THE MOLECULAR MECHANISMS INVOLVED IN THE GENETIC INSTABILITY OF THE CCTG•CAGG REPEATS ASSOCIATED WITH MYOTONIC DYSTROPHY TYPE 2

A Dissertation

by

RUHEE J. DERE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2006

Major Subject: Genetics

THE MOLECULAR MECHANISMS INVOLVED IN THE GENETIC INSTABILITY OF THE CCTG•CAGG REPEATS ASSOCIATED WITH MYOTONIC DYSTROPHY TYPE 2

A Dissertation

by

RUHEE J. DERE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Robert D. Wells
Committee Members,	James E. Womack
	Richard R. Sinden
	Richard H. Finnell
Chair of Genetics Faculty,	James R. Wild

May 2006

Major Subject: Genetics

ABSTRACT

The Molecular Mechanisms Involved in the Genetic Instability of the CCTG•CAGG Repeats Associated with Myotonic Dystrophy Type 2. (May 2006)

Ruhee J. Dere, B.S., University of Mumbai;

M.S., University of Mumbai

Chair of Advisory Committee: Dr. Robert D. Wells

Myotonic dystrophy type 2 (DM2) is caused by the extreme expansion (from < 30 repeats in normal individuals to ~ 11,000 for the full mutation in certain patients) of the repeating tetranucleotide CCTG•CAGG sequence in the intron of the zinc finger protein 9 (ZNF9) gene. The genetic instabilities of the CCTG•CAGG repeats were investigated to evaluate the molecular mechanisms responsible for these massive expansions. The effects of replication, recombination, repair and transcription on the genetic instabilities have been investigated in COS-7 cells and *E. coli* model systems. A replication assay was established in COS-7 cells wherein the CCTG•CAGG repeats cloned proximal to the SV40 origin of replication resulted in expansions and deletions in a length and orientation-specific manner, whereas the repeats cloned distal to the same origin were comparatively stable. These results fit with our data obtained from biochemical studies on synthetic oligonucleotides since these biochemical studies revealed that the $d(CAGG)_{26}$ oligomer had a marked propensity to adopt a hairpin structure as opposed to its complementary $d(CCTG)_{26}$ that lacked this capacity.

Furthermore, a genetic assay in *E. coli* was used to monitor the intramolecular frequency of recombination. This assay revealed that the tetranucleotide repeats were indeed hot spots for recombination. Moreover, studies conducted in SOS-repair mutants showed that recombination frequencies were much lower in a SOS^- strain as compared to a SOS^+ strain. However, experiments conducted to ascertain the level of induction of the SOS response revealed that the SOS pathway was not stimulated in our studies. These results revealed that although breaks may occur within the repeats, the damage is most likely repaired without induction of the SOS response contrary to previous beliefs.

Thus, a complex interplay of replication, recombination, and repair is likely responsible for the expansions observed in DM2.

DEDICATION

To my parents

ACKNOWLEDGMENTS

I would like to begin by thanking my advisor Dr. Robert D. Wells for the wisdom and guidance that he has provided me with, not only in regards to my work but in all aspects of life. He has always taken the time to talk and listen to me, and has been instrumental in guiding me in the right direction.

I also wish to express my gratitude to all the members of my committee who took time from their busy schedules to be a part of my advisory committee. Their suggestions and positive criticisms have enabled me to think beyond the most obvious implications of my experiments. I would especially like to thank Dr. Womack for driving down to Houston for most of my committee meetings and being there in person to partake in all the discussions regarding my project. I would also like to express my gratitude to Dr. Laura P. W. Ranum who was instrumental in providing us with the DM2 sequence that has been the focus of my work and for all the helpful suggestions regarding my project.

Sincere thanks to all of the past and present members of the Wells group including Drs. Albino Bacolla, Alexander Vetcher, Anna Pluciennik, Micheal Hebert, Beata Kosmider and Ms. Leslie Spitz-Son, Lili Mochmann, and Jackie Larson that have helped me in the laboratory not only with experimental techniques but also with advice and helpful discussions regarding my work.

I would like to express my heart felt gratitude to my parents and my sister for believing in me and giving me this wonderful opportunity to pursue my dreams. Last but not the least I would like to thank some very special people, Mr. Sohail Alam, Ms.

Mona Dhurandhar, Ms. Leslie Spitz-Son, Ms. Nicole Ramlachan, Mr. Joe Stafford and Dr. Kaytan Amrute for always being there for me, and cheering me on when the going got tough.

LIST OF ABBREVIATIONS

HNPCC	hereditary nonpolyposis colorectal carcinoma
TRS	trinucleotide repeat sequence
FRAXA	Fragile X syndrome
DM1	myotonic dystrophy type 1
FRDA	Friedreich's ataxia
HD	Huntington's disease
HDL2	Huntington disease-like 2
SCA	spinocerebellar ataxia
DRPLA	dentatorubral-pallidoluysian atrophy / Haw River syndrome
SBMA	spinal and bulbar muscular atrophy / Kennedy's disease
COMP	cartilage oligomeric matrix protein
SPD	synpolydactyly
HFGS	Hand–foot–genital syndrome
CCD	cleidocranial dysplasia
HPE	holoprosencephaly
XLMR+GHD	X-linked mental retardation with growth hormone deficiency
XLAG	X-linked mental retardation and abnormal genetalia
CCHS	congenital central hypoventilation, Haddad syndrome
OPMD	oculopharyngeal muscular dystrophy
BPEIS	blepharophimosis-ptosis-epicanthus inversus syndactyly

UTR	untranslated region
SSED	small-slipped expansions and deletions
MMR	methyl directed mismatch repair
NER	nucleotide excision repair
GFP	green fluorescent protein
PKD1	polycystic kidney disease
APRT	adenine phosphoribosyltransferase
CHO	Chinese hamster ovary
DUE	DNA unwinding element
CD	circular dichroism
DSB's	double strand breaks
FEN-1	flap endonuclease 1
HR	homologous recombination
SDSA	synthesis-dependent strand annealing
EPM1	progressive myoclonus epilepsy 1/Unverricht – Lundborg disease
CSTB	cystatin B
SCA10	spinocerebellar ataxia type 10
DM2	myotonic dystrophy type 2
ZNF9	zinc finger protein 9
DMPK	dystrophia myotonica protein kinase
PROMM	proximal myotonic myopathy
PDM	proximal myotonic dystrophy

CUG-BP	CUG binding protein
DMEM	Dulbecco's modified Eagle's media
OsO_4	osmium tetraoxide
$KMnO_4$	potassium permanganate
DEPC	diethyl pyrocarbonate
DMS	dimethyl sulfate
SDS	sodium dodecyl sulfate
ONPG	o-nitrophenyl-β-D-galactosidase
bp	base pairs
Kbp	kilo base pairs

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

A. Overview

The genetic instabilities associated with hereditary neurological diseases have been studied extensively, since the discovery, in 1991, of the CGG•CCG expansion causative of the Fragile X syndrome (1-3). Although, repetitive sequences are found dispersed all over the human genome only a handful of these repeat tracts at certain loci have been implicated in a disease phenotype. Analyses of these repeats has lead to the identification of several *cis*-elements that contribute to the instability of the repeat tracts including, repeat sequence, the length and purity of the repeat tracts, sequences flanking the repeats, DNA methylation and chromatin structure (4). Numerous groups have conducted experiments in several model systems to implicate cellular pathways including, DNA replication, repair, recombination, and transcription in generating the dynamic mutations observed with these microsatellite disorders (5-8). The ability of the repetitive DNA sequences, including CTG•CAG, CGG•CCG, GAC•GTC, GAA•TTC, CCTG•CAGG and ATTCT•AGAAT to adopt non-B DNA structures such as hairpins, triplexes and tetraplexes has been proposed to be a source for mutagenesis (5-8). Herein, the important aspects of hereditary neurological diseases have been reviewed.

This dissertation follows the style of the *Journal of Biological Chemistry*.

B. Microsatellite instabilities

Microsatellite sequences usually have tandem arrays ranging between one to six base pairs and are a part of a larger family of tandem repetitive sequences (9-11). These microsatellites are highly abundant, polymorphic and are widespread in all eukaryotic genomes from yeast to mammals (12). In mycoplasma and bacteria the genetic polymorphisms of these sequences play a vital role in phase variation that controls gene expression (13). In the past decade a new pathway for the development of cancers in humans has emerged, that is characterized by the inactivation of the DNA mismatch repair system leading to a hypermutable state in which microsatellites become unstable during DNA replication (14-17). Although this instability was first discovered in hereditary nonpolyposis colorectal carcinoma (HNPCC) (16,18-22), the number of diseases characterized by these microsatellite instabilities has increased to include gastric adenocarcinomas, pancreatic carcinomas, prostatic adenocarcinomas, small-cell lung carcinomas, renal carcinomas as well as carcinomas of the endometrium (23-28). Additionally, microsatellites are used extensively in genetic mapping (29-33), paternity testing (34) and forensic medicine (35,36). A growing number of neurological diseases result from the expansion of a particular class of microsatellites called trinucleotide repeats [reviewed in (4,6,37-39)]. More recently this class of neurological diseases has broadened to include a tetranucleotide, pentanucleotide and even a dodecamer repeating tract (4,40-42).

C. Trinucleotide repeat diseases

The expansion of several trinucleotide repeat sequences (TRS) has been associated with almost 30 hereditary neurological diseases including the Fragile X syndrome (FRAXA), myotonic dystrophy type 1 (DM1), Friedreich's ataxia (FRDA), Huntington's disease (HD) as well as the spinocerebellar ataxia's (SCAs) (4-6,38) (Fig. 1). Most of these diseases exhibit genetic anticipation, a term used to describe the increase in disease severity and an earlier age of onset from one generation to the next (5,38). Furthermore, the term 'dynamic mutation' is also used to describe these repetitive sequences as there is a greater propensity for the expanded triplet repeats to undergo further expansions in subsequent generations (43). Repeat instability can either occur throughout the genome or at a particular locus within the genome. Genome-wide instability occurs in several cancers (4) however, the instability associated with the triplet repeat disorders is locus-specific (44-46). Thus, the expansion of the repeats alters the expression levels or the biological activity of the gene at that particular locus. These expansions have been found in coding as well as non-coding regions and the threshold number of repeats beyond which the disease manifests itself is different for each disorder. Based on the location of the repeat within the disease gene, the triplet repeat diseases have been classified into two distinct categories.

1. Type I diseases. The type I diseases are characterized by the presence of the repeats within the coding region of a gene (Fig. 1). These repeats have been identified as CAG•CTG tracts encoding polyglutamine, and are associated with several

neurological diseases including HD, Haw River syndrome (DRPLA), Kennedy's disease (SBMA) as well as a number of spinocerebellar ataxia's (SCA1, 2, 3, 6, 7, and 17) (5,37-39,47,48). In addition to the polyglutamine diseases this category also consists of the GAC•GTC repeats found in the cartilage oligomeric matrix protein (COMP) gene associated with multiple skeletal dysplasia's (49-51). Recently, polyalanine tract expansions (encoded by the GCG•CGC repeats) have been shown to cause nine human conditions most of which occur in transcription factor genes (52-56). The expansions of the repeats in type I diseases are small, between 40 to a 100 repeats and are characterized by a progressive neuronal loss (5,37). Although the genetics of these polyglutamine diseases are well defined, the molecular mechanism itself is poorly understood and is currently being investigated. The mutated protein has been found to form insoluble aggregates in the cytoplasm and nucleus of neurons (57-60). These inclusions have been hypothesized to affect neurons through the alteration of gene transcription, steric interference with cellular activities and the activation of apoptotic or other signaling pathways (61,62).

2. Type II diseases. In these diseases, the repeat expansions occur in the non-coding regions of genes (Fig. 1) and are characterized by large expansions (up to several thousand repeats) in affected individuals. Most of these diseases show a premutation size before a full mutation occurs in subsequent generations (4,5,37). A CGG•CCG expansion in the 5` untranslated region (UTR) of the FMR1 gene is associated with the Fragile X syndrome (2). The mutational basis of this disorder involves the

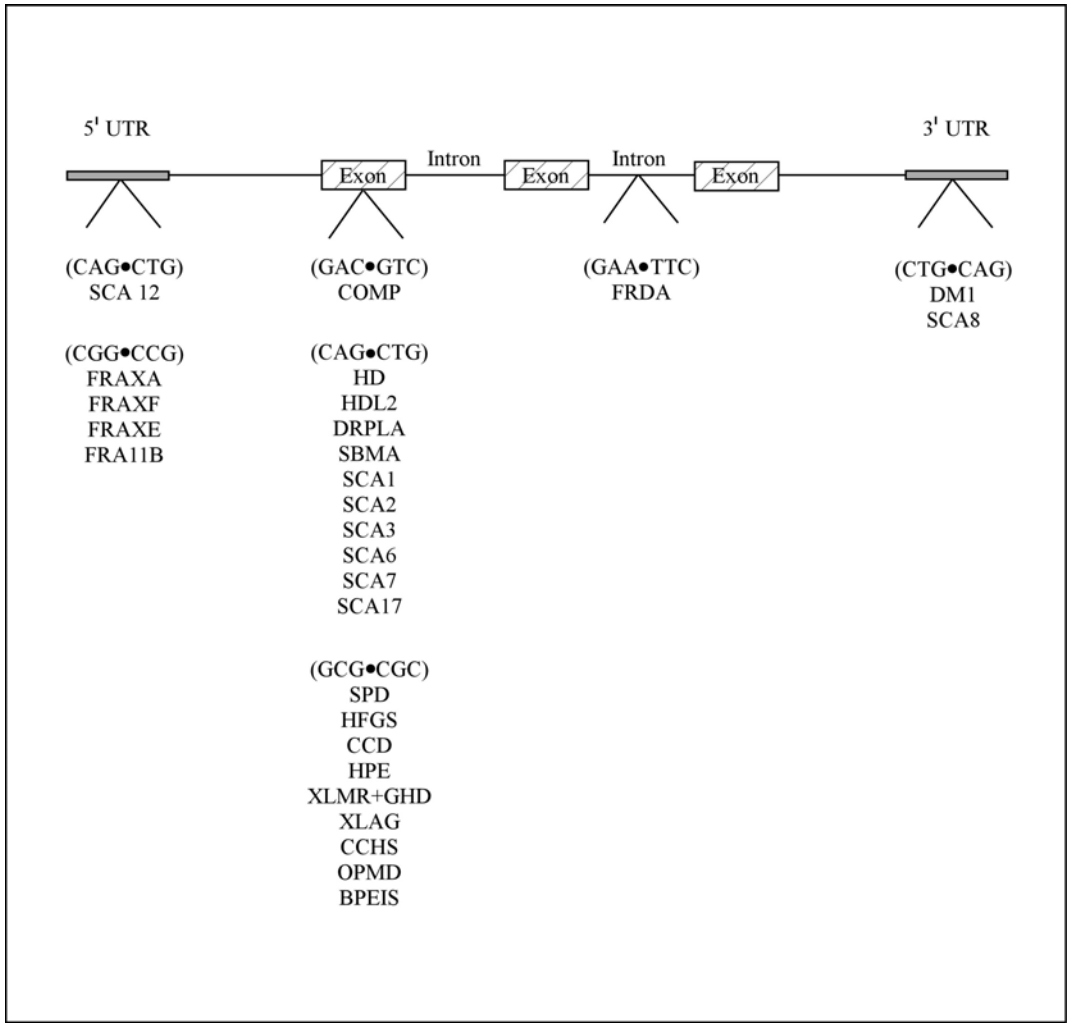


FIG. 1. **Genomic location of the repeating sequences.** The repeating tracts are found either in the coding (exons) or non-coding (introns, 5` - UTR or the 3` - UTR) regions of the gene. The repeating sequences causing each of the disorders are placed in parenthesis above all the disorders implicating that particular repeating tract.

hypermethylation of CpG islands in the promoter of the FMR1 gene leading to its transcriptional repression (63-65). The DM1 and the SCA8 CTG•CAG sequences expand in the 3` UTR of the dystrophia myotonica protein kinase (DMPK) and SCA8 gene, respectively (66-71). Both of these disorders are now hypothesized to be caused due to a toxic gain-of-function by the corresponding RNA transcripts (72-75).

Friedreich's ataxia caused by the expansion of a GAA•TTC repeat in intron 1 of the frataxin gene (76-78) is one of the triplet repeat disorders that does not exhibit anticipation, a characteristic of most of the other TRS diseases. The GAA•TTC expansion leads to insufficient production of the frataxin protein resulting in the accumulation of iron in the mitochondria, eventually causing neuronal death (79-81). Thus, the pathogenesis for each of the disorders in this category depends on a loss of function mutation for the protein or a toxic gain-of-function for the RNA transcript.

D. Types of genetic instability

Studies in several different systems including bacteria, yeast, mice as well as a number of eukaryotic cells in culture have investigated the ability of the triplet repeat sequences to either expand or delete in these models. These studies have been extremely useful in understanding the complex mechanisms and the array of factors involved in the genetic instability associated with the human disease [reviewed in (5-7,38)]. In addition to the expansions and deletions, similar to those observed in the human diseases, a number of other types of instabilities including small slipped-register instabilities,

tandem duplications and gross deletions/rearrangements have been observed in the various model systems and are discussed below.

1. Large expansions and deletions. The expansions of the repeat sequences result in the manifestation of the disease phenotype in humans. The repeats can expand to a premutation size and then into the full mutation in subsequent generations. These expansions are characteristic of these diseases and constitute the predominant type of instability in humans (5). Several model systems used to study the mechanisms of genetic instability have generated both expansions and deletions. Deletions of the expanded tracts (also called reverse mutations) have been documented for several triplet repeat disorders in humans. A number of cases of DM1 have shown a reduction in the number of repeats in transmission from an affected parent to offspring (82-86). Repeat deletions in sperm have also been observed for both FRDA (87) as well as SBMA (88). Additionally, CGG•CCG tract deletions have been found within or around the folate-sensitive fragile sites (89,90). Several *cis*-elements including repeat length, repeat purity, proximity to the origin of replication, direction of replication and chromosomal location have been implicated to play an important role in explaining the bias of expansions to deletions in humans. Moreover, several *trans*-factors such as various repair and replication proteins may act in concert with the *cis*-elements to enhance repeat instability (4).

2. Small slipped-register instabilities. Small expansions and deletions (in 3 base pair registers) were studied *in vivo* in *E. coli* (91). These studies revealed for the first time that large and small expansions and deletions occurred via different

mechanisms. A model involving strand misalignment, incision or excision followed by DNA synthesis and ligation was proposed to explain small slipped expansions and deletions (SSED) (91). The small slippage events were considered to be substrates for methyl-directed mismatch repair (MMR) and nucleotide excision repair (NER). This mechanism was proposed to explain the small expansions associated with the type I hereditary neurological diseases.

3. Tandem duplications. This form of triplet repeat instability was observed in *E. coli*, in which the entire block of repeats as well as 129 base pairs (bp) of non-repetitive flanking sequence was duplicated (92). The products from this duplication contained as many as 34 repeat units. This duplication required the presence of two or more TRS containing units in close proximity (170 bp) to the R6K γ origin of replication. The mechanism proposed to explain these duplications involved secondary structure formation, stalling of DNA synthesis and the slippage-mediated misalignment of complementary strands during DNA replication.

4. Gross deletions/rearrangements. Recently a new type of instability has been associated with triplet repeat sequences studied in *E. coli*. This instability is termed gross deletions/rearrangements wherein, a deletion occurs not only within the repeat tract but also in the region flanking the repeats (7,93). Studies using the CTG•CAG tracts showed that the presence of long tracts of the repeats promoted formation of inversions and long deletions that removed part or all of the repeats as well as the flanking green fluorescent protein (GFP) reporter gene (93). Shorter repeat lengths, as well as the GAA•TTC repeats were inert. The potential for the triplet repeats to be

mutagenic was investigated after observations of rearrangements occurring within the 2.5 kbp (purine – pyrimidine) tract from the human polycystic kidney disease 1 (PKD1) gene (94). These studies proposed the formation of alternative DNA structures to trigger genomic rearrangements by activation of the recombination-repair pathways (95,96). Similar to the PKD1 studies, the gross deletions associated with the DM1 sequence were predicted to be formed due to the presence of non-B DNA folded conformations found at or near the breakpoints for all rearrangements (93). Additionally, long CTG•CAG repeats were shown to induce deletions and rearrangements as a result of recombination at the adenine phosphoribosyltransferase (APRT) locus in Chinese hamster ovary (CHO) cells (97), further supporting the mutagenic role of the trinucleotide repeats.

E. Mechanisms involved in genetic instability

There are a number of mechanisms that work in concert with each other and contribute to the genetic instabilities of the triplet repeat disorders. The most important mechanisms include replication, repair, recombination, and transcription. The propensity of the various repeating sequences to form non-B DNA structures is believed to be involved in all of the mechanisms generating instabilities. In individuals with these hereditary neurological diseases, the expansions are locus-specific (4,5). However, tissue-specific somatic instability is observed in some cases (98-103). This somatic instability has been observed in both proliferating as well as non-proliferating tissues (99,104-106). A number of mechanisms have been implicated to influence repeat instability; these mechanisms can either work in concert with each other or may

represent conditions in a certain tissue, or at a particular locus, or even a specific stage of development.

1. Non-B DNA Structures. The repeating sequences associated with neurological diseases have been shown to form several secondary structures. These structures include hairpins, slipped-strand DNA, triplexes, tetraplexes, DNA unwinding elements (DUE's) and sticky DNA (Fig. 2). The various mechanistic models proposed to explain genetic instability implicate the formation of folded-back structures by the repeating sequences [reviewed in (4-7,38)].

