Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 3636-3641, April 1995

Review

Simple tandem DNA repeats and human genetic disease

(fragile X syndrome/trinucleotide repeats/chromosomal fragile sites/spinocerebellar ataxias/myotonic dystrophy)

Grant R. Sutherland and Robert I. Richards

Centre for Medical Genetics, Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, Adelaide 5006 Australia

ABSTRACT **The human genome contains many repeated DNA sequences that vary in complexity of repeating unit from a single nucleotide to a whole gene. The repeat sequences can be widely dispersed or in simple tandem arrays. Arrays of up to 5 or 6 nt are known as simple tandem repeats, and these are widely dispersed and highly polymorphic. Members of one group of the simple tandem repeats, the trinucleotide repeats, can undergo an increase in copy number by a process of dynamic mutation. Dynamic mutations of the CCG trinucleotide give rise to one group of fragile sites on human chromosomes, the rare folate-sensitive group. One member of this group, the fragile X (FRAXA) is responsible for the most common familial form of mental retardation. Another member of the group FRAXE is responsible for a rarer mild form of mental retardation. Similar mutations of AGC repeats give rise to a number of neurological disorders. The expanded repeats are unstable between generations and somatically. The intergenerational instability gives rise to unusual patterns of inheritance—particularly anticipation, the increasing severity and/or earlier age of onset of the disorder in successive generations. Dynamic mutations have been found only in the human species, and possible reasons for this are considered. The mechanism of dynamic mutation is discussed, and a number of observations of simple tandem repeat mutation that could assist in understanding this phenomenon are commented on.**

The human genome contains many nucleotide sequences that occur repeatedly. These repeat sequences vary in complexity from complete genes (such as the ribosomal RNA genes) down to simple sequences of one or a few base pairs. The physical organization of repeat sequences can vary from widely dispersed copies of a relatively long, complex sequence to tandem arrays of simple sequence composition. Among the simplest and most common repeats are the dinucleotide repeats, primarily having the bases AC on one DNA strand and GT on the other. There are other such simple tandem repeats (STRs), which involve mono-, tri-, tetra-, and pentanucleotide repeating units.

Many STR repeat sequences are polymorphic in copy number in human populations. These are therefore a rich source of DNA polymorphisms that have been exploited widely for studies of the human genome. Many of the STRs have imperfections in the repeating unit and, in general, the degree of instability of such repeats is directly related to the length of perfect repeat (1). No definite function has yet been ascribed to STRs although sequence-specific DNA-binding proteins have been identified for the di- and trinucleotide repeats (2), and one of the repeats can act as a preferential site of nucleosome assembly *in vitro* (3).

One feature of the trinucleotide STRs is their ability to undergo dynamic mutation. Dynamic mutation (4) is a process of change in genetic material that can occur over several generations. It is distinguished from conventional (or static) mutational events by a number of properties: (*i*) The product of a dynamic mutation has a different risk of undergoing further change than the original DNA sequence. (*ii*) The probability of dynamic mutation of an STR is a function of the number of perfect repeating units. (*iii*) The dynamic mutation of a DNA-repeat sequence from being a harmless copy-number polymorphism to a disease causing unstable DNA sequence is a process typically involving multiple (sometimes small) changes and is thus not a single event.

There are only two trinucleotide repeats (of the possible 10) that have been demonstrated to undergo dynamic mutation resulting in genetic disease. These are CCG and AGC repeats. (We designate STR sequences in the 5′–3′ direction of the DNA molecule and in alphabetical order. This minimizes the number of apparent STR sequences—without this designation a single trinucleotide unit can be written up to six different ways.) The CCG repeats are responsible for one group of fragile sites on chromosomes and can expand to very high copy numbers, in excess of 1000 copies. The AGC repeats are involved in a number of neurological disorders. They can also expand to high-copy numbers when in the untranslated region of a gene (as in myotonic dystrophy) but are mostly in coding regions where their copy number is usually <100. This review will discuss the trinucleotide expansions leading to disease, examine the situation in other species, and look at other changes to STRs that might shed some light on the significance of these repeats and the mechanisms by which they expand and give rise to disease.

Fragile Sites

The first genetic disease reported in association with the dynamic mutation of an STR was fragile X syndrome, the most common form of familial mental retardation (5, 6). This disease is due to the dynamic mutation of a CCG trinucleotide repeat in the 5′-untranslated region of the *FMR1* gene (7–9). This repeat is polymorphic with normal X chromosomes, having from 6 to ≈50 copies of the repeat, typically with one or two imperfections within the sequence (10–13). At a length above ≈50 copies the repeat has a readily observable rate of dynamic mutation and can increase via a number of steps, each of higher probability, through a clinically innocuous premutation phase where the repeat increases up to ≈230 copies. From within the premutation range the copy number can increase within a single generation to >230 copies and become a full mutation (14)—although the chance of such large expansion depends upon copy number such that carriers of high copy-number premutations are more likely to have affected children (5, 15). In a full mutation the CpG residues in the repeat itself and in an adjacent CpG island (which is part of the promoter region of the *FMR1* gene) become methylated (6). The consequences of this are that transcription of the *FMR1* gene ceases (16), males with this full mutation have the mental retardation and minor dysmorphisms that constitute fragile X syndrome, and a fragile site on the X chromosome at the CCG repeat can be demonstrated. Females with the full mutation have the phenotypic consequences of this ameliorated, probably because they also have a normal X chromosome producing functional *FMR1* gene product in those cells in which the fragile X chromosome is inactivated. Hence ≈60% of females with a

Abbreviations: STR, simple tandem repeat; HNPCC, hereditary nonpolyposis colon cancer; SCA, spinocerebellar ataxia.

full mutation have features of fragile X syndrome, but the degree of intellectual handicap is less severe than that of males with this disorder (17).

