CMLS, Cell. Mol. Life Sci. 57 (2000) 914–931
1420-682X/00/060914-18 $ 1.50 + 0.20/0

CMLS Cellular and Molecular Life Sciences

Review

DNA repeat expansions and human disease

K. Usdin* and E. Grabczyk

Section on Genomic Structure and Function, Laboratory of Molecular and Cellular Biology, National Institute of Diabetes and Kidney Diseases, National Institutes of Health, Bethesda (Maryland 20892-0830, USA), Fax + 1 301 402 0053, e-mail: ku@helix.nih.gov

Received 13 December 1999; received after revision 3 February 2000; accepted 7 February 2000

Abstract. The repeat expansion diseases are genetic disorders caused by intergenerational expansions of a specific tandem DNA repeat. These disorders range from mildly to severely debilitating or fatal, and all have limited treatment options. How expansion occurs and causes disease is only now beginning to be understood. Efforts to model expansion in mice have so far met with only limited success, perhaps due to a requirement for specific cis- or trans-acting factors. In vitro studies and data from bacteria and yeast suggest that in addition to secondary structures formed by the repeats, components of the DNA replication and recombination machinery are important determinants of instability. The consequences of expansion differ depending on where in the gene the repeat tract is located, and range from reduction of transcription initiation to protein toxicity. Recent advances are beginning to make rational approaches to the development of therapies possible.

Key words. Repeat expansion; DNA instability; DNA structure; triplet repeats; polyglutamine toxicity.

Introduction

The repeat expansion diseases are a group of genetic disorders that includes Huntington's disease, myotonic dystrophy (DM), fragile X syndrome and Friedreich ataxia (FRDA) (fig. 1, table 1). All disorders currently in this category are neurological or neuromuscular, but a wide range of disorders may turn out to have the same genetic basis [1]. These disorders result from instability at a single disease-associated tandem array. Pathogenic expansion involves an intergenerational increase in array length to a size that has pathogenic consequences. The risk of intergenerational expansion is correlated with the number of repeats in the parental allele. In general, in the children that inherit these expansions, the greater the number of repeats, the more severe the disorder, and in those disorders which are not congenital, the earlier the age at onset. There are two known exceptions: in progressive myoclonus epilepsy of the Unverricht-Lundborg type (EPM1), there is no correlation between repeat size and the age at onset [2], and in spinocerebellar ataxia type 8, alleles of intermediate size are pathogenic, but very long alleles are not [3]. The tendency of these tandem arrays to increase in size from generation to generation explains one of the hallmarks of these disorders, anticipation, which is the progressive increase in disease severity and decrease in age of onset seen with successive generations in an affected family.

The first disorders to be described involved tandem arrays with a triplet repeat unit. However, it is now known that expansion diseases can involve a larger repeat unit [4]. Moreover, unstable tandem arrays with a wide range of repeat unit sizes are known. The preponderance of disorders involving triplets probably reflects, at least in part, the fact that many of these disorders involve an expansion of an array located in a coding region that results in a dominant and readily

* Corresponding author.

recognizable phenotype. Furthermore, work to identify new members of the repeat expansion disease family has focused on finding expansions of repeats that are already known to be disease associated [5].

The first part of this review will discuss some of the biochemical properties of the disease-associated repeats. These properties may provide the molecular basis for expansion, while also accounting for some of the consequences of expansion in several of these disorders. The second part of the review will cover what is currently known about the mechanism of expansion. The final section will address the consequences of expansion as it applies to each disorder or group of disorders.

Nucleic acid structure

Only three triplets are currently known to be associated with an expansion disease. These are the triplets CGG·CCG, CTG·CAG, and GAA·TTC. The dodecamer repeat, $C_4GC_4GCG{\cdot}CGCG_4CG_4$, is the only non-triplet repeat currently known to be responsible for an expansion disease. These repeats can form a variety of non-canonical secondary structures depending on the sequence of the repeat, the number of repeats, the pH, ionic strength, DNA concentration, superhelical density, and whether or not the repeat is single stranded. A subset of these structures is shown in figure 2.

CTG·CAG and CGG·CCG tracts have a 20% increase in persistence length and a high degree of torsional flexibility [6]. Individual strands of these sequences form homoduplexes at high DNA concentrations when the number of repeats is small. These repeats also readily form hairpins which predominate over homoduplexes at repeat lengths above ten, or when the DNA concentration is low [reviewed in ref. 7 and illustrated in fig. 2a]. The hairpins differ in the nature and extent of base pairing and their subsequent stability. RNA containing CUG repeats can also form hairpins [8].

Longer CGG tracts can also form intramolecular [9, 10] and intermolecular G tetraplexes [11] (fig. 2b), which are four-stranded structures stabilized by G_4 tetrads,

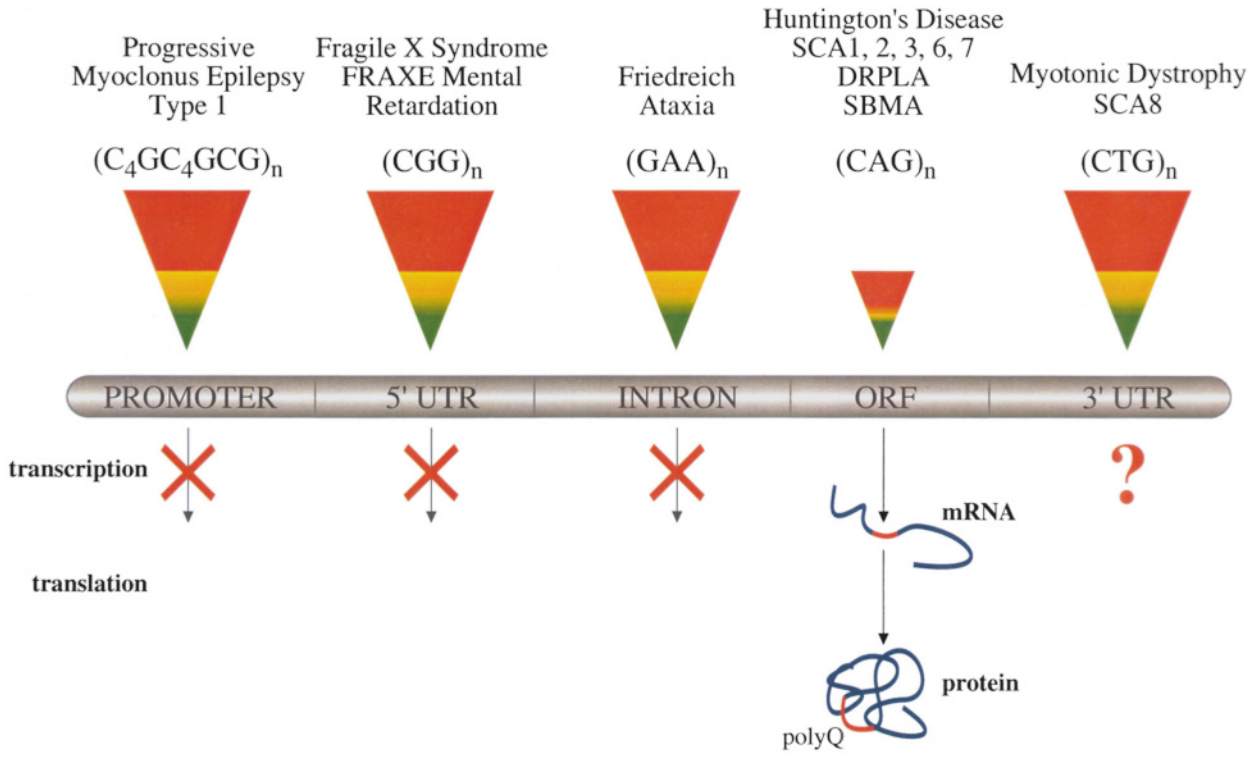

Figure 1. Graphic depiction of the repeat expansion diseases showing the location of all the known disease-associated expansions on a generalized representation of a gene. The green area of each triangle represents the number of repeats in normal individuals, the yellow area the repeat number in individuals at risk for transmission of the disorder (carriers of 'premutation' alleles), and the red area the number of repeats in clinically affected patients. The repeat unit associated with each disorder is shown above the triangles. SCA, spinocerebellar ataxia; DRPLA, dentatorubral pallidoluysian atrophy; SBMA, spinal and bulbar muscular atrophy. The effect of the expansion on transcription and translation, where known, is shown schematically at the bottom of the figure. The red crosses indicate the step affected by the expansion.

Table 1. Repeat expansion diseases.

Disorder	Affected protein/ location	Repeat	Repeat location	Normal	Premutation	Full mutation	Inheritance	References/ OMIM #
EPM1	cystatin B 21q22.3	$(C)_4G$ $(C)_4GCG$	promoter	2–3	12–17	30–75	AR	[4, 91] #254800
FRAXA	FMRP Xq27.3	CGG	5′ UTR	6–52	~60–200	~200–>2000	X-linked dominant ND	[32, 96, 97] #309550
FRAXE	FMR2 Xq28	CGG	5′ UTR	6–25		>200	X-linked maternal bias	[98] #309548
FRDA	frataxin 9q13	GAA	intron 1	7–22		200–>900	AD maternal bias	[111] #229300
HD	huntingtin 4p16.3	CAG	ORF	6–34		36–180	AD paternal bias	[159] #143100
DRPLA	atrophin 12p12	CAG	ORF	7–25		49–88	AD paternal bias	[160, 161] #125370
SBMA	androgen receptor Xq11–q12	CAG	ORF	11–24		40–62	X-linked recessive ND	[123] #313200
SCA1	ataxin-1 6p23	CAG	ORF	6–39		39–83	AD paternal bias	[162] #164400
SCA2	ataxin-2 12q24	CAG	ORF	15–29		34–59	AD paternal bias	[5] #183090
SCA3	ataxin-3 14q24.3–q31	CAG	ORF	13–36		55–84	AD paternal bias	[35] #109150
SCA6	P/Q calcium channel 19p13	CAG	ORF	4–16		21–30	AD ND	[124] #183086
SCA7	ataxin-7 3p21.1–p12	CAG	ORF	4–35	28–35	34–>300	AD paternal bias	[79] #164500
DM	DMPK 19q13.3	CTG	3′ UTR of DMPK	5–37	~50–180	~200–>2000	AD maternal bias	[33, 139–141] #160900
SCA8	SCA8 13q21	CTG	3′ UTR	6–37		~107–250*	AD maternal bias	[81] #603680

AD, autosomal dominant; AR, autosomal recessive; ND, not determined; UTR, untranslated region; ORF, open reading frame; OMIM, Online Mendelian Inheritance in Man.
* Longer alleles exist but are not pathogenic [3].

planar arrangements of four G residues in which each G residue is simultaneously hydrogen bonded to two other G residues. Since intrastrand tetraplexes can be thought of as folded hairpins or hairpin dimers (compare fig. 2a and 2b), some of the properties generally associated with hairpins may be relevant to tetraplexes as well. In addition, the extraordinary stability of the G-tetraplexes relative to hairpins may account for some of the unique properties associated with sequences with tetraplex-forming potential [10]. Denaturation and renaturation of CTG·CAG repeats or CGG·CCG repeats results in the formation of slipped DNA (S-DNA) in which the two strands have reannealed out of register forming a heterogenous group of molecules containing duplex regions interspersed with loops and folded structures [12]. GAA·TTC repeats are unique amongst the disease-associated repeats in being unable to form a hairpin at physiological temperatures [13]. However, GAA·TTC repeats can form three-stranded DNA structures that involve both the GAA and the TTC strands. These structures are purine·purine·pyrimidine (R·R·Y) triplexes [E. Grabczyk and K. Usdin, unpublished observations] (fig. 2d) and pyrimidine·purine·pyrimidine (Y·R·Y) triplexes [14] (fig. 2e). These triple-stranded structures may be intramolecular or intermolecular. The Y·R·Y triplex contains protonated cytosines and its formation is therefore favored by low pH. Two interacting R·R·Y triplexes can form a bimolecular complex, known as 'Sticky DNA' [15].

