Journal of Biochemistry and Molecular Biology, Vol. 31, No. 1, January 1998, pp. 2–19

JOURNAL OF
Biochemistry
and
Molecular Biology

DNA Structure, Triplet Repeats, and Hereditary Neurological Diseases

Robert D. Wells*

Institute of Biosciences and Technology, Texas A&M University, Texas Medical Center, 2121 W. Holcombe Blvd., Houston, Texas 77030-3303, USA

Received 15 December 1997

Genetic Instabilities of Repetitive DNA and Human Diseases

Repetitive DNA sequences are dispersed throughout natural genomes, both within and outside of known coding sequences (Charlesworth *et al.*, 1994; Tautz and Schlotterer, 1994). The varied distribution of particular repetitive sequences between different species suggests that they fulfill specific cellular requirements, but direct evidence for what these may be is lacking. However, any cellular functions are likely to be restricted to eukaryotes since long repeating sequences are absent from prokaryotic genomes.

Variations in the unit size and degree of repetition within DNA repeats has led to their division into a number of categories. In this review, we will concentrate on simple repeating DNA sequences, also termed microsatellites; these sequences are tandem (direct) repeats with a high degree of repetition and consists of 1–5 bp in their unit structure. More complex repeat units, known as minisatellites, are also found in eukaryotes and have been particularly well characterized in humans (Armour and Jeffreys, 1992).

Simple repeating sequences have an intrinsic genetic instability, manifested as frequent length changes due to insertions (also referred to as expansions) or deletions of repeat units. The rate of genetic change that occurs in these sequences is related to their copy number, and a mutated product therefore has a different potential for mutation compared to its predecessor; this phenomenon has been termed dynamic mutation (Richards and Sutherland, 1992; Sutherland and Richards, 1995).

The mechanisms for evolution of such sequences have been much debated (Dover, 1995; Hancock, 1996). The genetic instability associated with repeating sequences has been proposed to occur via a variety of pathways which are known to modify the genetic material. Several arguments suggest that changes in repeat number reflect DNA polymerase slippage (Levinson and Gutman, 1987b; Lustig and Petes, 1993; Hancock, 1996). The susceptibility of repetitive DNA sequences to slipped-strand mispairing provides a plausible mechanism for generating unusual DNA conformations (Sinden and Wells, 1992; Wells, 1996). This topic will be discussed in more detail below.

Slippage of the DNA strands at the replication fork is the favored mechanism to produce changes within simple repeats, but direct evidence that this occurs inside cells has not been obtained. Hence, it is possible that other mechanisms may be involved. Unequal exchange during recombination between homologous DNA sequences (Smith, 1973) is an attractive possibility, because it can produce deletions and insertions from one event. Gene conversion has been proposed to explain polymorphisms at minisatellites (Jeffreys *et al.*, 1994), although its relationship to the smaller repeat units of microsatellites is not known. It is clear that the evolution of repetitive sequences is complex, and it is likely that interactions occur between these different pathways.

Recently, genetic instabilities within repetitive DNA sequences have been linked to a variety of human diseases (Krontiris, 1995; Sutherland and Richards, 1995). These findings have produced a surge in interest across a range of scientific disciplines. Studies into the molecular mechanisms producing genetic instabilities of simple repeating sequences and their relationship to various hereditary disorders are the topic of this review.

Diseases Associated with Expansion of Triplet Repeat Sequences Unusual mutation events involving the expansion of specific triplet repeat sequences (TRS) have been linked to a number of human hereditary neuromuscular or neurodegenerative disorders (Caskey *et al.*, 1992; Willems 1994; Ashley and Warren, 1995; Paulson and Fischbeck, 1996; Warren, 1996). Other

* To whom correspondence should be addressed. Tel: 1-713-677-7651; Fax: 1-713-677-7689; E-mail: rwells@ibt.tamu.edu

candidate diseases are being studied, and it is likely that more will be related to this phenomenon. Most of these disorders show the unusual clinical behavior of anticipation, which is defined as the increased severity and/or decreased age of onset of a hereditary disease with progression through a pedigree. Analysis of TRS has identified the molecular basis for this clinical behavior: decreased age of onset or more severe symptoms of disease correlate with longer TRS, and anticipation occurs because genetic instabilities within TRS can produce longer repeats upon transmission to offspring. Recent reviews on myotonic dystrophy (Harris *et al.*, 1996), fragile X syndrome (Warren and Ashley, 1995) and Huntington's disease (HD) (Nasir *et al.*, 1996) have discussed the relationship of triplet repeat expansion to the clinical features of their respective diseases.

The events associated with expansion of TRS fall into two categories, termed types 1 and 2 (Paulson and Fischbeck, 1996; see Tables 1, 2). In type 1 expansions, the TRS is always CTG·CAG and is in a coding segment of the gene. Disorders in this category are essentially confined to the nervous system and comprise HD (Huntington's Disease Collaborative Research Group, 1993), dentatorubral–pallidoluysian atrophy (DRPLA; Koide *et al.*, 1994; Nagafuchi *et al.*, 1994), spinal and bulbar muscular atrophy (SBMA, also known as Kennedy's disease; La Spada *et al.*, 1991), and the spinocerebellar ataxias type 1 (SCA1; Orr *et al.*, 1993), type 2 (SCA2; Imbert *et al.*, 1996; Pulst *et al.*, 1996; *et al.*, Sanpei *et al.*, 1996), and type 3 (SCA3, also known as Machado–Joseph disease, MJD; Kawaguchi *et al.*, 1994). In these diseases, expansion events occur over a limited range; normal individuals have up to 30–40 repeat lengths which are stable, and afflicted individuals have 35–100 repeats which are genetically unstable (Tables 1, 2). Hence, there is a threshold length above which each sequence is unstable upon genetic transmission.

The net effect of type 1 expansions is to produce proteins with longer tracts of polyglutamines. The proteins associated with each disease are unrelated and probably have different functions. Thus, although the mutation mechanism producing these extended proteins may be similar, pathways producing the disease pathologies are likely to be different. The expansion events may create proteins with a gain of function that is particularly deleterious in neurons. Proteins have been identified that interact with some of the disease polypeptides via their glutamine tracts (Li *et al.*, 1995; Burke *et al.*, 1996). A stretch of glutamine residues may undergo a conformational change once expanded beyond a crucial threshold (Li *et al.*, 1995; Trottier *et al.*, 1995) and may thus alter the strength of any protein–protein interactions. Hence, it is possible that the selective expression of associated proteins provides the tissue specificity for the diseases.

The various type 1 diseases have quite different pathologies, and a more detailed examination of their

Table 1. Human type 1 diseases associated with expansion of trinucleotide repeats.

Type 1 diseases
Characteristics: CAG repeat in translated region of RNA
small expansions produce longer tracts of polyglutamines
specific neurons affected due to toxic gain of function of protein

Disorder	Gene locus	Repeat Length	
		Normal	Disease
Spinal and bulbar muscular atrophy	*AR*	11–34	38–66
Huntington's disease	*Hdh (IT15)*	10–35	36–121
Dentatorubral-pallidoluysian atrophy and Haw River syndrome	*DRPLA (B37)*	7–25	49–75
Spinocerebellar ataxias			
Type 1	*SCA1*	6–39	41–81
Type 2	*SCA2*	15–29	35–59
Type 3 (Machado–Joseph disease)	*SCA3 (MJD1)*	12–37	61–84

Table 2. Human type 2 diseases associated with expansion of trinucleotide repeats.

Type 2 disease
Characteristics: various repeats in untranslated region of RNA
very large expansions can occur
multi-system disorders caused by altered gene expression

Disorder	Gene locus	Repeat sequence	Repeat length				Repeat location
			Normal	Pre-mutation	Proto-mutation	Disease	
Myotonic dystrophy	*DMPK*	CAG	5–37	–	50–200	200–3000	3′-Untranslated
Fragile X syndrome	*FRAXA*	CGG	6–52	60–200	–	230–1000	5′-Untranslated
Friedreich's ataxia	*X25*	GAA	7–22	–	–	200–900	First intron

respective TRS expansions may provide clues to their disease pathways. For example, the identification of the gene associated with SCA2 showed that disease alleles had somewhat shorter lengths of triplet repeats compared to the other type 1 diseases. Thus it seems likely that expansions within the glutamine tract are tolerated to different levels for the various diseases. Comparison of the expansions associated with the various spinocerebellar ataxias supports this idea (Zoghbi, 1996) and shows that disease severity and age of onset are determined by factors other than the length of the polyglutamine tract.

Type 2 expansion events are associated with multisystem disorders and have a number of differences compared to type 1 (Table 2); various types of sequences are found, and the TRS is not located within a coding region of the gene. For example, a CGG·CCG repeat is found in the 5'-untranslated region of the gene in fragile X syndrome (Ashley *et al.*, 1993), a CTG·CAG repeat is found in the 3'-untranslated region of a gene linked with myotonic dystrophy (Fu *et al.*, 1992; Mahadevan *et al.*, 1992), and an AAG·CTT repeat occurs in the first intron of the gene associated with Friedreich's ataxia (Campuzano *et al.*, 1996). Extremely large expansions (hundreds of copies of the repeat) are required to produce type 2 diseases. The underlying mechanism of these diseases seems to be altered gene expression, and it is likely to occur at different stages for each disorder.

Although the mutation events can be characterized as those that undergo relatively short expansions (type 1) compared to those capable of very large expansions (type 2), type 2 disease alleles do exhibit short increases in length. An attractive hypothesis is that all triplet repeats can undergo short expansions during meiosis, and postmeiotic mechanisms produce the large expansions observed at type 2 loci (Ashley and Warren, 1995; Paulson and Fischbeck, 1996); possibly large expansions cannot occur for type 1 loci due to constraint produced by the translation of the CAG codon into a polypeptide.

The genetic instability of a specific TRS is dependent on its locus, and expansions of TRS are not due to a destabilization of the whole genome (Loeb, 1994; Sutherland and Richards, 1995). Therefore, it is apparent that the genetic instabilities associated with triplet repeat diseases do not reflect those observed at microsatellite sequences in general (see section on microsatellite instability and cancer). Clearly, there are many factors controlling the genetic stability of trinucleotide repeats, but a major determinant is the length of the repeat sequence. The majority of the population have alleles that are genetically stable, while disease families have longer alleles that are unstable. The existence of a critical threshold for disease formation is substantiated by studies which show that interruptions within the repeat sequence stabilize its genetic propagation (Chung *et al.*, 1993; Eichler *et al.*, 1994; Kunst and Warren, 1994). The correlation between increasing repeat size and disease severity is best illustrated by the type 1 disorders, although analysis of the spinocerebellar ataxias shows that each disease has specific characteristics (see above). Type 2 disorders have a wider range of observed repeat lengths, but the most severe diseases are still associated with the largest number of repeats. The existence of similar threshold sizes for the various triplet repeat diseases supports the notion that common mechanisms produce repeat tract instability in each disorder.

Examination of different tissues and cell lines has shown variations in the genetic instability of their TRS. All tissues examined for the HD repeat displayed repeat mosaicism, with the greatest instability in brain and sperm (Duyao *et al.*, 1993; Telenius *et al.*, 1994). Single cells from male myotonic dystrophy patients were found to have various sizes of TRS, with a directional bias towards increasing length in somatic tissues (Monckton *et al.*, 1995). Analysis of transgenic mouse lines with large repeats derived from the human myotonic dystrophy gene revealed a high degree of genetic instability in germline and somatic cells (Monckton *et al.*, 1997). Interestingly, germline variation of the myotonic dystrophy TRS included frequent deletions to the repeat (Monckton *et al.*, 1995, 1997). These observations of germline and somatic mosaicism within particular TRS suggest that cell-specific factors influence the genetic stability of these repeats.