The simplest structure that seems to be formed by a large number of repeating sequences is a hairpin-loop (Fig. 2A). A hairpin structure is formed by the folding back of single-stranded DNA on itself. Several biophysical, biochemical and theoretical experiments have shown the CTG, CAG, CGG, CCG, GTC, GAC, GAA, TTC, and CAGG strands, to all form hairpin structures (107-120). The stability of each of these hairpins varies depending on the stability of the mispairs within the structure. Additionally, the stability of the mispairs is sequence context dependant wherein, the nearest neighbor flanking the mismatch infuences its stability in a given non-B DNA structure (121-127). An order of stability for the DM1 and FRAXA sequences, CGG > CTG > CAG > CCG was established using CD, optical melting, differential scanning and calorimetry (112). This variation in the stabilities of the structures formed by each of the repeating sequences can potentially explain a strand bias, where one strand in a DNA duplex would be more likely to form a stable secondary structure as compared to its complement. This strand bias (differences between orientations I and II) was

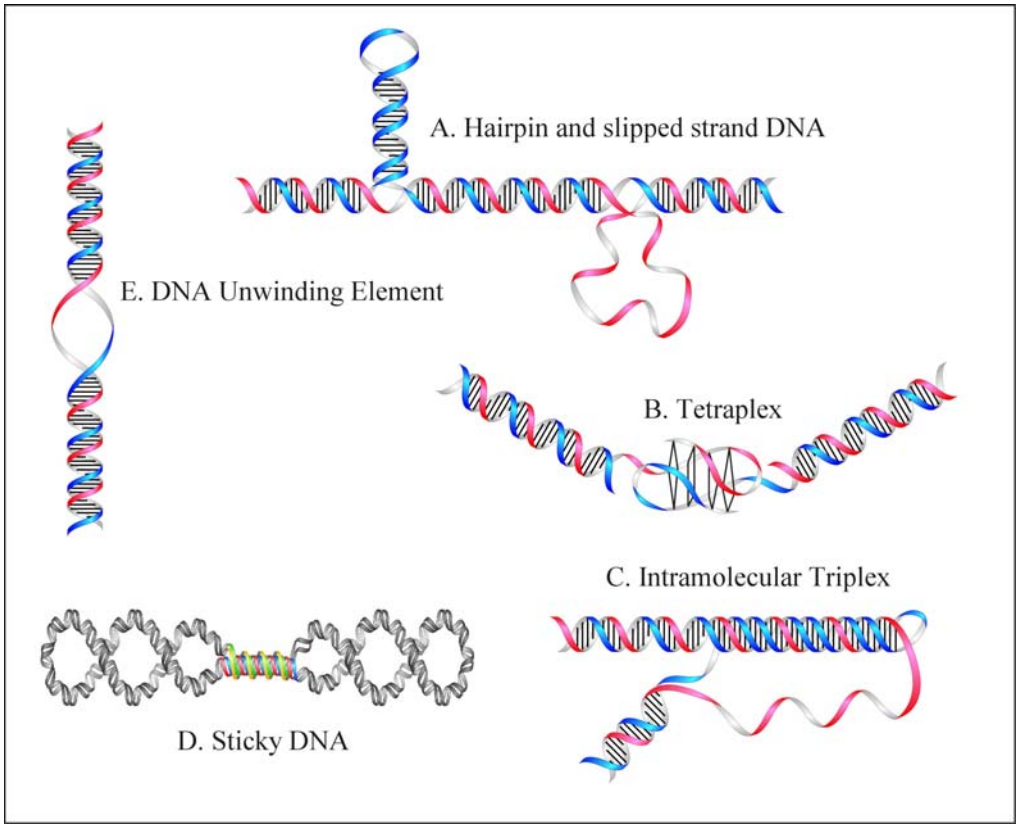


FIG. 2. **Non-B DNA structures formed by repeating sequences.**

observed in a number of experiments in *E. coli*, yeast as well as tissues culture models (120,128-137). Long CTG•CAG and CGG•CCG repeats form slipped-strand DNA (*in vitro*), where hairpins are formed on both the complementary strands (138-140). This occurs when the two strands are allowed to denature and are then re-annealed. Furthermore, studies conducted on hairpin/slipped structures have shown several repair proteins, such as UvrA and MSH2 to bind to them (141,142).

In addition to hairpins the CGG repeats associated with the Fragile X syndrome were hypothesized to form tetraplexes (143-145) (Fig. 2B). These tetraplexes occur at G-rich sequences such as telomeric DNA where the G-residues form a G-quartet (146,147). These structures can either be unimolecular or can result from the interaction between two hairpins (143-145). Tetraplexes occurring at the CGG•CCG repeats are believed to be responsible for the replication pausing observed *in vitro* which can be alleviated by addition of a helicase that unwinds the structure. These studies proposed that quadruplexes could potentially exist *in vivo* and contribute to the instability observed in the Fragile X syndrome (148-151). However, recent experiments by Fojtik *et al.* showed the reluctance of these repeats to form tetraplexes at physiological concentrations of sodium and potassium ions. These results revealed the improbability of the formation of tetraplexes by the CGG repeats *in vivo* (152).

Triplexes (Fig. 2C) are formed at long runs of polypurine•polypyrimidine mirror repeat sequences where the purine strand in a duplex DNA, Hoogsteen base pairs with a second purine or pyrimidine strand of another duplex to result in a three strand interaction (7,153-155). The GAA•TTC repeats associated with Friedreich's ataxia have

been shown to form both inter as well as intra-molecular triplexes (155-157). Additionally, the CTG•CAG and CGG•CCG repeats were postulated to from triplex DNA (158). Recently, a novel structure termed "sticky DNA" was proposed where two long tracts of GAA•TTC repeats within the same molecule associate with each other to form a dumbbell shaped conformation in bacterial plasmids (7,154,159) (Fig. 2D). Although the exact conformation of sticky DNA is not clearly elucidated, it has special requirements for formation and once formed does not dissociate easily (7,154,159,160). This structure has been implicated in the decrease of the recombination hotspot activity of the GAA•TTC repeats (161).

The pentanucleotide repeat sequence associated with spinocerebellar ataxia type 10 (SCA10) has recently been shown to be a DNA unwinding element (DUE) (41,162). DUE's are A+T rich sequences commonly found at replication origins (153) (Fig. 2E). The ATTCT•AGAAT sequence shows a high propensity to unwind and can act as an aberrant DNA origin. The DNA unwinding characteristic of this sequence and its function as an aberrant origin may contribute to the instability associated with the repeat tract (162).

The formation of all of these folded-back structures have been implicated in replication pausing, recombination-repair, generation of double-strand breaks (DSBs) and transcription inhibition [reviewed in (4,6,7)].

2. Replication. The first factor to be implicated in repeat instability, early in the history of triplet repeat disorders, was replication-slippage (128). Repeat instability was observed in, but not limited to, highly proliferative cells and during embryogenesis and

fetal development (100,106,163,164). Additionally, an orientation bias obtained in studies conducted in *E. coli* as well as yeast systems further implicated a slippage mechanism (128,130,131,133,135,136,165). The bias was attributed to the propensity of one of the two complementary strands of the repeat tract to form stable secondary structures. In the case the structure formed on the template strands, the subsequent daughter strands would be deleted. However, if slippage occurred on the nascent (newly synthesized) strand then an expansion would ensue. Recent studies on CTG•CAG and CCTG•CAGG repeats in COS cells, have shown similar results to those obtained in bacterial and yeast models (120,137).

Most models have proposed a greater opportunity for slippage and structure formation on the discontinuously synthesized daughter strand, due to its single-stranded nature as compared to the continuously synthesized leading strand. Thus, Okazaki fragment maturation has gained a lot of interest in recent years. The location of initiation of the Okazaki fragments within the repeat tract was proposed to influence repeat stability (4,120,137). Also, the human flap endonuclease 1 (FEN-1 and Rad27 in yeast) was shown to destabilize the repeats as a result of the abberant 5`-end processing of the Okazaki fragments by FEN-1/Rad27 (166-171). Mutations in FEN-1/Rad27 lead to expansions and this behavior was attributed to the ability of the 5` flaps to fold back into secondary structures that could not be processed by FEN-1/Rad27. The subsequent re-annealing of these unprocessed adjacent Okazaki fragments would lead to the observed expansions (165,166,168). These studies with flap processing have been conducted in bacteria and yeast cells. In mammals a complete loss of FEN-1 results in

embryonic lethality (172). However, transgenic mice either heterozygous or homozygous for FEN-1 were created harboring the HD CAG•CTG repeat sequence (173). The heterozygous mice showed an intergenerational instability of the repeats when compared to the homozygous wild-type mice. These results along with studies in human cell lines deficient in FEN-1 that gave rise to instability (173), substantiate the protective function of FEN-1 in repeat tract expansions.

In vitro studies using human, bacterial and phage polymerases (115,148,151,174,175) have shown both the CTG•CAG and the CGG•CCG repeats to cause a length-dependent replication pausing. Two-dimensional gel electrophoresis was used to demonstrate the accumulation of replication intermediates in *E. coli* (176). These studies implicated a length and sequence-dependent slowing or pausing of the replication fork that occurred specifically within the CTG•CAG and CGG•CCG triplet repeats. The pausing was dependent on the orientation, where the more stable structure-forming sequences, CTG and CGG, caused a more pronounced pausing effect than the complementary CAG and CCG repeats. Consistent with the bacterial data, studies conducted *in vivo* in yeast also showed replication pausing for CTG•CAG, CGG•CCG as well as the GAA•TTC repeat tracts (177,178). The pausing of the replication fork at these repeats was hypothesized to lead to the formation of double strand breaks (DSBs), the repair of which could lead to instability.

In addition to the direction of replication fork progression, Okazaki fragment maturation, and polymerase pausing, studies have revealed that the distance of the repeats from the replication origin also influences genetic instability (120,137). Repeats

cloned proximal to the origin were more unstable as compared to repeats cloned distal to the same origin. Also, studies conducted using mutants of DNA polymerases (179-183), proteins involved in lagging strand maturation and processing (179,183-187), and single-strand binding protein (188) have revealed a role for all of these proteins in promoting genetic instabilities.

3. Recombination. The formation of double strand breaks (DSBs) is thought to occur when the replication fork is challenged by the non-B DNA structures formed by the repeating tracts. Additionally, the repair of these structures may also result in DSBs or nicks within the repeats (4,6,189). The repair of the breaks occurs by a recombination mediated mechanism. Initial models of recombination based genetic instability were not supported by observations in humans, as recombination was defined as reciprocal crossing-over between sister or homologous chromosomes (45,190-192). However, data from patients implicated a gene-conversion (non-reciprocal) event (137,193). Therefore, DSB repair and recombination via gene conversion is proposed as another mechanism that might generate the genetic instability of trinucleotide disorders.

Recent studies using long CTG•CAG repeat tracts (up to 250 repeats), showed the induction of DSB repair and recombination on the yeast chromosome (165). However, previous work (136,170,179) in yeast model systems were unable to show this effect, and it was thought that the shorter repeat lengths used in these studies may not be efficient in generating DSBs, or that breaks that were indeed induced were repaired by a mechanism other than homologous recombination (HR). Freudenreich *et al.* determined that the CTG•CAG tracts present on the yeast chromosome induced DSBs in a length-

dependent manner (165). Additionally, expansions occurred at a higher frequency during meiosis when compared to similar constructs replicated in mitosis (194-198). The meiosis specific endonuclease SPO11 created DSBs, the repair of which lead to the instability of the CTG•CAG tracts (198). DSBs were also artificially induced in yeast *in vivo* by using the homing endonuclease, *I-SceI* (199). This work implicated long lengths of the CTG•CAG tracts to induce gene conversion mediated expansions. A synthesis-dependent strand annealing (SDSA) pathway, following DSBs, was proposed to explain both the meiosis and mitosis based gene conversion events leading to instability (200).

Studies in *E. coli* have also proposed recombination-mediated repair of DSBs as a plausible mechanism for the generation of large expansions associated with the repeating disorders. A gene conversion mechanism was first proposed to explain the expansions of the CTG•CAG repeats in an intermolecular assay (193). Another set of studies using both intermolecular and intramolecular assays, further showed a stimulation of the frequency of crossing-over between long tracts of CTG•CAG repeats (201,202). Similar work with the GAA•TTC repeats associated with Friedreich's ataxia showed these repeats to be recombination hotspots (161) however, the formation of sticky DNA reduced the hotspot activity of these repeats. Recent work on the CCTG•CAGG tetranucleotide repeats have also implicated a recombination-repair mechanism to generate the observed expansions (R. Dere and R. D. Wells, manuscript in preparation). Additionally, the transformation of break containing plasmids harboring several different lengths of the CTG•CAG repeats into *E. coli* showed a pronounced effect of the break on the stability of the repeats (203). Another genetic assay to

measure recombination rates revealed that mutations in *recA* and *recB* had a stabilizing effect on the CTG•CAG sequences (204). The stabilization was measured as a reduction in the amount of deletions obtained as a result of recombination.

The involvement of recombination-repair in mammalian systems has only recently been investigated. Studies in CHO cells revealed a role for long CTG•CAG tracts to influence recombination between two copies of the APRT gene (97). The rate of gene conversion was 3 – 4 fold lower and the rate of crossing-over was 2 – 3 fold higher when compared to the controls lacking the repeats. In another study, DNA breaks were artificially introduced into CTG•CAG tracts before transfection into COS-1 cells (205). The repair of these breaks yielded deletion products, which were attributed to formation and subsequent repair of secondary structures at the repeats. A DM1 mouse model was used to show the involvement of the single-strand annealing pathway in the CTG•CAG repeat instability in mice (206).

4. Repair. The errors introduced in the DNA, due to the presence of the triplet repeats, can be repaired by a variety of cellular pathways. Expansions of the repeat sequences could occur either due to a lack of repair function or even during the repair process itself. Several proteins from major repair pathways including methyl-directed mismatch repair (MMR) and nucleotide excision repair (NER) have been implicated in triplet repeat stability. Recently, the SbcCD and MRX complexes from *E. coli* and yeast, respectively have also been shown to prevent the expansions of the CTG•CAG tracts (207,208). Both of these complexes have been implicated in hairpin excision both *in vitro* and *in vivo* (209-212).

A number of experiments have been conducted to evaluate the role of MMR in trinucleotide repeat instability. Since the triplet repeats can form non-B DNA structures, especially hairpin loops and slipped-strands, the mismatches in the stem of the hairpins could lead to activation of the MMR pathway. Initial experiments revealed that mutations in the MMR proteins did not influence long tract changes but only caused small expansions and deletions (1 – 2 repeats) (91,213-217). Similar results were also observed in yeast (165,170,218-220). However, another set of experiments with pure tracts of CTG•CAG showed no influence of MMR on genetic instability (214). These inconsistencies were resolved a few years later when the length and sequence purity of the repeat tracts was proposed to affect the role of MMR in instability and the kind of instability (small expansions *vs.* large expansions) obtained (215,216). Studies in transgenic mice have revealed that the small expansions of the CAG•CTG repeats observed in HD mice require the presence of the MSH2 gene (192,221). The expansions were proposed to occur either due to the recognition and binding of the MSH2 to the small loops formed by the repeats thus protecting them from excision or due to the repair of the loops by the MMR pathway. Recently, a deletion of Msh3 but not Msh6 was found to repress somatic expansions in a DM1 knock-in mouse (222). The Msh2/Msh3 complex was proposed to be acting on the triplet repeats in the DM1 mice leading to the somatic instability. Additional support for the involvement of MMR in triplet repeat instability came from studies in which a MutL homologue, Pms2, was shown to be a genetic enhancer of CTG•CAG repeats in somatic mosaicism (223).

Another repair pathway that could be involved in triplet repeat instability is NER. The NER pathway is known to recognize damage that causes distortions in the DNA duplex. Secondary structures formed by the repeating sequences could thus be potent inducers of NER. Parniewski *et al.* showed that mutations in the UvrA protein increased the instability of the CTG•CAG tracts (but not the GAA•TTC and CGG•CCG repeats) compared to strains that carried mutations in UvrB (and an active UvrA) (224). These instabilities were further enhanced in the presence of transcription. In contrast another study showed a stabilization of the repeats in the absence of the UvrA protein (142). Furthermore, UvrA was shown to bind loops of 1, 2 or 17 CAG repeats *in vitro*. Although experiments in bacteria have implicated a role for NER in genetic instability, studies in yeast cells showed that the deletion of Rad1 (involved in NER) did not stimulate any instability (165). Thus, additional work would have to be conducted before the NER pathway can be implicated conclusively in repeat instability.

5. Transcription. Transcription is another factor that has been investigated as a potential mechanism of repeat instability. Bowater *et al.* showed that transcription destabilized the CTG•CAG repeats in *E. coli* (225). The repeats were found to be 5 – 20 fold more stable in the absence of active transcription. Another set of experiments showed that the NER pathway influenced the stability of the DM1 sequence in transcribed plasmids (224). These results indicated the involvement of repair during transcription as a plausible mechanism of genetic instability. Moreover, studies modulating transcription through plasmids harboring the GAC•GTC repeats implicated transcription as a mechanism to generate large deletion events whereas small deletions

and expansions were attributed to replication-slippage (226). An *in vitro* transcription system was used to show a sequence specific pausing of RNA polymerase II through CTG, CAG, CGG and CCG repeats (227). More recently, the effect of transcription on the genetic instability of the tetranucleotide CCTG•CAGG repeats was analyzed. Transcription was once again shown to influence the instability both in COS-7 cells and *E. coli* (R. Dere and R. D. Wells, manuscript in preparation).

F. Non-triplet repeat disorders

The expansions of repeats within the coding or non-coding region of a gene have been associated with several neurological diseases. Most of these disorders are caused by trinucleotide repeat sequences with the exception of three diseases caused by non-triplet sequences including a tetranucleotide, a pentanucleotide and dodecamer repeat sequences. The dodecamer sequence was found to be associated with progressive myoclonus epilepsy 1 (EPM1) (also called Unverricht-Lundborg disease), the pentanucleotide repeat, and the tetranucleotide repeat were associated with SCA10 and DM2, respectively. The dodecamer and pentanucleotide repeat disorders are described in this section whereas, the tetranucleotide repeat disease (DM2) has been described in the following section as it is the repeat sequence that is the focus of this dissertation.

1. Progressive myoclonus epilepsy 1. The Unverricht-Lundborg disease is an autosomal recessive neurological disorder caused by the expansion of a dodecamer sequence (CCCCGCCCCGCG•CGCGGGGCGGGG) in the 5` flanking region of the cystatin B (CSTB) gene (228). Normal alleles contain either 2 or 3 repeats whereas the

alleles of affected patients contain approximately 30 to 80 copies of the repeat (228-230). There does not seem to be any correlation between the age of onset and the severity of the disease (anticipation). The repeat expansion results in a marked decrease in the expression levels of CSTB. Studies have suggested a role for CSTB in the maintenance of normal neuronal structure however the relationship between the loss of CSTB and the EPM1 phenotype is still unclear. The mechanisms of repeat expansions in this disease are believed to be similar to the mechanisms for trinucleotide repeat expansions which include the proteins involved in replication and recombination (40). Additionally, the dodecamer repeat sequence has been shown to form secondary non-B DNA structures (231-234) that have been implicated in repeat instability.

2. Spinocerebellar ataxia type 10 (SCA10). SCA10 is an autosomal dominant disorder characterized by ataxia, and seizures. It is caused by the expansion of a pentanucleotide ATTCT•AGAAT repeat tract in intron 9 of a novel gene termed SCA10 (235). Normal individuals carry 10 – 22 uninterrupted ATTCT•AGAAT repeats and are predominantly heterozygous, whereas patients with SCA10 exhibit a loss of heterozygosity and carry expansions between 800 – 4000 repeats (41,235). Currently, the exact genetic or biochemical functions of the SCA10 gene are unknown however, this protein is expressed at high levels in the brain. The possible mechanism in SCA10 pathogenesis is hypothesized to be either a genomic disruption (alterations in chromatin structure) due to the large size of the repeat tract, an RNA gain-of-function as observed for DM1 and DM2 or simply a SCA10 loss-of-function (41). The structural properties of these pentanucleotide repeats were studied in plasmids and the results indicated that

the repeat containing strands of the DNA remained unpaired, and this unpaired structure functioned as an aberrant origin of replication in HeLa cell extracts (162).

G. Myotonic dystrophy type 2

Myotonic dystrophy type 2 (DM2) is caused by a CCTG•CAGG tetranucleotide repeat expansion in intron 1 of the zinc finger protein 9 (ZNF9) gene located on chromosome 3q (236). The clinical presentation of this neurological disease is strikingly similar to that of myotonic dystrophy type 1 (DM1), including myotonia, proximal weakness, frontal balding, cardiac arrhytmias, insulin-resistance associated diabetes mellitus, polychromatic cataracts and infertility (42,237). Although, the tetranucleotide repeat is dominantly inherited, DM2 patients with two mutant alleles were not reported until recently where, three homozygotes were found in a large consanguineous family from Afghanistan (238).

1. Historical perspective. In 1994, patients showing clinical similarity to DM1 but without a dystrophia myotonica protein kinase (DMPK) gene mutation were described (239,240). This condition was termed proximal myotonic myopathy (PROMM) and showed features of proximal weakness, cataracts and myotonia in the presence of a normal DMPK gene. Furthermore, in 1997, a Finnish family with PROMM and hearing loss, muscular dystrophy and hypogonadism was described and this condition was termed proximal myotonic dystrophy (PDM) (241). In 1998, further evidence for the clinical heterogeneity of non-DM1 myotonic dystrophy was provided to describe a large Minnesota kindred with distal weakness (242). This condition was

termed dystrophia myotonica type 2 (DM2). However, all three conditions of PROMM, PDM and DM2 were finally mapped to the same locus on chromosome 3 a year later (243,244). In 2001, Liquori *et al.* detected an unstable CCTG•CAGG repeat expansion to be responsible for DM2 (236). The CCTG•CAGG repeat tract is part of a complex motif with the overall configuration $(TG•CA)_n(TCTG•CAGA)_n(CCTG•CAGG)_n$. The expanded allele sizes ranged from 75 to 11,000 repeats, averaging at 5000 repeats. The largest normal allele was sequenced and shown to contain 26 repeats with two interruptions. However, the time-dependent somatic instability associated with this disease complicates the correlation of repeat length with the age of disease onset (anticipation) and the estimation of the smallest pathogenic size (236).

2. RNA pathogenesis. The CCTG•CAGG repeat expansion is located in intron 1 of the ZNF9 gene. This gene is highly expressed in the heart and skeletal muscles, the two tissues most affected in DM2 (42). The transcribed ZNF9 mutant RNA accumulates in numerous nuclear foci, similar to the mutant DMPK RNA foci observed in DM1 (236). Recent studies have revealed that the expanded mutations in both DM1 and DM2 exert their deleterious effects at the RNA level. The CUG (in DM1) and CCUG (in DM2) containing transcripts are thought to alter the regulation and localization of certain CUG-binding proteins (72,245-247). The CUG-BP (CUG binding protein) which is over-expressed in the myotonic dystrophies was shown to bind single-stranded CUG *in vitro* (72,248) but failed to co-localize with the ribonuclear foci (249,250). On the other hand, three muscleblind proteins – MBNL, MBLL and MBXL (homologues of the *Drosophila* muscleblind protein) co-localize with the nuclear foci (246,250,251). The

sequestration of these RNA-binding proteins is causative of dysregulation of alternative downstream splicing that has been shown to be characteristic of both DM1 and DM2. It is currently believed that the antagonistic effects of the increase in CUG-BP (which promotes the inclusions of exons normally favored during fetal development) and the depletion of the muscleblind proteins (which normally favor splice forms expressed in adults) affects alternative splicing (246,247). However, studies are still underway to establish a better understanding of the molecular mechanism of this RNA gain-of-function pathway.

CHAPTER II

HAIRPIN STRUCTURE-FORMING PROPENSITY OF THE CCTG•CAGG TETRANUCLEOTIDE REPEATS CONTRIBUTES TO THE GENETIC INSTABILITY ASSOCIATED WITH MYOTONIC DYSTROPHY TYPE 2*

A. Overview

The genetic instabilities of $(CCTG•CAGG)_n$ tetranucleotide repeats were investigated to evaluate the molecular mechanisms responsible for the massive expansions found in myotonic dystrophy type 2 (DM2) patients. DM2 is caused by an expansion of the repeat from the normal allele of 26 to as many as 11,000 repeats. Genetic expansions and deletions were monitored in an African green monkey kidney cell culture system (COS-7 cells) as a function of the length (30, 114 or 200 repeats), orientation, or proximity of the repeat tracts to the origin (SV40) of replication. As found for CTG•CAG repeats related to DM1, the instabilities were greater for the longer tetranucleotide repeat tracts. Also, the expansions and deletions predominated when

---

* Reproduced with permission from "Hairpin Structure-forming Propensity of the CCTG•CAGG Tetranucleotide Repeats Contributes to the Genetic Instability Associated with Myotonic Dystrophy Type 2" by Ruhee Dere, Marek Napierala, Laura P. W. Ranum and Robert D. Wells (2004) Journal of Biological Chemistry, Volume 279, 41715 – 41726. (See Appendix)

cloned in orientation II (CAGG on the leading strand template) rather than I and when cloned proximal rather than distal to the replication origin. Biochemical studies on synthetic $d(CAGG)_{26}$ and $d(CCTG)_{26}$ as models of unpaired regions of the replication fork revealed that $d(CAGG)_{26}$ has a marked propensity to adopt a defined base paired hairpin structure whereas the complementary $d(CCTG)_{26}$ lacks this capacity. The effect of orientation described above differs from all previous results with three triplet repeat sequences (TRS) (including CTG•CAG) which are also involved in the etiologies of other hereditary neurological diseases. However, similar to the TRS, the ability of one of the two strands to form a more stable folded structure, in our case the CAGG strand, explains this unorthodox "reversed" behavior.