TGG·CCA and AGG·CCT tracts are also hypervariable although not currently associated with a disease. The G-rich strands form G tetraplexes [9], while CCT tracts form i-tetraplexes [16]. An i-tetraplex contain two pairs of DNA strands each pair held together by hydrogen bonds between a protonated C and an unprotonated one. The two sets of DNA strand are held together by intercalation of the C^+·C pairs [16] (fig. 2c). Of the remaining triplets, only GTC and GAC tracts form stable structures [7]. However, they are less stable than those formed by disease-associated repeats. These repeats are also underrepresented in the human genome [7].

The myoclonus epilepsy repeat forms the most complex set of structures of all the disease associated repeats

including hairpins stabilized primarily by G·G base pairs, G-tetraplexes and i-tetraplexes, as well as both Y·R·Y and R·R·Y triplexes [K. Usdin, unpublished observations]. Other hypervariable sequences have also been shown to form hairpins, triplexes, tetraplexes, or i-tetraplexes, including the insulin promoter VNTR [17, 18], the mouse *Ms6-hm* microsatellite [18, 19], and the human D4S43 locus [18].

The fact that hypervariable sequences all form secondary structures has led to the suggestion that these structures may play a role in the instability of these sequences, and in some cases in disease pathology. The

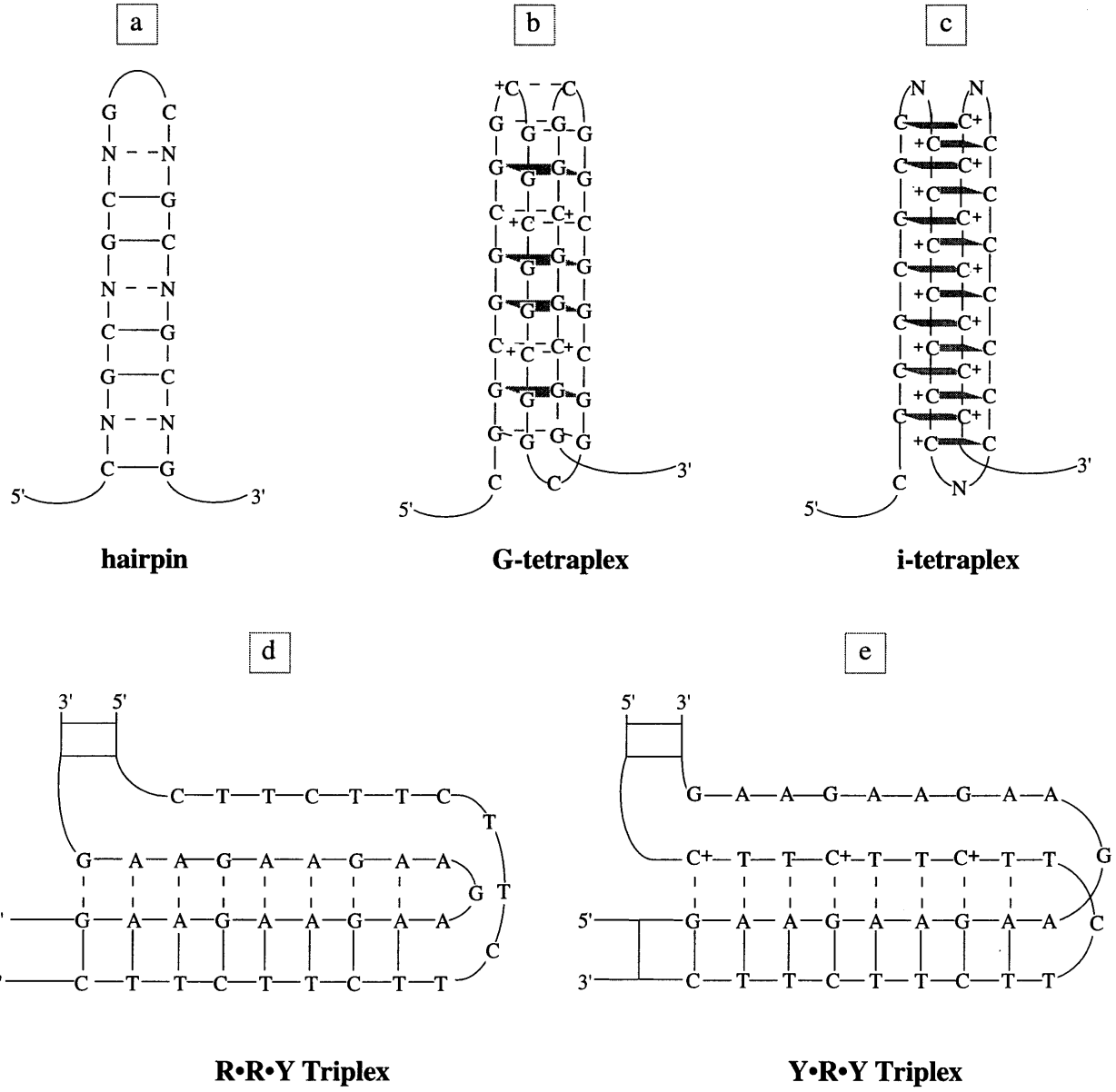

Figure 2. Schematic representation of some of the types of structures formed by disease-associated repeats. (*a*) Generic representation of the hairpins formed by repeats of the form CNG, where N is any nucleotide. Watson-Crick hydrogen bonds are indicated by solid lines, while non-Watson-Crick interactions or mispaired bases are indicated by dashed lines. Hairpins formed by the bottom strand of the EPM1 repeat form hairpins that have G·G base pairs and no Watson-Crick base pairs. (*b*) An intrastrand G tetraplex. A CGG tetraplex is shown as an example. The basic stabilizing unit in these tetraplexes is a G_4 tetrad containing four guanine residues each hydrogen bonded to two other guanine residues in the tetrad (shown as the dark-gray parallelograms). The Cs in this structure can also form $C{\cdot}C^+$ base pairs indicated by the dashed lines. (*c*) Generic i tetraplex that can be formed by C-rich repeats. The hemiprotonated C·C pairs are shown as light-gray parallelograms. (*d*) R·R·Y triplex. (*e*) Y·R·Y triplex.

properties of these structures in vitro suggest possible in vivo effects. For example, many of these sequences form a barrier to DNA or RNA polymerase. Tetraplexes form very stable blocks to DNA synthesis in vitro [10], while the structures formed by CTG·CAG [20] and GAA·TTC tracts [21] form weaker blocks. CGG·CCG and CAG·CTG tracts block replication in bacteria [22], while GAA·TTC tracts inhibit replication in mammalian cells [23]. GAA·TTC triplexes also block RNA polymerase in vitro [E. Grabczyk and K. Usdin, unpublished observations]. The single-stranded regions of these structures may also be prone to strand cleavage. Long CGG·CCG tracts exclude nucleosomes in vitro, while long CTG·CAG tracts are strong nucleosome-positioning signals [24, 25]. This not only illustrates how the protein-binding properties of these repeats may be unusual but specifically suggests their potential effect on local chromatin structure.

In both bacteria and yeast, the deletion frequency and pattern of repair of sequences that can form hairpins in vitro lend support to the idea that they can form in vivo [26, 27], and the existence of triplexes has been directly demonstrated in bacteria and insects [28, 29].

Mechanisms of expansion

Current models for expansion fall into two basic classes: those that are replication based, and those that involve recombination. Any model for expansion needs to account for the distinctive features of expansion seen in affected families. These include the correlation between the number of uninterrupted repeats in the array and the risk of expansion [30, 31], the large changes in allele size that are seen in some disorders [32, 33], whether the disorder shows paternal or maternal bias in transmission, and differences in the timing of expansion and the duration of instability [34–38].

Strand-slippage models

Bacterial and yeast studies have shown that while mutations that affect mismatch repair do not affect the frequency of large changes in repeat number in *Escherichia coli* [39], yeast [40] and humans [41], mutations that affect replication do. For example, mutations in the DnaE and DnaQ subunits of DNA polymerase III, the enzyme responsible for replicating the *E. coli* chromosome, and DnaB, the replication fork helicase, increase the instability of tandem repeats [42]. In yeast, mutations in DNA polymerase α and ∂, two of the three polymerases involved in genome replication in eukaryotes, and in Pol30, which encodes the replication processivity factor PCNA [43], all increase instability. Expansions in bacteria and yeast occur predominantly when the strand which forms the more stable secondary structure is the lagging daughter strand, while deletions occur predominantly when this strand is the template for lagging-strand DNA synthesis [37–40]. The influence of the orientation of the repeat relative to the origin of replication on instability supports a replication-based model and suggests that DNA structure may play a role in this process.

In simple strand-slippage models of instability, transient dissociation of the nascent strand from the template during DNA replication provides an opportunity for the strands to slip relative to one another [44] [step (i) in fig. 3b]. If the nascent strand slips backwards and DNA polymerase primes DNA synthesis from this position, a gain of repeat units or expansion may result. If on the other hand, the nascent strand slips forward, a deletion can result.

A model like this could account for small changes in repeat number, and could also explain how interruptions to the repeat motif stabilize the repeat tract, since they would increase the likelihood that mismatches would occur between the slipped strands. This would reduce the likelihood that priming would occur from the slipped position. For large increases in repeat number to occur via this mechanism, many slippages would be needed in a single round of replication. This may happen when the Okazaki fragments generated during lagging-strand DNA synthesis are entirely comprised of the repeat so that they are free to slip repeatedly during replication [45]. DNA secondary structures formed behind the polymerase may promote instability by stabilizing the strand-slippage intermediates thus favoring priming from the slipped position. Blocks to DNA synthesis formed on the template strand in front of the polymerase may promote instability by increasing the likelihood that strand-slippage will occur.

The FEN-1 interference model

Mutations in rad27, the yeast homolog of the human FEN-1 endonuclease gene, greatly increase the frequency of CTG·CAG and CGG·CCG expansions in yeast [46–48]. During lagging-stand DNA replication the 5′ end of the Okazaki fragment is displaced by an upstream helicase or polymerase generating a single-stranded flap [step (i) in fig. 3c]. FEN-1 then removes the flap by endonucleolytic cleavage before ligation of adjacent Okazaki fragments. The flap must be single-stranded for effective FEN-1 cleavage [49]. Formation of intrastrand folded structures may make them resistant to FEN-1 cleavage, thus preventing flap removal [50] [step (ii) in fig. 3c]. Ligation would then incorporate the unprocessed flap DNA into the nascent strand leading to an expansion [step (iii) in fig. 3c]. In the case of GAA·TTC repeats, the flap region might anneal with