The *FMR1* gene encodes an RNA-binding protein (18, 19) that is widely expressed in embryonic life and postnatally is strongly expressed in testis, uterus, and brain. Differential RNA splicing produces a number of different forms of the protein in different tissues. Why *FMR1* genes with full mutations become methylated and cease to express FMR1 protein and how lack of this protein causes fragile X syndrome is unclear. A mouse model of fragile X syndrome has been created by knockout of the *FMR1* gene, and the mice have phenotypic features consistent with human fragile X syndrome (20).

There are several requirements for the cytogenetic expression of the fragile site at the CCG repeat. These requirements include expansion of the repeat, methylation of its CpG residues, and culture of cells in medium relatively deficient in thymidine or deoxycytidine. It remains a mystery why the fragile site is rarely expressed in $>\approx 50\%$ of metaphases (and frequently in a much lower proportion), and it is also unclear why a relative deficiency of thymidine has an effect on a DNA sequence devoid of this base. It may be that the effect is due to nucleotide pool imbalances with the low level of thymidine producing a low level of guanosine (21). It is also unclear what the fragile site is at the molecular level. It represents DNA that is not packaged for mitosis, but it is not known whether this is due to under-replication, to a change in local protein-binding properties, or to some other factor.

The timing of amplification of the CCG repeat from its premutation size to a full mutation is unclear. In fragile X syndrome, males only transmit premutations in sperm (22), and these never amplify to full mutations postzygotically. The transition from pre to full mutation only occurs on female transmission of the sequence, but it is unclear whether this is a pre- or postzygotic event. It has been argued that this is a postzygotic event but that germline precursor cells are in some way protected from repeat expansion (23). Certainly the full mutation can be modified postzygotically, and many individuals with full mutations have different numbers of copies of the repeat in different cells within a single tissue (revealed as multiple bands or smears on Southern blot). The accumulated evidence suggests that somatic instability is limited to a short period very early in development (23). If the female transmits only a premutation in her ovum and it expands to a full mutation postzygotically, some form of imprinting must be invoked because male-transmitted premutations do not undergo such a process. The pattern of expansion seen in extraembryonic tissues is very similar to that in embryonic tissue, although methylation of the CpG residues occurs later in development in the extraembryonic tissues. This result indicates either that there is some factor within the cells which determines that the pattern of differential expansion of the repeat within each tissue will be similar or that the expansion has occurred before separation of the embryonic and extraembryonic cell lineages. If this latter situation, which seems most likely, has occurred, it is difficult to see how the gonadal precursor cells could escape such amplification.

The likely mechanism of inheritance of fragile X syndrome is that the change from pre to full mutation occurs during oogenesis, that the full mutation is transmitted via the ovum, and that length instability of full mutations occurs in the very early postzygotic cell divisions. Furthermore, in males the length of the repeat is reduced to or remains within the premutation range during spermatogenesis. This hypothesized mechanism of repeat expansion is consistent with what is known about the repeat sequence, and it does not require any imprinting mechanism to be invoked. Meiosis is a very different process in the two sexes, and these differences may be adequate to explain the different behaviors of the same sequence in these very different environments. Transmission of the autosomal fragile site *FRA16A* exhibits an important distinction from *FRAXA* in that stable transmission of the *FRA16A* full mutation by males is observed (24). The behavior of dynamic mutations of AGC repeats in other diseases appears to differ at meiosis in the two sexes (see below).

The properties of the fragile X dynamic mutation explain its unusual inheritance pattern. Males have been observed to only transmit premutations, and these do not usually change much in size. Hence all the daughters of male premutation carriers have a premutation about the same size as that of their father, but they never have full mutations or, consequently, features of fragile X syndrome. Sherman (25, 26, 86) delineated the paradox which bears her name—that the mothers and daughters of males with premutations have very different risks of having children with fragile X syndrome. This was inexplicable in terms of static mutation and classical genetics. It is now known that these two groups of women are at different stages of progression of the dynamic mutation. The mothers of the males with premutations have, on average, smaller premutations than the daughters of these males. The premutation usually increases in size when transmitted from the mothers to their sons but then goes relatively unchanged to the daughters. Because the risk of a premutation changing to a full mutation on transmission by a woman is a function of its size, the paradox is explained (5, 15). All full mutations observed have arisen from a carrier of a premutation. The transition from normal copy number to full mutation has never been recorded.

Three other folate-sensitive fragile sites (*FRAXE, FRAXF*, and *FRA16A*) have been characterized (24, 27, and 28) and are all dynamic mutations of polymorphic (CCG) trinucleotide repeats with very similar properties to the fragile X repeat. The genes (if any) associated with these fragile sites have not been identified. *FRAXE* appears to be associated with a mild form of mental retardation (29, 30). *FRAXF* is without phenotypic effect (28). *FRA16A* is also without phenotypic effect in heterozygotes; homozygotes have not been recorded.

It is of interest that *FRA16A* is in a region of the genome not normally subject to CpG methylation, yet the expanded repeat in those who express the fragile site is methylated. It had earlier been argued (31) that the *FRAXA* mutation was due to failure to erase the DNA methylation imprint that occurred as part of normal X chromosome inactivation. The studies on *FRA16A* indicate that the mutation mechanism involved in this process is not preceded by DNA methylation; instead, methylation is a consequence of the repeat expansion.