B. Introduction

Myotonic dystrophy type 2 (DM2) is a dominantly inherited neurological disease caused by the expansion of a CCTG•CAGG tetranucleotide repeat in intron 1 of the zinc finger protein 9 (ZNF9) gene (236). The characteristics of DM2 are very similar to those observed for DM1 and include myotonia, proximal weakness, frontal balding, cardiac arrhythmias, insulin-resistance associated diabetes mellitus, polychromatic cataracts and infertility (42,252-255). The largest normal allele was found to contain 26 CCTG•CAGG repeats, whereas the repeats were expanded to 75 – 11,000 (average of 5,000 repeats) in patients (236). This is the largest known repeat expansion associated with a disease; also, it is the first tetranucleotide repeat to be implicated in a hereditary neurological disease.

At least 15 hereditary neurological diseases (*i.e.* myotonic dystrophy, Fragile X syndrome and Friedreich's ataxia) are associated with the expansions of $(CTG•CAG)_n$, $(CGG•CCG)_n$ or $(GAA•TTC)_n$ repeat tracts, respectively (4,5,37,38,189,256,257). Additionally, two other non-triplet repeat neurological diseases, spinocerebellar ataxia type 10 (SCA10) (235) and progressive myoclonus epilepsy of Unverricht-Lundborg type (EPM1) (228), are caused by the expansions of pentanucleotide and dodecanucleotide repeating sequences, respectively.

Replication (5,38,128,130,131,137,165,174,176,177), recombination (161,193,201,202,258), and repair (38,165,205,213) were shown to be responsible for the instabilities of triplet repeat sequences (TRS). Slippage of the repeats (259-262) as promoted by non-B DNA structures (5,38,114,257,263,264) formed by these repeating sequences causes polymerases to pause during replication, as shown both *in vivo* as well as *in vitro* (115,117,129,148,174,176,178,265), thereby generating instabilities. Furthermore, these structures are also recognized by mismatch repair (MMR) (91,141,213,214,219,266) and nucleotide excision repair (NER) (142,224); both pathways have been implicated in the stability of the secondary structures, thus influencing the expansion and deletion processes. Also, double-strand breaks (DSB) caused by replication fork arrest or repair of the non-B DNA structures induces repair-mediated recombination which may participate in the expansions observed in both prokaryotic as well as eukaryotic model systems (136,165,196,203,205,267-269). Triplet repeat sequences are hotspots for recombination, which may account for the massive expansions found in certain diseases (161,193,201,202,258,270,271).

Herein, we show that the repeating tetranucleotide genetic instabilities associated with DM2 are caused by a structure-mediated replication-based slippage mechanism. The difference in the propensities of the CAGG and CCTG strands to transiently form quasistable DNA hairpin structures determines the type and level of instability. Furthermore, this instability is influenced by the length, orientation and position of the repeats with respect to the origin of replication.

C. Experimental procedures

Construction of the (CCTG•CAGG)$_n$ Containing Shuttle Vector – The CCTG•CAGG tetranucleotide repeats (236) were obtained from three pCR2.1TOPO derivatives containing either 30, 114 or 200 repeats. All three (CCTG•CAGG)$_n$ inserts (where n = 30, 114 or 200) were excised from their respective pCR2.1TOPO derivative plasmids using either an EcoRV/EagI digest or an EcoRI digest (all enzymes used in this study were purchased from New England Biolabs, Inc.). The CCTG•CAGG inserts are flanked on either side by vector sequences and lack any flanking non-repetitive human DM2 sequences. On excising the inserts using the EcoRV/EagI digest, in addition to the 30, 114 and 200 repeats there are 21 bp of vector flanking sequences on the EcoRV side of the insert and 24 bp of vector flanking sequence on the EagI side. The EcoRV/EagI fragments were filled-in using 1 unit of the Klenow fragment of *E. coli* DNA polymerase I (U.S.Biochemical Corp.), purified on a 5.5% polyacrylamide gel in TAE buffer (40mM Tris acetate, 1mM EDTA, pH 8) and blunt-end ligated into the SmaI site of pCDNA3.1 (Invitrogen). Alternately, the (CCTG•CAGG)$_n$ tracts were excised using an EcoRI

digest, purified on a 5.5% polyacrylamide gel and ligated into the MfeI site of the same shuttle vector. This fragment had 11 bp of pCR2.1TOPO vector flanking sequence on one side and 9 bp of vector flanking sequence on the other side of the repeating tract. The ligations were performed at 16°C for 16 h by the addition of 20 units of T4 DNA ligase (U.S.Biochemical Corp.) in the presence of 1mM ATP followed by transformation into *E. coli* HB101 (New England Biolabs, Inc.) (F^- *Δ(gpt-proA)62, leuB6, glnV44, ara14, galK2, lacY1, Δ(mcrC-mrr), rpsL20 (*Str^r*), xyl5, mtl-1, hsdS20 (*r_B^-*, *m_B^-*), recA13*) and plated on LB plates containing ampicillin (100 μg/ml). Individual colonies were grown in LB broth containing ampicillin (100 μg/ml) for 16 h at 37°C. Plasmids were isolated using the alkaline lysis procedure (Promega, Wizard Plus Miniprep DNA Purification System).

The plasmids carrying the $(CCTG•CAGG)_n$ inserts were characterized using restriction mapping and DNA sequencing. An EcoRI digest was used for repeat tracts cloned proximal (SmaI site) to the SV40 origin of replication and an AflIII/BglII digest was used for inserts cloned into the site distal (MfeI site) to the same origin of replication. The restriction fragments were end-labeled with [α-^{32}P] dATP and 1 unit of the Klenow fragment of *E. coli* DNA polymerase I and analyzed on 5.5% polyacrylamide gels. Furthermore, the plasmids containing the $(CCTG•CAGG)_n$ repeats were dideoxy sequenced on both strands to determine the length, purity and orientation of the repeats with respect to the origin of replication. The sequencing reactions were performed using the Thermo Sequenase Radiolabeled Terminator Cycle Sequencing Kit (U.S.Biochemical Corp.) according to the manufacturer's recommendations. The

products of the sequencing reactions were analyzed on 6% Long Ranger gels (FMC BioProducts) containing 7.5M urea in the glycerol tolerant gel buffer (1.78M Tris, 0.57M taurine and 0.01M EDTA) (U.S.Biochemical Corp.). The $(CCTG•CAGG)_{30}$ is a pure repeat (i.e. it contains no polymorphisms/interruptions) as determined by sequencing of the entire repeat containing tract. The $(CCTG•CAGG)_{114}$ and the $(CCTG•CAGG)_{200}$ carried a single bp interruption 11 repeats into the tract to give the sequence $(CCTG)_{11}CCTT(CCTG)_n$ where n = 102 and 188, respectively. The $(CCTG•CAGG)_n$ (where n = 114 or 200) repeats were sequenced using primers located at both ends of the repeating tract. In case of the $(CCTG•CAGG)_{114}$, sequencing enabled the determination of ~ 200 – 300 bps into the repeats from either side; however, on excising and analyzing the repeat containing fragment from the plasmid, an error of approximately ± 5 repeats was obtained. Similarly, on analyzing the excised $(CCTG•CAGG)_{200}$ fragment on polyacrylamide gels, the estimated repeat length varied by ± 5 repeats. Thus, the repeat lengths were estimated by both sequencing and fragment analyses. The plasmids carried the $(CCTG•CAGG)_n$ repeat tracts in both orientations with respect to the SV40 origin of replication. However, the SV40 origin is bi-directional (272) and is the operative origin in COS-7 cells (273). Orientation I refers to the orientation in which the CCTG repeats are on the leading strand template with respect to the SV40 origin of replication whereas the orientation in which the CCTG repeats are on the lagging strand template with respect to the same origin are referred to as orientation II.

Transfection of the (CCTG•CAGG)$_n$ Repeat Containing Plasmids into COS-7 Cells – The plasmids containing the (CCTG•CAGG)$_n$ tracts were transfected into COS-7 cells. The COS-7 cells were grown in DMEM media (Sigma) containing 10% fetal bovine serum (Gibco) on 10 cm diameter plates. The cells were 75% confluent when transfected with 2 μg of DNA using Lipofectamine 2000 (Invitrogen). The COS-7 cells were allowed to grow for 24 h after transfection before replacing the DMEM medium. The cells were then cultured for either 48 h or 2 weeks after transfection. The COS-7 cells cultured for 2 weeks were split every 36 – 48 h when the cells were almost 100% confluent. Antibiotic selection [geneticin (G418) – 400 μg/ml] (Invitrogen) was applied 48 h after transfection and continued thereafter for the entire 2-week period. The episomal DNA was isolated using alkaline lysis (Promega, Wizard Plus Miniprep DNA Purification System). This episomal DNA was then cleaved with 10 units of DpnI at 37°C for 2 h. DpnI cleaves only DNA which is methylated at its GATC recognition site by the DAM methylase. Thus, treatment of the episomal DNA preparation fragments only the unreplicated DNA which is methylated, leaving the newly replicated but unmethylated or hemimethylated DNA intact. The effectiveness of the DpnI cleavage was assessed by digesting the parental plasmids with DpnI, followed by transformation in *E. coli* HB101 and plating on LB plates containing ampicillin (100 μg/ml). The absence of colonies on LB plates confirmed the complete fragmentation of the plasmids by DpnI. The DNA after the DpnI digestion was then further purified by phenol-chloroform extractions and ethanol precipitation.

Genetic Instabilities by Individual Colony Analyses – The episomal DNA, obtained after transfection in COS-7 cells cultured for either 48 h or 2 weeks, was transformed into *E. coli* HB101 and plated on LB plates containing ampicillin (100 μg/ml). Individual colonies were then picked and grown in LB broth for 16 h at 37°C. The analyses of single colonies enabled the detection of individual events during replication of the plasmids in COS-7 cells. The plasmids were isolated using alkaline lysis and the DNA analyzed by restriction mapping. The $(CCTG•CAGG)_n$ inserts were excised using an EcoRI digest for plasmids carrying the repeats cloned into the SmaI site whereas an AflIII/BglII digest was used to excise the inserts cloned into the MfeI site. The restriction fragments were resolved on 5.5% polyacrylamide gels. The genetic instability was measured as a change in the length of the $(CCTG•CAGG)_n$ repeat containing fragments when compared to the size of the insert excised from the parental plasmid and the 1 Kbp DNA ladder size standard (Invitrogen). The lengths of each insert excised from individual colonies were determined using FluorChem version 3.04 (Alpha Innotech Corp.). Furthermore, ~ 20 random clones containing expansions were sequenced to confirm the lengths and sequences of the repeat containing inserts.

The genetic instability (expansions and deletions) of the $(CCTG•CAGG)_n$ repeats generated in COS-7 cells is calculated by subtracting the background instability of the repeat-containing plasmids in *E. coli* HB101. The genetic instability of the parental plasmids was measured by transforming these plasmids into *E. coli* HB101 and analyzing individual colonies using restriction mapping. The restriction digests used for these analyses were similar to those used for individual colony analyses of the episomal

DNA. The percentage of background instability [average of two experiments for the three tetranucleotide lengths in both orientations (total of six experiments)] varied from 2 to 7 % for repeats cloned proximal (SmaI site) and from 0 to 12 % for repeats cloned distal (MfeI site) (except for $(CCTG•CAGG)_{200}$ in orientation I which was 23 %) to the SV40 origin (data not shown). All statistical analyses were performed using SigmaStat version 2.03.

Substrate Preparation for Chemical Modification and Enzymatic Probing – In order to analyze the structural features of the CCTG•CAGG repeats, we used "single stranded" synthetic oligonucleotides as models of unpaired regions of the duplex repeat sequences for chemical and enzymatic modification studies. The individual oligonucleotides (Genosys), $d(CCTG)_{26}$ and $d(CAGG)_{26}$ were purified on a 6% denaturing polyacrylamide gel containing 7.5M urea in glycerol tolerant gel buffer (U.S.B. Corp.). The purified oligonucleotides were labeled at the 5' end with 15 units of T4 polynucleotide kinase (PNK) (U.S.B. Corp.) and [γ-^{32}P] ATP at 37°C for 1 h. The labeled oligonucleotides were purified on a 6% denaturing polyacrylamide gel. These purified and labeled oligonucleotides were used as substrates for the chemical and enzymatic probing reactions.

Chemical Modifications and Enzymatic Probing – Three chemical probes, osmium tetraoxide (OsO_4) (Aldrich), potassium permanganate ($KMnO_4$) (Fisher) and diethyl pyrocarbonate (DEPC) (Sigma), each were used to modify the $d(CCTG)_{26}$ oligonucleotide whereas the latter two chemicals were used to modify the $d(CAGG)_{26}$ oligonucleotide. The purified and labeled oligonucleotides ($4 – 5 \times 10^5$ cpm/reaction) in

10mM Tris, 40mM NaCl and 10mM $MgCl_2$ were denatured by heating at 80°C for 5 min followed by renaturation by gradually decreasing the temperature (2°C/min) to the indicated reaction temperature (274). The chemical and enzymatic probes were then added along with the carrier DNA (salmon sperm DNA - 1 μg/μl) (Invitrogen).

The DEPC modification of the oligonucleotides was carried out in a buffer containing 50mM sodium cacodylate, 40mM NaCl and 10mM $MgCl_2$ (pH 7.0) at 25°C for 60 min (115,275-280). The $KMnO_4$ reaction was performed at 20°C for 30 min in a buffer containing 50mM sodium cacodylate and 1mM EDTA (pH 7.0) (110,278,281-283). A final concentration of 8% DEPC and 0.1, 0.5 and 1mM $KMnO_4$ was used. The chemical modification with 1mM OsO_4 in the presence of 1mM 2,2'-bipyridine (Sigma) and 150mM sodium phosphate (pH 7.8) was carried out at 5°C for 30 min (277,278,280,284-286). The modified oligonucleotides were then ethanol precipitated and cleaved at the phosphodiester bonds using 1M piperidine (280,284,285,287,288). A series of ethanol precipitations and lyophilization steps then were used to ensure removal of the piperidine.

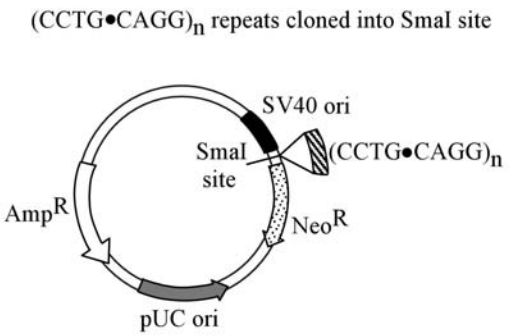
The enzymatic probes used included S1 nuclease (289,290) (Gibco BRL), P1 nuclease (109,110,154,278,291) (Gibco BRL) and mung bean nuclease (140,292,293) (New England Biolabs Inc.). All of the enzymatic probing reactions were carried out at 5°C for 60 min in a buffer containing 10mM Tris, 40mM NaCl and 10mM $MgCl_2$ (pH 7.2). 1mM $ZnCl_2$ was included in the buffer used for the S1 nuclease reactions (287,290). Concentrations ranging from 4.3 – 8.7U of S1 nuclease and 0.25 – 1.0U of mung bean nuclease were used in a 10μl total reaction volume. P1 nuclease was used in

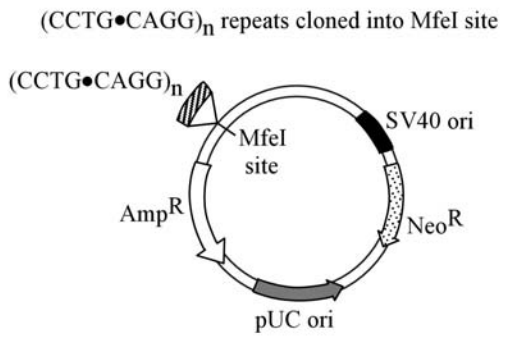
concentrations ranging from 0.05 – 0.25 μg/ml. The reactions were stopped by addition of a urea-EDTA-dye solution (274) followed by quick freezing on dry ice.

The products of the chemical and enzymatic reactions were then analyzed on 10% denaturing polyacrlyamide gels using glycerol tolerant gel buffer. The size marker was prepared using the Maxam-Gilbert sequencing reaction on each of the synthetic deoxyoligonucleotides with DMS (Sigma) followed by piperidine cleavage (115,276,287). The gels were analyzed using a Molecular Dynamics – Storm 820 (Amersham Biosciences) phosphor-imager and quantitative analyses were performed using ImageQuant version 5.1.

D. Results

Strategy of Study – The $(CCTG•CAGG)_n$ repeats were cloned either proximal (SmaI site) or distal (MfeI site) to the SV40 origin of replication in the pcDNA3.1 shuttle vector (Fig. 3). An African green monkey kidney cell culture system was used to determine the genetic instability of the $(CCTG•CAGG)_n$ repeats associated with myotonic dystrophy type 2 (DM2). COS-7 cells were transfected with the tetranucleotide repeat containing plasmids (Fig. 3) and were then cultured for either 48 h or 2 weeks (*i.e.* for approximately two and fourteen cell divisions, respectively) in order to determine the extent of genetic instability after several rounds of replication. The effects of length, orientation, and location of the repeats with respect to the origin of replication were determined.





Plasmid	Number of Repeats (n)	Orientation
pRW5110	30	II
pRW5111	30	I
pRW5112	114	II
pRW5113	114	I
pRW5114	200	II
pRW5115	200	I

Plasmid	Number of Repeats (n)	Orientation
pRW5116	30	II
pRW5117	30	I
pRW5118	114	II
pRW5119	114	I
pRW5120	200	II
pRW5121	200	I

FIG. 3. **Plasmids used in the study.** The $(CCTG•CAGG)_n$ repeats (where n = 30, 114 or 200) were cloned into the SmaI site (proximal to the SV40 origin of replication) or the MfeI site (distal to the SV40 origin of replication) of pcDNA3.1 in both orientations relative to the bidirectional SV40 origin of replication. Orientations I and II are defined under "Experimental Procedures".

To evaluate the effect of the length of the $(CCTG•CAGG)_n$ repeat tract on the genetic instability, repeats of 30, 114 or 200 CCTG•CAGG units were studied. Previous studies showed a marked effect of orientation of trinucleotide repeats (CTG•CAG, CGG•CCG and GAA•TTC) with respect to the origin of replication (38,128-132,134,165). Therefore, we studied plasmids carrying the CCTG•CAGG repeats in both orientations. The distance of the $(CCTG•CAGG)_n$ repeats from the origin of replication may also play an important role in determining the genetic instabilities of these repeats (R. R. Iyer and R. D. Wells, unpublished work) (137). Thus, the repeats were cloned either proximal [~ 74 bp from the center of the SV40 core origin (64 bp in length) (294) at the SmaI site (map position 2078)] or distal [~ 1842 bp from the center of the same core origin of replication at the MfeI site (map position 162)] to the SV40 origin of replication.

Genetic Instability of the $(CCTG•CAGG)_n$ Repeats – The plasmids carrying either 30, 114 or 200 CCTG•CAGG repeats cloned in both orientations were transfected into COS-7 cells (Figs. 3 and 4) and the cells were cultured. The episomal DNA was isolated using alkaline lysis and the DNA was digested with DpnI (see Experimental Procedures). In order to determine the genetic instability of the episomal DNA after 48 h cultures, Southern blot analyses were performed with the $(CCTG)_8$ oligonucleotide as a probe (data not shown), which confirmed the presence of the CCTG•CAGG repeats in the excised fragments. Since, no instability was observed, the more sensitive single colony analyses were performed.

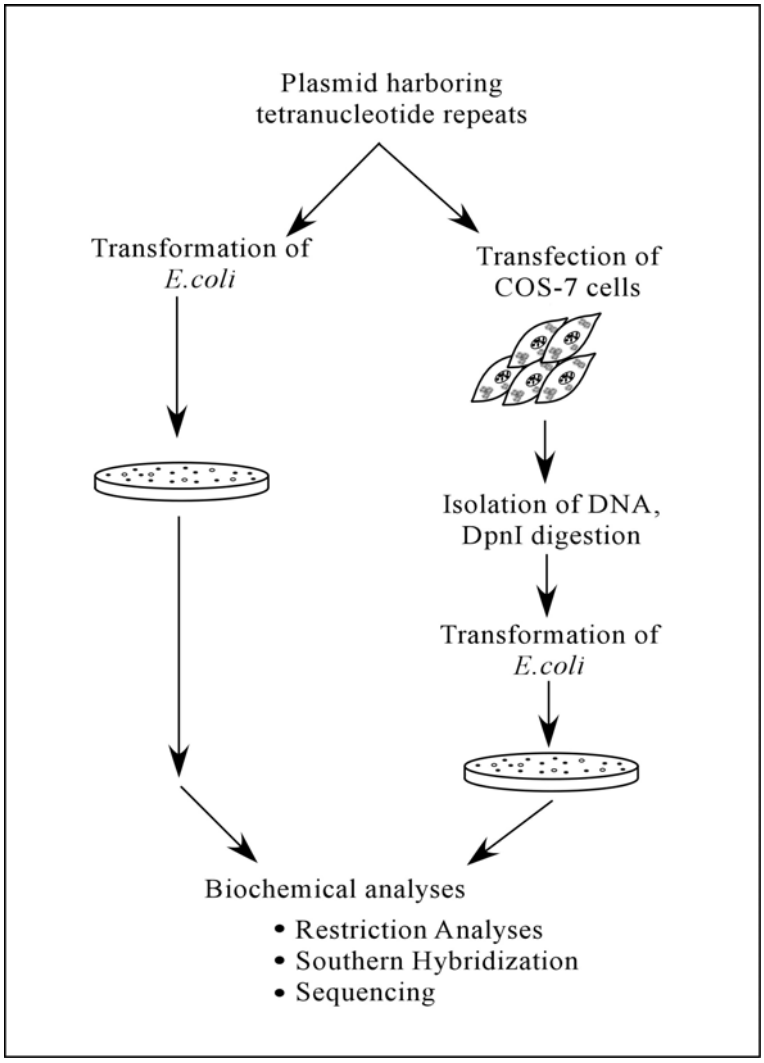


FIG. 4. **Experimental strategy using the mammalian cell culture assay.** The $(CCTG•CAGG)_n$ tracts (where n = 30, 114 or 200) were cloned either proximal to the origin of replication in the SmaI site (map position 2078) or distal to the origin of replication in the MfeI site (map position 162) (Fig. 3). These plasmids were then transfected into COS-7 cells, cultured for both 48 h and 2 weeks and the episomal DNA isolated using the alkaline lysis method. The episomal DNA was digested with DpnI to fragment the unreplicated DNA. The episomal DNA was transformed into *E. coli* HB101 and individual colonies were analyzed using biochemical analyses. Simultaneously, plasmids that were not replicated in COS-7 cells were also transformed into *E. coli* HB101 and individual colonies were subjected to similar biochemical analyses (see Experimental Procedures).