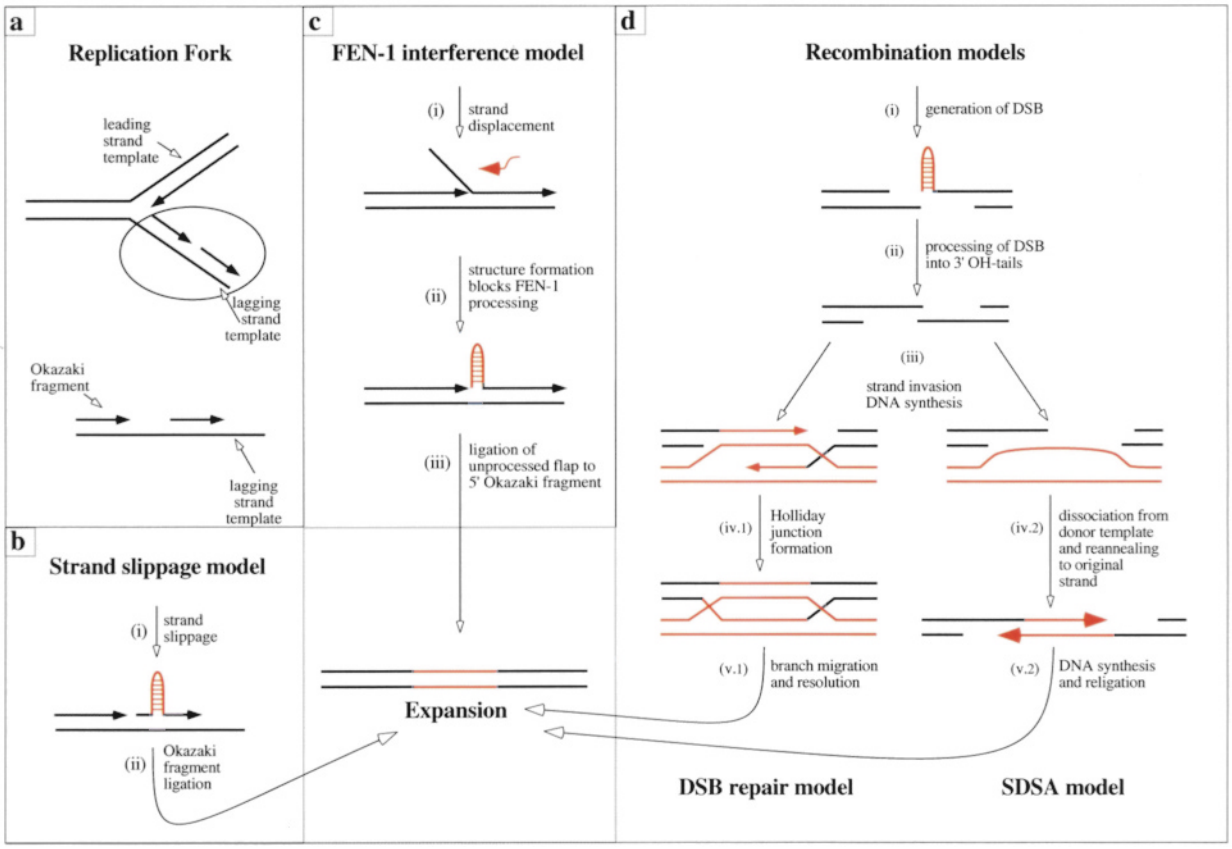

Figure 3. Potential expansion mechanisms. The original repeat tract is shown in black, repeat units gained or the source of these repeats (the donor strand) are indicated in red. Filled arrows indicate the DNA strands being synthesized with the arrowhead pointing in the direction of DNA polymerase movement. While some of these mechanisms could also result in deletions, only pathways leading to expansions are shown. (*a*) A schematic representation of a replication fork is shown at the top. The lagging-strand template and Okazaki fragments are circled. A simplified depiction of DNA synthesis on the lagging strand is shown below. (*b–d*) Models for expansion shown as occurring during lagging-strand DNA replication, although replication slippage (*b*) can in theory occur during both leading- and lagging-strand synthesis, and the double-strand breaks that initiate the recombination models (*d*, *e*) may be caused by events other than the failure to process Okazaki fragments properly as depicted here. (*b*) Simple strand-slippage model. (*c*) A FEN-1 interference model. (*d*) Two recombination models: variations of the double-strand-break repair model (left-hand side) and a synthesis-dependent strand-annealing model (right hand side).

repeats in the adjacent duplex region to form a triplex thereby also inhibiting FEN-1 processing. The FEN-1 model is appealing because it may account for the fact that for disease alleles of a certain size, expansion is more common than contraction.

Recombination models

Some repeat instability in bacteria is recA dependent, including that due to mutations in polymerase III or its associated proteins [42, 51]. This suggests that recombination might play a role in instability. Yeast rad27 strains also exhibit high levels of recombination, perhaps because the FEN-1-resistant flaps generated during lagging-strand synthesis can undergo cleavage to produce double-strand breaks [step (i) in fig. 3d]. DSBs may also be generated by stalled replication forks [52]. Double-strand breaks are processed exonucleolytically to generate single-stranded 3′ OH tails [step (ii) in fig. 3d], which can be processed in a number of different ways to generate expansions.

In the double-strand break repair-type model originally proposed for yeast [53], these tails invade the tandem array on a sister chromatid or sister chromosome that is aligned out of register with the first chromatid or chromosome [step (iii) in fig. 3d]. These tails prime DNA synthesis using the intact donor strand as template, forming a heteroduplex which can undergo branch migration before being converted to double Holliday junctions [step (iv.1) in fig. 3d]. The Holliday junctions are

then resolved to generate duplexes that may have aquired additional repeat units in the process [step (v.1) in fig. 3d]. This process may produce both gene conversions and crossover events.

Most homologous recombination in mammalian cells results in gene conversions rather than crossovers, and therefore models that do not involve Holliday junctions might be more likely to account for expansion. One such model is a modified synthesis-dependent strand-annealing (SDSA) model [47] shown on the right-hand side of figure 3d, in which strand invasion of the homologous template by one 3′ OH tail takes place [step (iii) in fig. 3d]. This is followed by DNA synthesis, and unwinding of the newly synthesized strand from the donor allele [step (iv.2) in fig. 3d]. This process of strand invasion, DNA synthesis, and unwinding of the donor and acceptor alleles could be repeated many times before both strands of acceptor allele reanneal [step (v.2) in Fig. 3d]. The donor allele may be either the same molecule, the sister chromatid, or the sister chromosome. The end result would be an expansion that leaves the donor allele unchanged. According to this model, the resultant expansion may contain a complex patchwork of information derived from both alleles, that may account for some of the complex expansion products seen in SCA8 pedigrees [54], and certain minisatellites [55].

Repair of double-strand breaks requires DNA synthesis. Strand-slippage or failure to process Okazaki fragments may thus also occur during recombination. In fact, strand-slippage might be more likely to occur at this time, since repair synthesis is thought to be less processive than normal replication [56].

The relative contribution of these different instability mechanisms is unclear. Those disorders that involve small changes in repeat number, particularly where the repeat unit is small, could occur mostly by replication slippage. In disorders such as fragile X syndrome, small alleles show small intergenerational changes, while larger 'at-risk' or 'premutation' alleles show much larger changes in repeat number [30]. In this case, replication slippage might account for the small expansions, while larger expansions may be due to some combination of replication and recombination-based mechanisms.

The role of cis- and trans-acting factors. A role for cis- and trans-acting factors is suggested by the fact that the stability of CTG·CAG tracts is different at different chromosomal loci in humans [57], and chromosomes with specific flanking haplotypes are predisposed to expansions in many disorders [58–61]. Furthermore, instability in fragile X syndrome and FRAXE mental retardation is often observed in conjunction with a second instability at an independent locus [62]. Expansion in many disorders also occurs in a polar fashion, with repeats being preferentially added to or lost from one end of the array, suggesting either an effect of flanking sequences or an inherent polarity in the expansion mechanism. Long repeat tracts are not sufficient to produce large expansions in mice, although a number of mice with long repeat tracts show high rates of relatively small changes in repeat number [63–66]. This includes mice with part of a human fragile X syndrome allele that has an $\sim 100\%$ risk of expansion during maternal transmission in humans [65], and mice with 109 CTG·CAG repeats inserted into the murine huntingtin gene [64]. Since mice show DNA instability at other loci [67], as well as folate-sensitive fragile sites [68], which in humans are associated with long CGG·CCG tracts [69], this may also reflect a requirement for particular cis- or trans-acting factors. These factors may include particular origins of replication or recombination signals as well as variants in one or more proteins involved in DNA replication, repair or recombination.

Timing of expansion. Disorders such as DM and FRDA show somatic instability that occurs throughout the lifespan of the individual. Other disorders such as fragile X syndrome show relatively little somatic instability, although it is unclear whether expansion is prezygotic [36] or postzygotic [70]. Fragile X monozygotic twins with expansions of different sizes suggest that some instability must occur postzygotically [71]. However, other monozygotic twins with similar size expansions [34], and the absence of significant tissue-to-tissue variation in repeat sizes of individuals with fragile X syndrome at autopsy [72] suggest that significant instability does not occur after early embryogenesis. This suggests that some cis- and/or trans-acting factors that are developmentally regulated may be important in some expansions.

By analogy with factors that affect repeat instability in bacteria, transcription through a given repeat region may be higher during gametogenesis or early embryogenesis, or differential use of origins of replication may affect which repeat strand is the lagging-strand template. Expanded fragile X alleles which are methylated in affected adults are less unstable than large unmethylated alleles [73]. This may be due to an effect on transcription, since methylation is known to silence the promoter immediately upstream [74]. The consequent reduction of transcription of the repeat may reduce instability, as seen in bacteria [75]. Blastocyst DNA is relatively undermethylated, and according to this hypothesis, more likely to show instability. Methylation of expanded fragile X alleles occurs early in development [34], which may impose limits on any further somatic instability.

Paternal or maternal bias in transmission. Some expansions are predominantly paternally inherited while oth-

ers show a maternal bias. Disorders involving small changes in repeat number are often paternally inherited, or show earlier age of onset and increased disease severity when affected alleles are paternally inherited [76–79]. In contrast, diseases involving very large expansions are most often associated with maternal transmission, e.g., most cases of fragile X syndrome [80], SCA8 [81], and congenital DM [82]. This may be due to differences in the numbers of cell divisions undergone by male and female gametes, or to differences in the selection pressures active in these cells. The high rate of small changes in sperm is probably related to the larger number of cell divisions undergone by sperm cells relative to oocytes: human oocytes undergo 22 mitotic and two meiotic divisions before fertilization, while sperm from a 35-year-old man might have undergone more than 540 cell divisions [83]. This idea is supported by evidence that the repeat lengths of children of the same Huntington's disease father increase with increasing paternal age [84].

The tendency of long alleles to contract rather than expand on paternal transmission may be due to counterselection against expansions during sperm maturation. In fragile X syndrome, this effect is not due to a requirement for the affected gene product during spermatogenesis, since mice containing targeted disruptions of this gene are fertile [85]. Sperm DNA is about sixfold more highly condensed than mitotic chromosomes [86]. Very long repeats might not be packaged effectively due to their unusual properties. Alternatively, difficulties replicating long repeat tracts, related perhaps to the formation of secondary structures, may have different outcomes in the male and female germline. A more stringent checkpoint mechanism is thought to operate during male meiosis, such that spermatocytes arrested in meiosis undergo rapid degeneration [87]. Spermatocytes containing long repeat tracts that have failed to complete replication, for example, might trigger the checkpoint mechanism and undergo degradation at a higher rate than those containing short repeats, while little if any selection against oocytes with long repeat tracts takes place. Difficulties in replicating long repeat tracts are consistent with the observation that the number of somatic cells or sperm with very large expansions decreases with age [88, 89].

Consequences of expansion

The consequences of expansion in each disease depend on the location of the expanded repeat tract in the affected gene (fig. 1). Expansions in the coding sequence may result in proteins with altered properties. Expansions in the non-coding portions of the gene may exert their effects at the level of transcription initiation, transcription elongation or posttranscriptionally due to altered processing or novel properties of the transcript. The individual expansion disorders will be described starting with those expansions involving the 5′ end of the gene.

Myoclonus epilepsy: a promoter disruption mutation?

Clinical features. Progressive myoclonus epilepsy of Unverricht-Lundborg type (EPM1) is an autosomal recessive disorder characterized by stimulus-sensitive myoclonus, tonic-clonic seizures, and ultimately cerebellar ataxia. The principal cytopathology is a loss of cerebellar granule cells. These cells frequently display changes characteristic of apoptosis including condensed nuclei, and fragmented DNA. Onset is between age 6–15 years, and shows a variable rate of progression.

Genetics. This disorder results from mutations in the cystatin B (CSTB) gene which encodes a non-caspase cysteine protease inhibitor [4, 90]. Although intragenic mutations are known, the most common mutation associated with this disorder is an expansion of a tandem array with a dodecamer repeat of the sequence $C_4GC_4GCG{\cdot}CGCG_4CG_4$. This repeat array is located approximately 66 and 77 bp 5′ of the two transcriptional start sites of the cystatin B gene [91]. Normal alleles contain 2 or 3 copies of this repeat. Expanded alleles contain more than 60 repeats and have reduced levels of CSTB messenger RNA in some cells. Since mice homozygous for a disruption in the cystatin B gene show similar symptoms to myoclonus epilepsy patients [92], EPM1 is likely due to the loss of cystatin B function.

No hypermethylation of the expanded promoter has been detected [91]. However, expanded repeats reduce the activity of the cystatin B promoter in certain cells, as does replacement of the repeat with non-repetitive DNA of normal G + C content, or deletion of DNA upstream or downstream of the repeat [93]. Expansion may thus affect gene expression by altering the spacing of regulatory elements in the promoter [93]. How reductions in the levels of cystatin B produce epilepsy is not clear at this point, but the identification of this gene as the causative agent provides a starting point for the rational design of treatments that may be used to ameliorate the symptoms of this disorder.