The fragile X mutation was shown to be in linkage disequilibrium with flanking DNA polymorphisms (haplotypes) (32). In classical terms, the identification of such "founder chromosomes" would indicate that the change from wild type to mutant was a very rare event and that the mutant chromosomes with the same haplotypes were all descended from a single mutational event. In terms of the mechanism of dynamic mutation the explanation could be that over-represented haplotypes are at greater risk of undergoing dynamic mutation. Investigation of this phenomenon by haplotype sequencing has shown most of the variation in CCG repeat length is at the 3′ end (with respect to transcription) of the sequence (i.e., it is polar). These observations were made possible because of occasional imperfections in the CCG motif occurring mainly toward the 5′ end of the repeat. The haplotypes that are over-represented among fragile X chromosomes, have, in their nonmutant state, longer stretches of perfect repeat, either because of their overall length or because they appear to have lost an imperfection in their repeat (10–13). A similar loss of repeat interruption has been found to be associated with increased instability of certain alleles of the *FRA16A* CCG repeat (33).

AGC Repeats

A group of human neurological disorders has been found to be due to dynamic mutation of the AGC trinucleotide repeat. In one of these disorders, myotonic dys-

trophy, the mutation is in the 3′-untranslated region of the myotonin gene on chromosome 19 (34). Spinobulbar muscular atrophy or Kennedy disease (35) was the first of a growing list of others (including those in Table 1) in which the mutations are in protein coding regions and are translated as polyglutamine tracts. Location within a coding region appears to constrain the extent of amplification. When the repeat is in an untranslated region, the constraints appear to be minimal, and several thousand copies of the repeat may be present. In the translated regions of the genes there are rarely more than 100 copies.

Myotonic Dystrophy. The finding that myotonic dystrophy was due to a dynamic mutation offered an explanation (and legitimization) of the phenomenon of anticipation (36). In this phenomenon a genetic disease becomes increasingly severe, or presents at an earlier age, in successive generations of a family. In myotonic dystrophy, the mutation carriers in the earliest generation of a family may have no discernible phenotypic abnormalities, or merely senile cataracts. However, in successive generations, there can be progressively earlier onset of muscle disease, culminating in children with congenital myotonic dystrophy. Such children are primarily born to women carriers of the myotonic dystrophy mutation, rather than to the wives of male carriers. Anticipation had been dismissed as an artefact of ascertainment bias (37) until dynamic mutation offered a molecular mechanism by which such apparently non-Mendelian genetics could occur (38).

An explanation, at least in part, for the congenital cases being the offspring of female mutation carriers lies in a difference in the behavior of the repeat during meiosis in the two sexes (39, 40). In females the size of the repeat can increase without apparent restraint from generation to generation until it reaches the point of genetic lethality (congenital myotonic dystrophy). In males the repeat reaches a maximum size of ≈1000 copies and then appears to mainly decrease in size upon transmission. Behavior of the repeat in the other disorders shows that increases in size during male transmission are much greater than during female transmission. Thus paternal anticipation is exhibited by Huntington disease, spinocerebellar ataxia (SCA) (type 1), and dentatorubral pallidoluysian atrophy.

In all the neurological disorders there is a general correlation between the degree of trinucleotide expansion and the severity (probably in myotonic dystrophy only) or age of disease onset. (This contrasts with the fragile X where there is a threshold effect—i.e., once the full mutation is present, full expression of the syndrome can occur.) However, the age range of onset for any given number of copies of the repeat is so wide that copy number cannot be used as a specific prognostic indicator in individuals. The most direct relationship between copy number and age of onset is shown for SCA type 1 (41). The mechanism by which increased AGC copy number results in disease is not clear. There have been conflicting reports of increased (42) and decreased (43) message in myotonic dystrophy.

Neurodegenerative Disorders. In the neurodegenerative diseases (Table 1) where the repeat codes for polyglutamine, it would appear that the protein has gained some disease-producing function (45). What is particularly surprising is that an increase of copy number by as little as 5% above a normal level can result in disease. Green (55) has put forward an intriguing hypothesis that can account for this effect and the relationship between copy number and age-of-onset. He hypothesized that the reiterated glutamines are involved in protein aggregation by crosslinking and that this is a slow process critically dependent upon the number of glutamine residues. Perutz *et al.* (56) have similarly suggested that protein aggregation could be the disease-causing process but hypothesize that the mechanism of aggregation is that the glutamine repeats act as polar zippers. Either process is likely to act in a dominant manner and may be quite independent of the normal function of the repeat-containing protein. If protein aggregates accumulate with time, then this could provide a molecular basis for the copy number/age-of-onset relationship.

An alternative hypothesis has been put forward (57) based on the observation that variation in length of polyglutamine tracts can have profound effects on the function of certain transcription factors. One of these, the androgen receptor, is the site of p(AGC) amplification in spinobulbar muscular atrophy (35). Similar gain-of-function for the other targets of polyglutamine expansion may become a clearer possibility with the characterization of the normal cellular role of these proteins.

Linkage disequilibrium has been shown for some of the dynamic mutations. The myotonic dystrophy mutation is in total linkage disequilibrium with a nearby insertion/deletion polymorphism. It would appear that an ancient mutational event on a chromosome containing the insertion took the number of copies on it from 5 to somewhere between 19 and 30 and that this group of chromosomes constitutes a reservoir for recurrent increases in copy number to generate myotonic dystrophy mutations (58). It is of interest to note that these expanded alleles do not exist in some African Black populations that are also free of myotonic dystrophy (59).

In the fragile sites there is a clear premutational phase in which the repeat is not methylated, fragile sites are not expressed, and (for the FRAXA group) there is no disease. In the neurological disorders there is less evidence for such premutations. In myotonic dystrophy, alleles with 50–80 copies of the repeat are associated with nonpenetrance or very minimal expression, and these have been referred to as a "protomutation" (60). In Huntington disease alleles of 30–37 repeats, which are above the top of the normal range but below the disease range, have been referred to as intermediate alleles (61). These alleles may not cause readily detectable disease but are liable to increase upon transmission by a few copies and produce affected individuals who might be regarded as having a "new" mutation.