The episomal DNA after treatment with DpnI from 48 h and 2 week cultures was transformed into *E. coli* HB101 and DNA from individual colonies was analyzed by restriction digestion and the products of cleavage were analyzed in high-resolution 5.5% polyacrylamide gels (Fig. 5). These analyses on individual colonies enable the detection of relatively small amounts of expansions and deletions which were not detected (see above) by the population studies by Southern blots. The instability observed (Fig. 4, right side) is the composite instability of the $(CCTG•CAGG)_n$ repeats in the eukaryotic and prokaryotic systems. In order to obtain the percentage of instability (expansions and deletions) in the mammalian system alone, the percentage of instability obtained from the prokaryotic cells was subtracted as the background (left side of Fig. 4). Thus, the percentages of both expansions and deletions as depicted in the bar graphs in Fig. 6 represent the net genetic instability of the $(CCTG•CAGG)_n$ tracts in the eukaryotic system.

Analyses of the digestion products from individual colonies by restriction digestion and polyacrylamide gels revealed four different types of products (Fig. 5): the unchanged starting length tetranucleotide repeats, expansions, deletions, and "rearrangements" which showed an alternate digestion pattern from that observed for full length, expansions and deletions. In some cases, the rearrangements showed a digestion pattern in which the vector backbone was altered such that an expected fragment was missing (*e.g.* Fig. 5B, lane 5). Previous studies (295-297) have shown these products to result from illegitimate recombination in vectors carrying the SV40 origin of replication. To confirm that the rearrangements arose from an inherent

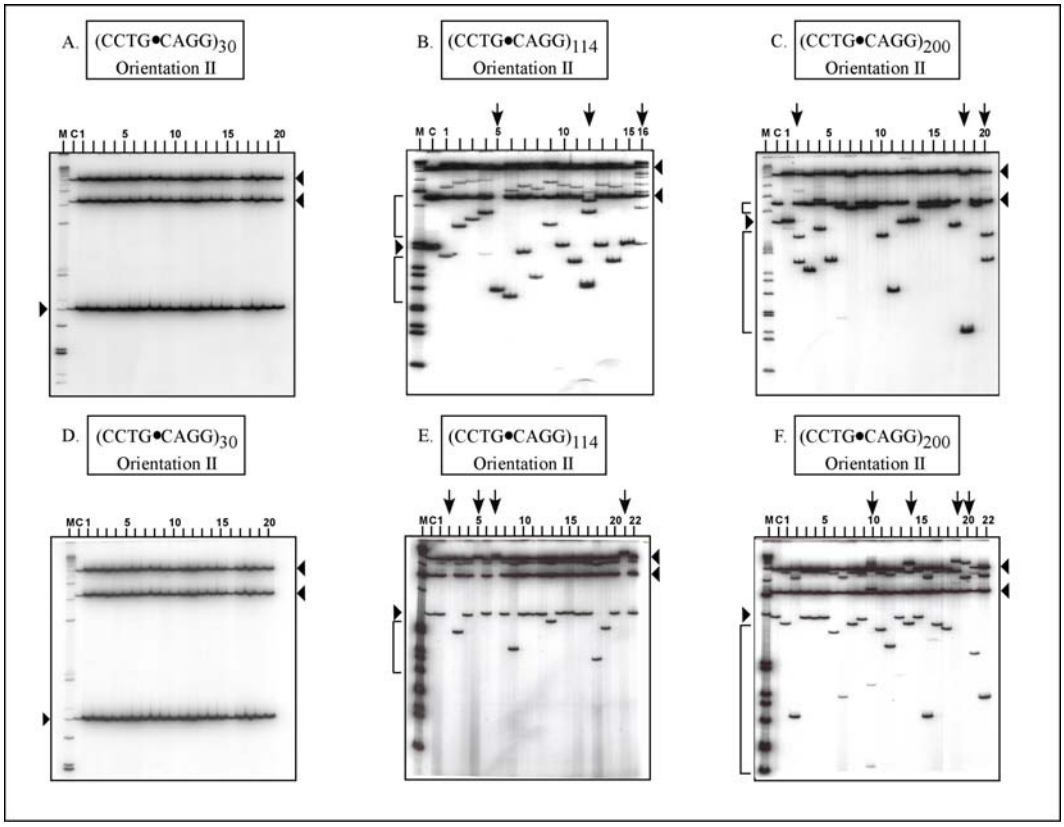


FIG. 5. **Genetic instability of the (CCTG•CAGG)$_n$ repeats.** The episomal DNA obtained from COS-7 cells cultured for 2 weeks was digested to excise the tetranucleotide repeats and the products were analyzed on 5.5% polyacrylamide gels. A-C, representative gels showing the various products of genetic instability of the (CCTG•CAGG)$_n$ repeats (A. pRW5110 B. pRW5112 C. pRW5114) cloned proximal to the SV40 origin of replication (SmaI site). The repeats were excised using an EcoRI digestion. D-F, representative gels showing the products of genetic instability of the tetranucleotide repeats (D. pRW5116 E. pRW5118 F. pRW5120) cloned distal to the SV40 origin of replication (MfeI site). The inserts were excised using an AflIII/BglII digest. Lane M, 1 Kbp ladder, lane C, (CCTG•CAGG)$_n$ repeats excised from the parental plasmids used as controls. The numbers above each gel indicate the various clones used. The arrowhead on the left side of each gel indicates the size of the full-length progenitor fragment whereas the brackets above and below the arrowheads indicate the expansions and deletions, respectively. The clones that contained 'rearrangements' (see Results) are designated by arrows above each of the gels. The arrowheads on the right side of each gel indicate the two fragments that result from the digestion of the vector and the faint bands between these two fragments result from an incomplete digestion, as seen in panels B, E and F.

property of the vector rather than a result of the presence of the $(CCTG•CAGG)_n$ inserts, the pcDNA3.1 vector (lacking the $(CCTG•CAGG)_n$ repeats) was transfected into COS-7 cells. The cells were cultured for 2 weeks and the episomal DNA isolated, DpnI cleaved and transformed into *E. coli* HB101. The DNA from individual colonies was analyzed on a 1% agarose gel. The number of rearrangements observed were similar to those obtained for the plasmids carrying the $(CCTG•CAGG)_n$ repeats. Furthermore, no rearrangements were observed on analyzing individual colonies obtained from transforming the parental plasmids into *E. coli* HB101. Likewise, rearrangements were not observed previously in other prior instability studies with triplet repeat inserts in *E. coli* HB101 (38,81,128-131,134,298,299). This confirms that the rearrangements were due to an intrinsic property of the pcDNA3.1 shuttle vector that carries the SV40 origin of replication.

Length-dependent Instability of the $(CCTG•CAGG)_n$ Tetranucleotide Repeats – For CTG•CAG, CGG•CCG, and GAA•TTC trinucleotide repeats, the number of repeat units plays an important role in genetic instability (38,81,129,130,165,300). In order to evaluate the effect of the length of the $(CCTG•CAGG)_n$ tetranucleotide repeats (where n = 30, 114 or 200) on the genetic instability, three different lengths of the repeats were studied. The percentage of expansions and deletions found for each of the $(CCTG•CAGG)_n$ repeats in COS-7 cells for a 2 week period is indicated in Fig. 6. In general, the longer the length of the CCTG•CAGG tract, the greater the observed instability. The $(CCTG•CAGG)_{30}$ repeats were extremely stable (lack of any genetic instability) when the plasmids harboring these repeats were cultured in COS-7 cells for

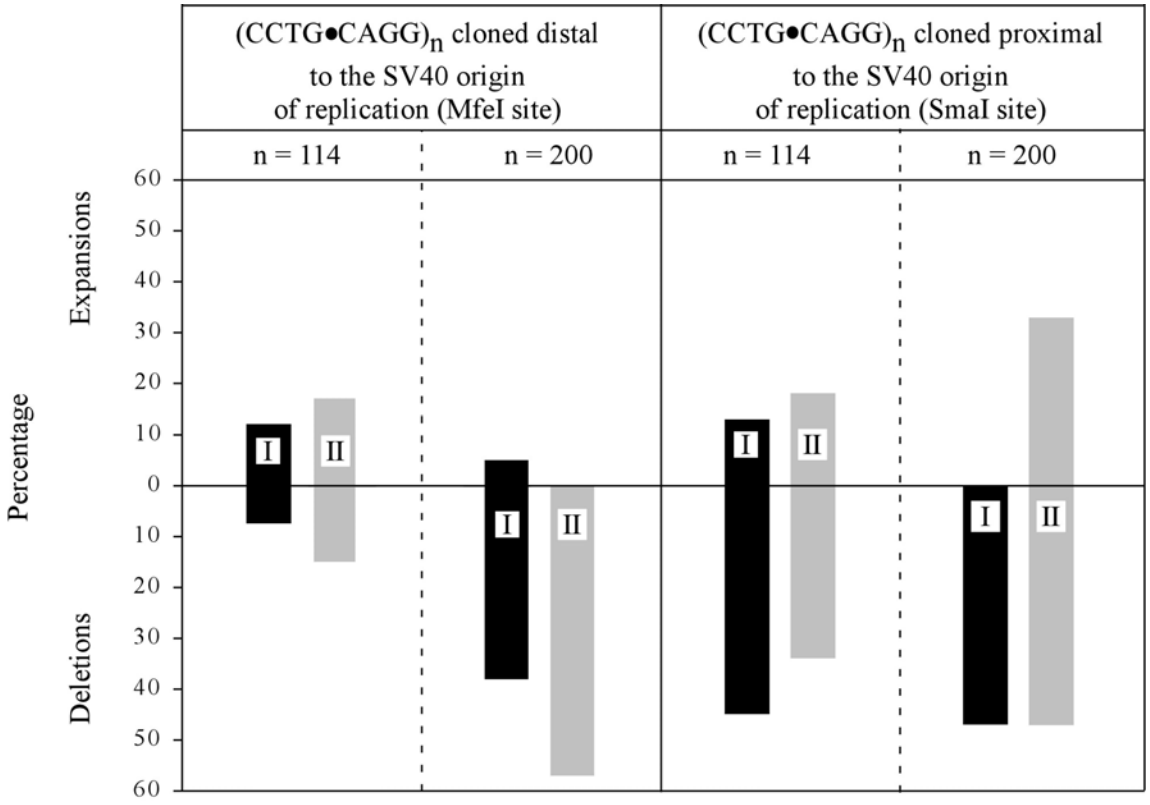


FIG. 6. **Relative amounts of expansions and deletions of $(CCTG•CAGG)_n$ repeats.** The episomal DNA was isolated from COS-7 cells, which were cultured for 2 weeks, and the lengths of the tetranucleotide repeats were determined as described (Experimental Procedures). The percentages of expansions and deletions of the tetranucleotide repeats were calculated by subtracting the background percentages obtained in *E. coli* HB101. The expansion and deletion percentages of plasmids carrying the $(CCTG•CAGG)_n$ tracts (where n = 114 and 200) proximal (SmaI site) and distal (MfeI site) to the SV40 origin of replication are indicated as bars above and below the line marked zero, respectively. The black bars indicate the percentage of expansions and deletions for inserts in orientation I whereas the grey bars indicate the percentage of expansions and deletions for inserts in orientation II. All data are the average of duplicate experiments with the appropriate background subtracted for each experiment.

both 48 h and 2 weeks (Fig. 5A and D). A statistically significant difference in the genetic instabilities was observed between pRW5114 [$(CCTG•CAGG)_{200}$ – orientation II] and pRW5112 [$(CCTG•CAGG)_{114}$ – orientation II] ($p = 0.003$). Also, similar statistically significant differences in the genetic instability of these repeats were observed on comparing the two pairs of plasmids pRW5121 and pRW5119 ($p = 0.012$) and pRW5120 and pRW5118 ($p = 0.010$). However, in the case of the plasmids carrying $(CCTG•CAGG)_{200}$ cloned in orientation I proximal (SmaI site) to the SV40 origin of replication, a statistically significant length-dependant effect was not observed.

Furthermore, the increase in genetic instability with an increase in the length of the $(CCTG•CAGG)_n$ tract was observed for repeats cloned both proximal (SmaI site) and distal (MfeI site) to the SV40 origin of replication. Thus, there is a positive correlation between the length of the $(CCTG•CAGG)_n$ tracts and their genetic instabilities.

Effect of Orientation – Prior genetic instability studies on microsatellites revealed the inequality of leading versus lagging strand DNA replication (i.e. the direction of replication) with respect to the types and amount of products (38,128,130,132,133,135-137). To determine if a similar orientation effect was observed for the tetranucleotide repeats, $(CCTG•CAGG)_n$ repeats were cloned in both orientations with respect to the bi-directional SV40 origin of replication (see Experimental Procedures). The percentage of instability (both expansions and deletions) obtained from plasmids cultured in COS-7 cells for 2 weeks (Fig. 6) showed that orientation II was more unstable than orientation I, especially when the inserts were cloned in the SmaI site. This effect was most pronounced for plasmids carrying the longest repeats. Thus, a statistically significant

difference in instability was observed when comparing pRW5115 and pRW5114 ($p = <0.001$). In experiments where the plasmids were cultured in COS-7 cells for 48 h, a similar effect was observed (data not shown) for $(CCTG•CAGG)_{200}$. Thus, once the tetranucleotide repeats reached a threshold of about 200 repeats, an orientation effect, where orientation II was significantly more unstable than orientation I, was observed. This was true of repeats cloned proximal (SmaI site) to the SV40 origin of replication. Alternatively, for repeats cloned distal (MfeI site) to the SV40 origin, the differences were not statistically significant, (Fig. 6), even for 48 h cultures (data not shown).

Furthermore, on analyzing the types of genetic instabilities *i.e.* expansions and deletions, it was observed that expansions in most cases predominated in orientation II compared to orientation I (Fig. 6); this behavior was especially pronounced for $(CCTG•CAGG)_{200}$ when cloned into the SmaI site. A 1.5 to at least 35 fold increase in expansions in orientation II compared to orientation I was observed which was dependant on the length of the $(CCTG•CAGG)_n$ repeats. Analyses of the instability products from plasmids cultured in COS-7 cells for 48 h also showed a similar trend where expansions predominated in orientation II, especially in the case of the repeats cloned proximal (SmaI site) to the SV40 origin of replication (data not shown). Deletions were present in both orientations. The percentage of deletions did in most cases increase as a function of length; however, an orientation effect was not observed.

Hence, the genetic instability was greater in the case of the repeats cloned in orientation II than in I and when the repeats were cloned proximal (SmaI site) rather than distal (MfeI site) to the SV40 origin. Furthermore, as the length of the repeats increased,

the orientation effect was more pronounced. Expansions predominated in orientation II compared to I, especially at the SmaI site.

Magnitude of Expansions and Deletions – The ranges of expansions and deletions of the $(CCTG•CAGG)_n$ repeats cloned proximal (SmaI site) and distal (MfeI site) to the SV40 origin of replication is shown in Table I. For the $(CCTG•CAGG)_n$ repeats cloned proximal (SmaI site) to the SV40 origin of replication (Table IA), the largest expansion product obtained from $(CCTG•CAGG)_{114}$ (pRW5112) was a $(CCTG•CAGG)_{210}$, showing a 1.84 fold increase in the number of repeats. pRW5114 expanded from 200 repeats to a maximum of 300 repeats, indicating a 1.5 fold increase in length. Both of these expansions were obtained for the repeats cloned in orientation II (Table IA). The deletions obtained for the $(CCTG•CAGG)_n$ repeats cloned proximal to the SV40 origin of replication ranged from a complete loss of the repeats to a reduction of 2 or 3 repeats (Table I). Plasmids harboring the shortest repeats (n = 30) cloned into either of the sites were completely stable in all experiments.

The largest percent expansion obtained for the $(CCTG•CAGG)_n$ repeats cloned distal (MfeI site) to the SV40 origin (Table IB) was 56 repeats or 1.5 fold for pRW5119. The largest increase for $(CCTG•CAGG)_{200}$ was 14 repeats. The deletions ranged from a complete loss of repeats to a loss of only 2 or 3 repeats similar to those observed for repeats cloned proximal to the SV40 origin (Table IA). Approximately 20 clones containing expansions were chosen at random and the inserts sequenced; all inserts contained the CCTG•CAGG repeat tracts within the limits of our detection.

TABLE I

Expansions and deletions of (CCTG•CAGG)$_n$ repeats

The episomal DNA was isolated from COS-7 cells cultured for 2 weeks and the sizes of the inserts were analyzed by restriction digestion and polyacrylamide gel electrophoresis (Experimental Procedures). The ranges of expansions and deletions of the (CCTG•CAGG)$_n$ repeats are shown; in general, the distribution of product sizes within these ranges was random. The bracketed values indicate the average of the observed values. pRW5110 and pRW5116 were extremely stable after 2 weeks; hence, the data for these DNAs were derived from 48 h culture studies.

Plasmid	Number of repeats	Orientation	Expansions	Deletions
A. Inserts cloned proximal to the SV40 origin of replication (SmaI site)				
pRW5115	200	I	284 (284)	0-197 (90)
pRW5114	200	II	202-300 (246)	29-197 (130)
pRW5113	114	I	117-190 (148)	13-112 (85)
pRW5112	114	II	116-210 (158)	0-112 (82)
pRW5111	30	I	0	0
pRW5110	30	II	0	0
B. Inserts cloned distal to the SV40 origin of replication (MfeI site)				
pRW5121	200	I	206-214 (210)	71-197 (139)
pRW5120	200	II	200 (200)	9-198 (130)
pRW5119	114	I	116-170 (133)	0-111 (85)
pRW5118	114	II	116-124 (119)	16-108 (63)
pRW5117	30	I	0	0
pRW5116	30	II	0	0

The plasmids cultured in COS-7 cells for a period of 48 h (data not shown) showed expansions ranging between 118 to 144 repeats for $(CCTG•CAGG)_{114}$ and between 202 to 370 repeats (up to1.85 fold) for $(CCTG•CAGG)_{200}$. The deletions were also within a similar range as observed for the plasmids cultured for 2 weeks. This range of expansions and deletions was similar for repeats cloned both proximal (SmaI site) and distal (MfeI site) to the SV40 origin.

In summary, a range of expansions and deletions were found. The larger expanded products were found for the repeats cloned in orientation II compared to I. Furthermore, the magnitude of expansions was much larger for repeats cloned proximal to the origin than for repeats cloned distal.

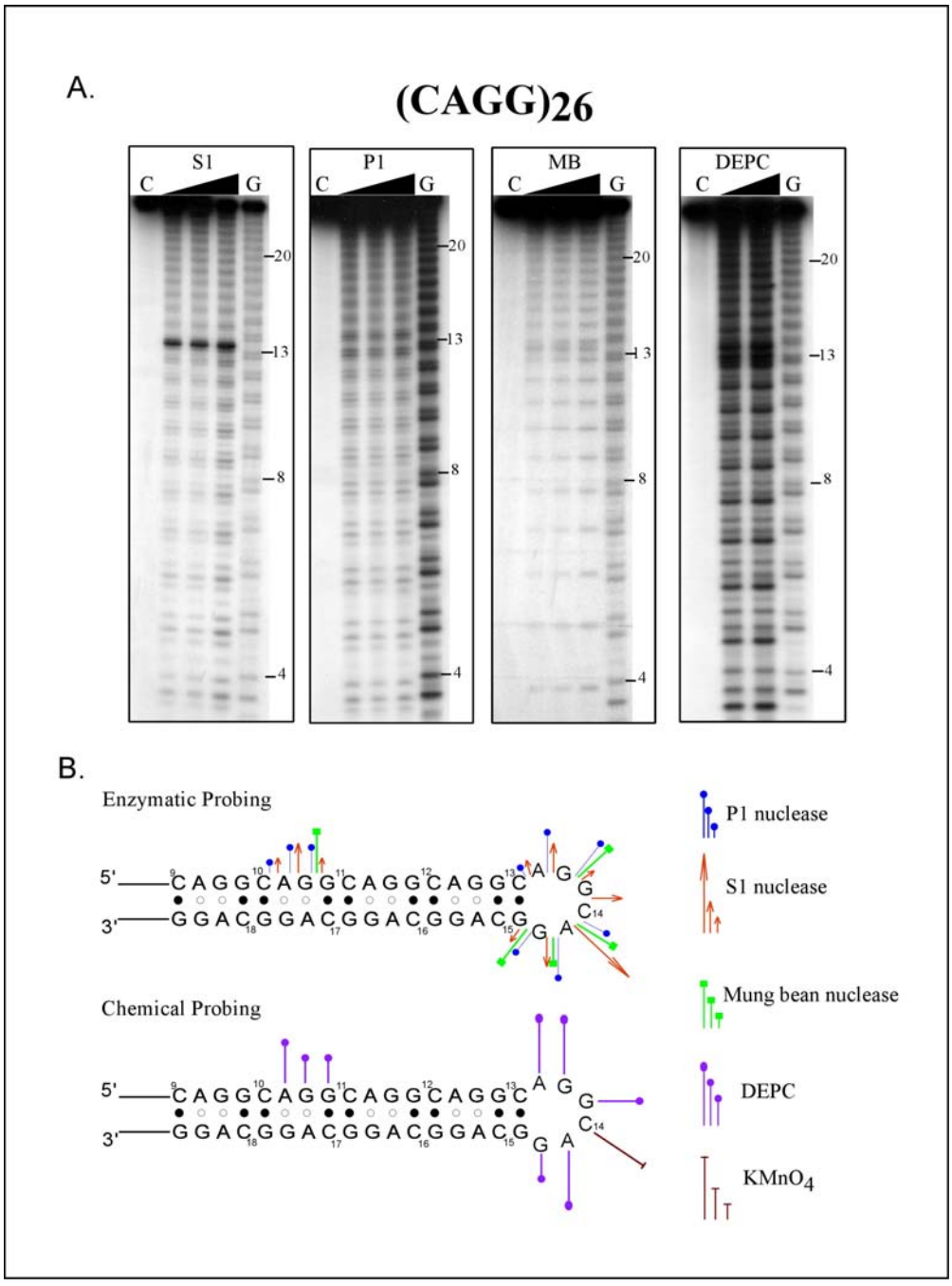
Oligonucleotide Model Studies: Enzymatic Probing – Two oligonucleotides, $d(CAGG)_{26}$ and $d(CCTG)_{26}$, were chemically synthesized to study their structural properties as related to the behavior of unpaired regions of the CCTG•CAGG repeats during replication and related processes that unwind the duplex. $d(CAGG)_{26}$ and $d(CCTG)_{26}$ were purified and labeled (Experimental Procedures). The labeled oligomers were then probed with conformation-sensitive enzymatic agents (S1 nuclease, P1 nuclease and mung bean nuclease) (110,140,154,278,289-292,301). All three enzymes have been used widely to evaluate ordered and H-bonded pseudo-duplex DNA regions within DNA tracts which are otherwise disordered, random-coil structures (109,110,278,290,301).