Fragile X syndrome and FRAXE mental retardation: expansion induced, methylation mediated, promoter silencing

Clinical features. Fragile X syndrome and FRAXE mental retardation are both mental retardation disorders resulting from mutations on the X chromosome (fig. 1, table 1). Fragile X syndrome is the most com-

mon heritable form of mental retardation. Retardation can range from moderate to profound. Males are more severely affected than females due to the normal mosaic pattern of X chromosome inactivation in females. While the major symptom is mental retardation, macro-orchidism, large ears, autistic behavior, a prominent jaw, and high-pitched speech are often observed. Dendrites in the cerebral cortex of fragile X knockout mice and human fragile X patients have long, thin dendritic spines resembling those seen during early brain development, and these are thought to reflect dendritic pruning deficits [94]. Premature ovarian failure is also associated with fragile X syndrome premutations [95]. Individuals with FRAXE mental retardation have a milder form of mental retardation including speech delay and problems with reading and writing.

Genetics. The gene affected in fragile X syndrome is known as FMR1 (fragile X mental retardation gene 1) [32, 96, 97]. The gene affected in FRAXE mental retardation, FMR2, is located 600 kb telomeric to FMR1 [98]. The most prevalent mutation in both disorders is the expansion of a CGG·CCG tract in the 5′-untranslated region of the affected gene. Intragenic FMR1 mutations with the same phenotype as CGG·CCG expansions, and an FMR1 knockout mouse with learning disabilities, indicate that fragile X syndrome is due to a loss of FMR1 function [85].

In both disorders, expansion is associated with hypermethylation of the repeat and the adjacent CpG-rich promoter [99], and decreased transcription. The demethylating agent 5-azacytidine increases FMR1 gene expression in tissue culture consistent with a role for hypermethylation in transcriptional silencing [100]. Mispaired or unpaired cytosines are good substrates for methylation by CpG methyltransferases perhaps because these residues resemble methylation intermediates in which the cytosine base is flipped out of the helix, or because less energy is required to form this intermediate [101]. The tetraplex structures formed by CGG repeats and the hairpin formed by the CCG repeats both contain mispaired cytosines, and thus may account for the methylation propensity of long CGG·CCG tracts.

The product of the FMR1 gene, FMRP, is an RNA-binding protein that binds a subset of brain mRNAs including its own message [102]. It has a nuclear localization signal and a nuclear export signal, and is found in actively translating polysomes particularly at the synapse where it is translated in response to stimulation of metabotropic glutamate receptors [103]. A missense mutation, I304N, that produces a protein that binds RNA but does not associate with polysomes results in a very severe phenotype [104]. In the current model for FMRP function, FMRP acts as a chaperone for specific transcripts to which it binds in the nucleus and escorts to the polysome in response to sensory input. Reduced levels of FMRP resulting from expansion-induced promoter silencing alter the affected individual's ability to regulate protein translation in response to synaptic stimulation. The severity of the I304N mutation may result from the sequestration of bound mRNAs in non-translatable mRNP particles [104].

Much less is known of the role of the product of the gene affected in FRAXE mental retardation, FMR2. The FMR2 protein is a nuclear protein expressed in the hippocampus, in Purkinje cells in the hindbrain and piriform cortex, in the amygdala, and the cingulate gyrus. It has homology to the human AF-4 and LAF-4 genes, two putative transcription factors [105].

Chromosome fragility. Both fragile X syndrome and FRAXE mental retardation are associated with folate-sensitive fragile sites that colocalize with the expansion on the X chromosome. Fragile sites are gaps or constrictions in the chromosome that resemble underreplicated chromosomes. Folate-sensitive fragile sites are seen either when cells are starved for folate or when very high levels of folate are present. Folate is a precursor for methylene tetrahydrofolate, a cofactor for thymidylate synthase which is required for pyrimidine nucleotide biosynthesis. Too little folate reduces the dTTP pool, while too much inhibits ribonucleotide reductase leading to a deficiency in dCTP. Folate is also an important source of methyl groups for CpG methylation. All folate sensitive fragile sites characterized thus far are long CGG·CCG tracts [reviewed in ref. 69].

Long CGG repeats, especially those that are methylated, exclude nucleosomes in vitro. Similar behaviour in vivo might provide the basis for fragility [24]. However, folate starvation decreases genomic methylation, and demethylation of genomic DNA by growth in 5-azacytidine, or by methionine starvation, does not reduce fragile site expression [106]. Moreover, folate-sensitive fragile sites are also strongly induced by fluorodeoxyuridine (Fudr) which inhibits thymidylate synthase [106]. This effect can be reversed by thymidine but not folic acid, suggesting that the effect of folate on fragile site expression is at the level of DNA replication. Moreover, agents that induce other types of fragile sites act primarily by interfering with DNA replication [69]. It is feasible then that difficulties in replication of the sequence at different fragile sites cause fragility. Low dNTP levels produced by folate stress may exacerbate the difficulties in replicating CGG·CCG tracts. Expanded alleles do complete replication later in the cell cycle than normal alleles as predicted by such a hypothesis, but the delay in replication involves a relatively large region of the X chromosome [107], and may result, at least in part, from the transcriptional silencing of the FMR1 gene, since such regions also show delayed replication [108].

Fragile sites have increased chromosome rearrangements, sister chromatid exchanges, and intrachromosomal gene amplification, suggesting high rates of recombination [69], but whether this is a cause or simply a consequence of fragility is not clear. In any event, these sites are frequent chromosome breakpoints in vivo. Chromosome breakage at the folate-sensitive fragile site, FRA11B, results in the loss of the end of the long arm of chromosome 11 in a subset of Jacobsen syndrome patients [109]. Translocations involving other fragile sites are responsible for a number of different malignancies [110]. Understanding the molecular basis of chromosome fragility might suggest ways in which breakage and translocations might be reduced.

Friedreich's ataxia: a transcription defect resulting in an iron distribution disorder

Clinical features. Friedreich's ataxia (FRDA) is the most common hereditary ataxia. It is an autosomal recessive disorder characterized by progressive unremitting ataxia, dysarthria, nystagmus, diminished or absent tendon reflexes, Babinski sign, and impairment of position and vibratory senses. Degeneration of the spinocerebellar tracts, dorsal columns, pyramidal tracts, and to a lesser extent the cerebellum and medulla, are seen, as are abnormalities in motor and sensory nerve conduction. Hypertrophic cardiomyopathy is seen in about 75% of affected individuals and cardiac failure is the cause of about half of FRDA-related deaths. There is a broad clinical spectrum for FRDA: the more typical form, with onset in early adolescence; Acadian FRDA, seen in a subset of the French Canadian population, which has a milder course that is not often associated with cardiomyopathy; a late-onset form (LOFA), and a form of FRDA with retained reflexes (FARR). FRDA has a worldwide distribution and occurs at a frequency of about 1/50,000 individuals.

Genetics. FRDA is caused by mutations in the X25 or frataxin gene. Ninety-eight percent of FRDA mutations are expansions of a GAA·TTC repeat array in the first intron [111]. There is significant somatic instability of the GAA·TTC tracts which may contribute to the clinical variability in FRDA patients. LOFA and FARR, which are generally associated with a milder clinical phenotype, have smaller expansions than the more typical form of FRDA [112]. The Acadian form, on the other hand, is not associated with significantly shorter expansions, suggesting that other factors influence disease severity [113].

Expansion reduces the amount of full-length transcript produced in affected individuals [114]. Transcripts truncated at the 3′ end of the GAA tract are seen when supercoiled templates are transcribed in vitro. This suggests that a supercoil-dependent secondary structure causes transcript truncation. The amount of full-length transcript is not reduced by lowering the pH or when large amounts of GAA containing RNA are added. This suggests that reduced transcript levels are not due to Y·R·Y triplex formation or formation of a triplex between the template and the RNA, but rather to a R·R·Y triplex that may act by trapping the RNA polymerase [E. Grabczyk and K. Usdin, unpublished observations], analogous to what occurs in other purine-rich sequences [115]. That long GAAGGA tracts which do not form triplexes are not pathogenic and do not show reduced levels of transcription in tissue culture [116] supports this idea.

Frataxin, the affected gene product in this disorder, is a nuclear-encoded mitochondrial protein [117]. A knockout of the yeast homolog of this gene, YFH1 (yeast frataxin homolog 1), shows increased mitochondrial iron accumulation at the expense of cytosolic iron, suggesting a role for frataxin in regulating mitochondrial iron transport [118]. An increased sensitivity to oxidative stress is also seen that suggests that FRDA may result from oxidative damage resulting from mitochondrial iron overload. Frataxin has recently been shown to bind iron directly and to form a high-molecular-weight complex that sequesters iron [119]. Iron deposits occur in the heart, in the dentate nucleus, and fibroblasts of affected individuals. Heart biopsies of FRDA patients show decreased activities of aconitase and respiratory chain complexes I, II, and III, consistent with oxidant stress [120]. Similarities between the symptoms of vitamin E deficiency and FRDA support the role of oxidant damage in FRDA, since vitamin E protects mitochondria from iron-induced free radical damage. Preliminary results from clinical trials suggest that idebenone, a free radical scavenger protects the heart muscle of FRDA patients from iron-induced injury [121].

CAG·CTG expansions

Disorders involving CTG·CAG expansions currently represent the largest category of expansion disorders. Within this group, the biggest subset involves expansions in a coding region. These disorders will be discussed as a group since there is reason to believe they all involve the same basic pathogenic mechanism. The remaining two disorders will be discussed separately.

The polyglutamine diseases: polyglutamine toxicity disorders

Clinical features. There are currently eight diseases known to belong to this category (fig. 1, table 1). All are progressive neurodegenerative disorders, and with the

exception of SBMA which is X-linked, are autosomal dominant. Huntington's disease (HD) is the most prevalent of this class of disorders, occurring in 4–10 per 100,000 Caucasians. Symptoms of HD include hyperkinetic involuntary movements, slowing of voluntary movements, as well as cognitive impairment and psychiatric disturbances. The age at onset is usually in middle age and progresses to death over 15–20 years. The region of the brain most severely affected in HD patients is the striatum and deep layers of the cerebral cortex. There is selective atrophy of the caudate and putamen of the striatum with reactive astrocytosis and loss of neurons, in particular the medium spiny projection neurons. DRPLA has symptoms that overlap with HD. This disorder, found predominantly in Japanese populations, is characterized clinically by a variable combination of myoclonus, epilepsy, mental retardation, cerebellar ataxia, choreoathetosis, euphoria and dementia, and pathologically by selective neuronal death in the dentatofugal and pallidofugal pathways. SBMA shows a pattern of progressive muscle atrophy and weakness due to degeneration of motor neurons. Symptoms include facial fasciculations, dysphagia, intention tremor, lower motor and primary sensory neuropathy, and androgen insensitivity. The spinocerebellar ataxias are characterized by ataxia, dysarthria, dysmetria, and intention tremor. Overlaps in the clinical presentation of these SCA subtypes is broad with all the SCAs showing predominant degeneration of the Purkinje cells of the cerebellar cortex, and gait abnormalities.

Genetics. These disorders all result from an increase in the number of CTG·CAG repeats in the coding sequence of the affected gene that extends the length of a polymorphic polyglutamine tract. The amount of mRNA and protein produced seems to be relatively unaffected. In most cases, pathology is seen when the number of glutamines exceeds ~ 35 (table 1), and 60–70 repeats results in a juvenile onset form of the disease. However, in SCA6, the pathological threshold is about 21, and in SCA3 it is closer to 60. While neuronal cell death is seen ultimately in all of these disorders, many of the symptoms are apparent well before cell death is observed [122], suggesting that neuronal dysfunction rather than neuronal death may be responsible for the observed phenotype.