Locus Heterogeneity in SCAs. One of the striking features of the dynamic mutation disorders had been their homogeneity. The vast majority of fragile X syndrome cases and all of the myotonic dystrophy, spinobulbar muscular atrophy, and Huntington disease cases are due to a single site of mutation in a single gene. The autosomal dominant ataxias exhibit clinical and locus heterogeneity (Table 1). The clinical definition of syndromes that fall in this category is not yet clear—the distinction between certain subtypes being difficult at best. With this caveat in mind there appear to be at least five loci for

Table 1. Locus and clinical heterogeneity in neurodegenerative disorders

Disease	Gene	AGC repeat	Location	Ref.
Kennedy disease	Androgen receptor	+	Xq12	35
Huntington disease	Huntingtin	+	4p16.3	44
DRPLA	Atrophin	+	12p12	45, 46
Haw River syndrome	Atrophin	+	12p12	47
DRPLA-like	?	?	14q24.3	48
SCAs				
SCA1	Ataxin	+	6p22-p23	41
SCA2	?	?	12q23-24	49, 50
SCA3	?	?	14q24.3-32.1	51
SCA4	?	?	16q24	52
SCA5	?	?	11cen	53
Machado Joseph disease	(ORF)	+	14q24.3-32.1	54

DRPLA, dentatorubral pallidoluysian atrophy.

SCA. *SCA1* has been characterized and found to consist of an unstable AGC repeat. The mutation causing an apparently distinct form of autosomal dominant ataxia, Machado Joseph disease, is also due to an unstable AGC repeat (54). This locus maps at the same 14q chromosomal region as *SCA3*, and it will be of interest to see whether both of these disorders are due to mutation in the same gene. A family exhibiting a dentatorubral pallidoluysian atrophy-like phenotype also maps to 14q24.3-qter, suggesting that clinical heterogeneity may extend even further (48), although the normally accepted statistical evidence of this map position was not achieved. Finally Haw River syndrome, which is characterized by ataxia, chorea, seizures, and dementia in a five-generation African-American family, has recently been shown to be associated with the same AGC repeat instability in the gene at 12p12-pter that causes dentatorubral pallidoluysian atrophy in the Japanese. Either the genetic background in which these mutations occur or polymorphisms in the coding regions of the disease-causing genes are responsible for the apparently distinct phenotypes. DNA-based tests offer a means of differential diagnosis in the SCAs. The cloning of the additional SCA loci is therefore eagerly awaited. Oddly enough, although the Machado Joseph disease gene was found to be a member of a multigene family, none of the other members of this family were found to map to any of the other known SCA loci (54).

Animal Models

There are no known animal equivalents of the human dynamic mutations of trinucleotide repeats. The mouse homologues of each of the human genes containing an unstable repeat are notable for their shorter length repeat and lower polymorphism. The mouse Huntington disease gene homologue encodes only seven glutamines in an interrupted $(CAG)_2CAA(CAG)_4$ repeat that is not polymorphic (62). Similarly the mouse ataxin homologue (*Sca1*) has only two CAG repeats, which suggests that the polyglutamine tract is not essential for the normal function of this protein (63). The apparent lack of unstable repeats in animal genomes and the greater stability of repeats at homologous loci could suggest that the mechanism of dynamic mutation may largely be restricted to the human genome. However, little is likely to be known about late-onset diseases in animal populations. On the assumption that intracellular metabolism varies little between species, the late-onset dynamic mutation diseases of humans may not have time to manifest in species with substantially shorter lifespans.

There may be, however, nonhuman examples of repeat sequence changes associated with phenotypic effect. The nonobese diabetic (NOD) mouse has the number of copies of glutamine (the AGC repeat again!) in the N-terminal region of interleukin 2 increased from 8 to 12. It is unclear whether this is simply a strain difference in a sequence that would be polymorphic in outbred populations or is related to the disease process. There are other differences between the interleukin 2 gene in the NOD and nondiabetic strains (64) that could also have accounted for the NOD phenotype.

The sex-determining gene (*Sry*) of the mouse contains an AGC repeat that codes for polyglutamine. Differences in the number of glutamine residues (over the range 11–13) are associated with different degrees of sex reversal when introduced into different genetic backgrounds (65). Whether these differences in repeat number are the cause of the sex reversal or are markers for other differences in the *Sry* gene remain to be elucidated (66).

Adaptive Mutation

Several years ago Cairns and Foster (67), in what became a controversial study, demonstrated that bacteria that were unable to digest lactose preferentially acquired mutations to enable them to become lactose utilizers. The molecular basis for this phenomenon was found to be a distinct form of mutation. The majority of mutations that appeared during selection by nutritional stress were due to 1-base deletions in runs of iterated bases. This mechanism has been referred to as adaptive mutation and is thought to function through polymerase errors that persist as a result of a deficiency in mismatch repair (68, 69). The similarity with the basis for mutations associated with hereditary nonpolyposis colon cancer (HNPCC) is striking (see below). Recently it has been noted that a proportion of mutations in the C1 inhibitor gene cluster at a short trinucleotide-repeat sequence (70). This sequence contains three direct repeats of the triplet CAA, and a similar slipped mismatch-repair mechanism has been proposed to account for the observed mutations in this repeat sequence. A common pathway of mutation may well be responsible for each of these observations (in HNPCC, dynamic mutation, and adaptive mutation). It could be hypothesized that cells that contain intragenic repeats which can be subjected to this mutational mechanism have a selective advantage in circumstances where mutation is required to allow for or enhance survival. In other words the burden of susceptibility to disease-producing mutation is offset by the ability to undergo adaptive mutation under conditions of environmental stress.