S1 nuclease from *Aspergillus oryzae*, a zinc requiring enzyme with an acidic pH optimum, cleaves single stranded and partially unpaired DNA and is not base-specific

(289,290,302). On probing $d(CAGG)_{26}$ with S1 nuclease (Fig. 7), the greatest reactivity was observed between the cytosine and adenine residues of the 14th CAGG repeat, followed by weaker cleavage between the adenine and guanine of the 14th CAGG repeat, as well as cleavage between the first and second guanine residues of the same repeat. Thus, these residues are more accessible to the enzymatic probe and may form a terminal loop of a hairpin structure. The stem of the hairpin probed with S1 nuclease showed cleavage of the phosphodiester bonds between residues CpA, ApG and GpG (Fig. 7) but the intensity of cleavage in the proposed stem is weaker than that found in the putative loop (Fig. 7A). Alternatively, on probing the labeled $d(CCTG)_{26}$ (Fig. 8), an equal intensity of cleavage was observed between the CpT and TpG residues, suggesting the lack of formation of a stable secondary structure.

P1 nuclease from *Penicillium citrium* is similar in its behavior to S1 nuclease in preferentially cleaving single-stranded, non-helical DNA tracts or regions that are transiently unpaired (109,110,154,278,291,302). However, P1 nuclease has a neutral pH optimum and lacks a requirement for zinc. Probing the $d(CAGG)_{26}$ substrate with P1 nuclease further supported the evidence for formation of a terminal loop as demonstrated by the hypersensitivity of the substrate to this enzyme between the residues ranging from the adenine to the guanine residues of the 13th CAGG repeat and between the adenine and guanine of the 14th CAGG repeat. Furthermore, cleavage was also observed between the cytosine and adenine residues of the 14th CAGG repeat (Fig. 7). Cleavage occurred between ApG, GpG and CpA with the strongest cleavage between the ApG and GpG residues in the stem (as illustrated for repeat 10 but also found for all other repeats

FIG. 7. **Analyses of the products of chemical and enzymatic probes on d(CAGG)$_{26}$.** The preparation and characterization of the labeled synthetic oligonucleotide (CAGG)$_{26}$ is described in Experimental Procedures. A, analyses of the probing for oligonucleotide secondary structure on 10% polyacrylamide gels with 7.5M urea. Data obtained from S1 nuclease (S1), P1 nuclease (P1), mung bean nuclease (MB) and diethylpyrocarbonate (DEPC) are shown. The numbers to the right of each of the panels indicate the number of tetranucleotide repeats (as designated on the second G residue of the CAGG repeats). Increasing concentrations of enzymatic probes (4.3, 6.5 and 8.7U of S1 nuclease, 0.15, 0.2 and 0.25 μg/ml of P1 nuclease and 0.25, 0.5 and 1.0U of mung bean nuclease) are indicated by a triangle above each of the panels. The triangle above the DEPC panel indicates an 8% DEPC solution used over two increasing time points of 30 min and 60 min. C - control (no probe added), G - size marker [Maxam-Gilbert sequencing with dimethyl sulfate (DMS)]. B, possible secondary structures formed by d(CAGG)$_{26}$. The reactivity of the oligonucleotide to the various probes has only been shown for the 10th CAGG repeat of the stem of the proposed hairpin loop structure for clarity; however, the same cleavage pattern was seen for all of the other CAGG repeats in the stem of the hairpin loop. Furthermore, the cleavage of the phosphodiester bonds in the terminal loop of the hairpin has also been indicated. The symbols corresponding to the various probes are shown to the right of the figure. The length of the symbols corresponds to the cleavage intensity as quantitated using ImageQuant version 5.1. The filled circles between the two DNA strands indicate Watson Crick pairing and the open circles denote non-Watson Crick pairing.


A.
(CAGG)26
S1
P1
MB
DEPC
C
G
20
13
8
4
B.
Enzymatic Probing
Chemical Probing
5'
3'
P1 nuclease
S1 nuclease
Mung bean nuclease
DEPC
KMnO4

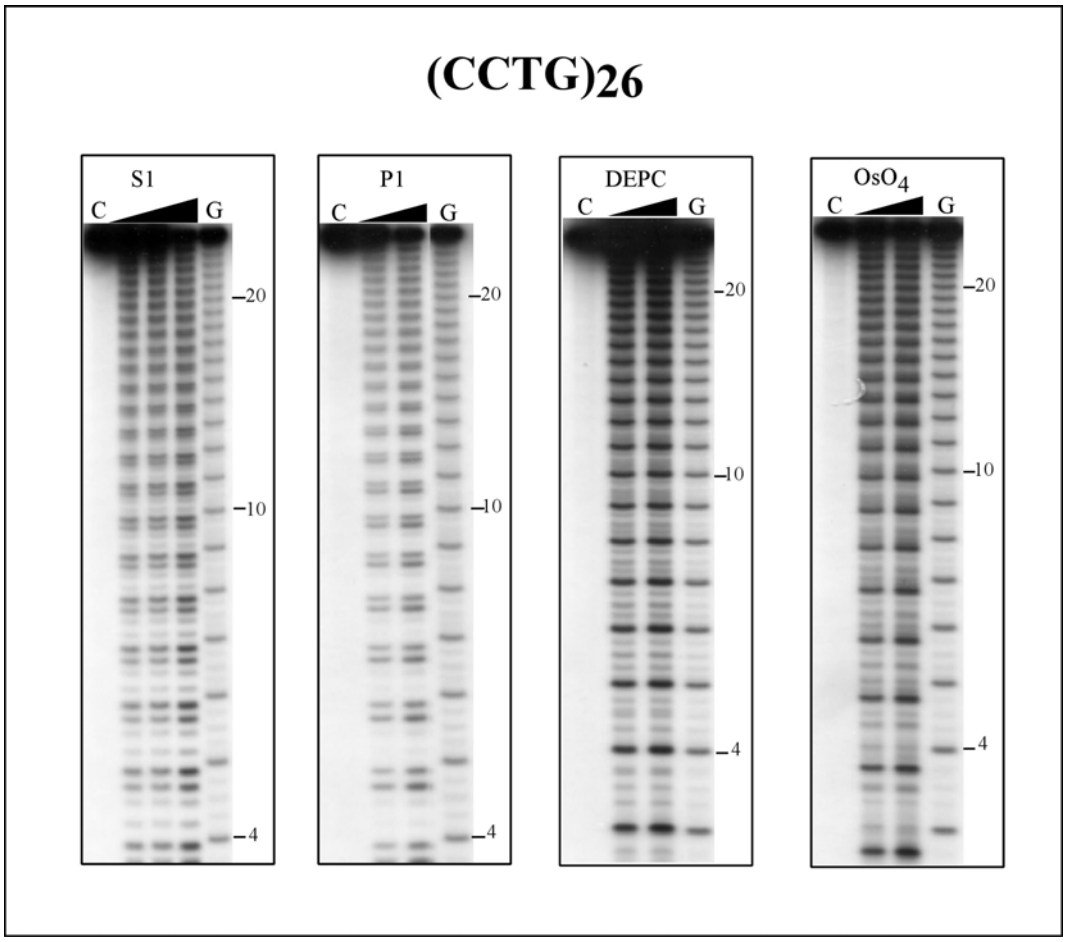


FIG. 8. **Analyses of the products of chemical and enzymatic probes on d(CCTG)$_{26}$.** The results of the probing for oligonucleotide secondary structure in d(CCTG)$_{26}$ using S1 nuclease (S1), P1 nuclease (P1), diethylpyrocarbonate (DEPC) and osmium tetraoxide (OsO_4) are shown. Analyses were performed in 10% polyacrylamide gels with 7.5M urea. Increasing concentrations of the enzymatic probes, (4.3, 6.5 and 8.7U of S1 nuclease, 0.05, 0.1 and 0.15 μg/ml of P1 nuclease) are indicated as triangles above each panel. The triangle above the panel for DEPC indicates two increasing time points of 30 and 60 min using an 8% DEPC solution. The triangle above the panel for OsO_4 indicates two increasing time points of 15 and 30 min using 1mM OsO_4 in the presence of 1mM 2,2'-bipyridine. The numbers to the right denote the number of tetranucleotide repeats (as designated on the second G residue of the CCTG repeats). C – control (no probe added), G – size marker [Maxam-Gilbert sequencing using DMS].

in the stem) (Fig. 7). $d(CCTG)_{26}$ showed a similar cleavage pattern to that observed with S1 nuclease when probed with P1 nuclease, where the phosphodiester bonds between the CpT and TpG residues were cleaved with an equal intensity (Fig. 8), again indicating that the $d(CCTG)_{26}$ oligonucleotide did not form a stable secondary structure.

Mung bean nuclease is also highly sensitive to variations in DNA structure and converts single stranded or unpaired DNA to mono- or oligonucleotides with 5′-phosphates (140,292,293,302). The probing of $d(CAGG)_{26}$ with mung bean nuclease showed predominant cleavage between the first and second guanines of the 13th CAGG repeat, the cytosine and adenine, as well as the first and second guanine residues of the 14th CAGG repeat. Less cleavage was observed between the adenine and guanine residues of the 14th CAGG repeat (Fig. 7). These residues form the proposed terminal hairpin loop. The bonds between the G's of each of the CAGG repeats (Fig. 7) were more resistant to cleavage than for the bonds in the putative loops; thus, we propose that they exist in the stem. Alternatively, the probing of $d(CCTG)_{26}$ with mung bean showed cleavage of the bonds between GpC and CpC of each of the CCTG repeats (data not shown), thus indicating the lack of a stable secondary structure in agreement with the results from S1 and P1 nucleases.

Hence, the enzymatic probing studies revealed the preferential formation of a hairpin structure by $d(CAGG)_{26}$ as opposed to $d(CCTG)_{26}$.

Chemical Probe Determinations – $d(CAGG)_{26}$ and $d(CCTG)_{26}$ were individually probed with OsO_4 or DEPC or $KMnO_4$. OsO_4 specifically reacts with thymines, and to a much lesser extent with cytosines, in single stranded oligonucleotides and single

stranded or distorted double stranded regions in DNA to yield mainly cis-thymine glycols (129,277,278,284-286). OsO_4/piperidine was used to probe $d(CCTG)_{26}$ (Fig. 8). An equal intensity of cleavage was observed at each of the thymine residues of the CCTG repeats. This indicates the equal accessibility of the pyrimidines and hence the oligonucleotide did not form any preferential secondary structure. This probe was not used for the complementary oligomer since it lacked thymine residues.

DEPC reacts with the N7 positions of the purines, both adenines and guanines in single stranded DNA, to form the corresponding ring-opened dicarbethoxylated derivatives (115,129,275-277,279,280). DEPC/piperidine was used to probe both $d(CAGG)_{26}$ and $d(CCTG)_{26}$. Similar to the enzymatic probing, the most effective modification occurred at the adenine and the first guanine residues of the 13^{th} CAGG repeat and at the adenine of the 14^{th} CAGG repeat (Fig. 7) indicating the formation of a terminal loop of the putative hairpin structure. The putative stem showed cleavage of the adenine residues to a greater extent than the guanine residues (Fig. 7) as previously stated. In the case of $d(CCTG)_{26}$, DEPC modified the G's of each of the CCTG repeats and an equal intensity of cleavage was observed at all of the G's (Fig. 8).

$KMnO_4$ preferentially oxidizes unpaired or unstacked thymines (and to a much lesser extent cytosines) resulting in strand cleavage upon subsequent treatment with piperidine (110,278,281,282,288). $KMnO_4$ showed reactivity at all residues in the oligomer but a very prominent cleavage was observed at the cytosine residue in the 14^{th} CAGG repeat of $d(CAGG)_{26}$ (data not shown). However, in the case of $d(CCTG)_{26}$,

KMnO4 equally modified the thymine residues of each CCTG repeat (data not shown) indicating again the lack of any specific stable secondary structure.

Thus, the chemical modification studies further confirmed the results obtained from the enzymatic probing studies which indicate that $d(CAGG)_{26}$, in contrast to $d(CCTG)_{26}$, preferentially forms a stable hairpin structure.

E. Discussion

DM2 is caused by the expansion of a CCTG•CAGG tetranucleotide repeat (236) from a normal range of ~ 26 to ~ 11,000 repeats. Using an African green monkey kidney cell system, we demonstrate that replication-based slippage contributes to the genetic instability of these repeats. The instability was dependent on the length of the repeats, their orientation, and their distance from the replication origin. Furthermore, synthetic oligonucleotides representing the unpaired repeat regions during replication were analyzed with chemical and enzymatic probes revealing the preferential formation of hairpin structures by the CAGG oligomer, further supporting our structure-mediated replication-slippage model.

Prior genetic instability studies on CTG•CAG, CGG•CCG and GAA•TTC triplet repeats (5,38,81,128-130,134,139,165,174,256,257,298,299) have shown a length-dependant effect on instability. Similarly, the longest CCTG•CAGG repeat containing plasmids were the most unstable, especially after 2 weeks of culturing in COS-7 cells, generating both expansions and deletions when cloned proximal to the SV40 origin. However, when the repeats were cloned distal to the SV40 origin, deletions were

predominant. Thus, for longer repeating sequences, an increase in the genetic instability was observed, as expected.

To study the propensity of the CCTG•CAGG repeats to form non-B DNA structures, we analyzed the synthetic oligonucleotides $d(CAGG)_{26}$ and $d(CCTG)_{26}$ with chemical and enzymatic probes. The modification patterns obtained for $d(CAGG)_{26}$ showed the formation of a folded back, hairpin structure. Several different fold-back structures are possible (Fig. 9). $d(CAGG)_{26}$ can form hairpin structures with either 6, 5, 4 or 3 unpaired residues in the terminal loop (Fig. 9A, B, C and D, respectively). This would result in different Watson-Crick and non-Watson-Crick pairing schemes in the putative stems. For the species shown in Fig. 9B and C, there would be no Watson-Crick pairing making them less stable than the structures shown in Fig. 9A and D. However, the structure formed with 6 unpaired residues in the terminal loop (Fig. 9A) is more stable than that with 3 residues in the loop (Fig. 9D) due to the two non-Watson-Crick A•G pairs that flank the Watson-Crick G•C pairs (Fig. 9A) compared to the A•A and G•G oppositions that flank the G•C pairs in structure 9D. Although the stability of the mispairs is context dependent, an approximation of the pairing stabilities is G•C > A•T > G•G > G•T ≈ G•A > A•C^+ > T•T ≈ A•A ≈ C•C^+ > T•C ≥ A•C ≥ C•C (121-127). Previous studies have also indicated the greater stability of the loop of a hairpin with an even number of residues compared to an odd-numbered loop (303,304) as well as the favored closing of the loop with a 5′ pyrimidine and a 3′ purine pair (303,304). Thus, our chemical and enzymatic probe results showed the formation of the structure (Fig. 9A) which the above stated rationale predicts to be most stable. Furthermore, DEPC

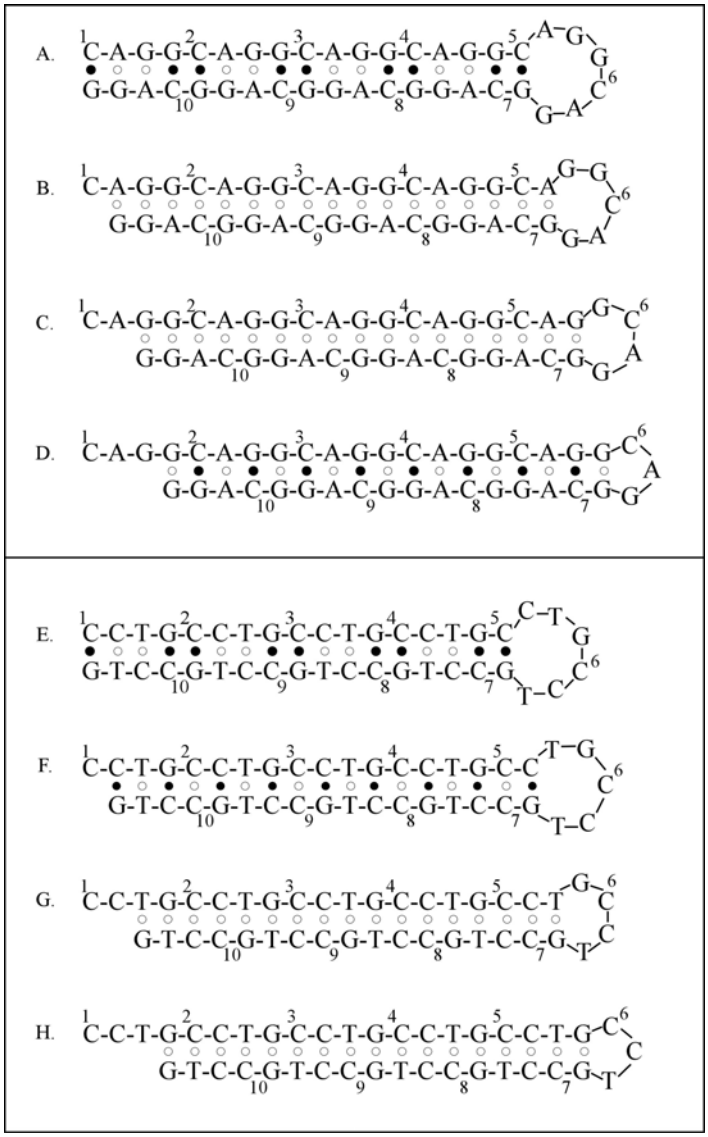


FIG. 9. **Theoretical hairpin structures formed by d(CAGG) and d(CCTG) oligonucleotides.** The various folded-back structures that can be formed by the d(CAGG) and d(CCTG) oligomers are shown in this figure. Hairpin structures with either 6,5,4 or 3 residues in the terminal loop formed by slippage and misalignment of 0, 1, 2 or 3 nucleotides are shown for the d(CAGG) (A – D) and d(CCTG) (E – H) oligonucleotides. The filled circles between the two DNA strands indicate Watson-Crick pairing and the open circles denote non-Watson-Crick pairing. Although only ten CAGG and CCTG repeats are shown for simplicity, the same types of loops and pairing arrangements would apply to oligonucloetides of any length.

modified both the adenines and guanines at the N7 positions showing that this position is not involved in hydrogen bonding and that the G•A pairs are of the G(anti)•A(anti) or A(anti)•G(anti) types (123,126,305).

Alternatively, $d(CCTG)_{26}$ did not show the formation of any stable secondary structure that could be detected under our analytical conditions. We believe this is due to the extremely unstable pairing schemes as shown in Fig. 9E – H. The structures with 4 and 3 unpaired residues in the terminal loops lack any Watson-Crick pairs (Fig. 9G and H, respectively) making them less stable than the structures with 6 and 5 residues in the loops (Fig. 9E and F, respectively). Furthermore, the T•T mismatches as well as the C•T mispairs are extremely destabilizing (125,306) as stated earlier. Thus, the possible base pairing arrangements reveal a greater probability that the CAGG strand forms a more stable structure than the CCTG strand.

A distinct orientation effect was observed for repeats cloned into the SmaI site where orientation II was significantly more unstable than orientation I. For the triplet repeat sequences, instability was defined primarily as the loss of the full-length progenitor fragment. Also, deletions were the predominant products of instability. In case of the tetranucleotide repeats however, instability refers to both expansions and deletions. Thus, upon initial consideration, our results appear to be similar to the results with the TRS, including CTG•CAG, CGG•CCG and GAA•TTC, in which orientation II was shown to be more unstable (5,38,128-137,165). However, the CCTG sequences are genetically unstable in the orientation prone to expand (orientation II) (Fig. 10) as compared to the TRS that are unstable in the orientation prone to delete (orientation II).

Fig. 10 shows a mechanism for the orientation-dependant instability of the (CCTG•CAGG) sequences. In 1995, Kang *et al.* (128) proposed the original model to explain the orientation-dependence of CTG•CAG sequences associated with DM1. For this triplet repeat sequence, the CTG strand forms a more stable secondary structure than the CAG strand. However, for the DM2 sequence, the CAGG strand (on either the nascent lagging strand or the lagging strand template) forms a more stable structure compared to the CCTG strand, generating expansions and deletions, respectively, as shown herein. Thus, the tetranucleotide repeats cloned in orientation II are prone to expand whereas those cloned in orientation I are prone to delete (Fig. 10). This is the first case of a "reversed" orientation behavior (5,38). However, similar to the triplet repeats, the capability of one of the two DNA strands to form a quasistable folded structure, in our case the ability of the CAGG strand to adopt a hairpin structure, explains this "reversed" behavior.

The distance of the triplet repeats from the origin of replication plays an important role in the genetic instabilities of these repeats (R. R. Iyer and R. D. Wells, unpublished work) (128,131,137,307). In our studies, the CCTG•CAGG repeats cloned proximal to the SV40 origin were more unstable than those cloned distal to the same origin. Several hypotheses were described to explain this effect. If an Okazaki initiation zone is a region of single-stranded template DNA, averaging 135 to 145 nucleotides (in eukaryotes) (137,308), where priming of the Okazaki fragments occurs, then depending on where the repeats fall within the Okazaki fragments (*i.e.* at the 5′ end or the 3′ end), their ability to form secondary structures would differ thereby influencing the amount of

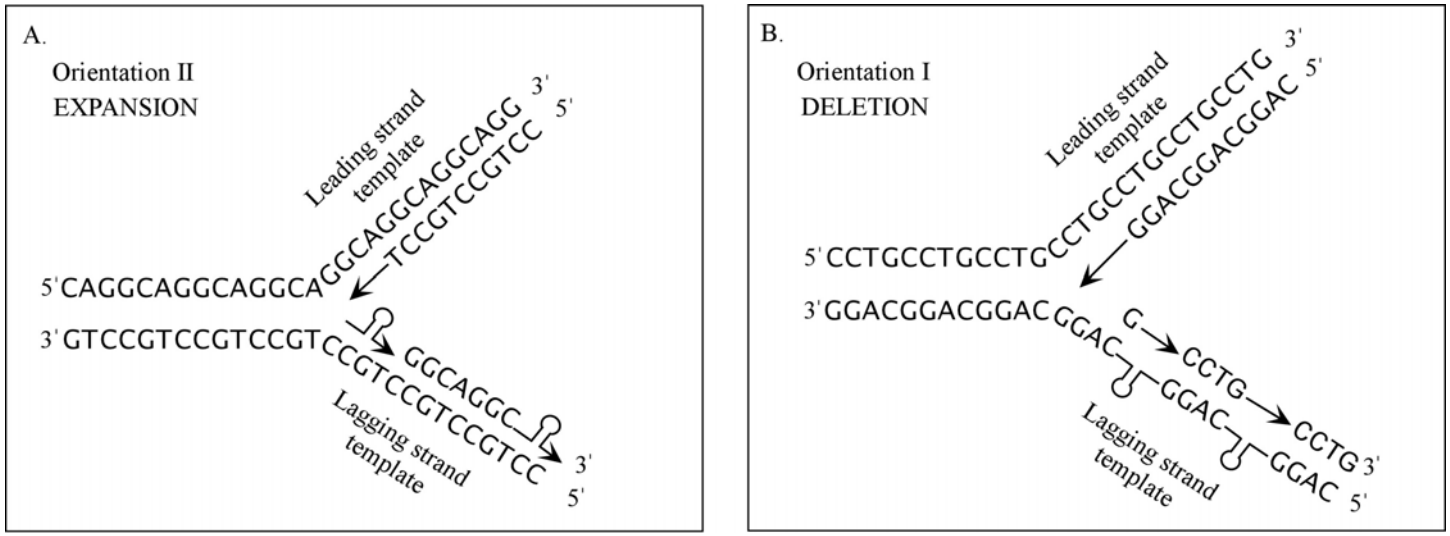


FIG. 10. **Model for the orientation dependent instability of (CCTG•CAGG)$_n$ repeats in mammalian cells.** A, the presence of the CAGG repeats on the leading strand template and on the newly synthesized products of the lagging strand template (orientation II) can give rise to expansions since the CAGG repeats on the nascent strand can form folded-back secondary structures by strand slippage and thus generate expansions. B, the presence of the CCTG repeats on the leading strand template and on the newly synthesized product of the lagging strand template (orientation I) preferentially gives rise to deletions since the CAGG repeats on the lagging strand template can form slipped structures which may be bypassed during synthesis (see Discussion).

instability observed (4,137,308). Alternatively, we speculate that due to almost continuous initiation events occurring at the origin, the repeats cloned close to the origin would be rendered single-stranded for a longer period of time thereby giving the repeats a greater opportunity to fold-back and form a stable secondary structure. Once replication has been initiated and the fork progresses away from the origin, the repeats lying distal to the origin do not have as great an opportunity to form these slipped structures on the leading strand. However, due to the single-stranded nature of the Okazaki fragments on the lagging strand, both expansions and deletions can occur depending on the stability of the hairpin structures formed by these repeats. The aberrant processing of the Okazaki fragments involving FEN-1 and DNA ligase has also been hypothesized to play an important role in generating genetic instabilities (165,168,169,171,220,309,310).