Apart from the polyglutamine tract, the affected proteins share no homology. The normal function of the affected gene product is only known for two disorders: in SBMA the affected gene product is the androgen receptor [123], and in SCA6, it is the P/Q calcium channel [124]. No cases of HD, DRPLA, SCA1, 2, 3, or 7 are known to be due to deletions or point mutations, suggesting that these disorders do not result from a loss of gene function. Point mutations in the androgen receptor produce testicular feminization but no neuronal degeneration, while SBMA expansions produce both androgen insensitivity and neuropathy. The androgen insensitivity may be due to the effect of the expansion on androgen receptor function, but the neuropathy is presumably due to a 'gain of function.' The situation in SCA6 is still unclear. Point mutations in the P/Q calcium channel gene cause the channel disorders episodic ataxia type 2 (EA-2), and familial hemiplegic migraine (FHM), and there is significant overlap between these disorders and SCA6 in their clinical presentations [125]. To date, no mice containing targeted disruptions of the affected genes show the progressive neurodegeneration that is typical of this group of disorders (no such model is available yet for SCA6). However, mice containing long polyglutamine tracts in the murine homologs of these proteins do [reviewed in ref. 126]. Moreover, truncated versions of these proteins produce a more severe disease pathology in transgenic mice than the full-length protein, and long polyglutamine tracts embedded in an unrelated protein, HPRT, mimic many of the clinical features of the polyglutamine diseases. This supports the idea that the expansion causes a pathogenic gain of function and suggests that all the polyglutamine diseases, with the possible exception of SCA6, are caused by a common underlying pathological mechanism. However, despite an overlapping pattern of expression of most of these proteins, each disorder involves the dysfunction of a specific subset of neurons. Therefore, specific modifiers of this mechanism must presumably operate to make the polyglutamine tract in these different proteins selectively deleterious.

With the exception of SCA6, some of the mutant protein is found in the nucleus even when the normal protein is not nuclear [see ref. 127, and references therein]. This nuclear localization seems to be important in SCA1, since mutations that abolish nuclear localization also prevent ataxia in transgenic mice [128]. Similarly, blocking the nuclear localization of mutant huntingtin reduces toxicity, while nuclear targeting increases toxicity [129]. The mutant proteins form aggregates or inclusions composed of a mixture of granules, filaments, and fibrils that resemble the β-amyloid fibrils seen in Alzheimer's disease [130]. The mechanism of aggregation is not known, but could result from the formation of hydrogen bonds between the main chain and side chain amides of two antiparallel beta strands forming a 'polar zipper' [131], or by the covalent linkage of glutamines to lysines by tissue transglutaminases [132]. In some cases these aggregates are ubiquitinated and sequester proteasome caps [133], suggesting that they accumulate due to abberrant folding and problems with proteolysis. However, suppressing the formation of nuclear inclusions in striatal cells transfected with mutant huntingtin increases cell death [134], and an ataxin-

1 deletion mutant which does not aggregate still causes ataxia in transgenic mice [122]. This suggests that aggregation is not pathogenic and may in fact have a protective effect.

Nuclear ataxin-1, unlike cytoplasmic ataxin-1, is not reactive with some antipolyglutamine antibodies even in the case of the non-aggregating mutant [128]. This suggests that the protein adopts a toxic conformation in the nucleus that precedes aggregation. Since shorter protein fragments are more toxic than full-length ones [126], perhaps because the additional flanking sequences in the intact protein constrain its folding into the toxic conformation, proteolysis may be a necessary prerequisite for toxicity. Huntingtin, the androgen receptor, ataxin-3 and atrophin-1 are substrates for caspases [135], and the expression of a dominant-negative caspase-1 mutant or intracerebroventricular administration of a caspase inhibitor delays the onset of symptoms in a mouse model for HD [136]. Neuronal cells with mutations in caspase-8 are resistant to polyglutamine-induced cell death in culture [137]. This suggests that the generation of proteolytic fragments containing the polyglutamine tract via a caspase-dependent apoptotic pathway may be important in the disease pathogenesis. In addition to causing cleavage of these proteins, interaction of the mutant protein with caspase-8 leads to caspase activation and recruitment into the nucleus, which may be another important component of disease pathology [137].

The pattern of selective cell toxicity in each disorder may reflect the intersection of the expression patterns of the polyglutamine-containing proteins and particular proteins with which they interact. Potential interacting proteins could be different proteases or other factors that allow the polyglutamine-containing protein to adopt its toxic conformation, or proteins that target the polyglutamine tract to its site of pathology in the cell. A large number of proteins that interact with the various affected proteins have been identified [see ref. 138 for examples], but their significance with respect to disease pathology is not known.

Similarities in the properties of the affected protein in this group of disorders and in other progressive neurological disorders such as Parkinson's disease, Alzheimer's disease and diseases resulting from prion proteins [130] may reflect a common mechanism whereby misfolded proteins produce neurodegeneration, and suggest potentially fruitful targets for drug design.

Myotonic dystrophy: a complex disorder with a complex etiology

Clinical features. Myotonic dystrophy (DM) is the most common inherited neuromuscular disorder in adults, with a global incidence of 1/8000. It is an autosomal dominant disorder with a highly variable clinical presentation involving multiple organ systems. The classic adult-onset form is progressive, and is characterized by distal muscle weakness and delayed muscle relaxation. Other symptoms include presenile cataracts, testicular atrophy, reduced fertility in females, diabetes, kidney failure, frontal balding in males, and cardiac muscle conduction abnormalities including a first-degree heart block. Atrophy of facial muscles produces a characteristic haggard appearance. Mental retardation as well as swallowing and speech difficulties are sometimes observed. Skeletal muscle biopsies show a marked atrophy and paucity of type I myofibers. Individuals with the congenital form of DM show delayed muscle maturation, have marked hypotonia, neonatal respiratory distress, and severe mental retardation.

Genetics. DM results from an expansion of the triplet CTG·CAG in the 3′-untranslated region of the DMPK gene which encodes a cAMP-dependent serine-threonine protein kinase [33, 139–141] However, DMPK homozygous knockout mice show only minor histopathological abnormalities late in life and heterozygous mice are phenotypically normal [142, 143]. Moreover, no point mutations in the DMPK gene that give rise to DM symptoms have been described to date. This suggests that simple DMPK haploinsufficiency cannot fully account for the complex range of symptoms seen in DM. Transgenic mice that overexpress a defective DMPK [63], DMPK knockout mice overexpressing normal DMPK [142], and mice with non-expressed CTG tracts [66] do not completely recapitulate the human disease either. It now seems likely that DM results from some combination of three distinct mechanisms, namely the reduction in DMPK protein, a transdominant defect in RNA metabolism, as well as an effect of the expansion on adjacent genes.

A role for DMPK in the cardiac muscle conduction abnormalities and muscle weakness is suggested by the fact that DMPK phosphorylates muscle-specific voltage-gated Na^+ channels and the voltage-dependent Ca^{2+} release channel, and modulates Ca^{2+} homeostasis in skeletal muscle [144], and substantiated by the finding that homozygous DMPK knockout mice show a first degree heart block [145].

An RNA-processing defect is suggested by the fact that while the total amount of both expanded and normal DMPK transcripts is similar to that in unaffected individuals, the amount of mature transcript produced from expanded alleles [146, 147], and perhaps also from normal alleles [147], is reduced in DM patients. The ability of the expanded allele to affect the normal allele in *trans* might explain some aspects of the dominant inheritance of DM. DM transcripts containing expanded triplet repeats are retained within the nucleus where they form

discrete foci containing between 15–230 RNA molecules [148]. The hairpins formed by CUG repeats [8] may affect processing and transport from the nucleus. A number of CUG-repeat-binding proteins have also been identified [see ref. 149 for a review]. It is possible that binding of one of these proteins to the RNA prevents DM transcripts from leaving the nucleus. This may result in the sequestration of not only the RNA, but bound protein(s) as well. Sequestration of transcript-processing proteins may account for some DM symptoms including the abnormal levels of insulin receptor mRNA and the abberant splicing of cardiac troponin T that is seen in some DM patients [150].

Overexpression of part of the 3′-untranslated region of a normal DMPK allele inhibits terminal muscle differentiation, strongly reminiscent of what is seen in congenital DM [151]. High levels of this transcript within the nucleus may affect terminal differentiation, perhaps by binding factors required for normal myogenesis. Nuclear retention of DMPK transcripts in DM patients may have the same effect.

A general effect on expansion on chromatin structure is suggested by the fact that CTG·CAG repeats associate strongly with histones [25, 152], and the observation that DM chromosomes lack a DNase I hypersensitive site 0.7 kb downstream of the CTG repeats that is seen in normal individuals [153]. In addition, expanded alleles are hypermethylated, a phenomenon usually associated with transcriptional silencing [154]. Both phenomena may affect the production of proteins from adjacent genes.

SIX5, formerly known as DMAHP (DM-locus-associated homeodomain protein), has a promoter that may overlap with the 3′-untranslated region of DMPK [155]. However, to date, studies of the effect of expansion on SIX5 expression have been contradictory. SIX5 is homologous to the *Drosophila* eye development gene *sine oculis* (*so*). The expression pattern for SIX5 in the normal adult eye matches the sites of the ocular pathology in DM, while DMPK is not expressed in adult lens. The SIX5 protein binds to the Na^+,K^+-ATPase $\alpha 1$ subunit gene, and by analogy with Nakano mice which show heriditary cataract development, SIX5 dysfunction may cause cataracts by affecting ion balance in the lens. SIX5 is not expressed in fetal eyes, which may explain why congenital DM patients do not show cataract development in infancy [155].

Gene 59 is located < 1 kb upstream of DMPK. It encodes a protein with WD repeat motifs normally associated with proteins involved in signal transduction or cell regulation. There is an inverse relationship between the size of the CTG·CAG expansion and levels of cytoplasmic polyA$^+$ mRNA from this gene in muscle biopsies of DM patients [156]. However, the mouse homolog of this gene, DMR-9, is highly expressed in brain and testis but not in skeletal muscle. Reduced expression of this gene may contribute to mental retardation and the reduced fertility seen in some DM patients.

Spinocerebellar ataxia type 8 (SCA8): an antisense RNA disease?

Clinical features. SCA8 is a rare autosomal dominant ataxia [81] that has an onset from age 18 to 65, with a mean of 39 years. Symptoms include cerebellar atrophy, spastic and ataxic dysarthria, nystagmus, limb and gait ataxia, limb spasticity, and diminished vibration perception. Progression is slow, but severely affected family members are non-ambulatory by 40 or 50.

Genetics. The most prevalent SCA8 mutation is a CTG repeat that maps to the 3′ region of a transcript with no open reading frames [81]. It is the first dominant spinocerebellar ataxia that is not caused by an expanded polyglutamine tract. The SCA8 gene has no homolog in mice. Its 5′ end overlaps with a gene called Kelch-like 1 (KLHL1), which is transcribed from the opposite strand and that has homology to the *Drosophila* actin-binding protein kelch [157]. KLHL1 is brain specific and is expressed in high levels in the cerebellum, frontal lobe, and subthalamic nuclei. SCA8 is expressed at highest levels in the cerebellum, the organ most affected in this disorder. The SCA8 transcript may act as an antisense RNA directed towards KLHL1. Transcripts containing expanded CUG tracts may be more stable than transcripts produced from normal alleles, perhaps because the long CUG tract forms a hairpin which protects the RNA from degradation by endogenous 3′–5′ exonucleases. This may effectively increase the antisense effect of SCA8. However, it is not known at this time whether SCA8 and KLHL1 are expressed in the same cells in the brain, or if SCA8 RNA accumulates to higher levels in affected individuals.

Very large SCA8 alleles are known. Curiously, these are either less pathogenic than shorter alleles or not pathogenic at all [3]. One explanation could be that SCA8 transcripts containing alleles in the pathological range can be exported to the cytoplasm where they exert their effect on the KLHL1 transcript, while transcripts with larger expansions are retained in the nucleus for reasons analogous to those proposed for DMPK transcripts in DM.

Conclusion

Many questions about the mechanism and the consequences of expansion remain unanswered. These questions are not simply of academic interest since their

answers may suggest therapeutic approaches to this group of disorders. In addition, a number of other diseases for which the genetic basis is not yet known may in fact turn out to be repeat expansion disorders. For example, the genes responsible for more than one-third of the cases of spinocerebellar ataxia are unknown. A single case of a sporadic-onset cerebellar ataxia accompanied by intellectual deterioration has recently been reported to be associated with an expansion of the polyglutamine tract in the TATA-binding protein (TBP) [158]. Further work is needed to assess whether the symptoms are indeed due to the TBP mutation, and whether expansions in TBP are a more widespread potential source of ataxias. Anticipation, one of the hallmarks of the repeat expansion disorders, has been reported in a large number of other diseases, including Crohn's disease, schizophrenia, and familial primary pulmonary hypertension [reviewed in ref. 1]. This category of disorders may thus increase in size and complexity as the genetic basis of additional diseases is elucidated.