Much effort has been devoted to the development of animal models of triplet repeat diseases in the hope that these will help us understand their molecular mechanisms. Inactivation of the gene associated with fragile X syndrome in humans (FMR1) caused mice to display abnormalities similar to those of human patients (Dutch–Belgian Fragile X Consortium, 1994) and thus may prove valuable in determining the physiological pathways involved in fragile X syndrome. Disruptions to the murine homolog of the HD gene showed that it was essential for normal embryonic development (Duyao *et al.*, 1995; Nasir *et al.*, 1995; Zeitlin *et al.*, 1995), and mice that transcribed a human HD cDNA with 44 CTG·CAG repeats did not develop HD symptoms (Goldberg *et al.*, 1996). Assuming that the mouse is an appropriate animal model in which to study HD, the lack of an HD phenotype in these studies suggests that the disorder is due to an altered interaction of the HD protein and that this may relate to a secondary function.

A number of studies have generated mice that are transgenic for proteins carrying expansions of polyglutamine tracts. The absence of phenotype in mice carrying the normal or expanded protein associated with SBMA may have been due to a nonphysiological pattern of expression (Bingham *et al.*, 1995). A neurological phenotype was observed in mice transgenic for SCA1 (Burright *et al.*, 1995) and SCA3 (MJD; Ikeda *et al.*, 1996). Recently, mice that were transgenic for the 5′ end

of the human HD gene have been shown to exhibit many of the features of HD (Mangiarini *et al.*, 1996). However, in transgenic mice studies in which repeat length was examined, the TRS were stable upon genetic transmission (Bingham *et al.*, 1995; Burright *et al.*, 1995; Goldberg *et al.*, 1996; Ikeda *et al.*, 1996). These studies provide further evidence that the relationship between repeat instability and disease formation is complex and is likely to involve the interaction of numerous cellular factors.

Microsatellite Instability and Cancer All organisms have biochemical systems that provide a high fidelity of genome replication and thus prevent the propogation of mutated phenotypes (Umar and Kunkel, 1996). Mechanisms are in operation to ensure high fidelity during DNA synthesis and also to correct mistakes that are made during replication. Deficiencies within these systems have the potential to produce many errors within genome sequences, and this situation has been termed the "mutator phenotype," since it would produce an organism prone to mutations. The progression of many human tumors through multiple stages (Vogelstein and Kinzler, 1993) suggests that their development requires a number of genetic alterations. The existence of a mutator phenotype would increase the potential for tumors to arise and therefore may be a primary event in tumor occurrence (Loeb, 1994).

Length changes are frequently observed within simple repeating sequences in all genomes. However, elevated frequencies of length changes occur under some conditions of reduced replication fidelity. A fundamental system involved in maintaining genomic integrity is that first identified in *E. coli* as methyl-directed mismatch repair (Modrich, 1991). A number of experiments have shown a similar mismatch repair system in eukaryotes, with a high degree of conservation throughout all organisms (Fishel and Kolodner, 1995; Kolodner, 1995; Modrich and Lahue, 1996; Umar and Kunkel, 1996). Upon inactivation of this system of DNA repair, increased heterogeneities have been observed at some unit lengths of simple repetitive DNA (e.g. mono- and dinucleotides) in bacterial systems (Levinson and Gutman, 1987a; Freund *et al.*, 1989) and in yeast (Strand *et al.*, 1993, 1995). Since there is an increased rate of mutation throughout the whole genome under these conditions. These observations suggest that the level of heterogeneity at simple repeating sequences is a good indicator of the overall level of DNA stability in cells.

Experimental evidence that deficient DNA repair could cause human tumors was obtained when cell lines from inherited and sporadic human cancers were shown to undergo an increased frequency of length changes in specific dinucleotide repeats (Aaltonen *et al.*, 1993; Ionov *et al.*, 1993; Thibodeau *et al.*, 1993). It was later shown that this increased genetic instability was due to defects in mismatch repair proteins (Fishel *et al.*, 1993; Leach *et al.*, 1993; Parsons *et al.*, 1993). Numerous studies have confirmed that defective mismatch repair is responsible for the elevated microsatellite instability and increased mutation rate in some cancers. For example, mice containing homozygous null mutations in some mismatch repair genes are viable, but develop tumors and exhibit elevated microsatellite instability (Baker *et al.*, 1995; de Wind *et al.*, 1995; Reitmair *et al.*, 1995).

The associations of defective mismatch repair and elevated microsatellite instability are particularly strong for hereditary nonpolyposis cancer, one of the most common inherited disorders known (de la Chapelle and Peltomaki, 1995; Marra and Boland, 1995). A number of recent reviews have discussed the relationship between microsatellite instability and cancer formation (Eshelmann and Markowitz, 1996; Kinzler and Vogelstein, 1996). Presently, it is not clear whether elevated microsatellite instability can be used as an absolute marker for defective mismatch repair. Although some cancers with elevated microsatellite instability do not carry a defect in the known mismatch repair genes, only a subset of genes involved in human mismatch repair have been identified (Kolodner, 1995; Eshelmann and Markowitz, 1996). Moreover, the effects of other DNA repair systems on microsatellite instability in humans is unknown, and it is conceivable that these will play a role in tumor development.

Genetic instabilities within mono- and dinucleotide repeats increase for longer runs of consecutive repeats and are therefore decreased by interruptions to the repeat sequence (Levinson and Gutman, 1987b; Umar and Kunkel, 1996). These observations are consistent with the hypothesis that slipped-strand mispairing during DNA synthesis generates misaligned intermediates (see above). These parameters are intrinsic to the DNA repeat, but it is also known that flanking sequences can influence the genetic stability of simple repeat sequences (Umar and Kunkel, 1996). As discussed below in relation to TRS, these observations suggest that many factors, including DNA repair, replication, and transcription, affect the genetic stability of microsatellite sequences.

Instabilities in a Genetically Defined Bacterial System

General comments on CTG·CAG As described above, progress has been made in our understanding of genetic instabilities from investigations with lower eukaryotes (yeast), cultured mammalian cells, transgenic mice models, and analyses of sperm. The optimum system would be a genetically controllable higher eukaryotic system. However, almost by definition, this is not available at the present time. Thus maximum information may be derived at present by investigations on simpler bacterial systems for evaluating the molecular processes that elicit expansions, whereas the more complex eukaryotic analyses

will provide useful information on the timing of events during development, the sex of parents determination of expansion, germline and somatic hypermutability, genome position effects, etc.

This laboratory has focused attention on investigations of molecular instabilities of CTG·CAG and CCG·CGG repeat sequences in a genetically and biochemically defined organism (*E. coli*; reviewed in Wells, 1996). Kang *et al.*, (1995a) established a defined genetic system which shows promise for the molecular dissection of this process. To study expansions, these workers determined whether a plasmid that contains $(CTG{\cdot}CAG)_{130}$ is completely homogeneous as a cloned molecule or whether deletions and expansions had occurred that gave rise to sequence heterogeneity, even in a tiny percentage of the molecules. The insert containing the triplet repeat was excised from the vector and separated by gel electrophoresis. The regions of the gel either above or below the insert band were eluted and "recloned"; recombinant plasmids were obtained that contained successively larger or smaller inserts, respectively. The family of inserts characterized by these methods contained repeat units ranging from 17 to 300. Hence, expansion and deletion occur in *E. coli*. This discovery lays the foundation for evaluating host cell genetic factors (e.g. replication, recombination, mismatch repair) that may elicit genetic instabilities.

The frequency of genetic expansions or deletions in *E. coli* depends on the direction of replication. Large expansions occur predominantly when the CTGs are in the leading template strand rather than the lagging strand. However, deletions are more prominent when the CTGs are in the opposite orientation. It should be noted that both deletions and expansions occur in both orientations. Most deletions generate products of defined size classes. Strand slippage coupled with nonclassical DNA structures probably accounts for these observations and relates to expansion/deletion mechanisms in eukaryotic chromosomes. DNA sequence analyses showed that expansion and contraction always occurred in multiple repeats of 3 bp. Prior investigations (Jaworski *et al.*, 1991) showed that deletions in dinucleotide repeat sequences occurred in multiple units of 2 bp.

A possible mechanism for the expansion and deletion behaviors has been proposed (Fig. 1; Kang *et al.*, 1995a). For expansion, a hairpin loop may form on the lagging strand nascent DNA (CTG strand). NMR investigations (Smith *et al.*, 1995) revealed that CTG oligomers form a stable antiparallel duplex with TT pairs, whereas the complementary CAG strand forms a metastable conformation. When CTG is the lagging strand template (orientation II), a loop may form on the lagging strand which will be bypassed during DNA synthesis to generate deletions. Multiple slippages (Wells and Sinden, 1993) may be promoted by an "idling polymerase" caused by a strong block such as a DNA structure or the presence of

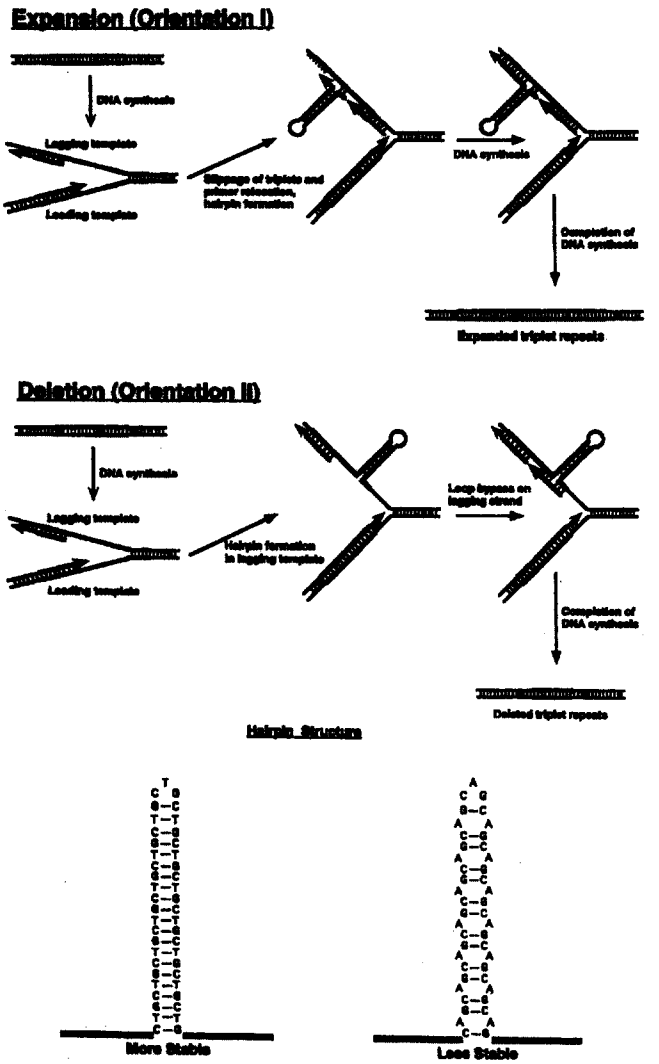

Fig. 1. A model for orientation-dependent instability of CTG repeats during replication. (a) Expansion (orientation I). (b) Deletion (orientation II). (c) Hairpin structure. *Left*, More stable; *right*, less stable. (Wells, 1996; with permission)

proteins, which causes continuous slippage (primer realignment), resulting in the expansion of larger sequences. Recent work in yeast with similar TRS has verified this model (Freudenreich *et al.*, 1997). Other workers (Jeffreys *et al.*, 1994) suggested that gene conversion events explain germline mutations at human minisatellites; however, these two mechanisms are not mutually exclusive.

Recent investigations (Bowater *et al.*, 1996) revealed the relationship between cell growth and deletions of CTG·CAG triplet repeats in plasmids. Long CTG·CAG repeats in plasmids can influence cell growth, which results in the observed expansions and deletions. During extended growth periods, the observed frequencies of deletion were dramatically increased if the cells passed

through stationary phase before subculturing. High frequencies of deletions were observed because of a growth advantage of cells containing plasmids with deleted triplet repeats. These observations (Bowater *et al.*, 1996) are the first to show a direct influence between a plasmid based DNA sequence or structure and factors controlling bacterial growth. Additional studies (Bowater *et al.*, 1997) showed that transcription promotes deletions of long CTG·CAG triplet repeats from human neuromuscular disease genes. However, a lower frequency of deletions is found also in the absence of transcription. These investigations were performed in recombinant plasmids that contained inducible promoters, and the frequency of deletions was monitored biochemically. These studies suggest a role for the involvement of transcription in DNA strand slippage which gives rise to deletions and expansions.