Furthermore, polymerase switching (PolI/PolIII) has been implicated in *E. coli* to contribute to the genetic instability of triplet repeats (128,214). Polymerase switching has also been proposed to occur during eukaryotic replication where following synthesis of the RNA-DNA primer by the polα-primase complex, replication is continued by polδ, both on the leading and lagging strands (294,311). Thus, polymerase switching may further contribute to the observed instability of the tetranucleotide repeats.

In the present study, we have assayed for replication-based instability; however, transcription through the repeats could also have an influence on the levels of observed instability (80,81,224-227,312). The $(CCTG•CAGG)_n$ repeats in our experiments, cloned into the SmaI site, fall within the promoter region of the neomycin resistance

gene that is transcribed. Since this gene is transcribed to elicit the drug resistance, a further increase in the amount of instability could be obtained for repeats cloned proximal to the SV40 origin. Also, the secondary structures formed by these repeats may cause both the transcription as well as the replication machinery to pause, as shown both *in vitro* and *in vivo* for triplet repeats (174,176,177,227). This in turn could lead to activation of repair (4,38,91,133,136,141,165,205,213,214,218,266,300,313,314) and recombination (161,193,201,202,258) processes which, in the case of triplet repeats, are known to generate instability.

Thus, a complex interplay of replication, repair, recombination and transcription may effect the massive expansions observed in DM2. In this study, we have specifically focused on replication and its role in generating the instabilities. However, ongoing work implicates recombination as a powerful mechanism that contributes to the instabilities of the tetranucleotide repeats (R. Dere and R. D. Wells, manuscript in preparation). These data provide the first insights into an understanding of the molecular mechanisms of the tetranucleotide instabilities. Our ultimate goal is to understand the molecular processes causing these expansions in order to develop therapeutic strategies.

CHAPTER III

THE EXPANSION PRONE RECOMBINATION HOTSPOT ACTIVITY OF THE CCTG•CAGG REPEATS IN *ESCHERICHIA COLI*

A. Overview

Myotonic dystrophy type 2 (DM2) is caused by the extreme expansion of the repeating tetranucleotide CCTG•CAGG sequence from < 30 repeats in normal individuals to ~ 11,000 for the full mutation in certain patients. This repeat is in intron 1 of the zinc finger protein 9 gene on chromosome 3q21. Since prior work demonstrated that CTG•CAG and GAA•TTC triplet repeats (responsible for DM1 and Friedreich's ataxia, respectively) can expand by genetic recombination, we investigated the capacity of the DM2 tetranucleotide repeats to also expand during this process. Both gene conversion and unequal crossing over are attractive mechanisms to effect these very large expansions. $(CCTG•CAGG)_n$ (where n = 30, 75, 114 or 160) repeats showed high recombination frequencies (up to 27 fold higher than the non-repeating control) in an intramolecular plasmid system in *E. coli*. Furthermore, a distinct orientation effect was observed where orientation II (CAGG on the leading strand template) was more prone to recombine. Expansions of up to double the length of the tetranucleotide repeats were found. Also, the repeating tetranucleotide sequence was more prone to expansions rather than deletions than observed for any of the three TRS. Thus, the genetic instabilities of the CCTG•CAGG repeats are mediated by a recombination-repair mechanism.

B. Introduction

The massive expansion (up to 11,000 copies) of CCTG•CAGG repeats mapped to intron 1 of the zinc finger protein 9 (ZNF9) gene on chromosome 3q21 (236) are associated with myotonic dystrophy type 2 (DM2). This is the first tetranucleotide repeat to be implicated in a hereditary neurological disease; almost twenty other neurological diseases are characterized by the expansions of triplet repeat sequences (TRS) (4-6,37,38,256). Additionally, a pentanucleotide repeat ATTCT•AGAAT and a dodecamer CCCCGCCCCGCG•CGCGGGGCGGGG repeat were shown recently to be associated with spinocerebellar ataxia type 10 (SCA10) (235) and progressive myoclonus epilepsy of Unverricht-Lundborg type (EPM1) (228), respectively. However, the largest expansions of all repeating sequences involved in these neurological diseases were observed for DM2. The longest normal allele sequenced was about 26 repeats, containing two interruptions, whereas the longest expansion in affected individuals was close to 11,000 repeats (236).

We have recently proposed a hairpin-structure mediated replication slippage model for the tetranucleotide genetic instabilities of DM2 (120). However, the large expansions associated with this disease are not easily explained solely by this replication-slippage model. Recombination is a powerful and plausible mechanism (38,161,193,201,202,258,315) for the generation of large expansions. Chromosomal DNA is continuously challenged by a variety of damaging agents (such ionizing radiation and free radicals) and errors including those resulting from replication that can lead to double strand breaks (DSBs) (203,316-322). These breaks are repaired via a

recombination-repair pathway, either by homologous recombination (HR), single-strand annealing (SSA) or non-homologous end joining (NHEJ) (321,323-330). Although homologous recombination is believed to be the major pathway in prokaryotes (321,331-335), there is evidence that SSA and NHEJ may also occur in these organisms (321,333,336-341). Furthermore, a large number of DSBs / lesions within the DNA can cause the induction of the SOS response in bacteria (332,334,342-346) which could further increase the instability of the repeating tracts.

Herein, we show that the CCTG•CAGG tetranucleotide repeats associated with DM2 are recombination hotspots. Analysis of the types of genetic instabilities in this intramolecular assay showed that expansions (lengths longer than a single tract) were the predominant products even with long lengths of the repeating tracts. This is contrary to previous results with the TRS (CTG•CAG, and GAA•TTC) where deletions (lengths shorter than a single tract) were the preferential products of recombination. The recombination frequencies were stimulated in the SOS^+ strains as compared to the SOS^- strain. Thus, the large expansions observed in DM2 can be explained by a recombination-repair mechanism.

C. Experimental procedures

Construction of the $(CCTG•CAGG)_n$ Containing Plasmids – The genetic assay used to study recombination was previously established to study the recombination behavior of CTG•CAG repeats (201). Briefly, the $(CCTG•CAGG)_n$ inserts (where n = 30, 75, 114 and 160) were cloned into the EcoRI/HindIII sites of pBR322. This tract is

referred to as the X insert. Another tract of the tetranucleotide repeats of the same length as the first tract was then cloned into the PvuII site of the pBR322 derivative carrying the X insert (referred to as the Y insert). Furthermore, a *GFPuv* gene from the pGFPuv vector (B.D. Biosciences Clontech) was cloned at the EcoRV/EagI sites in the vector carrying two CCTG•CAGG tracts.

The $(CCTG•CAGG)_n$ repeats were excised with a EcoRV/EagI restriction digestion from the pCR2.1TOPO derivatives (kind gifts of Dr. L. P. W. Ranum, University of Minnesota) as done for our previous studies (120). The insert containing fragments were blunt-end ligated into the EcoRI/HindIII site and PvuII site of pBR322 as described above. This was the cloning strategy for all the plasmids used in this study except for pRW5230. In this case, problems were encountered during the cloning of the repeats and thus a slightly modified strategy was used. The $(CCTG•CAGG)_{160}$ insert was blunt-end ligated into the EcoRI/HindIII site of pBR322. The clone containing the repeat tract in orientation I was chosen as the desired clone. Simultaneously, the same insert was also cloned into the PvuII site of pBR322. Again, the selected clone carried the repeat tract in orientation I. The two clones thus generated were then digested with NdeI/BsgI and the two repeat containing fragments ligated to give rise to the plasmid carrying the two CCTG•CAGG tracts. The *GFPuv* gene was cloned into this plasmid as described above for all the remaining plasmids.

The plasmids carrying the $(CCTG•CAGG)_n$ repeats were characterized using restriction mapping and DNA sequencing to determine the length and orientation of the repeats. The fragments obtained after restriction mapping were end-labeled with [α-^{32}P]

dATP and 1 unit of the Klenow fragment of *E. coli* DNA polymerase I and analyzed on 5.5% polyacrylamide gels in TAE buffer (40mM Tris acetate, 1mM EDTA, pH 8). The sequencing reactions were performed using the Thermo Sequenase Radiolabeled Terminator Cycle Sequencing Kit (U. S. Biochemical Corp.) according to the manufacturer's recommendations. The products of the sequencing reactions were analyzed on 6% Long Ranger gels (FMC BioProducts) containing 7.5M urea in the glycerol tolerant gel buffer (1.78M Tris, 0.57M taurine and 0.01M EDTA) (U. S. Biochemical Corp.). The $(CCTG•CAGG)_{30}$ repeat tract is pure and uninterrupted as determined by DNA sequencing. The $(CCTG•CAGG)_{75}$ is also a pure repeat tract, however, in addition to the 75 tetranucleotide repeats there are 46 bp of human non-repeating ZNF9 DNA flanking the repeats on one end and 140 bp on the other end. The 140 bp flanking the repeats included a $(TG)_{16}(TCTG)_{10}•(CAGA)_{10}(CA)_{16}$ tract immediately adjacent to the CCTG•CAGG repeats (236). The $(CCTG•CAGG)_n$ tracts, where n = 114 or 160, carried an interruption 11 repeats into the tract to give the sequence $(CCTG)_{11}CCTT(CCTG)_n$ where n = 102 and 148, respectively. Thus, the 114 and 160 tetranucleotide repeat tracts are also pure except for a TA pair in place of a GC pair in the twelfth repeat unit. As stated in our previous replication studies with the CCTG•CAGG repeats (120), the long tracts of the tetranucleotide repeats were sequenced using primers located on either end of the repeating tract. The estimation of the fragment length using restriction mapping had an error of ± 5 repeats for both $(CCTG•CAGG)_{114}$ and $(CCTG•CAGG)_{160}$.

For all four repeat lengths, the CCTG•CAGG tracts were cloned in the direct repeat orientation with respect to each other. Orientation I is defined as the orientation in which the CCTG repeats are on the template for leading strand DNA synthesis and when the CCTG repeats are on the lagging strand template it is referred to as orientation II. A single plasmid (pRW5232) was created with two tracts of $(CCTG•CAGG)_{160}$ in the inverted repeat orientation with respect to each other, i.e., the X tract was in orientation II whereas the Y tract was in orientation I. Additionally, a plasmid carrying two tracts of a 354 bp non-repeating region of the DMPK gene (in the direct repeat orientation) was used as a control (201). This plasmid was originally named pRW4871gfp (201); however, for our studies this plasmid was renamed pRW5233. To evaluate the recombination properties of plasmids harboring two tetranucleotide tracts, the plasmids were maintained in *E. coli* HB101 which is Rec A^-. Previous studies (201,347-351) showed that intramolecular recombination occurs even in the absence of the RecA protein. Therefore, the parental plasmids were separated and purified on agarose gels as described earlier (161,201).

Bacterial Strains – The plasmids used in this study were maintained in *E. coli* HB101 (Invitrogen) *(mcrB, mmr, hsdS20 (r_B^-, m_B^-), recA1, supE44, ara14, galK2, lacY1, proA2, rplS20 (Sm^R), xyl5, λ^-, leuB6, mtl-1)*. Two parental *E. coli* strains were used in our studies, AB1157 and KMBL1001. AB1157 [*thr-1, ara-14, leuB6, D(gpt-proA)62, lacY1,tsx-33, qsr[prime]-, glnV44(AS), galK2, l, arc-hisG4(Oc), rfbD1, mgl-51, rpsL31(sm^R), kdgK51, xylA5, mtl1, argE3(Oc), thi-1*] was obtained from the *E. coli* Genetic Stock Center at Yale University (New Haven, CT). KMBL1001 (no known

mutations) was obtained from Dr. Nora Goosen (Leiden Institute of Chemistry, The Netherlands). Strains JJC510 (parental strain for the SOS response – SOS^+) [*GY4786; λ[p(sfiA::lacZ) CIind⁻] Δlac-pro, rpsL*], JJC523 (constitutive induction of the SOS response – SOS^+) [*GY5425; recA441, sulA II, ΔlacI169, thi, leuB6, his4, argE3, ilvTS, galK2, rpsL37, lexA71::Tn5*] and JJC123 (deficient for the SOS response – SOS^-) [*GY6781; λ[sfiA::lacZ], Δpro-lac, gal⁺, rpsL, mal::Tn9, LexAind1*] were kind gifts from Dr. Benedicte Michel (Institut National de la Recherche Agronomique, France). *E. coli* JH139 [dinD1::MudI1734 kan^R lac] contains a *dinD1::LacZ* fusion which activates the β-galactosidase gene when the SOS response is induced and was a kind gift from Dr. Joseph Heitman (Duke University Durham, North Carolina).

Analyses of the Products of Recombination – The $(CCTG•CAGG)_n$ containing plasmids were transformed into the *E. coli* strains and the transformation mixtures were plated on LB plates containing ampicillin (100 μg/ml) as in our previous studies (201). Approximately 40 white colonies were picked at random from every plasmid transformed into every strain used. Liquid cultures were obtained from each of the individual colonies and DNA was isolated using alkaline lysis (Promega, Wizard Plus Miniprep DNA Purification System) and the supercoiled DNA was analyzed on 1% agarose gels in TAE buffer. The products of recombination were then cleaved with AatII and NdeI and the fragments labeled using [α-^{32}P] dATP and 1 unit of the Klenow fragment of *E. coli* DNA polymerase I. The fragments were then analyzed on 5.5% polyacrylamide gels. The 1 Kbp DNA ladder size standard (Invitrogen) was used as the size marker. The lengths of each insert excised from individual colonies were

determined using FluorChem version 3.04 (Alpha Innotech Corp.). Furthermore, ~ 100 recombinants were sequenced using primers from both ends of the CCTG•CAGG tetranucleotide repeats to verify the lengths of the repeat tracts deduced from the restriction studies. As stated previously, lengths of about 70 repeats were sequenced from both ends to enable the precise determination of the lengths; however, the longer repeat lengths which were determined by gel electrophoresis measurements gave an error of ± 5 repeats.

Assay to Evaluate the Induction of the SOS Response – In order to evaluate the role of the SOS repair in our studies, we tested the induction of this repair pathway using a β-galactosidase activity assay (352). The plasmids harboring $(CCTG•CAGG)_{160}$ in both orientations were transformed into *E. coli* JH139 and grown in K-media overnight at 37˚C. The JH139 strain contains a *dinD1::LacZ* fusion that activates the β-galactosidase gene on induction of the SOS response. The overnight culture was diluted (1/100) the next day in fresh media and the cells were grown to mid-log phase. The cultures were then chilled on ice to stop growth. 2 mL of the culture was then centrifuged at 4˚C and the cell pellet resuspended in an equal volume of chilled Z-buffer. The OD_{600} of these cells was measured. 0.5 mL of the culture was then added to the same volume of Z-buffer and the cells permeabilized by addition of 100 µl of chloroform and 50 µl of 0.1% SDS (sodium dodecyl sulfate) followed by vortexing for 10 s. The reaction was started by addition of 200 µl of ONPG (o-nitrophenyl-β-D-galactosidase, 4 mg/ml). The time of addition of the ONPG was recorded and the cells were then incubated at room temperature until the yellow color was produced. The reaction was

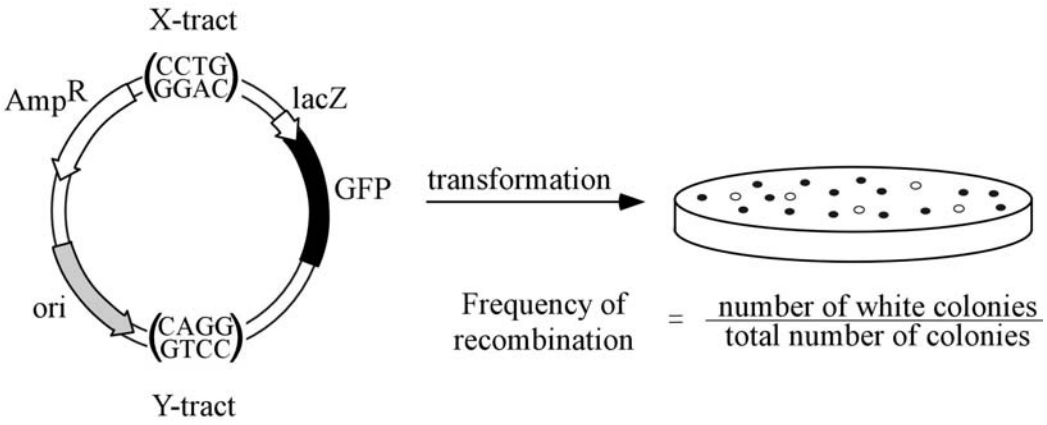
stopped by addition of 500 µl of 1M Na_2CO_3. The time of addition was again noted and the OD_{420} and OD_{550} measured for each sample. The Miller units of activity were calculated using the formula; 1000 x [(OD_{420} – 1.75 x OD_{550})] / (T x V x OD_{600}) where T is the time of the reaction in minutes and V is the volume of culture used in the assay in mLs. The pGEM vector (Promega) was used as the negative control lacking any repeating tracts.

D. Results

Strategy of Study – The intramolecular genetic assay used for our studies was previously established (161,201). The $(CCTG•CAGG)_n$ tetranucleotide repeats (where n = 30, 75, 114 and 160) were used to construct the family of molecules shown in Fig. 11. All of these plasmids contain a GFP gene cloned between two tracts of the repeat sequence in the direct repeat orientation with respect to each other. The presence of the GFP gene results in green colonies. Recombination between the two repeating tracts leads to the loss of the GFP gene which gives rise to white colonies thus, providing a useful green-white screen for measuring recombination events. The frequency of recombination was calculated as a ratio of the number of white colonies to the total number of colonies (Fig. 11A) (201). Additional controls, including plasmid copy number determination, growth advantage control and plasmid establishment control were previously conducted and shown to not influence the frequency of intramolecular

FIG. 11. **Plasmids used in the study.** A, the $(CCTG•CAGG)_n$ containing plasmids were transformed into the *E. coli* strains used in our study and plated to obtain green colonies (black spots) and white colonies (white spots). The frequency of recombination was calculated as a fraction of the number of white colonies over the total number of colonies. B, the $(CCTG•CAGG)_n$ repeats (where n = 30, 75, 114, and 160) were cloned in the direct repeat orientation into the EcoRI/HindIII site of the pBR322 vector (X tract) and into the PvuII site of the same derivative (Y tract) (see Experimental Procedures). The repeats were cloned in both orientations I and II. Orientations I and II are defined in the Experimental Procedures. Additionally, two control plasmids, one with $(CCTG•CAGG)_{160}$ cloned as indirect repeats (pRW5232) and the other carrying two tracts of the 375 bp DMPK DNA cloned as direct repeats (pRW5233), were also prepared for our studies. All plasmids contain the GFP gene.

A.



B.

Plasmid Name	No. of (CCTG•CAGG) repeats		Orientation	
	X Tract	Y Tract	X Tract	Y Tract
pRW5224	30	30	I	I
pRW5225	30	30	II	II
pRW5226	75	75	I	I
pRW5227	75	75	II	II
pRW5228	114	114	I	I
pRW5229	114	114	II	II
pRW5230	160	160	I	I
pRW5231	160	160	II	II
pRW5232	160	160	II	I
pRW5233	375 bp of DMPK DNA	375 bp of DMPK DNA	Direct	Direct

recombination (161,201). A control plasmid (pRW5233) carrying two 354 bp of random DMPK DNA in the direct repeat orientation was also used in these studies. All statistical analyses were perfomed using SigmaStat version 2.03. Statistical differences (p-values) were calculated using the z-test.

CCTG•CAGG Repeats Are Recombination Hotspots – Two parental *E. coli* strains, AB1157 and KMBL1001, were used to study the recombination behavior of the tetranucleotide repeats. The plasmids carrying the CCTG•CAGG repeats gave high frequencies of recombination, in comparison to the controls in both *E. coli* strains (Table II). In all cases, the frequencies obtained for plasmids transformed into *E. coli* KMBL1001 were lower than those obtained for plasmids transformed into *E. coli* AB1157. However, the general trends remained the same in both strains indicating that the genetic backgrounds of the host strains could potentially influence the frequency of recombination. The effect of different genetic backgrounds on the genetic instabilities of triplet repeat sequences (TRS) has been reported (353). A control plasmid (pRW5233) carrying two tracts of a 354 bp random DMPK DNA sequence was used in our studies, which gave frequencies of 0.5% and 0.2%, respectively, when transformed into *E. coli* AB1157 and KMBL1001 (Table II). Also, an additional control (pRW5232) carrying 160 CCTG•CAGG repeats in the inverted repeat orientation with respect to each other, when transformed into *E. coli* AB1157 resulted in no white colonies, as previously reported (201). Other workers (347,354,355) also demonstrated the necessity of repeats in the direct orientation for successful recombination.

Two major factors influenced the frequencies of recombination, the length of the repeats and their orientation with respect to the unidirectional ColE1 origin of replication. A distinct length effect was observed in both strains, where the frequency of recombination increased in a linear fashion with an increase in the length of the repeating CCTG•CAGG sequence (Table II). Thus, $(CCTG•CAGG)_{30}$ showed a 2 fold increase in frequency when compared to the control plasmid (pRW5233) that further increased in a linear fashion to a 27 fold increase for the $(CCTG•CAGG)_{160}$ containing plasmid in *E. coli* AB1157. This relationship has been reported previously for the CTG•CAG sequence and for the GAA•TTC sequence (in the absence of the sticky DNA forming propensity of the GAA•TTC sequence) (161,201). A similar trend was also observed for the plasmids transformed into *E. coli* KMBL1001 (Table II). Although the frequencies were slightly lower in KMBL1001, the fold difference between the frequencies for the various lengths of the CCTG•CAGG repeats remained almost the same for both parental strains. For example, there is a 13.7 fold increase in the frequency of recombination for pRW5231 [$(CCTG•CAGG)_{160}$, orientation II] when compared to pRW5225 [$(CCTG•CAGG)_{30}$, orientation II] in *E. coli* AB1157 which is similar to the 14.6 fold increase seen between the same plasmids in *E. coli* KMBL1001.