Acknowledgements. The authors wish to thank Stylianos E. Antonarakis, Walter Doerfler, Mark Hirst, and Robert D. Wells for sharing some of their data prior to publication. We also wish to thank Anthony Furano, Laura Ranum, Martine Usdin, John Hanover, and Nancy Nossal for helpful discussions.

1 Paterson A. D., Naimark D. M. J., Vincent J. B., Kennedy J. L. and Petronis A. (1998) Genetic anticipation in neurological and other disorders. In: Genome Instabilities and the Hereditary Neurological Diseases, pp. 413–428, Wells R. D. and Warren S. T. (eds), Academic Press, San Diego
2 Lalioti M. D., Scott H. S., Genton P., Grid D., Ouazzani R., M'Rabet A. et al. (1998) A PCR amplification method reveals instability of the dodecamer repeat in progressive myoclonus epilepsy (EPM1) and no correlation between the size of the repeat and age at onset. Am. J. Hum. Genet. **62:** 842–847
3 Ranum L. P. W., Moseley M. L., Leppert M. F., van den Engh G., La Spada A. R., Koob M. D. et al. (1999) Massive CTG expansions and deletions may reduce penetrance of spinocerebellar ataxia type 8. Am. J. Hum. Genet. **65:** A466
4 Lafreniere R. G., Rochefort D. L., Chretien N., Rommens J. M., Cochius J. I., Kalviainen R. et al. (1997) Unstable insertion in the 5′ flanking region of the cystatin B gene is the most common mutation in progressive myoclonus epilepsy type 1, EPM1. Nat. Genet. **15:** 298–302
5 Sanpei K., Takano H., Igarashi S., Sato T., Oyake M., Sasaki H. et al. (1996) Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. Nat. Genet. **14:** 277–284
6 Bacolla A., Gellibolian R., Shimizu M., Amirhaeri S., Kang S., Ohshima K. et al. (1997) Flexible DNA: genetically unstable CTG.CAG and CGG.CCG from human hereditary neuromuscular disease genes. J. Biol. Chem. **272:** 16783–16792
7 Mitas M. (1997) Trinucleotide repeats associated with human disease. Nucleic Acids Res. **25:** 2245–2254
8 Napieraa M. and Krzyosiak W. J. (1997) CUG repeats present in myotonin kinase RNA form metastable 'slippery' hairpins. J. Biol. Chem. **272:** 31079–31085
9 Usdin K. (1998) NGG-triplet repeats form similar intrastrand structures: implications for the triplet expansion diseases. Nucleic Acids Res. **26:** 4078–4085
10 Usdin K. and Woodford K. J. (1995) CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. Nucleic Acids Res. **23:** 4202–4209
11 Fry M. and Loeb L. A. (1994) The fragile X syndrome $d(CGG)_n$ nucleotide repeats form a stable tetrahelical structure. Proc. Natl. Acad. Sci. USA **91:** 4950–4954
12 Pearson C. E., Wang Y. H., Griffith J. D. and Sinden R. R. (1998) Structural analysis of slipped-strand DNA (S-DNA) formed in $(CTG)_n.(CAG)_n$ repeats from the myotonic dystrophy locus. Nucleic Acids Res. **26:** 816–823
13 Gacy A. M., Goellner G., Juranic N., Macura S. and McMurray C. T. (1995) Trinucleotide repeats that expand in human disease form hairpin structures in vitro. Cell **81:** 533–540
14 Hanvey J. C., Shimizu M. and Wells R. D. (1988) Intramolecular DNA triplexes in supercoiled plasmids. Proc. Natl. Acad. Sci. USA **85:** 6292–6296
15 Sakamoto N., Chastain P. D., Parniewski P., Ohshima K., Pandolfo M., Griffith J. D. et al. (1999) Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia. Mol. Cell **3:** 465–475
16 Leroy J. L. and Gueron M. (1995) Solution structures of the i-motif tetramers of d(TCC), d(5methylCCT) and d(T5methylCC): novel NOE connections between amino protons and sugar protons. Structure **3:** 101–120
17 Catasti P., Chen X., Moyzis R. K., Bradbury E. M. and Gupta G. (1996) Structure-function correlations of the insulin-linked polymorphic region. J. Mol. Biol. **264:** 534–545
18 Weitzmann M. N., Woodford K. J. and Usdin K. (1997) DNA secondary structures and the evolution of hypervariable tandem arrays. J. Biol. Chem. **272:** 9517–9523
19 Weitzmann M. N., Woodford K. J. and Usdin K. (1998) The mouse Ms6-hm hypervariable microsatellite forms a hairpin and two unusual tetraplexes. J. Biol. Chem. **273:** 30742–30749
20 Kang S., Ohshima K., Shimizu M., Amirhaeri S. and Wells R. D. (1995) Pausing of DNA synthesis in vitro at specific loci in CTG and CGG triplet repeats from human hereditary disease genes. J. Biol. Chem. **270:** 27014–27021
21 Gacy A. M., Goellner G. M., Spiro C., Chen X., Gupta G., Bradbury E. M. et al. (1998) GAA instability in Friedreich's ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. Mol. Cell **1:** 583–593
22 Samadashwily G. M., Raca G. and Mirkin S. M. (1997) Trinucleotide repeats affect DNA replication in vivo. Nat. Genet. **17:** 298–304
23 Ohshima K., Montermini L., Wells R. D. and Pandolfo M. (1998) Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo. J. Biol. Chem. **273:** 14588–14595
24 Wang Y. H. and Griffith J. (1996) Methylation of expanded CCG triplet repeat DNA from fragile X syndrome patients enhances nucleosome exclusion. J. Biol. Chem. **271:** 22937–22940
25 Wang Y. H., Amirhaeri S., Kang S., Wells R. D. and Griffith J. D. (1994) Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene. Science **265:** 669–671
26 Sinden R. R., Zheng G. X., Brankamp R. G. and Allen K. N. (1991) On the deletion of inverted repeated DNA in *Escherichia coli*: effects of length, thermal stability, and cruciform formation in vivo. Genetics **129:** 991–1005
27 Moore H., Greenwell P. W., Liu C. P., Arnheim N. and Petes T. D. (1999) Triplet repeats form secondary structures that escape DNA repair in yeast. Proc. Natl. Acad. Sci. USA **96:** 1504–1509

28 Kohwi Y. and Panchenko Y. (1993) Transcription-dependent recombination induced by triple-helix formation. Genes Dev. **7:** 1766–1778

29 Burkholder G. D., Latimer L. J. and Lee J. S. (1991) Immunofluorescent localization of triplex DNA in polytene chromosomes of *Chironomus* and *Drosophila*. Chromosoma **101:** 11–18

30 Snow K., Tester D. J., Kruckeberg K. E., Schaid D. J. and Thibodeau S. N. (1994) Sequence analysis of the fragile X trinucleotide repeat: implications for the origin of the fragile X mutation. Hum. Mol. Genet. **3:** 1543–1551

31 Eichler E. E., Holden J. J., Popovich B. W., Reiss A. L., Snow K., Thibodeau S. N. et al. (1994) Length of uninterrupted CGG repeats determines instability in the FMR1 gene. Nat. Genet. **8:** 88–94

32 Fu Y. H., Kuhl D. P., Pizzuti A., Pieretti M., Sutcliffe J. S., Richards S. et al. (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell **67:** 1047–1058

33 Aslanidis C., Jansen G., Amemiya C., Shutler G., Mahadevan M., Tsilfidis C. et al. (1992) Cloning of the essential myotonic dystrophy region and mapping of the putative defect. Nature **355:** 548–551

34 Devys D., Biancalana V., Rousseau F., Boue J., Mandel J. L. and Oberle I. (1992) Analysis of full fragile X mutations in fetal tissues and monozygotic twins indicate that abnormal methylation and somatic heterogeneity are established early in development. Am. J. Med. Genet. **43:** 208–216

35 Cancel G., Abbas N., Stevanin G., Durr A., Chneiweiss H., Neri C. et al. (1995) Marked phenotypic heterogeneity associated with expansion of a CAG repeat sequence at the spinocerebellar ataxia 3/Machado-Joseph disease locus. Am. J. Hum. Genet. **57:** 809–816

36 Moutou C., Vincent M. C., Biancalana V. and Mandel J. L. (1997) Transition from premutation to full mutation in fragile X syndrome is likely to be prezygotic. Hum. Mol. Genet. **6:** 971–979

37 Reyniers E., Martin J. J., Cras P., Van Marck E., Handig I., Jorens H. Z. et al. (1999) Postmortem examination of two fragile X brothers with an FMR1 full mutation. Am. J. Med. Genet. **84:** 245–249

38 MacDonald M. E., Barnes G., Srinidhi J., Duyao M. P., Ambrose C. M., Myers R. H. et al. (1993) Gametic but not somatic instability of CAG repeat length in Huntington's disease. J. Med. Genet. **30:** 982–986

39 Schumacher S., Fuchs R. P. and Bichara M. (1998) Expansion of CTG repeats from human disease genes is dependent upon replication mechanisms in *Escherichia coli*: the effect of long patch mismatch repair revisited. J. Mol. Biol. **279:** 1101–1110

40 Schweitzer J. K. and Livingston D. M. (1997) Destabilization of CAG trinucleotide repeat tracts by mismatch repair mutations in yeast. Hum. Mol. Genet. **6:** 349–355

41 Fulchignoni-Lataud M. C., Olchwang S. and Serre J. L. (1997) The fragile X CGG repeat shows a marked level of instability in hereditary non-polyposis colorectal cancer patients. Eur. J. Hum. Genet. **5:** 89–93

42 Morag A. S., Saveson C. J. and Lovett S. T. (1999) Expansion of DNA repeats in *Escherichia coli*: effects of recombination and replication functions. J. Mol. Biol. **289:** 21–27

43 Schweitzer J. K. and Livingston D. M. (1999) The effect of DNA replication mutations on CAG tract stability in yeast. Genetics **152:** 953–963

44 Streisinger G., Okada Y., Emrich J., Newton J., Tsugita A., Terzaghi E. et al. (1966) Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31:** 77–84

45 Richards R. I. and Sutherland G. R. (1994) Simple repeat DNA is not replicated simply. Nat. Genet. **6:** 114–116

46 White P. J., Borts R. H. and Hirst M. C. (1999) Stability of the human fragile X $(CGG)_{(n)}$ triplet repeat array in *Saccharomyces cerevisiae* deficient in aspects of DNA metabolism. Mol. Cell. Biol. **19:** 5675–5684

47 Schweitzer J. K. and Livingston D. M. (1998) Expansions of CAG repeat tracts are frequent in a yeast mutant defective in Okazaki fragment maturation. Hum. Mol. Genet. **7:** 69–74

48 Freudenreich C. H., Kantrow S. M. and Zakian V. A. (1998) Expansion and length-dependent fragility of CTG repeats in yeast. Science **279:** 853–856

49 Murante R. S., Rust L. and Bambara R. A. (1995) Calf 5′ to 3′ exo/endonuclease must slide from a 5′ end of the substrate to perform structure-specific cleavage. J. Biol. Chem. **270:** 30377–30383

50 Gordenin D. A., Kunkel T. A. and Resnick M. A. (1997) Repeat expansion – all in a flap? Nat. Genet. **16:** 116–118

51 Jakupciak J. P. and Wells R. D. (1999) Genetic instabilities in (CTG.CAG) repeats occur by recombination. J. Biol. Chem. **274:** 23468–23479

52 Michel B., Ehrlich S. D. and Uzest M. (1997) DNA double-strand breaks caused by replication arrest. EMBO J. **16:** 430–438

53 Szostak J. W., Orr-Weaver T. L., Rothstein R. J. and Stahl F. W. (1983) The double-strand-break repair model for recombination. Cell **33:** 25–35

54 Moseley M. L., Koob M. D. and Ranum L. P. W. (1999) Frequent duplications of triplet interruptions within a seven generation SCA8 family. Am. J. Hum. Genet. **65:** A462