Mutational Mechanisms

The molecular mechanism of simple repeat-sequence instability is complex and may differ in detail from one repeat to the next; however, recent observations of this form of instability have started to shed some light on common aspects of the process and therefore on its molecular basis. In tumor tissue from patients with HNPCC there is instability of many of the STRs with the presence of novel alleles resulting from increases and decreases in copy number (71–73). This instability is "genome-wide" and is evident when the copy number of repeat sequences in malignant tissue is compared with that of normal tissue in affected individuals. HNPCC loci were mapped genetically to chromosomes 2 and 3 (74, 75). At about this time it was also found that mutations in yeast *MLH1* and *MSH2* genes (homologues of the bacterial *mutL* or *mutS* genes that are normally involved in mismatch repair) severely affect instability of repeat sequences (76). By positional cloning, the HNPCC gene on chromosome 2 was found to be h*MSH2*, the human homologue of *mutS* (77, 78). With this information (and in one instance cDNA data-base searches) it was possible to pinpoint the chromosome 3 HNPCC gene as h*MLH1*, the human homologue of *mutL* (79, 80). Whether this form of mutation has specific repeat sequence-containing targets (associated with the transformation process) or whether a higher general rate of mutation results in the nonspecific activation of oncogenes and/or inactivation of tumor suppressors is not yet clear. One implication of these findings is that it is the various components of mismatch repair that are key trans-acting factors in the dynamic mutation process. The HNPCC mutations allow normal flaws in the replication of simple tandem-repeat sequences to manifest as changes in repeat copy number through failure to repair slippage-induced mismatch. One of the unusual aspects of the HNPCC mutations is their apparent tissue specificity (81). Tissue-specific repeat instability is also exhibited by the myotonic dystrophy (82) and Huntington disease (83) loci. In both cases the target tissues for these disorders exhibit increased repeat copy number compared with other tissues.

In addition to the trans-acting factors associated with simple repeat sequence instability, the observation of founder chromosome effects is strong evidence that the repeat sequences themselves (or sequences closely linked to them) play a cis-acting role in the mutation process. For the AC dinucleotide repeats the instability of these sequences (manifest as copy-number polymorphism) is proportional to their perfect repeat length (1). At the *SCA1* locus the unstable longer copy-number repeats (associated with the dis-

ease phenotype) exhibit loss of repeat-sequence interruption (84). Similar findings at the *FRAXA* (10–13) and *FRA16A* (33) loci suggest that this is a common determinant of repeat-sequence instability.

A simple slip-sliding mechanism of repeat-copy-number mutation can account for the cis and trans components of the mutation process and the very rapid increase in expansion that occurs when the repeat length exceeds that of an Okazaki fragment (85). Such a fragment, when completely contained within the repeat, could become an unanchored primer for extended polymerization, resulting in the massive increases in copy number associated with the fragile sites and the myotonic dystrophy expanded repeat.

Conclusions

The study of DNA repeat sequences, especially in humans, has in recent years provided a rich source of polymorphisms for linkage studies, a means of individual identification, and various approaches to loss-of-gene function in tumors. Furthermore, a group of human diseases has been shown to result from a dynamic mutation mechanism that was previously unknown. Dynamic mutation has been found to have a fundamental role in human genetics, accounting for what appeared to be non-Mendelian patterns of inheritance. Understanding the molecular basis of repeat-sequence instability will hopefully enable rational intervention in the process to overcome dysfunction.

We thank our colleagues Drs. Eric Haan, Graeme Suthers, and Duncan Hewett for helpful comments on this manuscript. R.I.R. thanks Shelley Richards for support and encouragement. G.R.S. is an International Research Scholar of the Howard Hughes Medical Institute. Other support for this work came from the Australian National Health and Medical Research Council.