Furthermore, single-stranded DNA-binding protein enhances the stability of CTG·CAG triplet repeats in *E. coli* (Rosche *et al.*, 1996). Studies were conducted with mutants that lack SSB protein in order to evaluate the possible involvement of this protein, which is an important component in DNA replication, repair, and recombination. SSB can prevent the formation of DNA secondary structures. Replication can pause at sites of potential DNA secondary structure, and pause sites are associated with template misalignment mutagenesis. The potential for slippage and for DNA secondary structure formation within TRS may contribute to the instabilities of these sequences observed in individuals with triplet repeat diseases (described above). With a biochemical assay for stability, Rosche *et al.*, (1996) showed that the absence of single-stranded DNA-binding protein leads to an increase in the frequency of large deletions within the TRS.

Preferential expansion of CTG·CAG Ohshima *et al.*, (1996a) discovered that the CTG·CAG triplet repeat is the dominant genetic expansion product. This extraordinary discovery was made possible by the successful cloning and characterization of all ten TRS (Ohshima *et al.*, 1996b, c). The relative capacity of the ten TRS to be expanded in *E. coli* (Kang *et al.*, 1995a) was explored with a competition study. Surprisingly, the CTG·CAG triplet repeat was expanded at least nine times more frequently than any of the other nine triplets (Ohshima *et al.*, 1996a). Low levels of expansion were found also for GTG·CAC, GTC·GAC, CGG·CCG, and GAA·TTC. Thus the structure of the CTG·CAG repeat and/or its utilization by the DNA synthetic systems *in vivo* must be quite different from the other triplets. The surprising discovery that CTG·CAG triplet repeats are the dominant expansion products, as found in clinical samples from human hereditary diseases (reviewed in Wells, 1996), suggests the importance of DNA structural properties (Wells, 1988, 1996; Sinden, 1994;). Other investigations have revealed that duplex CTG·CAG and CGG·CCG repeats have unorthodox properties, including nucleosome assembly (Wang *et al.*, 1994, 1996), the capacity to cause DNA polymerases to pause within the repeat sequences (Kang *et al.*, 1995b; Ohshima and Wells, 1997), and conformational features as revealed by helical repeat and polyacrylamide gel migrations (Chastain *et al.*, 1995; Bacolla *et al.*, 1997). Further elucidation of the CTG·CAG repeat structural features along with the genetic factors responsible for expansion may explain why most triplet repeat hereditary disease genes contain CTG·CAG repeats (reviewed in Wells, 1996). Although other triplet repeats are found in the human genome (Gastier *et al.*, 1995), the lengths are shorter (generally fewer than 15 repeats) than found for these disease genes.

Both CTG·CAG and CGG·CCG have topological properties of writhe and flexibility that have not been described for other DNA (see section on flexible and writhed CTG·CAG and CGG·CCG). In addition, the former TRS is unusual in its high binding affinity for nucleosomes and is stabilized by mismatch repair-deficient cells. The latter two properties are not shared with CGG·CCG. Future work will be required to evaluate the role of these behaviors in the preferential expansion of CTG·CAG.

Site of expansion Kang *et al.*, (1996) described an investigation aimed at identifying the region of CTG·CAG triplet repeats that is preferentially expanded. Interestingly, the repeats are expanded distal to the replication origin (ColE1) as a single large event. Analysis of expanded regions using the interrupting CTA triplet sequence as a location marker within the CTG·CAG tract revealed that the expansion of large CTG·CAG repeats is one event rather than an accumulation of multiple small expansions and that the expansions occur more frequently in the region distal from the replication origin. In addition, we showed that a loss of interruptions increases the expansion frequency. Thus the instability of large triplet repeats in hereditary diseases occurs by a mechanism different from the instability in microsatellite sequences caused by defects in mismatch repair systems for certain sporadic cancers and hereditary nonpolyposis colorectal cancers.

Mismatch repair Mismatch repair deficient *E. coli* (Modrich and Lahue, 1996) were studied in order to further elucidate the factors involved in genetic instabilities as well as DNA structural issues *in vivo* (Jaworski *et al.*, 1995). Long CTG·CAG repeats are stabilized in ColE1-derived plasmids in *E. coli* containing mutations in the methyl-directed mismatch repair genes (*mutS*, *mutL*, or *mutH*). When plasmids containing $(CTG\cdot CAG)_{180}$ were grown for about 100 generations in *mutS*, *mutL*, or *mutH* strains, 60–85% of the plasmids contained a full-length repeat, whereas in the parent strain only about 20% of the

plasmids contained the full-length repeat. The deletions occur only in the $(CTG \cdot CAG)_{180}$ insert, and not in DNA flanking the repeat. While many products of the deletions are heterogeneous in length, preferential deletion products of about 140, 100, 60, and 20 repeats were observed. The *E. coli* mismatch repair proteins apparently recognize three-base loops formed during replication and then generate long single-stranded gaps where stable hairpin structures may form; these can be bypassed by DNA polymerase during the resynthesis of duplex DNA (Fig. 2). Direct experiments will be required to test the veracity of this model. Similar studies were conducted with plasmids containing CGG·CCG repeats; no stabilization of these triplets was found in the mismatch repair mutants. The reason for this was unclear, but may be due to the rate of formation and the stability of hairpins in CTG·CAG and CGG·CCG repeats (designated by question marks in Fig. 2). Since prokaryotic and human mismatch repair proteins are similar (Modrich and Lahue, 1996), and since several carcinoma cell lines which are defective in mismatch repair show instability of simple DNA microsatellites (Jaworski *et al.*, 1995; Modrich and Lahue, 1996), these mechanistic investigations in a bacterial cell may provide insights into the molecular basis for some human genetic diseases.

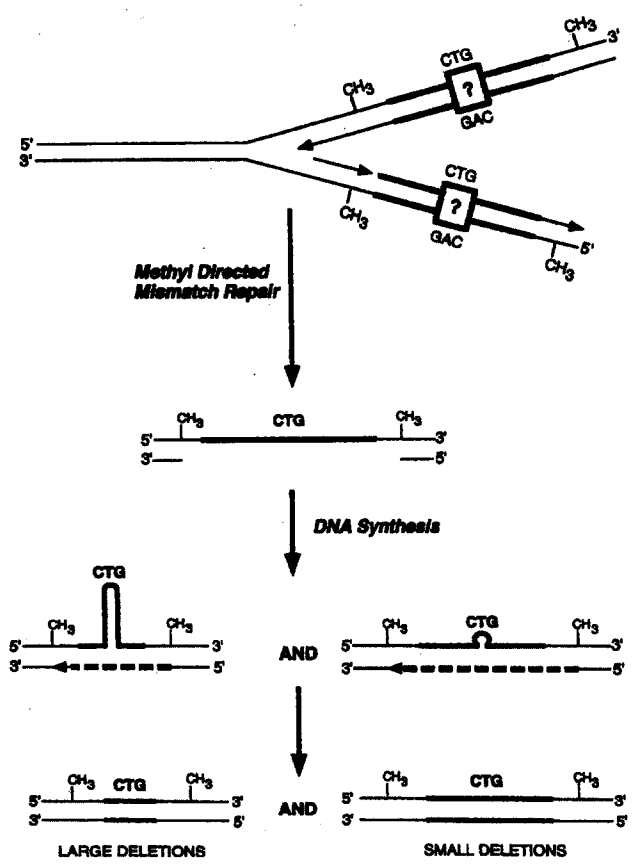

Fig. 2. Models for the involvement of *E. coli* methyl directed mismatch repair proteins in the enhancement of destabilization of $(CTG)_n$ triplet repeats *in vivo*. (Jaworski *et al.*, 1995; with permission).

Fragile X CGG·CCG A series of inserts containing 6–240 copies of CGG·CCG were stably cloned in plasmids. Several factors influence the stability (deletions and expansions) of the inserts: repeat length, the presence of interruptions, the orientation of the insert relative to the unidirectional replication origin, *E. coli* host strains, the location of the insert, and the copy number of the vector. The instability varies strongly with the length of the insert; longer tracts of CGG·CCG repeats show a greater degree of instability compared to shorter inserts. Furthermore, the effect of the length of DNA polymerase pausing was also observed during synthesis of the repeat *in vitro* when the Klenow fragment of DNA polymerase I was used; lengths of greater than 61 repeats showed stronger pausing sites, occurring at repeat number 30 (away from the CGG·CCG start site), when CCG was the template strand. This phenomenon was also observed with CTG·CAG triplet repeats (Kang *et al.*, 1995b). These results suggest that, at a critical length, the CGG sequence adopts a non-B conformation or conformations (see section on DNA polymerase pausing) which block DNA polymerase progression, leading to the idling and subsequent slippage to give expanded products and hence provide the molecular basis for this non-Mendelian genetic process.

The canonical human FMR-1 repeat carries 30 CGG·CCG triplets interrupted by two AGG triplets at the tenth and 20th repeat. Fragile X carriers carry longer repeats (50–200) containing long stretches of uninterrupted CGG·CCG triplets which predispose this sequence to hyperexpansion in successive generations. Affected individuals have longer methylated repeats (230–2000; Warren and Nelson 1994). Our results indicate that the presence of interruptions greatly enhances the stability of the CGG·CCG tract in *E. coli*. Other studies on the alleles derived from human patients show the presence of stable and unstable CGG·CCG triplets of similar size, suggesting that a feature other than length, but intrinsic to the repeat, was responsible for stability. This supported the observations made by Eichler *et al.*, (1995), who found that lengths of more than 33 uninterrupted CGG·CCG triplets showed marked instability, regardless of total repeat length, suggesting that loss of the AGG interruptions is an important mutational event in the generation of alleles predisposed to the fragile X syndrome.

As mentioned above, another important factor dictating stability is the orientation of the CGG·CCG-containing insert. Our results indicated that the triplet repeat was stably maintained in vectors if the CGG strand was in the leading template strand (orientation I) with respect to the origin of replication. However, if CCG fell in the leading template strand (orientation II), the insert was highly

destabilized (depending on the length), undergoing deletions and expansions. As in the case of CTG·CAG-repeating sequences, the frequency of expansion and deletion of the CGG·CCG triplet repeats is influenced by the direction of replication (Kang *et al.*, 1995a), which involves an asymmetric DNA polymerase complex that simultaneously replicates both the leading and the lagging strand (Wells and Sinden, 1993). Replication-dependent deletion between direct repeats occurs preferentially in the lagging strand due to the unequal probability of forming hairpins (Trinh and Sinden, 1991). Therefore, the deletion of the insert (in orientation II) can be explained by the propensity of the CGG template strand to form a stable hairpin (Fry and Loeb, 1994; Chen *et al.*, 1995; Gacy *et al.*, 1995; Mitas *et al.*, 1995; Mitchell *et al.*, 1995) which is bypassed by the replication machinery during resynthesis of the DNA. On the other hand, expansions within the tract are likely due to strand realignment through slippage of the complementary strands during pausing (described above) to generate a folded and elongated nascent DNA on the leading strand (Kang *et al.*, 1995a). Deletions were the most abundant species detected, but expansions were also visible when pRW3024, i.e. $(CGG \cdot CCG)_{24}$ in orientation II, was propagated in *E. coli* DH5a; the bands differed from each other by one repeating CGG·CCG unit, suggesting the involvement of slipped structures during replication. This method allowed the cloning of the expanded and deleted products (6 to 49 repeats) in orientation I and their propagation in *E. coli* SURE to give a stable DNA preparation.