55 Tamaki K., May C. A., Dubrova Y. E. and Jeffreys A. J. (1999) Extremely complex repeat shuffling during germline mutation at human minisatellite B6.7. Hum. Mol. Genet. **8:** 879–888

56 Paques F., Leung W. Y. and Haber J. E. (1998) Expansions and contractions in a tandem repeat induced by double-strand break repair. Mol. Cell. Biol. **18:** 2045–2054

57 Brock G. J., Anderson N. H. and Monckton D. G. (1999) Cis-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. Hum. Mol. Genet. **8:** 1061–1067

58 Neville C. E., Mahadevan M. S., Barcelo J. M. and Korneluk R. G. (1994) High resolution genetic analysis suggests one ancestral predisposing haplotype for the origin of the myotonic dystrophy mutation. Hum. Mol. Genet. **3:** 45–51

59 Squitieri F., Andrew S. E., Goldberg Y. P., Kremer B., Spence N., Zeisler J. et al. (1994) DNA haplotype analysis of Huntington disease reveals clues to the origins and mechanisms of CAG expansion and reasons for geographic variations of prevalence. Hum. Mol. Genet. **3:** 2103–2114

60 Takiyama Y., Igarashi S., Rogaeva E. A., Endo K., Rogaev E. I., Tanaka H. et al. (1995) Evidence for inter-generational instability in the CAG repeat in the MJD1 gene and for conserved haplotypes at flanking markers amongst Japanese and Caucasian subjects with Machado-Joseph disease. Hum. Mol. Genet. **4:** 1137–1146

61 Cossee M., Schmitt M., Campuzano V., Reutenauer L., Moutou C., Mandel J. L. et al. (1997) Evolution of the Friedreich's ataxia trinucleotide repeat expansion: founder effect and premutations. Proc. Natl. Acad. Sci. USA **94:** 7452–7457

62 Murray A., Macpherson J. N., Pound M. C., Sharrock A., Youings S. A., Dennis N. R. et al. (1997) The role of size, sequence and haplotype in the stability of FRAXA and FRAXE alleles during transmission. Hum. Mol. Genet. **6:** 173–184

63 Gourdon G., Radvanyi F., Lia A. S., Duros C., Blanche M., Abitbol M. et al. (1997) Moderate intergenerational and somatic instability of a 55-CTG repeat in transgenic mice. Nat. Genet. **15:** 190–192

64 Wheeler V. C., Auerbach W., White J. K., Srinidhi J., Auerbach A., Ryan A. et al. (1999) Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. Hum. Mol. Genet. **8:** 115–122

65 Lavedan C., Grabczyk E., Usdin K. and Nussbaum R. L. (1998) Long uninterrupted CGG repeats within the first exon of the human FMR1 gene are not intrinsically unstable in transgenic mice. Genomics **50:** 229–240

66 Monckton D. G., Coolbaugh M. I., Ashizawa K. T., Siciliano M. J. and Caskey C. T. (1997) Hypermutable myotonic dystrophy CTG repeats in transgenic mice. Nat. Genet. **15:** 193–196

67 Bois P., Williamson J., Brown J., Dubrova Y. E. and Jeffreys A. J. (1998) A novel unstable mouse VNTR family expanded from SINE B1 elements. Genomics **49:** 122–128

68 Djalali M., Adolph S., Steinbach P., Winking H. and Hameister H. (1987) A comparative mapping study of fragile sites in the human and murine genomes. Hum. Genet. **77:** 157–162

69 Sutherland G. R. and Richards R. I. (1995) The molecular basis of fragile sites in human chromosomes. Curr. Opin. Genet. Dev. **5:** 323–327

70 Reyniers E., Vits L., De Boulle K., Van Roy B., Van Velzen D., Graaff E. de et al. (1993) The full mutation in the FMR-1 gene of male fragile X patients is absent in their sperm. Nat. Genet. **4:** 143–146

71 Helderman-van den Enden A. T., Maaswinkel-Mooij P. D., Hoogendoorn E., Willemsen R., Maat-Kievit J. A., Losekoot M. et al. (1999) Monozygotic twin brothers with the fragile X syndrome: different CGG repeats and different mental capacities. J. Med. Genet. **36:** 253–257

72 Tassone F., Hagerman R. J., Gane L. W. and Taylor A. K. (1999) Strong similarities of the FMR1 mutation in multiple tissues: postmortem studies of a male with a full mutation and a male carrier of a premutation. Am. J. Med. Genet. **84:** 240–244

73 Wohrle D., Salat U., Glaser D., Mucke J., Meisel-Stosiek M., Schindler D. et al. (1998) Unusual mutations in high functioning fragile X males: apparent instability of expanded unmethylated CGG repeats. J. Med. Genet. **35:** 103–111

74 Sutcliffe J. S., Nelson D. L., Zhang F., Pieretti M., Caskey C. T., Saxe D. et al. (1992) DNA methylation represses FMR-1 transcription in fragile X syndrome. Hum. Mol. Genet. **1:** 397–400

75 Bowater R. P., Jaworski A., Larson J. E., Parniewski P. and Wells R. D. (1997) Transcription increases the deletion frequency of long CTG·CAG triplet repeats from plasmids in *Escherichia coli*. Nucleic Acids Res. **25:** 2861–2868

76 Nagafuchi S., Yanagisawa H., Ohsaki E., Shirayama T., Tadokoro K., Inoue T. et al. (1994) Structure and expression of the gene responsible for the triplet repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA). Nat. Genet. **8:** 177–182

77 Kremer B., Almqvist E., Theilmann J., Spence N., Telenius H., Goldberg Y. P. et al. (1995) Sex-dependent mechanisms for expansions and contractions of the CAG repeat on affected Huntington disease chromosomes. Am. J. Hum. Genet. **57:** 343–350

78 Duyao M., Ambrose C., Myers R., Novelletto A., Persichetti F., Frontali M. et al. (1993) Trinucleotide repeat length instability and age of onset in Huntington's disease. Nat. Genet. **4:** 387–392

79 David G., Abbas N., Stevanin G., Durr A., Yvert G., Cancel G. et al. (1997) Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. Nat. Genet. **17:** 65–70

80 Sherman S. L., Morton N. E., Jacobs P. A. and Turner G. (1984) The marker (X) syndrome: a cytogenetic and genetic analysis. Ann. Hum. Genet. **48:** 21–37

81 Koob M. D., Moseley M. L., Schut L. J., Benzow K. A., Bird T. D., Day J. W. et al. (1999) An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). Nat. Genet. **21:** 379–384

82 Redman J. B., Fenwick R. G. Jr, Fu Y. H., Pizzuti A. and Caskey C. T. (1993) Relationship between parental trinucleotide GCT repeat length and severity of myotonic dystrophy in offspring. JAMA **269:** 1960–1965

83 Vogel F. and Rathenberg R. (1975) Spontaneous mutation in man. Adv. Hum. Genet. **5:** 223–318

84 Leeflang E. P., Zhang L., Tavare S., Hubert R., Srinidhi J., MacDonald M. E. et al. (1995) Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: quantification of the mutation frequency spectrum. Hum. Mol. Genet. **4:** 1519–1526

85 The Dutch-Belgian Fragile X Consortium (1994) Fmr1 knockout mice: a model to study fragile X mental retardation. Cell **78:** 23–33

86 Balhorn R. (1982) A model for the structure of chromatin in mammalian sperm. J. Cell. Biol. **93:** 298–305

87 LeMaire-Adkins R., Radke K. and Hunt P. A. (1997) Lack of checkpoint control at the metaphase/anaphase transition: a mechanism of meiotic nondisjunction in mammalian females. J. Cell. Biol. **139:** 1611–1619

88 De Michele G., Cavalcanti F., Criscuolo C., Pianese L., Monticelli A., Filla A. et al. (1998) Parental gender, age at birth and expansion length influence GAA repeat intergenerational instability in the X25 gene: pedigree studies and analysis of sperm from patients with Friedreich's ataxia. Hum. Mol. Genet. **7:** 1901–1906

89 Ashley-Koch A. E., Robinson H., Glicksman A. E., Nolin S. L., Schwartz C. E., Brown W. T. et al. (1998) Examination of factors associated with instability of the FMR1 CGG repeat. Am. J. Hum. Genet. **63:** 776–785

90 Lalioti M. D., Mirotsou M., Buresi C., Peitsch M. C., Rossier C., Ouazzani R. et al. (1997) Identification of mutations in cystatin B, the gene responsible for the Unverricht-Lundborg type of progressive myoclonus epilepsy (EPM1). Am. J. Hum. Genet. **60:** 342–351

91 Lalioti M. D., Scott H. S., Buresi C., Rossier C., Bottani A., Morris M. A. et al. (1997) Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. Nature **386:** 847–851

92 Pennacchio L. A., Bouley D. M., Higgins K. M., Scott M. P., Noebels J. L. and Myers R. M. (1998) Progressive ataxia, myoclonic epilepsy and cerebellar apoptosis in cystatin B-deficient mice. Nat. Genet. **20:** 251–258

93 Lalioti M. D., Scott H. S. and Antonarakis S. E. (1999) Altered spacing of promoter elements due to the dodecamer repeat expansion contributes to reduced expression of the cystatin B gene in EPM1. Hum. Mol. Genet. **8:** 1791–1798

94 Comery T. A., Harris J. B., Willems P. J., Oostra B. A., Irwin S. A., Weiler I. J. et al. (1997) Abnormal dendritic spines in fragile X knockout mice: maturation and pruning deficits. Proc. Natl. Acad. Sci. USA **94:** 5401–5404

95 Macpherson J., Murray A., Webb J. and Jacobs P. (1999) Fragile X syndrome: of POF and premutations. J. Med. Genet. **36:** 171–172

96 Kremer E. J., Pritchard M., Lynch M., Yu S., Holman K., Baker E. et al. (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence $p(CCG)_n$. Science **252:** 1711–1714

97 Verkerk A. J., Pieretti M., Sutcliffe J. S., Fu Y. H., Kuhl D. P., Pizzuti A. et al. (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell **65:** 905–914

98 Knight S. J., Flannery A. V., Hirst M. C., Campbell L., Christodoulou Z., Phelps S. R. et al. (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell **74:** 127–134

99 Gecz J., Oostra B. A., Hockey A., Carbonell P., Turner G., Haan E. A. et al. (1997) FMR2 expression in families with FRAXE mental retardation. Hum. Mol. Genet. **6:** 435–441

100 Chiurazzi P., Pomponi M. G., Willemsen R., Oostra B. A. and Neri G. (1998) In vitro reactivation of the FMR1 gene involved in fragile X syndrome. Hum. Mol. Genet. **7:** 109–113

101 Smith S. S., Laayoun A., Lingeman R. G., Baker D. J. and Riley J. (1994) Hypermethylation of telomere-like foldbacks at codon 12 of the human c-Ha-ras gene and the trinucleotide repeat of the FMR-1 gene of fragile X. J. Mol. Biol. **243:** 143–151

102 Ashley C. T., Wilkinson K. D. Jr, Reines D. and Warren S. T. (1993) FMR1 protein: conserved RNP family domains and selective RNA binding. Science **262:** 563–566

103 Weiler I. J. and Greenough W. T. (1999) Synaptic synthesis of the fragile X protein: possible involvement in synapse maturation and elimination. Am. J. Med. Genet. **83:** 248–252

104 Feng Y., Absher D., Eberhart D. E., Brown V., Malter H. E. and Warren S. T. (1997) FMRP associates with polyribosomes as an mRNP, and the I304N mutation of severe fragile X syndrome abolishes this association. Mol. Cell **1:** 109–118

105 Gecz J., Bielby S., Sutherland G. R. and Mulley J. C. (1997) Gene structure and subcellular localization of FMR2, a member of a new family of putative transcription activators. Genomics **44:** 201–213

106 Glover T. W. and Howard-Peebles P. N. (1983) The combined effects of FUdR addition and methionine depletion on the X-chromosome fragile site. Am. J. Hum. Genet. **35:** 117–122

