce their phenotypic effects are not clear. In fragile X syndrome the full mutation results in inactivation of the FMR1 gene (for a putative RNA binding protein) with cessation of transcription in lymphocytes. In DM there are conflicting reports of decreased²⁷ and increased²⁸ steady state mRNA levels. In the fragile sites and DM the repeat is in a non-coding part of the gene. The functional role of these sequences is therefore unrelated to the encoded protein. One possibility is that they are recognition sites for RNA or DNA binding proteins. Nuclear binding proteins for each of the di- and trinucleotide repeats (including various forms of the fragile X p(CCG)n repeat) have been reported.²⁹ In the other disorders, where the repeat codes for polyglutamine, the

expansion may confer some gain of function to the protein involved.

The finding³⁰ that FMR1 is a member of the KH domain family of RNA binding proteins is intriguing given the known roles of other members of this family. One of these, the yeast protein MER1, is a splicing regulator of the RNA transcript of another gene, MER2. The function of MER1 is crucial for meiosis.³¹ FMR1 appears to be essential for sperm viability (all X bearing sperm of fragile X affected males have a p(CCG)n copy number permissive for FMR1 expression) and therefore a role in male meiosis would be consistent with both these in vivo observations of FMR1 and the function of its yeast relative MER1. Assessment of spermatogenesis in males with deletion or point mutation in FMR1 would be of interest.

Unstable repeat sequences in other diseases

Two additional instances of changes in repeat copy number associated with disease warrant discussion. The first of these is in hereditary non-polyposis colon cancer (HNPCC). In this disorder tumour tissue exhibits genome wide instability of simple tandem repeat sequences.³² This instability is associated with genetic predisposition to HNPCC which maps to chromosome 2.³³ At this stage it is not clear whether the chromosome 2 gene is a cause of, or an 'at risk' target for, this instability. Instability appears to be restricted to the malignant cells of affected subjects in either sporadic or familial HNPCC.

The second instance of repeat copy number associated disease involves a 24 base pair repeat coding for an octapeptide in the prion protein (PrP) gene.³⁴ A variety of mutations in this gene cause Creutzfeldt-Jakob disease and Gerstmann-Straussler-Scheinker syndrome. The repeat copy number is normally four with insertions of four copies or deletions of one copy having no phenotypic effect; however, insertions of five or more extra copies of the repeat result in neurological disease of variable severity. No instability of the repeats has been noted within affected pedigrees although the sample size at this stage is very small.

While it might be premature to include the phenomena observed in HNPCC and in the PrP gene under the heading of dynamic mutation, it is important to be open to the possibility that this mutation mechanism is not restricted to just trinucleotide repeats.

How common are dynamic mutations?

There are now enough dynamic mutations for a number of general properties of this type of mutation to have become apparent. (1) The severity of the disease, which includes age of onset, can be related to the number of copies of the repeat and is highly variable. (2) Departures from mendelian inheritance are common. (3) Founder chromosomes which reflect either very low rates of mutation or haplotypes which are predisposed to mutation are seen. Certainly, new mutations are very rare. (4)

The diseases are virtually or completely mutationally homogeneous. (5) Parent of origin effects are common.

The six known dynamic mutations have all been recognised within the last three years. The common properties of these mutations, especially the departures from mendelian inheritance and the highly variable phenotypes with wide ranging ages of onset, may give some clues to other diseases which have similar mutational mechanisms. Because anticipation as a legitimate form of inheritance was so effectively discredited by Penrose,¹¹ it has rarely been mentioned in genetic publications since. A careful study, especially of clinically variable diseases, perhaps concentrating on those with a neurological abnormality since all the dynamic mutations so far have been in this region, looking for anticipation may give further clues. Care will need to be taken in such studies not to ignore Penrose completely,¹¹ since ascertainment bias may in some instances mimic anticipation. Similarly environmental effects which could possibly produce anticipation, especially in psychiatric disorders, will need to be taken into account. The discovery of dynamic mutation should now give rise to careful segregation analyses which in turn should provide new disorders which are candidates for dynamic mutation.