1. Weber, J. L. (1990) *Genomics* **7,** 524–530.
2. Richards, R. I., Holman, K., Yu, S. & Sutherland, G. R. (1993) *Hum. Mol. Genet.* **2,** 1429–1435.
3. Wang, Y.-H., Amirhaeri, S., Kang, S., Wells, R. D. & Griffith, J. D. (1994) *Science* **265,** 669–671.
4. Richards, R. I. & Sutherland, G. R. (1992) *Cell* **70,** 709–712.
5. Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, J. C., Warren, S. T., Schlessinger, D., Sutherland, G. R. & Richards, R. I. (1991) *Science* **252,** 1179–1181.
6. Oberlé, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M. F. & Mandel, J.-L. (1991) *Science* **252,** 1097–1102.
7. Verkerk, A. J. M. H., Pieretti, M., Sutcliffe, J. S., Fu, Y.-H., Kuhl, D. P. A., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F., Eussen, B. E., van Ommen, G.-J., Blonden, L. A. J., Riggins, G. J., Chastain, J. L., Kunst, C. B., Galjaard, H., Caskey, C. T., Nelson, D. L., Oostra, B. A. & Warren, S. T. (1991) *Cell* **65,** 905–914.
8. Kremer, E. J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S. T., Schlessinger, D., Sutherland, G. R. & Richards, R. I. (1991) *Science* **252,** 1711–1714.
9. Yu, S., Mulley, J., Loesch, D., Turner, G., Donnelly, A., Gedeon, A., Hillen, D., Kremer, E., Lynch, M., Pritchard, M., Sutherland, G. R. & Richards, R. I. (1992) *Am. J. Hum. Genet.* **50,** 968–980.
10. Kunst, C. B. & Warren, S. T. (1994) *Cell* **77,** 853–861.
11. Eichler, E. E., Holden, J. J. A., Popovich, B. W., Reiss, A. L., Snow, K., Thibodeau, S. N., Richards, C. S., Ward, P. A. & Nelson, D. L. (1994) *Nat. Genet.* **8,** 88–94.
12. Snow, K., Tester, D. J., Kruckeberg, K. E., Schaid, D. J. & Thibodeau, S. N. (1994) *Hum. Mol. Genet.* **3,** 1543–1551.
13. Hirst, M. C., Grewal, P. K. & Davies, K. E. (1994) *Hum. Mol. Genet.* **3,** 1553–1560.
14. Reiss, A. L., Kazazian, H. H., Krebs, C., McAughan, A., Boehm, C. D., Abrams, M. T. & Nelson, D. L. (1994) *Hum. Mol. Genet.* **3,** 393–398.
15. Fu, Y.-H., Kuhl, D. P. A., Pizzuti, A., Pieretti, M., Sutcliffe, J. S., Richards, S., Verkerk, A. J. M. H., Holden, J. J., Fenwick, R. G., Warren, S. T., Oostra, B. A., Nelson, D. L. & Caskey, C. T. (1991) *Cell* **67,** 1047–1058.
16. Pierreti, M., Zhang, F., Fu, Y.-H., Warren, S. T., Oostra, B. A., Caskey, C. T. & Nelson, D. L. (1991) *Cell* **66,** 817–822.
17. Rousseau, F., Heitz, D., Biancalana, V., Blumenfeld, S., Kretz, C., Boue, J., Tommerup, N., Van Der Hagen, C., DeLozier-Blanchet, C., Croquette, M.-F., Gilgenkrantz, S., Jalbert, P., Voelckel, M.-A., Oberlé, I. & Mandel, J.-L. (1991) *N. Engl. J. Med.* **325,** 1673–1681.
18. Ashley, C. T., Wilkinson, K. D., Reines, D. & Warren, S. T. (1993) *Science* **262,** 563–566.
19. Siomi, H., Siomi, M. C., Nussbaum, R. L. & Dreyfuss, G. (1993) *Cell* **74,** 291–298.
20. Dutch-Belgian Fragile X Consortium (1994) *Cell* **78,** 23–33.
21. Kunz, B. A. (1982) *Environ. Mutagen.* **4,** 695–725.
22. Reyniers, E., Vits, L., De Boulle, K., Van Roy, B., Van Velzen, D., de Graaff, E., Verkerk, A. J. M. H., Jorens, H. Z. J., Darby, J. K., Oostra, B. & Willems, P. J. (1993) *Nat. Genet.* **4,** 143–146.
23. Wohrle, D., Hennig, I., Vogel, W. & Steinbach, P. (1993) *Nat. Genet.* **4,** 140–142.
24. Nancarrow, J. K., Kremer, E., Holman, K., Eyre, H., Doggett, N. A., Le Paslier, D., Callen, D. F., Sutherland, G. R. & Richards, R. I. (1994) *Science* **264,** 1938–1941.
25. Sherman, S. L., Morton, N. E., Jacobs, P. A. & Turner, G. (1984) *Ann. Hum. Genet.* **48,** 21–37.
26. Sherman, S. L., Jacobs, P. A., Morton, M. E., Froster-Iskenius, U., Howard-Peebles, P. N., Nielsen, K. B., Partington, N. W., Sutherland, G. R., Turner, G. & Watson, M. (1985) *Hum. Genet.* **69,** 289–299.
27. Knight, S. J. L., Flannery, A. V., Hirst, M. C., Campbell, L., Christodoulou, Z., Phelps, S. R., Pointon, J., Middleton-Price, H. R., Barnicoat, A., Pembrey, M. E., Holland, J., Oostra, B. A., Bobrow, M. & Davies, K. E. (1993) *Cell* **74,** 127–134.
28. Parrish, J. E., Oostra, B. A., Verkerk, A. J. M. H., Richards, C. S., Reynolds, J., Spikes, A. S., Shaffer, L. G. & Nelson, D. L. (1994) *Nat. Genet.* **8,** 229–235.
29. Hamel, B. C. J., Smits, A. P. T., de Graaff, E., Smeets, D. F. C. M., Schoute, F., Eussen, B. H. J., Knight, S. J. L., Davies, K. E., Assman-Hulsmans, C. F. C. H. & Oostra, B. A. (1994) *Am. J. Hum. Genet.* **55,** 923–931.
30. Mulley, J. C., Yu, S., Loesch, D. Z., Hay, D. A., Donnelly, A., Gedeon, A. K., Carbonell, P., Lopez, I., Glover, G., Gabarron, J., Yu, P. W. L., Baker, E., Haan, E. A., Hocket, A., Knight, S. J. L., Davies, K. E., Richards, R. I. & Sutherland, G. R. (1995) *J. Med. Genet.*, in press.
31. Laird, C. D. (1987) *Genetics* **117,** 587–599.
32. Richards, R. I., Holman, K., Friend, K., Kremer, E., Hillen, D., Staples, A., Brown, W. T., Goonewardena, P., Tarleton, J., Schwartz, C. & Sutherland, G. R. (1992) *Nat. Genet.* **1,** 257–260.
33. Nancarrow, J. K., Holman, K., Mangelsdorf, M., Hori, T., Denton, M., Sutherland, G. R. & Richards, R. I. (1995) *Hum. Mol. Genet.*, in press.
34. Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J.-P., Hudson, T., Sohn, R., Zemelman, B., Snell, R. G., Rundle, S. A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P. S., Shaw, D. J. & Housman, D. E. (1992) *Cell* **68,** 799–808.
35. La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E. & Fischbeck, K. H. (1991) *Nature (London)* **352,** 77–79.
36. Sutherland, G. R., Haan, E. A., Kremer, E., Lynch, M., Pritchard, M., Yu, S. & Richards, R. I. (1991) *Lancet* **358,** 289–292.
37. Penrose, L. (1948) *Ann. Eugen.* **14,** 125–132.
38. Harper, P. S., Harley, H. G., Reardon, W. & Shaw, D. J. (1992) *Am. J. Hum. Genet.* **51,** 10–16.
39. Lavedan, C., Hofmann-Radvanyi, H., Shelbourne, P., Rabes, J.-P., Duros, C., Savoy, D., Dehaupas, I., Luce, S., Johnson, K. & Junien, C. (1993) *Am. J. Hum. Genet.* **52,** 875–883.
40. Mulley, J. C., Staples, A., Donnelly, A., Gedeon, A. K., Hecht, B. K., Nicholson, G. A., Haan, E. A. & Sutherland, G. R. (1993) *Lancet* **341,** 236–237.
41. Orr, H. T., Chung, M.-Y., Banfi, S., Kwiatkowski, T. J., Jr., Servadio, A., Beaudet, A. L., McCall, A. E., Duvick, L. A., Ranum, L. P. W. & Zoghbi, H. Y. (1993) *Nat. Genet.* **4,** 221–226.
42. Sabouri, L. A., Mahadevan, M. S., Narang, M., Lee, D. S. C., Surh, L. C. & Korneluk, R. G. (1993) *Nat. Genet.* **4,** 233–238.
43. Fu, Y.-H., Friedman, D. L., Richards, S., Pearlman, J. A., Gibbs, R. A., Pizzuti, A., Ashizawa, T., Perryman, M. B., Scarlato, G., Fenwick, R. G. & Caskey, C. T. (1993) *Science* **260,** 235–238.