Furthermore, an orientation effect was also observed where the frequency of recombination was higher for plasmids carrying the CCTG•CAGG tetranucleotide repeats in orientation II (Table II). In this orientation, the CCTG repeats lie on the template for lagging strand synthesis. In *E. coli* AB1157, a significant difference was observed between the orientations for plasmids carrying repeat lengths of 75, 114 and

TABLE II

Frequency of recombination in parental strains E. coli AB1157 and KMBL1001

The CCTG•CAGG tetranucleotide repeat containing plasmids were transformed into the parental strains, *E. coli* AB1157 and KMBL1001. The frequency of recombination was obtained by dividing the number of white colonies by the total number of colonies. The frequencies for each of the plasmids used in our studies in both orientations are shown. The frequency of recombination for pRW5232, carrying the two tracts of CCTG•CAGG as inverted repeats, was obtained only for *E. coli* AB1157 and not for KMBL1001 (as indicated by the dash).

Plasmid Name	Number of repeats	Frequency of Recombination (%)	
		AB1157	KMBL1001
Orientation I			
pRW5224	30	0.8	0.5
pRW5226	75	4.5	2.0
pRW5228	114	6.5	3.6
pRW5230	160	10.8	6.2
Orientation II			
pRW5225	30	1.0	0.6
pRW5227	75	5.6	2.8
pRW5229	114	10.4	5.6
pRW5231	160	13.7	8.8
pRW5232	160	< 0.1	—
pRW5233	None	0.5	0.2

160 ($p = <0.001$). However, there was no statistically significant difference ($p = 0.221$) for $(CCTG•CAGG)_{30}$ repeat containing plasmids. In the case of *E. coli* KMBL1001, a difference between the orientations was observed in all cases ($p = <0.001$) including the 30-mer. All the plasmids carrying the repeats cloned in orientation II showed a 1.2 – 1.6 fold higher frequency of recombination in both *E. coli* strains.

Thus, both the length and orientation of the CCTG•CAGG repeats influenced the frequency of recombination; longer repeats cloned in orientation II show a higher frequency compared to shorter lengths in orientation I.

Recombinants Are Prone to Expand – The recombinants (white colonies) from each of the plasmids transformed into the parental *E. coli* strains were analyzed using agarose gel electrophoresis. The supercoiled DNA isolated from individual white colonies migrated faster than the parental plasmids at a distance representing approximately 2 kbp on the gel (data not shown), indicating that smaller plasmids were formed due to recombination between the two homologous CCTG•CAGG tracts and the loss of the intervening vector sequence including the GFP gene. Furthermore, the DNA was subjected to restriction mapping and polyacrylamide gel electrophoresis to estimate the length of the single recombined tetranucleotide tract in the recombinants. The expanded clones carry CCTG•CAGG lengths longer than a single tract of the parental plasmid whereas the deleted clones have fewer than the number of repeats in a single tract of the starting parental plasmids. The expanded clones never carried more repeats than the sum of the two tracts (see below).

The plasmids transformed into both *E. coli* strains showed a high propensity for expansions, especially when transformed into *E. coli* KMBL1001 (Fig. 12). In the case of $(CCTG•CAGG)_{30}$ (pRW5224), 80% of the recombinants in *E. coli* AB1157 had expanded tetranucleotide repeats. On increasing the length of the repeating tract, the percentage of expanded clones still remained high with $(CCTG•CAGG)_{160}$ (pRW5231) showing 59% of the recombinants with expansions. There was no statistically significant difference between the two orientations for the percentages of expanded and deleted clones. This was true for all plasmids except pRW5228 [$(CCTG•CAGG)_{114}$, orientation I] and pRW5229 [$(CCTG•CAGG)_{114}$, orientation II] transformed into *E. coli* AB1157 where a significant difference was observed; orientation I showed a higher percentage of deleted clones and orientation II showed a higher percentage of expanded clones (Fig. 12). Although this fits well with the model proposed in our previous studies (120) where the CCTG•CAGG repeats in orientation I are prone to delete and those in orientation II are prone to expand, the reason for this difference is uncertain. In *E. coli* KMBL1001, no difference was observed in the percentage of expanded and deleted clones between the two orientations, except for $(CCTG•CAGG)_{30}$ and $(CCTG•CAGG)_{75}$ where the percentage of deleted clones was significantly higher in orientation I and II, respectively (Fig. 12).

The high percentage of expanded clones observed in our studies is unique to the CCTG•CAGG tetranucleotide repeats. Previous studies with TRS including CTG•CAG and GAA•TTC showed a significant decrease in the percentage of expanded clones with

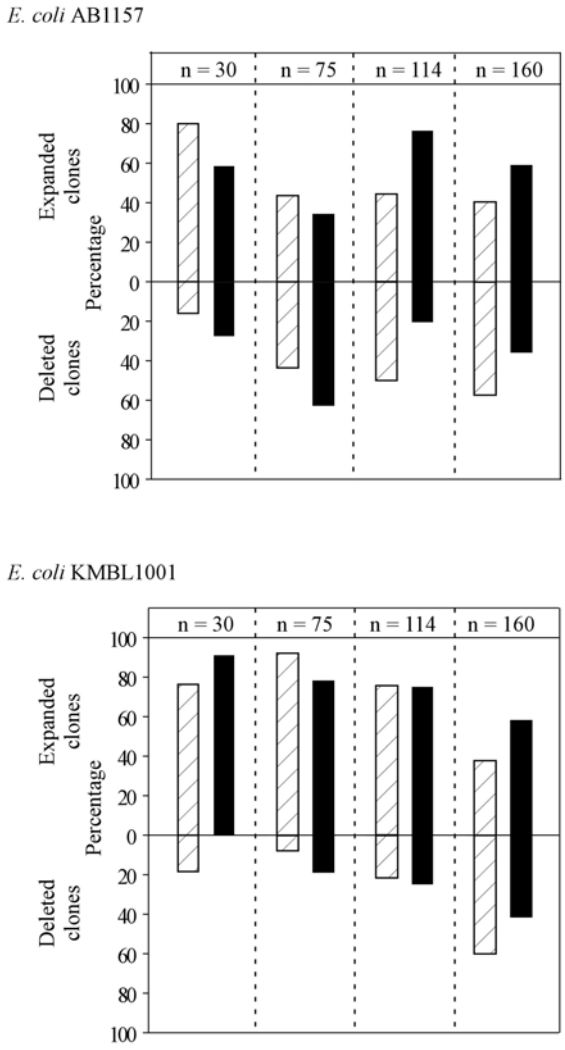


FIG. 12. **Analyses of the lengths of the recombination products.** The $(CCTG•CAGG)_n$ inserts (where n = 30, 75, 114 and 160) were excised from the recombinants (white colonies) and analyzed for their lengths using restriction mapping (see Experimental Procedures). The percentages of expanded and deleted clones are shown for each of the tetranucleotide lengths. The bars above zero percentage indicate expanded clones whereas the bars below zero show the deleted clones. The crosshatched bars indicate the percentage of expansions and deletions for inserts cloned in orientation I whereas the filled bars indicate the percentage of mutations for inserts in orientation II. The data for each bar was collected from the analyses of 40 colonies from duplicate experiments.

an increase in the number of repeats (161,201). Conversely, a high proportion of deleted clones were associated with an increase in the length of the repeating tract. Herein, we found that even with long lengths of the $(CCTG•CAGG)_n$ tract, a large number of recombinants are expanded. As shown in Fig. 13 the percentage of expanded clones in the case of $(CCTG•CAGG)_{160}$ (640 bp) is as high as 50%. In case of the triplet repeat CTG•CAG, the percentage of expanded clones is highest (31%) for the plasmid carrying 67 CTG•CAG repeats. In terms of bp, 294 bp (98 repeats) of CTG•CAG repeats show 20% expanded clones whereas, 300 bp (75 repeats) of the CCTG•CAGG repeats show 39%, almost twice the percentage of expanded clones. Increasing the number of bp to 495 for the triplet repeat (165-mer) decreases the percentage of expanded recombinants to 10% whereas for a corresponding 456 bp of the tetranucleotide repeat sequence (114-mer), the percentage is 60 (Fig. 13). Similarly, 495 bp of CTG•CAG shows 85% of deleted clones whereas a similar length of 456 bp of the CCTG•CAGG sequence shows 36%, almost half the number of deleted clones (Fig. 13).

Thus, the CCTG•CAGG repeats show a high frequency of recombination to yield a large percentage of expanded clones, emphasizing the unique ability of this repeat sequence to expand.

Magnitude of Expanded and Deleted Clones – The length ranges of expanded and deleted products for the plasmids (Fig. 11B) transformed into the parental *E. coli* strains is shown in Table III. The largest increase in the number of repeats was the sum of the two repeat tracts seen for pRW5229 [$(CCTG•CAGG)_{114}$, orientation II] and pRW5230 [$(CCTG•CAGG)_{160}$, orientation I] in the case of *E. coli* AB1157. However,

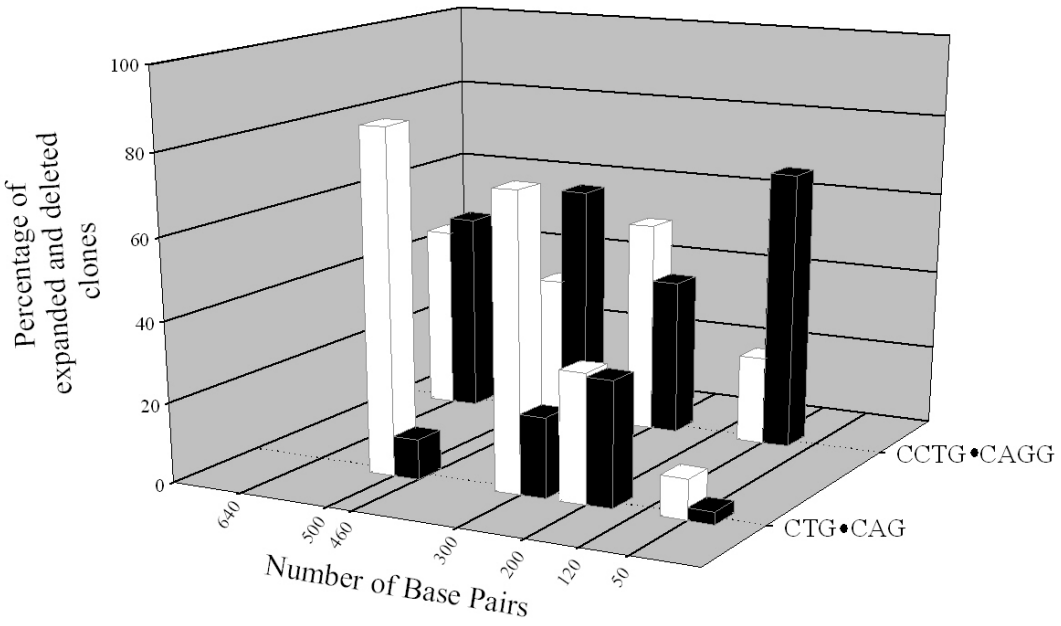


FIG. 13. **A comparative analysis of the expansions associated with CCTG•CAGG repeats as compared to CTG•CAG repeats.** The recombinants obtained from experiments with the CTG•CAG containing plasmids (201) and the CCTG•CAGG plasmids were analyzed on polyacrylamide gels. The data obtained from both orientations I and II was pooled since there was no significant difference between the two orientations. The filled bars indicate the percentage of expanded clones, whereas the white bars indicate the percentage of deleted clones (Z-axis). The number of bp of repeats in both the trinucleotide and tetranucleotide containing plasmids is indicated on the Y-axis. The two rows show the $(CCTG•CAGG)_n$ repeats (where n = 30, 75, 114 and 160 corresponding to 120, 300, 456 and 640 bp) and the $(CTG•CAG)_n$ repeats (where n = 17, 67, 98 and 165 corresponding to 51, 201, 294 and 495 bp) (X-axis).

most of the other plasmids also showed between a 1.5 – 1.9 fold increase in the number of repeats when compared to the starting length (the length of a single tract) (Table III). *E. coli* KMBL1001 showed expanded recombinants carrying a tetranucleotide repeat tract up to twice the length of the tract in the parental plasmids (pRW5225, pRW5226, pRW5228 and pRW5230). The largest deletion was the complete loss of the CCTG•CAGG repeats observed for pRW5224 [$(CCTG•CAGG)_{30}$, orientation I] in *E. coli* AB1157. The smallest deletion was the loss of only 2 – 3 repeats. pRW5225 [$(CCTG•CAGG)_{30}$, orientation II] did not yield any deleted clones when transformed into *E. coli* KMBL1001 (Table III).

Almost 40 recombinants from both *E. coli* strains were randomly selected and sequenced to confirm the lengths and determine the sequence of the repeating tract (data not shown). The sequence analysis showed the presence of one CCTG•CAGG tract flanked on either side by the vector sequence as seen in the case of our previous studies (161,201). Lengths of approximately 70 – 80 repeats were sequenced to precisely estimate the number of repeats, however longer lengths of the repeating tract were sequenced using primers from both ends of the repeating tract. The lengths of these longer tracts were estimated using electrophoretic techniques as stated in the above section, giving an error of ± 5 repeats.

Thus, a range of expansions and deletions were found in both orientations for plasmids carrying the CCTG•CAGG repeats.

TABLE III

Magnitude of expanded and deleted products from the recombination of the (CCTG•CAGG)$_n$ repeats in the parental E. coli strains

The plasmids were isolated from the recombinant (white) colonies and the lengths of the CCTG•CAGG inserts were analyzed by restriction digestion and polyacrylamide gel electrophoresis (see Experimental Procedures). The ranges of the number of repeats for different plasmids in the expanded and deleted products are shown for each of the two *E. coli* parental strains, AB1157 and KMBL1001. The estimated repeat lengths varied by ± 5 repeats for plasmids carrying 75, 114 and 160 repeats. The inserts from the plasmids carrying 30 repeats were sequenced to obtain the precise length of the inserts in the recombinants. The recombinants obtained from transformation of pRW5225 into KMBL1001 yielded no deletion products. The numbers in brackets following the names of the plasmids are the numbers of the repeats in the parental plasmids.

Plasmid Name	*E. coli* Strains	
	AB1157	KMBL1001
Expanded Products (Number of Repeats)		
pRW5224 (30)	33 - 49	32 - 46
pRW5225 (30)	32 - 44	35 - 60
pRW5226 (75)	77 - 132	86 - 149
pRW5227 (75)	77 - 123	77 - 112
pRW5228 (114)	116 - 198	118 - 228
pRW5229 (114)	117 - 224	122 - 222
pRW5230 (160)	165 - 319	170 - 315
pRW5231 (160)	167 - 301	167 - 299
Deleted Products (Number of Repeats)		
pRW5224 (30)	0 - 26	19 - 28
pRW5225 (30)	16 - 28	30
pRW5226 (75)	5 - 73	38 - 64
pRW5227 (75)	44 - 71	41 - 73
pRW5228 (114)	35 - 111	45 - 99
pRW5229 (114)	31 - 93	54 - 111
pRW5230 (160)	38 - 158	30 - 151
pRW5231 (160)	39 - 147	19 - 156

Stimulation of Recombination Frequencies in Strains with an Active or Inducible SOS response – Previous studies have shown intramolecular recombination frequencies to be elevated by, but not dependent on, the presence of RecA (201,347-351). The RecA protein plays a crucial role not only in homologous recombination (HR) but also in the SOS-repair pathway. In this pathway, RecA in its active form cleaves the LexA repressor, thus regulating all the genes under the SOS regulon. In order to determine the extent of involvement of the SOS-repair pathway in our intramolecular recombination assay, we used three SOS-mutant *E. coli* strains, JJC510 (parental for the SOS response), JJC523 (which constitutively expresses the SOS response) and JJC123 (SOS^{-}).

The frequencies of recombination for plasmids transformed into the *E. coli* SOS mutant strains are shown in Fig. 14. The repeat tract length and orientation influenced the recombination frequencies, similar to our results with *E. coli* AB1157 and KMBL1001. Moreover, the frequencies obtained in the parental JJC510 strain were similar to those obtained with *E. coli* AB1157 (Table II). Although high frequencies were observed for *E. coli* JJC510, the frequencies fell significantly in *E. coli* JJC123 (non-inducible SOS response) (Fig. 14). In orientation II, a statistically significant difference ($p = < 0.001$) in frequencies was obtained for plasmids carrying 114 and 160 repeats in JJC123 (SOS^{-}) (2.9 and 4.7 %, respectively) when compared to the parental strain JJC510 (6.6 and 15 %, respectively). Furthermore, all plasmids transformed into the strain in which the SOS-repair is constitutively expressed (JJC523) showed a higher frequency than those transformed into JJC123 ($p = < 0.001$) (Fig. 14). Similar results were obtained for plasmids carrying the repeats in orientation I (data not shown), where

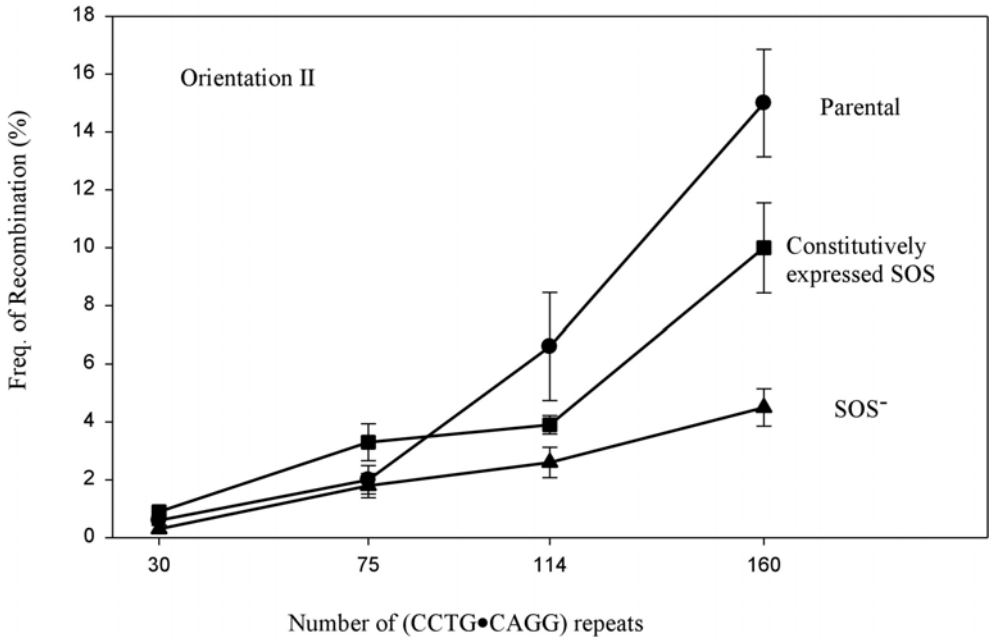


FIG. 14. **Frequency of recombination in the SOS mutant strains.** The frequency of recombination as a percentage was plotted as a function of the length of the $(CCTG•CAGG)_n$ repeat containing plasmids. The $(CCTG•CAGG)_n$ containing plasmids (where n = 30, 75, 114 and 160) were transformed into *E. coli* JJC510 (parental strain for SOS repair), JJC523 (SOS^+) and JJC123 (SOS^-). The frequency of recombination for plasmids carrying the CCTG•CAGG repeats cloned in orientation II (filled symbols) is shown. The circles, squares and triangles represent plasmids transformed into the SOS mutant strains JJC510, JJC523 and JJC123, respectively. The frequencies of recombination obtained from the control plasmid pRW5233 transformed into JJC510, JJC523 and JJC123 are 1.2, 0.4 and 0.1 percent, respectively. The standard deviations for the three experiments performed with each of the plasmids are indicated as error bars.

significant differences were observed between the strains harboring plasmids with all three (75, 114 and 160) repeat lengths.

Thus, there is a significant decrease in the frequency of recombination for pRW5231 [$(CCTG•CAGG)_{160}$, orientation II] in *E. coli* JJC123 compared to JJC510 (a 3.2-fold decrease) and JJC523 (a 2.5-fold decrease).

The SOS-repair Pathway is Not Induced by the CCTG•CAGG Repeats – To explain the high frequencies of recombination in the SOS^+ strains, we determined the level of induction of the SOS response by the repeat containing plasmids. *E. coli* JH139 carries a *dinD1::LacZ* fusion protein that activates the β-galactosidase gene if the SOS response is, in fact, induced. The plasmids harboring the longest CCTG•CAGG repeats were transformed into this strain and a β-galactosidase activity assay was performed to measure the amount of expression (Experimental Procedures). A pGEM plasmid lacking the repeats was used as a negative control. On measuring the levels of β-galactosidase produced in the assay, we determined that there was no difference in the amounts obtained from the control and the experimental plasmids harboring the repeats (data not shown). Furthermore, the amount of expressed protein was equivalent to the amounts obtained in the absence of induction. Thus, the differences in the frequencies of recombination obtained in the SOS^- strain, the parental (JJC510) and constitutively on (JJC523) strain may be due to differences in these isogenic strains (Discussion).

Repeat Tracts in Recombinants from the SOS^+ and SOS^- Strains Are Similar in Length – Analyses of expanded and deleted clones obtained from plasmids in the SOS-mutant strains revealed similar results to those seen in *E. coli* AB1157. A large number

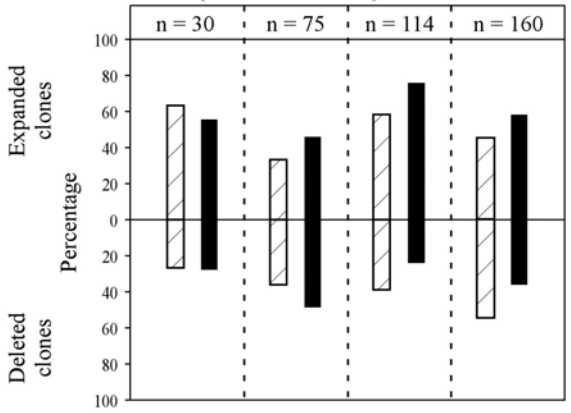
of expanded clones were observed in all three *E. coli* strains; JJC510 is the parental strain for the SOS response (SOS^+), JJC523, constitutively induced SOS (SOS^+), and JJC123 (SOS^-) (Fig. 15). The largest percentage of expanded clones was 76%, observed for pRW5229 [$(CCTG•CAGG)_{114}$, orientation II] in *E. coli* JJC510 (Fig. 15A). pRW5227 [$(CCTG•CAGG)_{75}$, orientation II] carried 81% deleted clones in *E. coli* JJC523 (Fig. 15B). On analyzing the lengths of the inserts (within the limits described above), a range of both expansions and deletions were obtained (Table IV). Again, the ranges were similar to those found with *E. coli* AB1157 and KMBL1001 (Table III). All the expanded clones showed a 1.5 – 2.0 fold increase in the length of the tract. Thus, the percentages of expanded and deleted products remained the same in all three mutant strains although the frequencies of recombination were significantly different.