DNA polymerase pausing The pausing of DNA synthesis *in vitro* at specific loci in double stranded CTG·CAG and CGG·CCG triplet repeats was found serendipitously (Kang *et al.*, 1995b). The DNA syntheses of CTG·CAG triplets ranging from 17 to 180 and CGG·CCG repeats ranging from 9 to 160 repeats in length were studied *in vitro*. Primer extensions using the Klenow fragment of DNA polymerase I, the modified T7 DNA polymerase (Sequenase), or the human DNA polymerase β paused strongly at specific loci in the CTG·CAG repeats. The pausings were abolished by heating at 70°C. As the length of the triplet repeats in duplex DNA, but not in single-stranded DNA, was increased, the magnitude of pausings increased. CGG·CCG triplet repeats also showed similar, but not identical patterns of pausings. These results indicate that appropriate lengths of the triplets adopt a non-B conformation or conformations that block DNA polymerase progression; the resultant idling polymerase may catalyze slippages to give expanded sequences and hence provide the molecular basis for this non-Mendelian genetic process. In addition, *in vivo* replication studies in *E. coli* (S.M. Mirkin, personal communication) with plasmids containing the CGG·CCG repeat revealed length-dependent pause sites; the nature of the DNA conformation which generates these pauses *in vivo* has not been characterized. Additional recent *in vitro* investigations (Ohshima and Wells, 1997) proved that the product of pausing was a hairpin structure caused by primer realignment, loop formation with reannealing of the nascent DNA to itself, and synthesis on the newly formed strand. In summary, the *in vitro* replication behavior of TRS is fascinating; our recent work with the Friedreich's ataxia GAA·TTC sequence shows the pausing behavior, presumably due to triplex formation (Ohshima *et al.*, 1996b).

Ohshima *et al.*, (1997) have isolated and analyzed the products of paused synthesis found at approximately 30–40 triplets from the beginning of the TRS. DNA sequence analyses revealed that the paused products contained short tracts of homogeneous TRS (6–12 repeats) in the middle of the sequence corresponding to the flanking region of the template–primer system. The sequence at the 3′ side terminated at the end of the primer, indicating that the primer molecule had served as template. In addition, chemical probe and polyacrylamide gel electrophoretic analyses revealed that the paused products existed in hairpin structures. We postulate (Fig. 3) that paused products for the DNA polymerases are caused by the existence of an unusual DNA conformation or conformations within the TRS (see section on flexible and writhed CTG·CAG and CGG·CCG) during *in vitro* DNA synthesis, enhancing the DNA slippages and the hairpin formation in the TRS due to primer realignment. The consequence of these steps is DNA synthesis to the end of the primer and termination. Primer realignment, including hairpin formation, may play an important intermediate role in the replication of TRS *in vivo* to elicit genetic expansions.

A logical reason for the impediment in long CTG·CAG and CGG·CCG sequences is flexible and writhed DNA (Bacolla *et al.*, 1997; Gellibolian *et al.*, 1997; see section on flexible and writhed CTG·CAG and CGG·CCG). However, we have not directly proven that this structural feature is responsible for the DNA polymerase pausing.

Molecular similarities between humans and *Escherichia coli* The studies described above on a genetically and biochemically tractable system for elucidating the molecular mechanisms responsible for expansion, and thus anticipation, represent a significant advance. Several remarkable molecular similarities exist, including the following (reviewed in Wells, 1997):

- TRS (CTG·CAG, CGG·CCG, or AAG·CTT) are genetically unstable (expansions and deletions).
- Longer repeats are more unstable than shorter sequences.
- CTG·CAG is preferentially expanded in *E. coli*; this repeat sequence was found in six of the nine triplet repeat diseases.

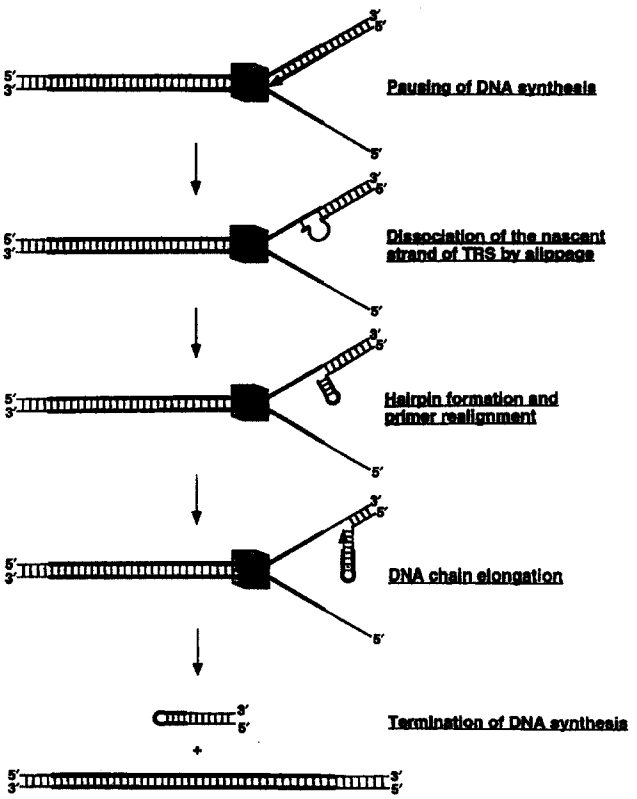

Fig. 3. Model for formation of hairpin structures at sites of paused DNA synthesis followed by strand slippage and primer realignment. (a) Pausing of DNA synthesis. (b) Dissociation of the nascent strand of triplet repeat sequences by slippage. (c) Hairpin formation and primer realignment. (d) DNA chain elongation. (e) Termination of DNA synthesis. (Ohshima and Wells, 1997; with permission)

- Repeat sequence imperfections (polymorphisms) stabilize long tracts of TRS.
- Similar types of imperfections (polymorphisms) are found (e.g. the polypurine·polypyrimidine motif is maintained in the Friedreich's ataxia AAG·CTT repeat sequence).
- The lengths of the smallest deletion products in *E. coli* (10–20 triplet repeats) approximate the lengths found in normal humans.
- DNA polymerases from humans and *E. coli* pause in long CTG·CAG, CGG·CCG, and AAG·CTT sequences, thus rendering them susceptible to mutations.

Hence, certain features of the molecular processes related to the involvement of TRS in human hereditary diseases may be elucidated effectively in simple cellular systems. Obviously, a number of other developmental and neurobiological questions can only be solved in higher eukaryotic cells. Thus some features of the concept of "unstable genes, unstable mind" may be tractable in genetically defined systems in mice, microbes, and molecules.

Summary of factors influencing genetic instability As elaborated above, several factors are known to influence the stability of long TRS in plasmids in *E. coli*. These factors include the following:

- *Genetic makeup of host cells.* The absence of recA is vital for the cloning of long TRS and, as discussed above, the presence of SSB protein is significant (Rosche *et al.*, 1996).
- *Growth conditions.* Several factors related to growth conditions including media and not permitting the cells to go through stationary phase, are also critical (Bowater *et al.*, 1996, 1997).
- *Generations of cells.* Since the deletion and expansion behavior appears to be due to DNA replication, the number of generations of cells (Kang *et al.*, 1995a) is important.
- *Transcription.* Active transcription through the TRS appears to enhance deletions (Bowater *et al.*, 1997).
- *MutS*, *MutH*, and *MutL*. The absence of these mismatch repair functions enhances stability, since it is likely that short loops which are formed by DNA slippage are recognized and cleaved with deletions as the products.
- *Nucleotide excision repair.* It is possible that nucleotide excision repair may also be involved in the observed instabilities (P. Parniewski and R. D. Wells, unpublished).
- *Vector and cloning location.* The type of vector and the location of cloning of a TRS in the vector is important and may relate to copy number. The instabilities may be due, in part, to the proximity to DNA polymerase I–III switch sites (Jaworski *et al.*, 1995).
- *Sequence of insert.* Certain types of sequences are more unstable than others. For example, the fragile X CGG·CCG sequence appears to be quite unstable, whereas several other sequences are somewhat more stable in the form of long tracts (Ohshima *et al.*, 1996b, c; Shimizu *et al.*, 1996).
- *Length of insert.* The length of the insert is critical; shorter lengths (30–50 TRS) are rather more stable, but longer sequences (>200 TRS) are quite unstable.
- *Interruptions in insert.* The presence of interruptions, or polymorphisms as recognized in human genetics, has a stabilizing effect in long TRS. This behavior has been observed with virtually all of the ten TRS studied to date (Shimizu *et al.*, 1996; Ohshima *et al.*, 1996b, c).
- *Orientation of insert.* The orientation of the insert is important for favoring deletions or expansions, as described above.

Flexible and Writhed CTG·CAG and CGG·CCG

Whereas the results described above implicate DNA structural features in the process of destabilization of the TRS, analyses conducted with chemical, enzymatic, and structural probes that detect perturbed or single-stranded regions failed to demonstrate that CGG·CCG and CTG·CAG adopt any of the non-B DNA structures identified so far, which include left-handed Z-DNA, inter and intramolecular triplexes, cruciforms, tetraplexes, and nodule DNA (Wells, 1988, 1996; Bacolla *et al.*, 1997). On the other hand, linear fragments containing sufficiently long CTG·CAG (and to a lesser extent CGG·CCG) migrated faster than expected on polyacrylamide gels (Chastain *et al.*, 1995; Bacolla *et al.*, 1997), indicating that these sequences possess peculiar structural properties. Such properties were elucidated from experiments on circularization kinetics (CK) and apparent helical repeat determination.

CK measures the rate of ring closure of linear DNA fragments that have complementary, single-stranded ends and enables the flexibility of DNA to be determined, whereas the determination of the apparent helical repeat gives a qualitative indication of writhe, i.e. the ability of the DNA helix to deflect from planarity in three-dimensional space. Before describing these results, we will outline the relevance of DNA flexibility in general biological processes.

Flexibility of DNA and general biological processes The flexibility of DNA is essential for the organization of genetic information. Linear duplex DNA is assembled into high-order chromatin structures in both prokaryotic and eukaryotic cells, which enables selective (e.g. tissue- and development-specific) gene expression. In addition, a high level of condensation is achieved in the eukaryotic nuclei during the segregation of metaphase chromosomes. Here, a 2.5-μm-long chromatid contains a single copy of a DNA molecule about 30 mm in length when fully extended (DuPraw, 1974). The required compaction factor — from 30 mm to 2.5 μm — of 12000-fold, obviously involves dramatic bending of the DNA molecule.

Binding between protein complexes distally located along the DNA provides a means for transcriptional control. A clear example is the regulation of the arabinose operon in *E. coli*. An interaction is necessary between two araC proteins, one bound to *araI*, located just upstream from the initiation of transcription, and the other to $araO_2$, situated 211 bp further upstream. In the absence of arabinose, araC bound to *araI* and $araO_2$, which stably dimerize as a result of DNA looping; this complex leads to repression of transcription. The presence of arabinose weakens the protein contacts, disfavors DNA looping, and results in high levels of transcription (Schleif, 1992). Similar looped structures are also thought to constitute a mechanism by which enhancer-mediated stimulations of transcription operate (Schleif, 1992).

Initiation of DNA replication necessitates unwinding of the double helix. Studies in viruses, bacterial cells and yeast cells all reveal that initiation of replication is contingent upon the presence of a DNA unwinding element (DUE; DePamphilis, 1993). The function of DUE is to provide a sequence in which the separation (melting) of the two strands is thermodynamically favorable. A region of nonhelical, single-stranded DNA is required for the onset and initial progression of a replication fork (DePamphilis, 1993). This melting operation involves unwinding (untwisting) of the helix, which is accomplished by one or two cooperative activities: DNA helicase and negative supercoiling. DNA helicases physically separate the two DNA strands, whereas negative supercoiling (which is essential in certain systems; Alfano and McMacken, 1988) lowers the activation energy of the enzymatic process.

Thus supercoiling is an important feature of DNA. Regardless of whether the linear or circular chromosomal DNA is involved, rotation of the strands about a free end (such as a nick or the chromosomal ends) is much restricted *in vivo*. It follows that unwinding activities, such as the one on DUE, force adjacent segments of the DNA to overwind, which causes the duplex DNA to supercoil. Due to a continuous requirement for unwinding, processes such as replication and transcription are associated with the constant accumulation of supercoils in front of the respective enzymatic complexes, which may prevent further enzyme translocation. These supercoils are normally removed by topoisomerases (Wang, 1996). Studies on topoisomerases have led to the appreciation that supercoiling plays both a favorable and an unfavorable role in general cellular functions, including replication, transcription, recombination, excision repair, chromatin organization, and genome stability (Wang, 1996).