44. Huntington's Disease Collaborative Research Group (1993) *Cell* **72,** 971–983.
45. Nagafuchi, S., Yanagisawa, H., Sato, K., Shirayama, T., Ohsaki, E., Bundo, M., Takeda, T., Tadokoro, K., Kondo, I., Murayama, N., Tanaka, Y., Kikushima, H., Umino, K., Kurosawa, H., Furukawa, T., Nihei, K., Inoue, T., Sano, A., Komure, O., Takahashi, M., Yoshizawa, T., Kanazawa, I. & Yamada, M. (1994) *Nat. Genet.* **6,** 14–18.
46. Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hayashi, T., Saito, M., Tomoda, A., Miike, T., Naito, H., Ikuta, F. & Tsuji, S. (1994) *Nat. Genet.* **6,** 9–13.
47. Burke, J. R., Wingfield, M. S., Lewis, K. E., Roses, A. D., Lee, J. E., Hulette, C., Pericak-Vance, M. A. & Vance, J. M. (1994) *Nat. Genet.* **7,** 521–524.
48. Cancel, G., Durr, A., Stevanin, G., Chneiweiss, H., Duyckaerts, C., Serdaru, M., de Toffol, B., Agid, Y. & Brice, A. (1994) *Nat. Genet.* **6,** 8.
49. Gispert, S., Twells, R., Orozco, G., Brice, A., Weber, J., Heredero, L., Scheufler, K., Riley, B., Allotey, R., Nothers, C., Hillermann, R., Lunkes, A., Khati, C., Stevanin, G., Hernandez, A., Magarino, C., Klockgether, T., Durr, A., Chneiweiss, H., Enczmann, J., Farrall, M., Beckmann, J., Mullan, M., Wernet, P., Agid, Y., Freund, H.-J., Williamson, R., Auburger, G. & Chamberlain, S. (1993) *Nat. Genet.* **4,** 295–299.
50. Lopes-Cendes, I., Andermann, E., Attig, E., Cendes, F., Bosch, S., Wagner, M., Gerstenbrand, F., Andermann, F. & Rouleau, G. A. (1994) *Am. J. Hum. Genet.* **54,** 774–781.
51. Stevanin, G., Le Guern, E., Ravise, N., Chneiweiss, H., Durr, A., Cancel, G., Vignal, A., Boch, A.-L., Ruberg, M., Penet, C., Pothin, Y., Lagroua, I., Haguenau, M., Rancurel, G., Weissenbach, J., Agid, Y. & Brice, A. (1994) *Am. J. Hum. Genet.* **54,** 11–20.
52. Gardner, K., Alderson, K., Galster, B., Kaplan, C., Leppert, M. & Ptacek, L. (1994) *Neurology* **44,** Suppl. 2, A361.
53. Ranum, L. P. W., Schut, L. J., Lundgren, J. K., Orr, H. T. & Livingston, D. M. (1994) *Nat. Genet.* **8,** 280–284.
54. Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M., Akiguchi, I., Kimura, J., Narumiya, S. & Kakizuka, A. (1994) *Nat. Genet.* **8,** 221–228.
55. Green, H. (1993) *Cell* **74,** 955–956.
56. Perutz, M. F., Johnson, T., Suzuki, M. & Finch, J. T. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 5355–5358.
57. Gerber, H.-P., Seipel, K., Georgiev, O., Hofferer, M., Hug, M., Rusconi, S. & Schaffner, W. (1994) *Science* **263,** 808–811.
58. Imbert, G., Kretz, C., Johnson, K. & Mandel, J.-L. (1993) *Nat. Genet.* **4,** 72–76.
59. Goldman, A., Ramsay, M. & Jenkins, T. (1994) *J. Med. Genet.* **31,** 37–40.
60. Barcelo, J. M., Mahadevan, M. S., Tsilfidis, C., MacKenzie, A. E. & Korneluk, R. G. (1993) *Hum. Mol. Genet.* **2,** 705–709.
61. Goldberg, Y. P., Kremer, B., Andrew, S. E., Theilmann, J., Graham, R. K., Squitieri, F., Telenius, H., Adam, S., Sajoo, A., Starr, E., Heiberg, A., Wolff, G. & Hayden, M. R. (1993) *Nat. Genet.* **5,** 174–179.
62. Lin, B., Nasir, J., MacDonald, H., Hutchinson, G., Graham, R. K., Rommens, J. M. & Hayden, M. R. (1994) *Hum. Mol. Genet.* **3,** 85–92.
63. Banfi, S., Servadio, A., Chung, M.-y., McCall, A. E., Orr, H. T. & Zoghbi, H. Y. (1994) *Am. J. Hum. Genet.* **55,** A17.
64. Ghosh, S., Palmer, S. M., Rodrigues, N. R., Cordell, H. J., Hearne, C. M., Cornall, R. J., Prins, J.-B., McShane, P., Lathrop, G. M., Peterson, L. B., Wicker, L. S. & Todd, J. A. (1993) *Nat. Genet.* **4,** 404–409.