107 Hansen R. S., Canfield T. K., Fjeld A. D., Mumm S., Laird C. D. and Gartler S. M. (1997) A variable domain of delayed replication in FRAXA fragile X chromosomes: X inactivation-like spread of late replication. Proc. Natl. Acad. Sci. USA **94:** 4587–4592

108 Yeshaya J., Shalgi R., Shohat M. and Avivi L. (1998) Replication timing of the various FMR1 alleles detected by FISH: inferences regarding their transcriptional status. Hum. Genet. **102:** 6–14

109 Jones C., Slijepcevic P., Marsh S., Baker E., Langdon W. Y., Richards R. I. et al. (1994) Physical linkage of the fragile site FRA11B and a Jacobsen syndrome chromosome deletion breakpoint in 11q23.3. Hum. Mol. Genet. **3:** 2123–2130

110 Glover T. W. (1998) Instability at chromosomal fragile sites. Recent Results Cancer Res. **154:** 185–199

111 Campuzano V., Montermini L., Molto M. D., Pianese L., Cossee M., Cavalcanti F. et al. (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science **271:** 1423–1427

112 Coppola G., De Michele G., Cavalcanti F., Pianese L., Perretti A., Santoro L. et al. (1999) Why do some Friedreich's ataxia patients retain tendon reflexes? A clinical, neurophysiological and molecular study. J. Neurol. **246:** 353–357

113 Montermini L., Richter A., Morgan K., Justice C. M., Julien D., Castellotti B. et al. (1997) Phenotypic variability in Friedreich ataxia: role of the associated GAA triplet repeat expansion. Ann. Neurol. **41:** 675–682

114 Campuzano V., Montermini L., Lutz Y., Cova L., Hindelang C., Jiralerspong S. et al. (1997) Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. Hum. Mol. Genet. **6:** 1771–1780

115 Grabczyk E. and Fishman M. C. (1995) A long purine-pyrimidine homopolymer acts as a transcriptional diode. J. Biol. Chem. **270:** 1791–1797

116 Ohshima K., Sakamoto N., Labuda M., Poirier J., Moseley M. L., Montermini L. et al. (1999) A nonpathogenic GAAGGA repeat in the Friedreich gene: implications for pathogenesis. Neurology **53:** 1854–1857

117 Koutnikova H., Campuzano V., Foury F., Dolle P., Cazzalini O. and Koenig M. (1997) Studies of human, mouse and yeast homologues indicate a mitochondrial function for frataxin. Nat. Genet. **16:** 345–351

118 Radisky D. C., Babcock M. C. and Kaplan J. (1999) The yeast frataxin homologue mediates mitochondrial iron efflux: evidence for a mitochondrial iron cycle. J. Biol. Chem. **274:** 4497–4499

119 Isaya G., Adamec J., Rusnak F., Owen W. G., Naylor S. and Benson L. M. (1999) Frataxin is an iron-storage protein. Am. J. Hum. Genet. **65:** A33

120 Rotig A., Lonlay P. de, Chretien D., Foury F., Koenig M., Sidi D. et al. (1997) Aconitase and mitochondrial iron-sulphur protein deficiency in Friedreich ataxia. Nat. Genet. **17:** 215–217

121 Rustin P., Kleist-Retzow J. C. von, Chantrel-Groussard K., Sidi D., Munnich A. and Rotig A. (1999) Effect of idebenone on cardiomyopathy in Friedreich's ataxia: a preliminary study. Lancet **354:** 477–479

122 Clark H. B., Burright E. N., Yunis W. S., Larson S., Wilcox C., Hartman B. et al. (1997) Purkinje cell expression of a mutant allele of SCA1 in transgenic mice leads to disparate effects on motor behaviors, followed by a progressive cerebellar dysfunction and histological alterations. J. Neurosci. **17:** 7385–7395

123 La Spada A. R., Wilson E. M., Lubahn D. B., Harding A. E. and Fischbeck K. H. (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature **352:** 77–79

124 Zhuchenko O., Bailey J., Bonnen P., Ashizawa T., Stockton D. W., Amos C. et al. (1997) Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. Nat. Genet. **15:** 62–69

125 Jen J. C., Yue Q., Karrim J., Nelson S. F. and Baloh R. W. (1998) Spinocerebellar ataxia type 6 with positional vertigo and acetazolamide responsive episodic ataxia. J. Neurol. Neurosurg. Psychiatry **65:** 565–568

126 Bates G. P. and Davies S. W. (1997) Transgenic mouse models of neurodegenerative disease caused by CAG/polyglutamine expansions. Mol. Med. Today **3:** 508–515

127 Ross C. A., Margolis R. L., Becher M. W., Wood J. D., Engelender S., Cooper J. K. et al. (1998) Pathogenesis of neurodegenerative diseases associated with expanded glutamine repeats: new answers, new questions. Prog. Brain Res. **117:** 397–419

128 Klement I. A., Skinner P. J., Kaytor M. D., Yi H., Hersch S. M., Clark H. B. et al. (1998) Ataxin-1 nuclear localization and aggregation: role in polyglutamine-induced disease in SCA1 transgenic mice. Cell **95:** 41–53

129 Peters M. F., Nucifora F. C., Kushi J. Jr, Seaman H. C., Cooper J. K., Herring W. J. et al. (1999) Nuclear targeting of mutant huntingtin increases toxicity. Mol. Cell. Neurosci. **14:** 121–128

130 Scherzinger E., Sittler A., Schweiger K., Heiser V., Lurz R., Hasenbank R. et al. (1999) Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology. Proc. Natl. Acad. Sci. USA **96:** 4604–4609

131 Perutz M. F. (1996) Glutamine repeats and inherited neurodegenerative diseases: molecular aspects. Curr. Opin. Struct. Biol. **6:** 848–858

132 Kahlem P., Green H. and Djian P. (1998) Transglutaminase action imitates Huntington's disease: selective polymerization of huntingtin containing expanded polyglutamine. Mol. Cell **1:** 595–601

133 Stenoien D. L., Cummings C. J., Adams H. P., Mancini M. G., Patel K., DeMartino G. N. et al. (1999) Polyglutamine-expanded androgen receptors form aggregates that sequester heat shock proteins, proteasome components and SRC-1, and are suppressed by the HDJ-2 chaperone. Hum. Mol. Genet. **8:** 731–741

134 Saudou F., Finkbeiner S., Devys D. and Greenberg M. E. (1998) Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. Cell **95:** 55–66

135 Wellington C. L., Ellerby L. M., Hackam A. S., Margolis R. L., Trifiro M. A., Singaraja R. et al. (1998) Caspase cleavage of gene products associated with triplet expansion disorders generates truncated fragments containing the polyglutamine tract. J. Biol. Chem. **273:** 9158–9167

136 Ona V. O., Li M., Vonsattel J. P., Andrews L. J., Khan S. Q., Chung W. M. et al. (1999) Inhibition of caspase-1 slows disease progression in a mouse model of Huntington's disease. Nature **399:** 263–267

137 Sanchez I., Xu C. J., Juo P., Kakizaka A., Blenis J. and Yuan J. (1999) Caspase-8 is required for cell death induced by expanded polyglutamine repeats. Neuron **22:** 623–633

138 Reddy P. H., Williams M. and Tagle D. A. (1999) Recent advances in understanding the pathogenesis of Huntington's disease. Trends Neurosci. **22:** 248–255

139 Harley H. G., Brook J. D., Rundle S. A., Crow S., Reardon W., Buckler A. J. et al. (1992) Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. Nature **355:** 545–546
140 Mahadevan M., Tsilfidis C., Sabourin L., Shutler G., Amemiya C., Jansen G. et al. (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. Science **255:** 1253–1255
141 Buxton J., Shelbourne P., Davies J., Jones C., Van Tongeren T., Aslanidis C. et al. (1992) Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. Nature **355:** 547–548
142 Jansen G., Groenen P. J., Bachner D., Jap P. H., Coerwinkel M., Oerlemans F. et al. (1996) Abnormal myotonic dystrophy protein kinase levels produce only mild myopathy in mice. Nat. Genet. **13:** 316–324
143 Reddy S., Smith D. B., Rich M. M., Leferovich J. M., Reilly P., Davis B. M. et al. (1996) Mice lacking the myotonic dystrophy protein kinase develop a late onset progressive myopathy. Nat. Genet. **13:** 325–335
144 Benders A. A., Groenen P. J., Oerlemans F. T., Veerkamp J. H. and Wieringa B. (1997) Myotonic dystrophy protein kinase is involved in the modulation of the Ca^{2+} homeostasis in skeletal muscle cells. J. Clin. Invest. **100:** 1440–1447
145 Berul C. I., Maguire C. T., Aronovitz M. J., Greenwood J., Miller C., Gehrmann J. et al. (1999) DMPK dosage alterations result in atrioventricular conduction abnormalities in a mouse myotonic dystrophy model. J. Clin. Invest. **103:** R1–7
146 Krahe R., Ashizawa T., Abbruzzese C., Roeder E., Carango P., Giacanelli M. et al. (1995) Effect of myotonic dystrophy trinucleotide repeat expansion on DMPK transcription and processing. Genomics **28:** 1–14
147 Wang J., Pegoraro E., Menegazzo E., Gennarelli M., Hoop R. C., Angelini C. et al. (1995) Myotonic dystrophy: evidence for a possible dominant-negative RNA mutation. Hum. Mol. Genet. **4:** 599–606
148 Taneja K. L., McCurrach M., Schalling M., Housman D. and Singer R. H. (1995) Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. J. Cell. Biol. **128:** 995–1002
149 Timchenko L. T. (1999) Myotonic dystrophy: the role of RNA CUG triplet repeats. Am. J. Hum. Genet. **64:** 360–364
150 Philips A. V., Timchenko L. T. and Cooper T. A. (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. Science **280:** 737–741
151 Sabourin L. A., Tamai K., Narang M. A. and Korneluk R. G. (1997) Overexpression of 3′-untranslated region of the myotonic dystrophy kinase cDNA inhibits myoblast differentiation in vitro. J. Biol. Chem. **272:** 29626–29635
152 Godde J. S. and Wolffe A. P. (1996) Nucleosome assembly on CTG triplet repeats. J. Biol. Chem. **271:** 15222–15229
153 Otten A. D. and Tapscott S. J. (1995) Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. Proc. Natl. Acad. Sci. USA **92:** 5465–5469
154 Steinbach P., Glaser D., Vogel W., Wolf M. and Schwemmle S. (1998) The DMPK gene of severely affected myotonic dystrophy patients is hypermethylated proximal to the largely expanded CTG repeat. Am. J. Hum. Genet. **62:** 278–285
155 Winchester C. L., Ferrier R. K., Sermoni A., Clark B. J. and Johnson K. J. (1999) Characterization of the expression of DMPK and SIX5 in the human eye and implications for pathogenesis in myotonic dystrophy. Hum. Mol. Genet. **8:** 481–492
156 Eriksson M., Ansved T., Edstrom L., Anvret M. and Carey N. (1999) Simultaneous analysis of expression of the three myotonic dystrophy locus genes in adult skeletal muscle samples: the CTG expansion correlates inversely with DMPK and 59 expression levels, but not DMAHP levels. Hum. Mol. Genet. **8:** 1053–1060
157 Koob D., Moseley M. L., Benzow K. A., Johnson C. M., Ranum L. P. W. and Day J. W. (1999) The SCA8 transcript is an antisense RNA to a brain-specific transcript encoding a novel actin-binding protein (KLHL1). Am. J. Hum. Genet. **65:** A30
158 Koide R., Kobayashi S., Shimohata T., Ikeuchi T., Maruyama M., Saito M. et al. (1999) A neurological disease caused by an expanded CAG trinucleotide repeat in the TATA-binding protein gene: a new polyglutamine disease? Hum. Mol. Genet. **8:** 2047–2053
159 The Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell **72:** 971–983
160 Koide R., Ikeuchi T., Onodera O., Tanaka H., Igarashi S., Endo K. et al. (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nat. Genet. **6:** 9–13
161 Nagafuchi S., Yanagisawa H., Sato K., Shirayama T., Ohsaki E., Bundo M. et al. (1994) Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. Nat. Genet. **6:** 14–18
162 Orr H. T., Chung M. Y., Banfi S., Kwiatkowski T. J. Jr, Servadio A., Beaudet A. L. et al. (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat. Genet. **4:** 221–226