Determination of DNA flexibility The determination of flexibility is a problem of physical chemistry and entails evaluating the forces that oppose bending and twisting of the double helix. Four techniques have provided estimates of such forces (moduli) for B-DNA of random composition: (1) transient electric birefringence and dichroism (TEB/TED), (2) electron microscopy (EM), (3) CK, and (4) Monte Carlo simulation (MC), a computational approach (reviewed by Diekmann *et al.*, 1982; Hagerman, 1988; Hagerman and Ramadevi, 1990; Taylor and Hagerman, 1990; Kahn *et al.*, 1994; Bednar *et al.*, 1995; Hodges-Garcia and Hagerman, 1995). These studies monitor the behavior of short (few hundred bp long) linear DNA fragments under different conditions and quantitate the moduli by applying theoretical treatments to experimentally measured parameters. Of these, CK has

become increasingly popular due to its simple execution (Shore *et al.*, 1981; Shore and Baldwin, 1983a) and to the development of comprehensive analytical solutions to the underlying statistical mechanical theory (Shimada and Yamakawa, 1984, 1985). CK measures the fraction of linear molecules that are converted into circles as a function of time by the action of T4 DNA ligase on complementary single-stranded ends. We have employed this method to measure the bending and twisting forces of DNA composed primarily of $(CTG \cdot CAG)_n$ and $(CGG \cdot CCG)_n$ repeats (Bacolla *et al.*, 1997).

Bending and twisting forces of $(CTG \cdot CAG)_n$ and $(CGG \cdot CCG)_n$

Theory of ring closure DNA is in constant motion in solution due to thermal energy. At any time, the end of a linear molecule may either encounter the end of a different molecule (bimolecular encounter) or its other end (cyclization encounter). *Bimolecular encounter* depends on temperature (higher temperature favors more encounters) and the total concentration of ends (more ends, more encounters), whereas *cyclization encounter* is dependent on temperature, but is independent of end concentration. In kinetic terms, the former process is second order, whereas the latter is first order (Jacobson and Stockmayer, 1950; Wang and Davidson, 1966). Additionally, cyclization encounter is affected by the length of the molecule, its stiffness, and the fractional helical turn. These factors are considered separately.

Length The two ends of a 10-bp-long DNA molecule will not circularize; it is like trying to bend a 1-cm-long nail into a circle. In a 1000-bp-long molecule, the chances are that the two ends will meet sooner or later; analogously, we can bend into a circle a 10-m-long rod made of the same material as our previous 1-cm-long nail. Therefore, very short molecules do not cyclize efficiently, whereas longer molecules do. However, the consideration that "the longer the molecule, the more frequent the cyclization" is only valid within a fixed range of length. In fact, if we keep one end of the molecule fixed at one position and follow the trajectory of the other end, we would see it traveling through a space volume whose outer border is determined by the extended length of the molecule. If we consider the probability that the traveling end will come in contact with the fixed end, two effects counteract each other. The longer the molecule, the more movement is allowed, which increases cyclization. However, the longer the molecule, the greater the volume "explored" by the free end, which reduces cyclization. This problem, which concerns the behavior of any elastic rod, has been studied both theoretically and computationally (Shimada and Yamakawa, 1984). Curve A in Fig. 4 shows how the probability of encounter of two ends of DNA

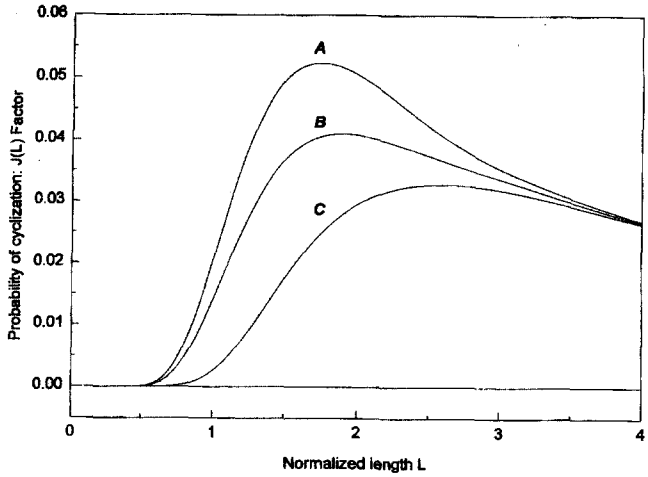

Fig. 4. Theoretical probability of circularization for a twisted wormlike chain. The curves represent the theoretical length-dependent, stiffness-dependent, and twist-dependent probability of ring closure for linear DNA molecules as calculated according to Eqs. (1), (37)–(39), (60), (69), (70), (73), (C1)–(C3) in Shimada and Yamakawa (1984). L indicates the reduced contour length of the chains, i.e. the ratio between the length of the Kuhn segment λ^{-1} and the contour length of the DNA (number of bp × 3.4 Å). The J(L) factor is the Jacobson and Stockmayer (1950) J factor, which is related to the ring closure probability by Eq. (1) in Shimada and Yamakawa (1984). The values of the Poisson's ratio s and the fractional twist r are: A, $\sigma = -0.2$, $r = 0$; B, $\sigma = 0.8$, $r = 0$; C, $\sigma = -0.2$, $r = 0.5$. A and C correspond to the theoretical curves for random B-DNA (Shimada and Yamakawa, 1984; Bacolla *et al.*, 1997) and indicate the upper and lower boundaries of the twist-dependent J factor. With $\lambda^{-1} = 950$ Å (persistence length of 475 Å) for random B-DNA, $L = 1$ corresponds to 279 bp (950 Å/3.4 Å) for curves A and C.

varies with the length of the DNA. At short lengths, the probability increases with length, but beyond a certain limit it decreases. This decrease is due to the "takeover" effect by the sphere volume.

Stiffness Stiffness increases the average end-to-end distance of an elastic rod and therefore decreases the probability of cyclization. Curve B in Fig. 4 shows how this probability varies for molecules that are twice as stiff as A. Whereas the short members do indeed circularize less efficiently than A, the sphere volume effect becomes dominant at longer lengths, to the point of abolishing the differences between A and B. This phenomenon is important and shows that, in order to study the flexibility of DNA, the length of the fragments used for analyses should not extend far beyond the peak of circularization probability.

Fractional Helical Turn Consider two DNA molecules, one with an integral number of helical turns and

the other 5 bp longer. During cyclization encounters, the free end of the first molecule is properly oriented for circle formation with its fixed end, whereas the free end of the second molecule is now turned 180° away from the fixed end. Thus, in contrast to the first molecule, circularization of the second molecule requires that the free end both collides with and rotates relative to the fixed end. Curve C in Fig. 4 shows the probability of ring closure for molecules as stiff as A, but whose free ends need to rotate by 180°. The differences between A and C are remarkable at short lengths, the values being much lower for C. Thus both the bending and the twisting forces may be evaluated by CK. Once again, the differences vanish at long lengths, further stressing the remarks pointed out previously. Since with real DNA molecules the fractional helical turn oscillates from 0 to 0.5 with length, the resulting probability follows an oscillatory pathway between an upper (A) and a lower (C) boundary.

Cyclization kinetics on DNA fragments containing $(CTG{\cdot}CAG)_n$ and $(CGG{\cdot}CCG)_n$ Linear DNA fragments with uninterrupted $(CTG{\cdot}CAG)_n$ or $(CGG{\cdot}CCG)_n$ triplet repeats were constucted for CK experiments (Bacolla *et al.*, 1997). Bimolecular and cyclization events were scored by sealing the hybridized ends with T4 DNA ligase in reaction mixtures where the radioactive fragments were reacted with the enzyme for varying time. These samples were electrophoresed through polyacrylamide gels to separate the linear monomers, the linear dimers, and the circles. The concentration of each species was calculated, and the molar J factor was computed for each fragment. The molar J factor corresponds to the ratio of the first-order kinetics of cyclization to the second-order kinetics of bimolecular reaction (Bacolla *et al.*, 1997) and is the primary quantity needed to derive the elastic forces. The molar J factors were then plotted as a function of length, and the formulas used to construct the curves in Fig. 4 were applied so as to find the best fit for the experimental data. The results are reported in Table 3 along with the values determined previously for DNA of random sequence (Shore *et al.*, 1981; Shore and Baldwin, 1983a; Shimada and Yamakawa, 1984; Bacolla *et al.*, 1997). The most significant result is that the bending moduli for the TRS were much lower (approximately 40%) than for random DNA. We interpreted this finding to mean that DNA composed of $(CTG{\cdot}CAG)_n$ or $(CGG{\cdot}CCG)_n$ bends more easily than the bulk of chromosomal DNA. On the other hand, both the torsional forces and the helical repeats (i.e. the number of base pairs per turn of double helix) of these TRS were identical to those of DNA of random sequence.

In addition to bending, we pointed out the importance of supercoiling in general biological processes. The amount of stress that is required to supercoil a DNA molecule depends on its flexibility. Therefore, one would expect that the TRS supercoils more easily than bulk chromosomal DNA. Experimental support of this contention was also found. First, the electrophoretic migration of supercoiled plasmids was greatly influenced by the length of an inserted fragment when it contained a TRS sequence. Since the effect was not observed when fragments of random DNA sequence were cloned instead of the TRS, we concluded that the altered migration reflected changes in the topological shape of the plasmids caused by the TRS. Specifically, under the stress of supercoiling, the TRS were more contorted (with a greater writhe) than the rest of the plasmid. Second, circles of fragments containing $(CTG{\cdot}CAG)_{64}$, $(CGG{\cdot}CCG)_{70}$, and $(CGG{\cdot}CCG)_{71}$ (224–245 bp in length) migrated on polyacrylamide gels as two or more species, indicating that these molecules contained a different number of helical turns. Since topologically isomeric circles also form with random DNA, but only at longer lengths, the formation of multiple circular species at such short lengths must have been due to the greater flexibility of the TRS as compared to random DNA.

The bending and torsional moduli also enable the application of analytical equations that predict the energy required to supercoil DNA. These theoretical calculations yield a parameter (K) that expresses the amount of free energy stored at each base pair at every superhelical turn (Shimada and Yamakawa, 1985). K was computed for DNA of $(CTG{\cdot}CAG)_n$ and $(CGG{\cdot}CCG)_n$ composition up to 10,000 bp (Gellibolian *et al.*, 1997). The results showed that the values for the TRS were lower than those for random B-DNA, confirming that less energy is required to supercoil a TRS than a random B-DNA. However, the differences in these free energies of supercoiling between TRS and random B-DNA varied with length. They first

Table 3. Elastic constants and helical repeat for $(CTG{\cdot}CAG)_n$ and $(CGG{\cdot}CCG)_n$

Parameter	Bending modulus $\alpha(\times 10^{-19}$ erg·cm)	Torsional modulus $\beta(\times 10^{-19}$ erg·cm)	Helical repeat h° (bp/turn)
Random DNA	1.92	2.4	10.46
$(CTG{\cdot}CAG)_n$	1.13	2.3	10.41
$(CGG{\cdot}CCG)_n$	1.27	2.4	10.35

The bending modulus α, the twisting modulus β, and the helical repeat h^0 were derived from the values of Poisson ratio σ, the length of the Kuhn segment λ^{-1}, and constant torsion τ_0 used to fit the molar J factors. The values were: $(CTG{\cdot}CAG)_n$, $\sigma = -0.51$, $\lambda^{-1} = 556$ Å (which corresponds to a persistence length of 278 Å), $\tau_0 = 0.1775$; $(CGG{\cdot}CCG)_n$, $\sigma = -0.46$, $\lambda^{-1} = 630$ Å (which corresponds to a persistence length of 315 Å), $\tau_0 = 0.1786$. The conversion factors were: $\alpha = +(k_B T\lambda^{-1})$; $\beta = \alpha/(1+\sigma)$; $h^0 = 2\pi/\tau_0 l_{bp}$; where k_B is the Boltzmann constant, T the absolute temperature, and l_{bp} the distance between adjacent base pairs, 3.4 Å.

increased, reached a maximum at approximately 520 bp, and then gradually decreased to steady levels. This unexpected result reveals a close correspondence between the ability of the TRS to supercoil and the length of 180–200 triplet repeats (540–600 bp) that separates the premutation range from the full mutation range in the fragile X and myotonic dystrophy pathologies. Therefore, these analyses reveal that supertwisting the highly flexible and writhed TRS promotes biochemical events that contribute deleteriously to the stability of these sequences.