- 1 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.
- 2 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.
- 3 Yu S, Pritchard M, Kremer E, *et al.* Fragile X genotype characterised by an unstable region of DNA. *Science* 1991;252:1179-81.
- 4 Richards RI, Sutherland GR. Dynamic mutations: a new class of mutations causing human disease. *Cell* 1992;70:709-12.
- 5 Sutherland GR, Haan EA, Kremer E, *et al.* Hereditary unstable DNA: a new explanation for some old genetic questions. *Lancet* 1991;338:289-92.
- 6 Aslanidis C, Jansen G, Amemiya C, *et al.* Cloning of the essential myotonic dystrophy region and mapping of the putative defect. *Nature* 1992;355:548-51.
- 7 Buxton J, Shelbourne P, Davies J, *et al.* Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature* 1992;355:547-8.
- 8 Harley HG, Brook JD, Rundle SA, *et al.* Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* 1992;355:545-6.
- 9 Harper PS, Harley HG, Reardon W, Shaw DJ. Anticipation in myotonic dystrophy: new light on an old problem. *Am J Hum Genet* 1992;51:10-16.
- 10 Sutherland GR, Richards RI. Anticipation legitimised: unstable DNA to the rescue. *Am J Hum Genet* 1992;51:7-9.
- 11 Penrose LS. The problem of anticipation in pedigrees of dystrophia myotonica. *Ann Eugen* 1948;14:125-32.
- 12 Knight SJL, Flannery AN, Hirst MC, *et al.* Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell* 1993;74:127-34.
- 13 Snell RG, Doucette-Stamm LA, Gillespie KM, *et al.* The isolation of cDNAs within the Huntington disease region by hybridisation of yeast artificial chromosomes to a cDNA library. *Hum Mol Genet* 1993;2:305-9.
- 14 Reyniers E, Vits L, De Boulle K, *et al.* The full mutation in the FMR-1 gene of male fragile X patients is absent in their sperm. *Nature Genet* 1993;4:143-6.
- 15 Duyao M, Ambrose C, Myers R, *et al.* Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature Genet* 1993;4:387-92.
- 16 Lavedan C, Hofmann-Radvanyi H, Shelbourne P, *et al.* Myotonic dystrophy: size- and sex-dependent dynamics of CTG meiotic instability and somatic mosaicism. *Am J Hum Genet* 1993;52:875-83.
- 17 Mulley JC, Staples A, Donnelly A, *et al.* Explanation for exclusive maternal origin for the congenital form of myotonic dystrophy. *Lancet* 1993;341:236-7.
- 18 Jansen G, Bartolomei M, Kalscheuer V, *et al.* No imprinting involved in the expression of DM-kinase mRNAs in mouse and human tissues. *Hum Mol Genet* 1993;2:1221-7.
- 19 Orr HT, Chung M, Banfi S, *et al.* Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nature Genet* 1993;4:221-5.
- 20 Harley HG, Rundle SA, Reardon W, *et al.* Unstable DNA sequence in myotonic dystrophy. *Lancet* 1992;339:1125-8.
- 21 Hunter A, Tsilfidis C, Metther G, *et al.* The correlation of age of onset with CTG trinucleotide repeat amplification in myotonic dystrophy. *J Med Genet* 1992;29:774-9.
- 22 Andrew SE, Goldberg YP, Kremer B, *et al.* The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nature Genet* 1993;4:398-403.
- 23 La Spada A, Roling DB, Harding AE, *et al.* Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. *Nature Genet* 1992;2:301-4.
- 24 Richards RI, Holman K, Friend K, *et al.* Evidence of founder chromosomes in fragile X syndrome. *Nature Genet* 1992;1:257-60.
- 25 Imbert G, Kretz C, Johnson K, Mandel JL. Origin of the expansion mutation in myotonic dystrophy. *Nature Genet* 1993;4:72-6.
- 26 MacDonald ME, Ambrose CM, Duyao MP, *et al.* A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993;72:971-83.
- 27 Fu YH, Friedman DL, Richards S, *et al.* Decreased expression of myotonin-protein kinase messenger RNA and protein in adult form of myotonic dystrophy. *Science* 1993;260:235-8.
- 28 Sabouri LA, Mahadevan MS, Narang M, *et al.* Effect of the myotonic dystrophy (DM) mutation on mRNA levels of the DM gene. *Nature Genet* 1993;4:233-8.
- 29 Richards RI, Holman K, Yu S, Sutherland GR. Fragile X syndrome unstable element, p(CCG)_n, and other simple tandem repeat sequences are binding sites for specific nuclear proteins. *Hum Mol Genet* 1993;2:1429-35.
- 30 Mattaj JW. RNA recognition: a family matter? *Cell* 1993;73:837-40.
- 31 Engelbrecht J, Voelkel-Meiman K, Roeder GS. Meiosis-specific RNA splicing in yeast. *Cell* 1991;66:1257-68.
- 32 Ionov Y, Peinado MA, Malkhosyan S, *et al.* Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* 1993;363:558-61.
- 33 Peltomaki P, Aaltonen LA, Sistonen P, *et al.* Genetic mapping of a locus predisposing to human colorectal cancer. *Science* 1993;260:810-12.
- 34 Palmer MS, Collinge J. Mutations and polymorphisms in the prion protein gene. *Hum Mutat* 1993;2:168-73.