65. Coward, P., Nagai, K., Chen, D., Thomas, H. D., Nagamine, C. M. & Lau, Y.-F. C. (1994) *Nat. Genet.* **6,** 245–250.
66. Eicher, E. M. (1994) *Nat. Genet.* **6,** 221–223.
67. Cairns, J. & Foster, P. L. (1991) *Genetics* **128,** 695–702.
68. Rosenberg, S. M., Longerich, S., Gee, P. & Harris, R. S. (1994) *Science* **265,** 405–407.
69. Foster, P. L. & Trimarchi, J. M. (1994) *Science* **265,** 407–409.
70. Bissler, J. J., Cicardi, M., Donaldson, V. H., Gatenby, P. A., Rosen, F. S., Sheffer, A. L. & Davis, A. E. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 9622–9625.
71. Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D. & Perucho, M. (1993) *Nature (London)* **363,** 558–561.
72. Aaltonen, L. A., Peltomaki, P., Leach, F. S., Sistonen, P., Pylkkanen, L., Mecklin, J.-P., Jarvinen, H., Powell, S. M., Jen, J., Hamilton, S. R., Peterson, G. M., Kinzler, K. W., Vogelstein, B. & de la Chapelle, A. (1993) *Science* **260,** 812–816.
73. Thibodeau, S. N., Bren, G. & Schaid, D. (1993) *Science* **260,** 816–819.
74. Peltomaki, P., Aaltonen, L. A., Sistonen, P., Pylkkanen, L., Mecklin, J.-P., Jarvinen, H., Green, J. S., Jass, J. R., Weber, J. L., Leach, F. S., Petersen, G. M., Hamilton, S. R., de la Chapelle, A. & Vogelstein, B. (1993) *Science* **260,** 810–812.
75. Lindblom, A., Tannergard, P., Werelius, B. & Nordenskjold, M. (1993) *Nat. Genet.* **5,** 279–282.
76. Strand, M., Prolla, T. A., Liskay, R. M. & Petes, T. D. (1993) *Nature (London)* **365,** 274–276.
77. Fishel, R., Lescoe, M. K., Rao, M. R. S., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M. & Kolodner, R. (1993) *Cell* **75,** 1027–1038.
78. Leach, F. S., Nicolaides, N. C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., *et al.* (1993) *Cell* **75,** 1215–1225.
79. Bronner, C. E., Baker, S. M., Morrison, P. T., Warren, G., Smith, L. G., Lescoe, M. K., Kane, M., Earabino, C., Lipford, J., Lindblom, A., Tannergard, P., Bollag, R. J., Godwin, A. R., Ward, D. C., Nordenskjold, M., Fishel, R., Kolodner, R. & Liskay, R. M. (1994) *Nature (London)* **368,** 258–261.
80. Papadopoulos, N., Nicolaides, N. C., Wei, Y.-F., Ruben, S. M., Carter, K. C., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M., Adams, M. D., Venter, J. C., Hamilton, S. R., Petersen, G. M., Watson, P., Lynch, H. T., Peltomaki, P., Mecklin, J.-P., de la Chappelle, A., Kinzler, K. W. & Vogelstein, B. (1994) *Science* **263,** 1625–1629.
81. Shibata, D., Peinado, M. A., Ionov, Y., Malkhosyan, S. & Perucho, M. (1994) *Nat. Genet.* **6,** 273–281.
82. Anvret, M., Ahlberg, G., Grandell, U., Hedberg, B., Johnson, K. & Edstrom, L. (1993) *Hum. Mol. Genet.* **2,** 1397–1400.
83. Telenius, H., Kremer, B., Goldberg, Y. P., Theilmann, J., Andrew, S. E., Zeisler, J., Adam, S., Greenberg, C., Ives, E. J., Clarke, L. A. & Hayden, M. R. (1994) *Nat. Genet.* **6,** 409–414.
84. Chung, M.-y., Ranum, L. P. W., Duvick, L. A., Servadio, A., Zoghbi, H. Y. & Orr, H. T. (1993) *Nat. Genet.* **5,** 254–258.
85. Richards, R. I. & Sutherland, G. R. (1994) *Nat. Genet.* **6,** 114–116.
86. Sherman, S. L., Jacobs, P. A., Morton, N. E., Froster-Iskenius, U., Howard-Peebles, P. N., Nielsen, K. B., Partington, M. W., Sutherland, G. R., Turner, G. & Watson, M. (1985) *Hum. Genet.* **71,** 184–186.