DNA flexibility as a source of genetic instability Both the accumulation of supercoiling and the induced sequence instability may occur through multiple mechanisms. Figure 5 shows the preferential partitioning of writhe (supercoiling) within a TRS in chromosomal DNA. A linear DNA segment composed of 50% TRS (left) and 50% random B-DNA (right) is flanked by two protein complexes (rectangles). If the filled rectangle rotates while the open rectangle remains static, the TRS and B-DNA overwind or unwind, depending on the sense of rotation of B, and therefore writhe. However, since the TRS is more flexible than B-DNA, more writhe accumulates within the TRS than within B-DNA. Furthermore, the difference in partitioning of supercoiling depends on the length of the TRS and B-DNA, reaching a maximum when both TRS and B-DNA are 520 bp long.

Translocation of both RNA and DNA polymerase complexes induces overwinding and unwinding, and topoisomerases relieve or induce supercoiling (Wang, 1996). Therefore, these activities may lead to the preferential accumulation of supercoiling within the TRS in the chromosome. Alternative base pair associations, such as slippage-associated hairpins, are favored for unwound DNA, and these may lead to genetic instability (Wells and Sinden 1993). The transcription-induced instability of the TRS is consistent with this mechanism (Bowater *et al.*, 1997). *In vitro*, DNA polymerases stall during replication through long TRS, suggesting that these sequences form tertiary structures incompatible with polymerase progression (Kang *et al.*, 1995b; Ohshima and Wells, 1997). It is possible that the unremoved supercoils interfere with the progression of translocating protein complexes and favor the formation of alternative DNA interactions that lead to sequence instability. The large expansions observed in type 2 disease alleles and the size of approximately 180–200 repeats that demarkates the boundary between the premutation and the full-mutation range (Ashley and Warren, 1995; Paulson and Fischbeck, 1996) correlate with the increased ability of the TRS to supercoil at this length. Whether this correlation is significant with respect to the disease pathologies remains to be elucidated. Lastly, high levels of supercoiling bring multiple segments of DNA in contact with one another. Due to the repetitive nature of the TRS, these close proximities may induce interstrand interactions that lead to destabilization of the duplex DNA and, ultimately, to genetic instability.

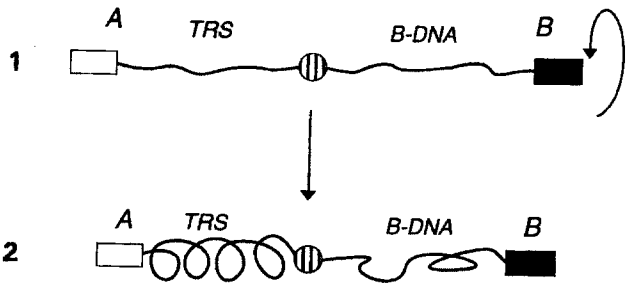

Fig. 5. The preferential partitioning of writhe within a region of DNA composed of random-sequence DNA adjoining a triplet repeat sequence. The two rectangles *A* and *B* represent protein complexes bound to the DNA, where *A* is static (its rotation is hindered), whereas *B* can rotate clockwise or counterclockwise (*top*). *TRS* [$(CTG \cdot CAG)_n$ or $(CGG \cdot CCG)_n$] and *B-DNA* indicate two segments of duplex DNA of equal length, and the *circle* represents a junction intersection that prevents supercoiling from diffusing between the two DNA segments. Rotation of *B* clockwise or counterclockwise overwinds or unwinds, respectively, the two DNA segments (*bottom*), whose helices adopt supercoils (writhe). However, since the TRS is more flexible than B-DNA, writhe will accumulate to a greater extent within the TRS region than within the B-DNA segment (Gellibolian *et al.*, 1997). The quantitative calculations of the free energies of supercoiling were performed under the assumption that the DNA possesses only one bending modulus and one torsional modulus. This is obviously not the case when a TRS is embedded in random sequence DNA. The statistical mechanical calculations of such a mixed DNA are greatly complicated by the presence of two very different bending moduli, and we have not attempted such calculations at the present time. As a consequence, the two domains (B-DNA and TRS) are purposely separated. In addition, the conformation shown at the bottom is a schematic representation of the writhe which is not intended to differentiate between plectonemic and solenoidal or between right-handed and left-handed supercoiling.

In summary, it is likely that the flexible and highly writhed duplex conformations of $(CTG \cdot CAG)_n$ and $(CGG \cdot CCG)_n$ play an important role in the strand slippage, hairpin-loop formation, DNA polymerase pausing, daughter strand hairpin formation, and possibly other events that are seminal to the expansion/deletion processes that give rise to anticipation.

Prospects for the Future

Enormous progress has been made in our understanding of the genetic basis of several human neurological diseases. In addition, we have been able to glean a great deal about the molecular mechanisms of the processes which are responsible for genetic instabilities from simpler systems. Future work will focus on simultaneous advances in both

complex human genetic systems and bacteria. Furthermore, we may expect to see substantial progress with investigations on yeast and other lower eukaryotes as well as animal model systems such as transgenic mice. Unfortunately, the phenotypes of transgenic mice do not always correlate with those of human patients. Thus it is possible that the nature of mammalian development may be too complex to allow complete identification and molecular pathways, at least in the near future.

The fundamental nature of the processes involved (transcription, replication, repair) suggests that these instabilities are likely to be conserved through evolution, although the details may vary in different systems. The observation that more than one pathway produces instabilities in long TRS in *E. coli* (Bowater *et al.*, 1997) may relate to different mechanisms for type 1 and type 2 disease. In all situations in which the expansion of a TRS is associated with a human disease, the TRS is located within a gene (Ashley and Warren, 1995; Paulson and Fischbeck, 1996). Thus there is potential for transcription-associated events to influence the stability of TRS in all disorders. The regulation of transcription is specific for each promoter, and this may provide one mechanism by which local phenomena alter the stability of a particular TRS in a manner that is dependent on its position.

Despite the optimism expressed above, numerous fundamental questions remain which will require innovative strategies for future experimentation. The degree of suffering encountered in patients with these diseases serves as an impetus for our best efforts to develop rational therapeutic strategies that will benefit mankind.

Acknowledgments This work was supported by grants from the National Institutes of Health (GM52982) and the Robert A. Welch Foundation. The authors wish to thank our colleagues for providing preprints of manuscripts for this review, and we express our appreciation to our colleagues who performed the original experiments from this as well as our collaborator's laboratories. These individuals include the following: Drs. S. Amirhaeri, P. Chastain, E. Eichler, R. Gellibolian, D. Giedroc, J. Griffith, S. C. Harvey, A. Jaworski, S. Kang, S. Kramer, J. Larson, S. Levene, D. Nelson, K. Ohshima, B. A. Oostra, P. Parniewski, W. Rosche, M. Shimizu, R. Sinden, B. D. Stollar, and Y.-H. Wang.

References

Aaltonen, L. A., Peltomaki, P., Leach, F. S., Sistonen, P., Pylkkanen, L., Mecklin, J. P., Jarvinen, H., Powell, S. M., Jen, J., Hamilton, S. R., Petersen, G. M., Kinzler, K. W., Vogelstein, B., and de la Chapelle, A. (1993) Clues to the pathogenesis of familial colorectal cancer. *Science* **260**, 812–816.

Alfano, C. and McMacken, R. (1988) The role of template superhelicity in the initiation of bacteriophage l DNA replication. *Nucleic Acids Res.* **16**, 9611–9630.

Armour, J. A. and Jeffreys, A. J. (1992) Biology and applications of human minisatellite loci. *Curr. Opin. Genet. Dev.* **2**, 850–856.

Ashley, C. T. and Warrens S. T. (1995) Trinucleotide repeat expansion and human disease. *Annu. Rev. Genet.* **29**, 703–728.

Ashley, C. T., Sutcliffe, J. S., Kunst, C. B., Leiner, H. A, Eichler, E. E., Nelson, D. L., and Warren, S. T. (1993) Human and murine FMR-1: alternative splicing and translational initiation downstream of the CGG-repeat. *Nat. Genet.* **4**. 244–251.

Bacolla, A., Gellibolian, R., Shimizu, M., Amirhaeri, S., Kang, S., Ohshima, K., Larson, J. E., Harvey, S. C., Stollar, B. D., and Wells, R. D. (1997) Flexible DNA: genetically unstable CGG and CTG from human hereditary neuromuscular disease genes. *J. Biol. Chem.* **272**, 16783–16792.

Baker, S. M., Bronner, C. E., Zhang, L., Plug, A. W., Robatzek, M., Warren, G., Elliott, E. A., Yu, J., Ashley, T., Arnheim, N., Flavell, R. A., and Liskay, R. M. (1995) Male mice defective in the DNA mismatch repair gene PMS2 exhibit abnormal chromosome synapsis in meiosis. *Cell* **82**, 309–319.

Bednar, J., Furrer, P., Katrich, V., Stasiak, A. Z., Dubochet, J., and Stasiak, A. (1995) Determination of DNA persistence length by cryo-electron microscopy. Separation of the static and dynamic contributions to the apparent persistence length of DNA. *J. Mol. Biol.* **254**, 579–594.

Bingham, P. M., Scott, M. O., Wang, S., McPhaul, M. J., Wilson, E. M., Garbern J. Y., Merry, D. E., and Fischbeck, K. H. (1995) Stability of an expanded trinucleotide repeat in the androgen receptor gene in transgenic mice. *Nat. Genet.* **9**, 191–196.

Bowater, R. P., Rosche, W. A., Jaworski, A., Sinden, R. R., and Wells, R. D. (1996) Relationship between *Escherichia coli* growth and deletions of CTG·CAG triplet repeats in plasmids. *J. Mol. Biol.* **264**, 82–96.

Bowater, R. P., Jaworski, A., Larson, J. E., Parniewski, P., and Wells, R. D. (1997) Transcription increases the deletion frequency of CTG·CAG triplet repeat sequences from human neuromuscular disease genes in *E. Coli*. *Nucleic Acids Res* **25**, 2861–2868.

Burke, J. R., Enghild, J.J., Martin, M. E., Jou, Y. S., Myers, R. M., Roses, A. D., Vance, J. M., and Strittmatter, W. J. (1996) Huntingtin and DRPLA proteins selectively interact with the enzyme GAPDH. *Nat. Med.* **2**, 347–350.

Burright, E. N., Clark, H. B., Servadio, A., Matilla, T., Feddersen, R. M., Yunis, W. S., Duvick, L. A., Zoghbi, H. Y., and Orr, H. T. (1995) SCA1 transgenic mice: a model for neurodegeneration caused by an expanded CAG trinucleotide repeat. *Cell.* **82**, 937–948.

Campuzano, V., Montermini, L., Molto, M. D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., Zara, F., Canizares, J., Koutnikova, H., Bidichandani, S. I., Gellera, C., Brice, A., Trouillas, P., de Michele, G., Filla, A., de Frutos, R., Palau, F., Patel, PI., di Donato, S., Mandel, J-L., Cocozza, S., Koenig, M., and Pandolfo, M. (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science.* **271**, 1423–1427.

Caskey, C. T., Pizzuti, A., Fu, Y. H., Fenwick, R. G. Jr, and Nelson, D. L. (1992) Triplet repeat mutations in human disease. *Science.* **256**, 784–789.

Charlesworth, B., Sniegowski, P., and Stephan, W. (1994) The evolutionary dynamics of repetitive DNA in eucaryotes. *Nature.* **371**, 215–220.

Chastain, P. D., Eichler, E. E., Kang, S., Nelson, D. L., Levene, S. D., and Sinden, R. R. (1995) Anomalously rapid electrophoretic mobility of DNA containing triplet repeats associated with human disease genes. *Biochemistry* **34**, 16125–16131.

Chen, X., Mariappan, S. V. S., Catasti, P., Ratliff, R., Moyzis, K., Laayoun, A., Smith, S. S., Bradbury, E. M., and Gupta, G. (1995) Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. *Proc. Natl. Acad. Sci. USA* **92**, 5199–5203.

Chung, M-Y., Ranum, L. P. W., Duvick, L. A. , Servadio, A., Zoghbi, H. Y., and Orr, H. T. (1993) Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nat. Genet.* **5**, 254–258.

de la Chapelle, A. and Peltomaki, P. (1995) Genetics of hereditary colon cancer. *Annu. Rev. Genet.* **29**, 329–348.

DePamphilis, M. L. (1993) Eukaryotic DNA replication: anatomy of an origin. *Annu. Rev. Biochem.* **62**, 29–63.

de Wind, N., Dekker, M., Berns, A., Radman, M., and de Riele, H. (1995) Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. *Cell.* **82**, 321–330.

Diekmann, S., Hillen, W., Morgeneyer, B., Wells, R. D., and Pörscke, D. (1982) Orientation relaxation of DNA restriction fragments and the internal mobility of the double helix. *Biophys. Chem.* **15**, 263–270.

Dover, G. (1995) Slippery DNA runs on and on and on ... *Nat. Genet.* **10**, 254.

DuPraw, E. J. (1974) Quantitative constraints in the arrangement of human DNA. *Symp. Quant. Biol.* **38**, 87–98. Cold Spring Harbor.

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

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

Duyao, M. P., Auerbach, A. B., Ryan, A., Persichetti, F., Barnes, G. T., McNeil, S. M., Ge, P., Vonsattel, J.-P., Gusella, J. F., Joyner, A. L., and MacDonald, M. E. (1995) Inactivation of the mouse Huntington's disease gene homologue Hdh. *Science* **269**, 407–410.

Eichler, E. E., Holden, J. J., Popovich, B. W., Reiss, A. L., Snow, K., Thibodeau, S. N., Richards, C. S., Ward, P. A., and Nelson, D. L. (1994) Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nat. Genet.* **8**, 88–94.

Eichler, E. E., Hammond, H. A., Macpherson, J. N., Ward, P. A., and Nelson, D. L. (1995) Population survey of the human FMR1 CGG repeat substructure suggests biased polarity of the loss of AGG interruption. *Hum. Mol. Genet.* **4**, 2199–2208.

Eshelmann, J. R. and Markowitz S. D. (1996) Mismatch repair defects in human carcinogenesis. *Hum. Mol. Genet.* **5**, 1489–1494.

Fishel, R. and Kolodner, R.D. (1995) Identification of mismatch repair genes and their role in the development of cancer. *Curr. Opin. Genet. Dev.* **5**, 382–395.

Fishel, R., Lescoe, M. K., Rao, M. R. S., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M., and Kolodner, R. (1993) The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell.* **75**, 1027–1038.

Freudenreich, C. H., Stavenhagen, J. B., and Zakian, V. A. (1997) Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol. Cell. Biol.* **17**, 2090–2098.

Freund, A. M., Bichara, M., and Fuchs, R. P. (1989) Z-DNA-forming sequences are spontaneous deletion hot spots. *Proc. Natl. Acad. Sci. USA* **86**, 7465–7469.

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.

Fu, Y.-H., Kuhl, D. P. A., Pizzuti, A., Pieretti, M., Sutcliffe, J. S., Richards, S., Verkerk, A. J. M. H., Holden, J. J. A., Fenwick, R. G. Jr, Warren. S. T., Oostra, B. A., Nelson, D. L., and Caskey, C. T. (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell.* **67**, 1047–1058.

Fu, Y.-H., Pizzuti, A., Fenwick, R. G. Jr, King, J., Rajnarayan, S., Dunne, P. W., Dubel, J., Nasser, G. A., Ashizawa, T., de Jong, P., Wieringa, B., Korneluk, R., Perryman, M. B., Epstein, H. F., and Caskey, C. T. (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science.* **255**, 1256–1258.

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.

Gastier, J. M., Pulido, J. C., Sunden, S., Brody, T., Buetow, K. H., Murray, J. C., Weber, J. L., Hudson, T. J., Sheffield, V. C., and Duyk, G. M. (1995) Survey of trinucleotide repeats in the human genome: assessment of their utility as genetic markers. *Hum. Mol. Genet.* **4**, 1829–1836.

Gellibolian, R., Bacolla, A., and Wells, R. D. (1997) Triplet repeat instability and DNA topology: a theoretical analysis. *J. Biol. Chem.* **27**, 16793–16797.

Goldberg, Y. P., Kalchman, M. A., Metzler, M., Nasir, J., Zeisler, J., Graham, R., Koide, H. B., O'Kusky, J., Sharp, A. H., Ross, C. A., Jirik, F., and Hayden, M. R. (1996) Absence of disease phenotype and intergenerational stability of the CAG repeat in transgenic mice expressing the human Huntington disease transcript. *Hum. Mol. Genet.* **5**, 177–185.

Hagerman, P. J. (1988) Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* **17.** 265–286.

Hagerman, P. J. and Ramadevi, V. A. (1990) Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure. I. Computational analysis. *J. Mol. Biol.* **212**. 351–362.

Hancock, J. M. (1996) Simple sequences and the expanding genome. *Bioessays.* **18**, 421–425.

Harris, S., Moncrieff, C., and Johnson, K. (1996) Myotonic dystrophy: will the real gene please step forward! *Hum. Mol. Genet.* **5**, 1417–1423.

Hodges-Garcia, Y. and Hagerman, P. (1995) Investigation of the influence of cytosine methylation on DNA flexibility. *J. Biol. Chem.* **270**, 197–201.

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.

Ikeda, H., Yamaguchi, M., Sugai, S., Aze, Y., Narumiya, S., and Kakizuka, A. (1996) Expanded polyglutamine in the Machado-Joseph disease protein induces cell death *in vitro* and *in vivo*. *Nat. Genet.* **13**, 196–202.

Imbert, G., Saudou, F., Yvert, G., Devys, D., Trottier, Y., Garnier, J.-M., Weber, C., Mandel, J.-L., Cancel, G., Abbas, N., Durr, A., Didierjean, O., Stevanin, G., Agid, Y., and Brice, A. (1996) Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nat Genet* **14**, 285–291.

Ionov, Y., Peinado, M., Malkhosyan, S., Shibata, D., and Perucho, M. (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature.* **363**, 558–561.

Jacobson, H. and Stockmayer, W. H. (1950) Intramolecular reaction in polycondensations. I. The theory of linear systems. *J. Chem. Phys.* **18**, 1600–1606.

Jaworski, A., Higgins, N. P., Wells, R. D., and Zacharias, W. (1991) Topoisomerase mutants and physiological conditions control supercoiling and Z-DNA formation *in vivo*. *J. Biol. Chem.* **266**, 2576–2581.

Jaworski, A., Rosche, W. A., Gellibolian, R., Kang, S., Shimizu, M., Sinden, R. R., and Wells, R. D. (1995) Mismatch repair in *Escherichia. coli.* enhances instability *in vivo* of $(CTG)_n$ triplet repeats from human hereditary diseases. *Proc. Natl. Acad. Sci. USA* **92**, 11019–11023.

Jeffreys, A. J., Tamaki, K., MacLeod, A., Monckton, D. G., Neil, D. L., and Armour, J. A. L. (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nat. Genet.* **6**, 136–145.

Kahn, J. D., Yun, E., and Crothers, D. M. (1994) Detection of localized DNA flexibility. *Nature.* **368**, 163–166.

Kang, S., Jaworski, A., Ohshima, K., and Wells, R. D. (1995a) Expansion and deletion of CTG triplet repeats from human disease genes are determined by the direction of replication. *Nat. Genet.* **10**, 213–218.

Kang, S., Ohshima, K., Shimizu, M., Amirhaeri, S., and Wells, R. D. (1995b) Pausing of DNA synthesis *in vitro* at specific loci in CTG and CGG triplet repeats from human hereditary diseases. *J. Biol. Chem* .**270**, 27014–27021.

Kang, S., Ohshima, K., Jaworski, A., and Wells, R. D. (1996) CTG triplet repeats from the myotonic dystrophy gene are expanded in *E. coli.* distal to the replication origin as a single large event. *J. Mol. Biol.* **258**, 543–547.

Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M., Akiguchi, I., Kimura, J., Narumiya, S., and Kakizuka, A. (1994) CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat. Genet.* **8**, 221–228.

Kinzler, K. W. and Vogelstein, B. (1996) Lessons from hereditary colorectal cancer. *Cell.* **87**, 159–170.

Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hayashi, T., Saito, M., Tomoda, A., Miike, T., Naito, H., Ikuta, F., and Tsuji, S. (1994) Unstable expansion of CAG repeat in hereditary dentatorubal-pallidoluysian atrophy (DRPLA). *Nat. Genet.* **6**, 9–13.

Kolodner, R. D. (1995) Mismatch repair: mechanisms and relationship to cancer susceptibility. *Trends. Biochem. Sci.* **20**, 397–401.

Krontiris, T. G. (1995) Minisatellites and human disease. *Science.* **269**, 1682–1683.

Kunst, C. B. and Warren, S. T. (1994) Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell.* **77**, 853–861.

LaSpada, 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.

Leach, F. S., Nicolaides, N. C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomaki, P., Sistonen, P., Aaltonen, L. A., Nystrom-Lahti, M., Guan, X-Y., Zhang, J., Meltzer, P. S., Yu, J.-W., Kao, F.-T., Chen, D. J., Cerosaletti, K. M., Fournier, R. E. K., Todd, S., Lewis, T., Leach, R. J., Naylor, S. L., Weissenbach, J., Mecklin, J. P., Jarvinen, H., Petersen, G. M., Hamilton, S. R., Green, J., Jass, J., Watson, P., Lynch, H. T., Trent, J. M., de la Chapelle, A., Kinzler, K. W., and Vogelstein, B. (1993) Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell.* **75**, 1215–1225.

Levinson, G. and Gutman, G. A. (1987a) High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res.* **15**, 5323–5338.

Levinson, G. and Gutman, G. A. (1987b) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**, 203–221.

Li, X.-J., Li, S.-H., Sharp, A. H., Nucifora, F. C. Jr, Schilling, G., Lanahan, A., Worley, P., Snyder, S. H., and Ross, C. A. (1995) A huntingtin-associated protein enriched in brain with implications for pathology. *Nature* **378**, 389–402.

Loeb, L. A. (1994) Microsatellite instability: marker of a mutator phenotype in cancer. *Cancer Re.s* **54**, 5059–5063.

Lustig, A. J. and Petes, T. D. (1993) Genetic control of simple sequence stability in yeast, in *Genome Rearrangement and Stability*, Davies, K. E. and Warren, S. T. (eds.) vol. 7. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 79–106.

Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'Hoy, K., Leblond, S., Earle-Macdonald, J., de Jong, P., Wieringa, B., and Korneluk, R. G. (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. *Science* **255**, 1253–1255.

Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetheringthon. C., Lawton, M., Trottier, Y., Lehrach, H., Davies, S. W., and Bates, G. P. (1996) Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell.* **87**, 493–506.

Marra, G. and Boland, C. R. (1995) Hereditary nonpolyposis colorectal cancer: the syndrome, the genes, and historical perspectives. *J. Natl. Can. Inst.* **87**,1114–1125.

Mitas, M., Yu, A., Dill, J., and Haworth, I. S. (1995) The trinucleotide repeat sequence $d(CGG)_{15}$ forms a heat-stable hairpin containing $G^{syn} \cdot G^{anti}$ base pairs. *Biochemistry* **34**, 12803–12811.

Mitchell, J. E., Newbury, S. F., and McClellan, J. A. (1995) Compact structures of $d(CNG)_n$ oligonucleotides in solution and their possible relevance to fragile X and related human genetic diseases. *Nucleic Acids Res.* **23**, 1876–1881.

Modrich, P. (1991) Mechanisms and biological effects of mismatch repair. *Annu. Rev. Genet.* **25**, 229–253.

Modrich, P. and Lahue, R. (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.* **65**, 101–133.

Monckton, D., Coolbaugh, M. I., Ashizawa, K. T., Sicilano, M. J., and Caskey, C. T. (1997) Hypermutable myotonic dystrophy CTG repeat mouse transgenes. *Nat. Genet.* **15**, 193–196.

Monckton, D. G., Wong, L. J., Ashizawa, T., and Caskey, C. T. (1995) Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Hum. Mol. Genet.* **4**, 1–8.

Nagafuchi, S., Yanagisawa, H., Sato, K., Shirayama, T., Ohsaki, E., Bundo, M., Takeda, T., Tadokoro, K., Kondo, I., Murayama, N., Tanaka, Y., Kikushima, H., Umino, K., Kurosawa, H., Furukawa, T., Nihei, K., Inoue, T., Sano, A., Komure, O., Takahashi, M., Yoshizawa, T., Kanazawa, I., and Yamada, M. (1994) Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nat. Genet.* **6**, 14–18.

Nasir, J., Floresco, S. B., O'Kusky, J. R., Diewert, V. M., Richman, J. M., Zeisler, J., Borowski, A., Marth, J. D., Phillips, A. G., and Hayden, M. R. (1995) Targeted disruption of the Huntington's disease gene results in embryonic lethality and behavioral and morphological changes in heterozygotes. *Cell.* **81**, 811–823.

Nasir, J., Goldberg, Y. P., and Hayden, M. R. (1996) Huntington disease: new insights into the relationship between CAG expansion and disease. *Hum. Mol. Genet* **5**, 1431–1435.

Ohshima, K., Kang, S., and Wells, R. D. (1996a) CTG triplet repeats from human hereditary disease are dominant genetic expansion products in *E. coli*. *J. Biol. Chem.* **271**, 1853–1856.

Ohshima, K., Kang, S., Larson, J. E., and Wells, R. D. (1996b) Cloning, characterization, and properties of seven triplet repeat DNA sequences. *J. Biol. Chem.* **271**, 16773–16783.

Ohshima, K., Kang, S., Larson, J. E., and Wells, R. D. (1996c) TTA·TAA triplet repeats in plasmids form a non-hydrogen bonded structure. *J. Biol. Chem.* **271**, 16784–16791.

Ohshima, K. and Wells, R. D. (1997) Hairpin formation during DNA synthesis primer realignment *in vitro* in triplet repeat sequences from human hereditary disease genes. *J. Biol. Chem.* **272**, 16798–16806.

Orr, H. T., Chung, M. Y., Banfi, S., Kwiatkowski, T. J. Jr, Servadio, A., Beaudet, A. L., McCall, A. E., Duvick, L. A., Ranum, L. P., and Zoghbi, H. Y. (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat. Genet.* **4**, 221–226.

Parsons, R., Li, G.-M., Longley, M. J., Fang, W.-h., Papadopoulos, N., Jen, J., de la Chapelle, A., Kinzler, K. W., Vogelstein, B., and Modrich, P. (1993) Hypermutability and mismatch repair deficiency in RER^+ tumor cells. *Cell.* **75**, 1227–1236.

Paulson, H. L. and Fischbeck, K. H. (1996) Trinucleotide repeats in neurogenetic disorders. *Annu. Rev. Neurosci.* **19**, 79–107.

Pulst, S.-M., Nechiporuk, A., Nechiporuk, T., Gispert, S., Chen, X.-N., Lopes-Cendes, I., Pearlman, S., Starkman, S., Orozco-Diaz, G., Lunkes, A., DeJong, P., Rouleau, G. A., Auberger, G., Korenberg, J. R., Figueroa, C., and Sahba, S. (1996) Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat. Genet.* **14**, 269–276.

Reitmair, A. H., Schmits, R., Ewel, A., Bapat, B., Redston, M., Mitri, A., Waterhouse, P., Mittrucker, H. W., Wakeham, A., Liu, B., Thomason, A., Griesser, H., Gallinger, S., Ballhausen, W. G., Fishel, R., and Mak, T. W. (1995) Msh2 deficient mice are viable and susceptible to lymphoid tumours. *Nat. Genet.* **11**, 64–70.

Richards, R. I. and Sutherland, G. R. (1992) Dynamic mutations: a new class of mutations causing human disease. *Cell.* **70**, 709–712.

Rosche, W. A., Jaworski, A., Kang, S., Kramer, S. F., Larson, J. E., Giedroc, D. P., Wells, R. D., and Sinden, R. R. (1996) Single strand DNA binding protein enhances the stability of CTG triplet repeats in *Escherichia coli*. *J. Bacteriol.* **178**, 5042–5044.

Sanpei, K., Takano, H., Igarashi, S., Sato, T., Oyake, M., Sasaki, H., Wakisaka, A., Tashiro, K., Ishida, Y., Ikeuchi, T., Koide, R., Saito, M., Sato, A., Tanaka, T., Hanyu, S., Takiyama, Y., Nishizawa, M., Shimizu, N., Nomura, Y., Segawa, M., Iwabuchi, K., Eguchi, I., Tanaka, H., Takahashi, H., and Tsuji, S. (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.

Schleif, R. (1992) DNA looping. *Annu. Rev. Biochem.* **61**, 199–223.

Sherman, S. L., Jacobs, P. A., Morton, N. E., Froster-Iskenius, U., Howard-Peebles, P. N., Nielsen, K. B., Partington, M. W., Sutherland, G. R., Turner, G., and Watson, M. (1985) Further segregation analysis of the fragile X syndrome with special reference to transmitting males. *Hum. Genet.* **69**, 289–299.

Shimada, J. and Yamakawa, H. (1984) Ring-closure probability for twisted wormlike chains. Application to DNA. *Macromolecules* **17**, 689–698.

Shimada, J. and Yamakawa, H. (1985) Statistical mechanics of DNA topoisomers. The helical worm-like chain. *J. Mol. Biol.* **184**, 319–329.

Shimizu, M., Gellibolian, R., Oostra, B. A., and Wells, R. D. (1996) Cloning, characterization, and properties of plasmids containing CGG triplet repeats from the fragile X gene. *J. Mol. Biol.* **258**, 614–626.

Shore, D. and Baldwin, R. L. (1983a) Energetics of DNA twisting. I. Relation between twist and cyclization probability. *J. Mol. Biol.* **170**, 957–981.

Shore, D. and Baldwin, R. L. (1983b) Energetics of DNA twisting. II. Topoisomer analysis. *J. Mol. Biol.* **170**, 983–1007.

Shore, D., Langowski, J., and Baldwin, R. L. (1981) DNA flexibility studied by covalent closure of short fragments into circles. *Biochemistry* **78**, 4833–4837.

Sinden, R. R. (1994) *DNA Sructure and Function*. Academic Press, San Diego, California.

Sinden, R. R. and Wells, R. D. (1992) DNA structure, mutations, and human genetic disease. *Curr. Opin. Biotechnol.* **3**, 612–622.

Smith, G. K., Jie, J., Fox, G. E., and Gao, X. (1995) DNA CTG triplet repeats involved in dynamic mutations of neurologically related gene sequences form stable duplexes. *Nucleic. Acids. Res.* **23**, 4303–4311.

Smith, G. P. (1973) Unequal crossover and the evolution of multigene families. *Symp. Quant. Biol.* **38**, 507–513. Cold Spring Harbor.

Strand, M., Prolla, T. A., Liskay, R. M., and Petes, T. D. (1993) Destabilisation of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274–276.

Strand, M., Earley, M. C., Crouse, G. F., and Petes, T. D. (1995) Mutations in the msh3 gene preferentially lead to deletions within tracts of simple repetitive DNA in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **92**, 10418–10421.

Sutherland, G. R. and Richards, R. I. (1995), Simple tandem DNA repeats and human genetic disease. *Proc. Natl. Acad. Sci.* USA **92**, 3636–3641.

Tautz, D. and Schlotterer, C. (1994) Simple sequences. *Curr. Opin. Genet. Dev* .**4**, 832–837.

Taylor, W. H. and Hagerman, P. J. (1990) Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure. II. NaCl-dependence of DNA flexibility and helical repeat. *J. Mol. Biol.* **212**, 363–376.

Telenius, H., Kremer, B., Goldberg, Y. P., Theilmann, J., Andrew, S. E., Zeisler, J., Adam, S., Greenberg, C., Ives, E. J., Clarke, L. A., and Hayden, M. R. (1994) Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nat. Genet.* **6**, 409–414.

Thibodeau, S. N., Bren, G., and Schaid, D. (1993) Microsatellite instability in cancer of the proximal colon. *Science* **260**, 816–819.

Trinh, T. Q. and Sinden, R. R. (1991) Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*. *Nature* **352**, 544–547.

Trottier, Y., Lutz, Y., Stevanin, G., Imbert, G., Devys, D., Cancel, G., Saudou, F., Weber, C., David, G., Tora, L., Agid, Y., Brice, A., and Mandel, J.-L. (1995) Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias. *Nature* **378**, 403–406.

Umar, A. and Kunkel, T. A. (1996) DNA-replication fidelity, mismatch repair and genome instability in cancer cells. *Eur. J. Biochem.* **238**, 297–307.

Vogelstein, B. and Kinzler, K. W. (1993) The multistep nature of cancer. *Trends Genet.* **9**, 138–141.

Wang, J. C. (1996) DNA topoisomerases. *Annu. Rev. Biochem.* **65**, 635–692.

Wang, J. C. and Davidson N (1966) On the probability of ring closure of lambda DNA. *J. Mol. Biol.* **19**, 469–482.

Wang, Y.-H., Amirhaeri, S., Kang, S., Wells, R. D., and Griffith, J. (1994) DNA triplet repeats from the myotonic dystrophy gene are preferential nucleosome assembly sites *in vitro*. *Science* **265**, 669–671.

Wang, Y.-H., Gellibolian, R., Shimizu, M., Wells, R. D., and Griffith, J. (1996) Long repeating CCG triplet repeat blocks exclude nucleosomes: a possible mechanism for the nature of fragile sites in chromosomes. *J. Mol. Biol.* **263**, 511–516.

Warren, S. T. (1996) The expanding world of trinucleotide repeats. *Science* **271**, 1374–1375.

Warren, S. T. and Ashley, J. C. T. (1995) Triplet repeat expansion mutations: the example of fragile X syndrome. *Annu. Rev. Neurosci.* **18**, 77–99.

Warren, S. T. and Nelson, D. L. (1994) Advances in molecular analysis of fragile X syndrome. *J. Am. Med. Assoc.* **271,** 536–542.

Wells, R. D. (1988) Unusual DNA structures. *J. Biol. Chem.* **263**, 1095–1098.

Wells, R. D. (1996) Molecular basis of genetic instability of triplet repeats. *J. Biol. Chem.* **271**, 2875–2878.

Wells, R. D. (1997) Triplet repeat diseases studied in man, microbes, and molecules. *Am. J. Psychiatry* **154**, 887.

Wells, R. D. and Sinden, R. R. (1993) Defined ordered sequence DNA, DNA structure, and DNA-directed mutation, in *Genome Analysis, Genome Rearrangement and Stability*, Davis, K. and Warren, S. (eds.) Vol 7. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 107–138.

Willems, P. J. (1994) Dynamic mutations hit double figures. *Nat. Genet.* **8**, 213–215.

Zeitlin, S., Liu, J. P., Chapman, D. L., Papaioannou, V. E., and Efstratiadis, A. (1995) Increased apoptosis and early embyonic lethality in mice nullizygous for the Huntington's disease gene homologue. *Nat. Genet.* **11**, 155–163.

Zoghbi, H. Y. (1996) The expanding world of ataxins. *Nat Genet* **14**, 237–238.