E. Discussion

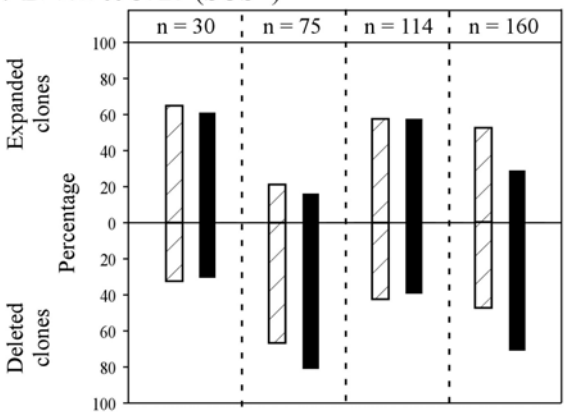
Myotonic dystrophy type 2 (DM2) is caused by a very large expansion of the CCTG•CAGG repeats (by an average of ~ 5000 repeats) in intron 1 of the zinc finger protein 9 (ZNF9) gene on chromosome 3q21 (236). Our recent work has shown a hairpin-structure forming propensity of these tetranucleotide repeats to lead to the observed genetic instabilities due to replication slippage (120). However, the massive expansions (up to ~ 11,000 repeats) associated with DM2 (236) are not easily explained solely by a slippage based mechanism. Herein, we showed the CCTG•CAGG repeats to be recombination hotspots which are prone to expand; this tendency towards expansions

FIG. 15. **Analyses of the lengths of the recombination products.** The $(CCTG•CAGG)_n$ inserts (where n = 30, 75, 114 and 160) were excised from the recombinants (white colonies) derived from the experiments described in Fig. 14 and were analyzed for their lengths using restriction mapping (see Experimental Procedures). The percentages of expanded and deleted clones are shown for each of the tetranucleotide lengths. The bars above zero percentage indicate expanded clones whereas the bars below zero show the deleted clones. The crosshatched bars indicate the percentage of expansions and deletions for inserts cloned in orientation I whereas the filled bars indicate the percentage of expansions and deletions for inserts in orientation II. The data for each bar was collected from the analyses of 40 colonies. A, *E. coli* JJC510, B, *E. coli* JJC523, and C, *E. coli* JJC123.

A. *E. coli* JJC510 (Parental SOS$^+$)



B. *E. coli* JJC523 (SOS$^+$)



C. *E. coli* JJC123 (SOS$^-$)

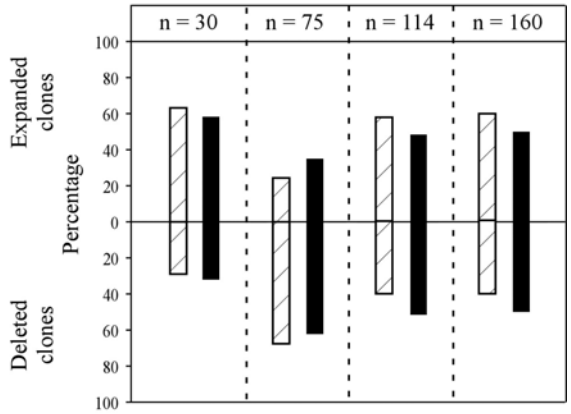


TABLE IV

Magnitude of expanded and deleted products from the recombination of the (CCTG•CAGG)$_n$ repeats in the SOS mutant strains

The plasmids were isolated from the recombinant (white) colonies and the lengths of the CCTG•CAGG inserts were analyzed by restriction digestion and polyacrylamide gel electrophoresis (see Experimental Procedures). The ranges of the number of repeats for different plasmids in the expanded and deleted products are shown for each of the three SOS mutant strains, JJC510 (parental), JJC523 (SOS $^+$) and JJC123 (SOS$^-$). The estimated repeat lengths varied by ± 5 repeats for plasmids carrying 75, 114 and 160 repeats. The inserts from the plasmids carrying 30 repeats were sequenced to precisely obtain the length of the inserts in the recombinants. The numbers in brackets following the names of the plasmids are the numbers of the repeats in the parental plasmids.

Plasmid Name	*E. coli* Strains		
	JJC510	JJC523	JJC123
Expanded Products (Number of Repeats)			
pRW5224 (30)	32 - 47	32 - 45	32 - 59
pRW5225 (30)	32 - 54	32 - 48	32 - 46
pRW5226 (75)	79 - 125	77 - 125	82 - 114
pRW5227 (75)	77 - 123	96 - 141	82 - 110
pRW5228 (114)	118 - 185	122 - 163	117 - 228
pRW5229 (114)	118 - 212	119 - 228	116 - 207
pRW5230 (160)	163 - 320	166 - 320	167 - 320
pRW5231 (160)	164 - 301	163 - 314	162 - 285
Deleted Products (Number of Repeats)			
pRW5224 (30)	0 - 27	0 - 28	0 - 28
pRW5225 (30)	16 - 28	8 - 28	0 - 28
pRW5226 (75)	30 - 72	29 - 73	19 - 73
pRW5227 (75)	24 - 71	37 - 72	17 - 70
pRW5228 (114)	37 - 107	40 - 107	27 - 112
pRW5229 (114)	60 - 111	44 - 112	38 - 110
pRW5230 (160)	45 - 152	58 - 152	44 - 140
pRW5231 (160)	53 - 155	51 - 158	45 - 150

in contrast to deletions is unique compared to all previous studies on CTG•CAG and GAA•TTC repeats (161,201). Recombination is a plausible mechanism (193,258,315) to mediate expansions. Furthermore, we evaluated the involvement of the SOS repair pathway in this recombination assay and found that although the recombination frequencies were significantly higher in the SOS^+ strain compared to the SOS^- strain, the SOS response itself was not induced.

The intramolecular assay used for our studies was established previously to study the recombination behavior of the CTG•CAG (201) and GAA•TTC (161) repeats. Using this assay we showed that the frequency of recombination of the CCTG•CAGG repeats increased with an increasing length of the repeats. The frequency of recombination was 27 fold higher for pRW5231 [$(CCTG•CAGG)_{160}$, orientation II] than for the control containing two 354 bp DMPK DNA fragments (pRW5233). Similar studies with the DM1 sequence showed only a 13 fold increase for the longest CTG•CAG tract (165-mer, orientation II) (201). In terms of bp, the $(CCTG•CAGG)_{114}$ is closest to $(CTG•CAG)_{165}$ and showed a 21 fold increase. Thus, the CCTG•CAGG repeats are indeed recombination hotspots and recombined giving higher percentages of recombination frequencies than their DM1 counterparts (201).

A distinct orientation effect was observed where orientation II showed higher frequencies of recombination as compared to orientation I. Similar studies conducted with the CTG•CAG repeats also showed high frequencies in orientation II (201). In the case of the TRS, this orientation effect was attributed to the ability of the CTG tracts to form a stable folded-back structure (as opposed to the complementary CAG strand) on

the lagging strand template, giving rise to deletions. Our previous studies (120) on the DM2 sequence showed the repeats to be unstable in the orientation prone to expand (orientation II), due to the propensity of the CAGG strand (on the nascent lagging strand) to adopt a stable hairpin loop. The complementary CCTG strand did not exhibit any structure-forming capability in our studies. We believe that the differences between the stabilities of the hairpin-loops formed by each of these repeats may contribute to the observed differences in orientation.

A model in which breaks occurring at the repeat tracts can lead to homologous recombination (HR) is proposed in Fig. 16A. We hypothesize that initiation of replication from the origin leads to the opening of the DNA duplex within the CCTG•CAGG repeats (step A). The tetranucleotide repeats in orientation II (where CAGG is on the newly synthesized lagging strand) have the opportunity to form hairpin-loop structures on the lagging strand (step B). The potential repair of these structures may generate double strand breaks (DSBs) within the CCTG•CAGG repeats of the Y tract. DSBs in prokaryotes are mainly repaired by HR (step D) and the required homology is provided by the X tract in our studies. The recombination between the two tracts can lead to either gene conversion or cross-over events (97,189,193,258,319,321,338) (Fig. 16A). The gene conversion events (step E) would not be detected in our assay since the colonies are still green due to the presence of the GFP gene. Alternatively, the recombination event between the two CCTG•CAGG tracts can lead to a cross-over (step F), which leads to a deletion of the intervening sequences

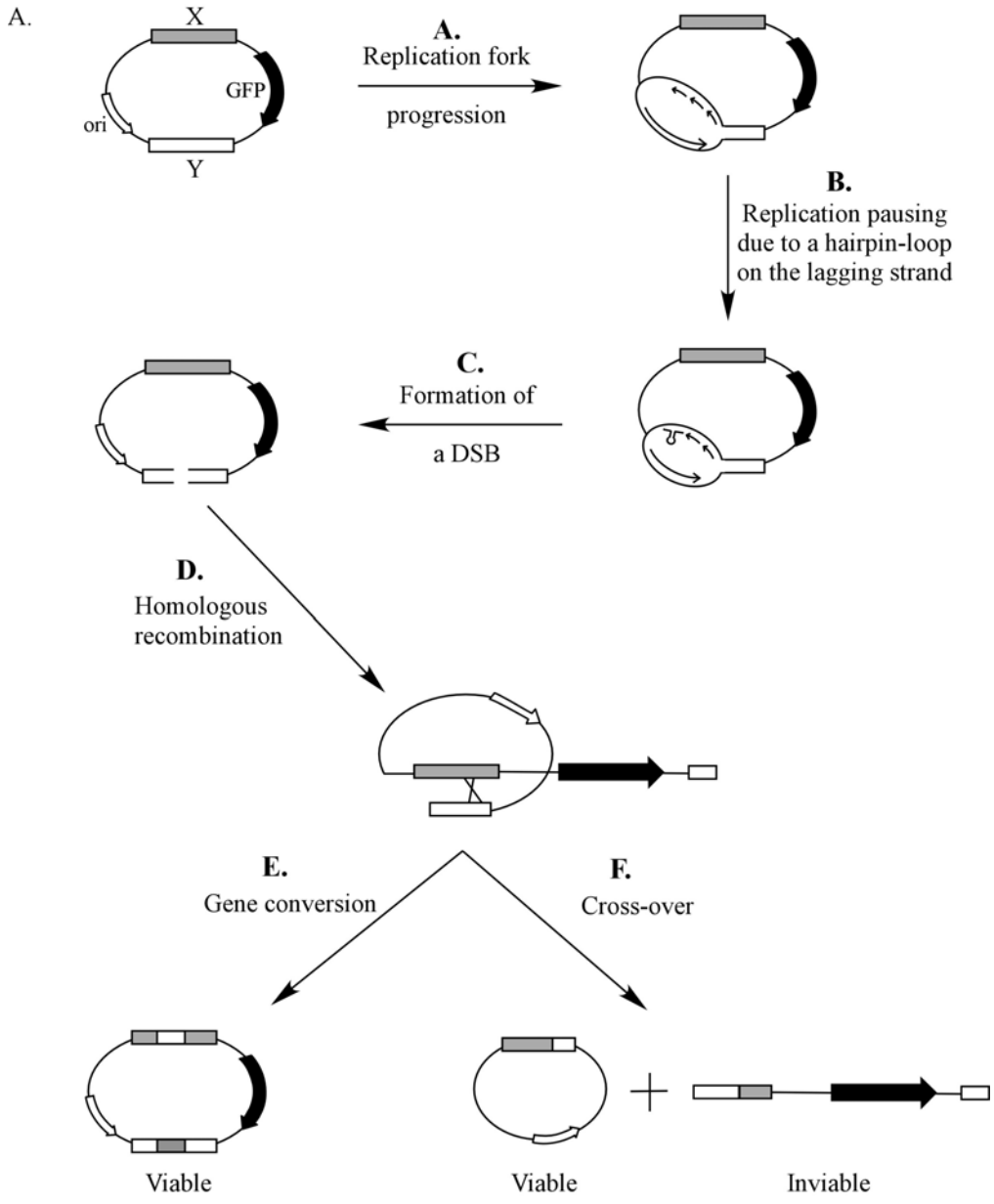


FIG. 16. **Model for the recombinational repair of the CCTG•CAGG repeats.** A, the replication process causes duplex unwinding that gives the CCTG•CAGG repeats the opportunity to form hairpin loops that could lead to replication pausing and eventually double strand breaks (DSBs). These breaks could be processed by homologous recombination (HR) as shown to give rise to either gene conversion or crossover events depending on the resolution of the Holliday junctions. The crossover that is observed in our system gives rise to a smaller plasmid that would carry the origin of replication (ori) and the ampicillin resistance gene (Amp^R) along with the single recombined CCTG•CAGG tract.

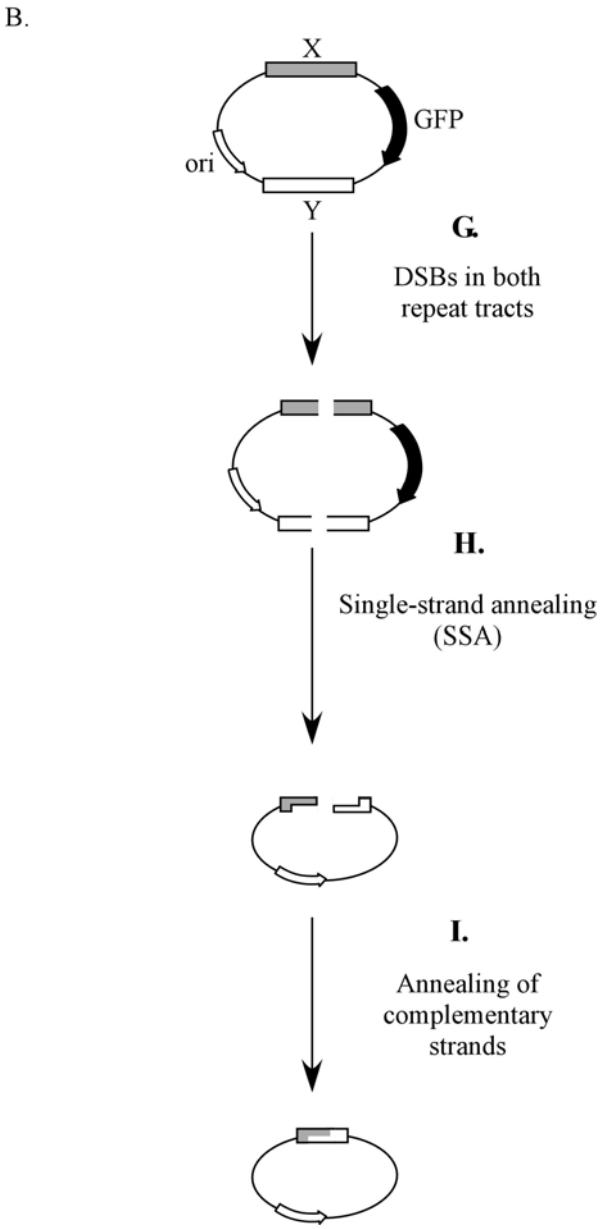


FIG. 16. Continued. B, the replication process causes duplex unwinding that gives the CCTG•CAGG repeats the opportunity to form hairpin loops that could lead to replication pausing and eventually double strand breaks (DSBs). DSBs could be repaired by single-strand annealing (SSA) if two breaks occur simultaneously in both repeat tracts as shown.

between direct repeats. This would result in the formation of two products (Fig. 16A); one is a smaller recombined plasmid carrying a single recombined CCTG•CAGG tract along with the origin of replication and the resistance gene, and the other product is a fragment carrying the GFP gene and probably some CCTG•CAGG repeats depending on where the crossover occurred, which is inviable due to a lack of the origin and the drug resistance gene. These are the products detected in our studies.

Analyses of the recombinants revealed that a high proportion of the clones carried expansions (lengths longer than the length of a single tract in the original plasmid). This is quite different from the results with the CTG•CAG and GAA•TTC repeats (161,201), where longer (up to 495 bp) repeats showed a decreased percentage of expanded clones compared to shorter repeats (up to 51 bp). Herein, even the longest tetranucleotide repeat carrying 160 repeats (640 bp) still showed 50 % expanded clones. Moreover, a range of expanded clones was observed, including some clones which had expansions almost twice the size of the starting length, i.e. 320 repeats (1280 bp).

We used the Mfold program (http://www.bioinfo.rpi.edu/applications/mfold) to determine the theoretical stabilities of the CTG, CAG, CCTG and CAGG hairpin loops. The CTG hairpin was found to be most stable with a ΔG value almost 3.5 times more negative than the ΔG value obtained for the CAGG repeats of similar length. We propose that this lower stability of the structure formed by the CAGG repeats contributes to its expansion behavior. The tetranucleotide sequence may fold-back into a hairpin-loop but may collapse before the structure can be detected and repaired by the repair pathways. However, the transient formation of these structures may still be sufficient to

result in expansions either during replication by causing impediments to the replication fork or due to a less efficient repair pathway. This could lead to breaks which could then be repaired by recombination-repair. In the case of the CTG•CAG repeats, the repair of the stable structure formed by the repeats could be more efficient leading to deletions. Additionally, this lack of stability may also indicate why the triplet repeats are more prone to delete in subsequent rounds of replication (91,128,134,213) as opposed to the CCTG•CAGG repeats that are stably maintained (unpublished data). Furthermore, the MMR pathway has been shown in previous work (213,215,216,219,221,223) to be involved in generating the instabilities of the CTG•CAG sequence; however, the MMR proteins do not influence the instabilities of the CCTG•CAGG repeats (unpublished data). This could again be suggestive of a repair-escape mechanism by the relatively transient and unstable structure formed at the tetranucleotide repeats.

Studies using intramolecular assays for measuring recombination events have shown this process to be independent of RecA, although recombination was stimulated to give high frequencies in its presence (201,347-351). Previous studies on double-strand breaks (DSBs) have shown that breaks within the repeat tracts can lead to instability [reviewed in (4,6,7)]. If multiple breaks occur during the cell cycle, various checkpoint pathways may be activated to repair the damage before the cell can divide. The SOS response in bacteria is one such mechanism that controls the expression of almost 40 genes, most of which are involved in the repair of damage (344). To evaluate the involvement of the SOS-repair pathway, we determined recombination frequencies in three SOS mutant strains. The SOS parental strain was *E. coli* JJC510. The JJC523

and JJC123 strains contain LexA variants that enable constitutive induction of the SOS pathway or prevent RecA-dependent cleavage thus inactivating the SOS response, respectively. Our results showed that high recombination frequencies were obtained for pRW5231 [$(CCTG•CAGG)_{160}$, orientation II] in JJC510 and JJC523 (15 and 9 %, respectively). On the contrary, the recombination frequency dropped significantly (to 4 %) in the SOS^- *E. coli* JJC123. Thus, a dramatic and significant reduction in recombination frequency was obtained for the different lengths of the CCTG•CAGG repeats in the SOS-deficient strain.

In order to evaluate the levels of induction of the SOS response, we assayed for β-galactosidase activity. These data showed that the SOS response was not induced. This led to the question of why the recombination frequencies were so dramatically different in the isogenic strains used in our assay. The differences between the SOS^+ strains and the SOS^- strain may potentially be due to different basal levels of some gene products under SOS regulation. The lack of these basal levels in the SOS deficient strain may be responsible for the lower levels of recombination observed in this strain. The frequencies of recombination obtained in JJC523 (constitutively expressed SOS strain) were lower than those observed in the parental strain which may suggest a suppressive effect on the recombination pathway, being exerted by one of the many proteins expressed constitutively in this strain. Thus, the complexity of the SOS response and the multiple pathways that it could influence makes it difficult to pinpoint any single gene product or pathway responsible for the observed differences in recombination frequencies between the three SOS mutant strains.

The model proposed in Fig. 16A hypothesis the formation of a break in a single tract (Y tract) and its subsequent repair by a homologous recombination mediated mechanism. Another possible model (Fig. 16B) proposes the formation of DSBs in both the repeat tracts (step G). The plasmids used in our studies are transcribed and contain one repeat tract downstream of the transcribed GFP gene and the other upstream of the same gene. The waves of negative supercoiling behind (upstream of) the RNA polymerase may influence the formation of the non-B DNA structures *in vivo* (153,356). The formation of these structures upstream of the progressing transcription apparatus could lead to formation of breaks in the X-tract. Thus two DSBs could form simultaneously in both the X and Y tracts (Fig. 16B). A mechanism that is known to act on DSBs in plasmids is single strand annealing (SSA) (321,326,332,333,336-339) (steps H and I). This is a non-conservative repair mechanism where the DSB is processed leading to the annealing of the complementary strands. This is another possible mechanism as shown in Fig. 16B that could lead to the products observed in our studies.

Thus, the CCTG•CAGG repeats are recombination hotspots prone to expand. The large expansions associated with DM2 could occur as a result of the DNA conformational properties of these repeats. Studies aimed at inhibiting the formation of secondary structures by manipulating the conditions favoring their formation may provide an encouraging therapeutic strategy in the future.

CHAPTER IV

SUMMARY AND CONCLUSIONS

The triplet repeat sequences and the mechanisms of expansions associated with these diseases have been studied extensively for the past decade and a half (4-7). The experiments designed herein, were aimed at understanding the molecular mechanisms of DM2, which is the first tetranucleotide to cause a hereditary neurological disorder. DM2 is characterized by the massive expansion of the CCTG•CAGG repeats from < 30 repeats to almost 11,000 repeats (236). This is the largest expansion observed in the category of repeating neurological disorders. By using both the simple and well defined *E. coli* model, as well as a more complex African green monkey kidney cell culture (COS-7) system we have proposed a number of mechanisms including non-B DNA structures, replication-slippage, and recombination-repair to influence the genetic instabilities of DM2.

Replication-slippage in previous studies with the TRS has been proposed as a potential mechanism of genetic instability [reviewed in (6,7,38)]. Our experiments in COS-7 cells showed that replication-based slippage of the CCTG•CAGG repeats could occur in this system, to yield both expansions and deletions (Chapter II). This instability was found to be dependent on the length of the repeats, orientation and proximity of the repeat tracts to the SV40 origin of replication. Additionally, biochemical studies on synthetic oligonucleotides revealed the structure-forming propensity of the $d(CAGG)_{26}$

oligomer as compared to its complementary $d(CCTG)_{26}$ oligonucleotide that did not form any preferential structure in our studies.

Although replication-slippage yielded both expansions and deletions, it would be rather difficult to explain the massive expansions associated with DM2, solely by this mechanism. Using an intramolecular genetic assay in *E. coli* we showed the CCTG•CAGG repeats to be recombination hotspots (Chapter III). The recombination frequencies obtained with the tetranucleotide repeats were up to 27-fold higher than the non-repeating control. Additionally, on analyzing the products of recombination a large percentage of expanded clones were obtained contrary to the large number of deleted clones obtained in similar work with the TRS (CTG•CAG and GAA•TTC). Moreover, we determined that although high recombination frequencies were obtained in SOS^+ strains as opposed to the SOS^- strain, the SOS response itself was not being induced in our studies.

The molecular mechanisms investigated herein, will be important to develop future therapeutic strategies aimed at this disease. The ability of the CCTG•CAGG repeats to adopt non-B DNA conformations seems to play an important role in generating the instabilities associated with DM2 in our model systems. A better understanding of the molecular mechanisms causing such expansions, including the conditions favoring the formation of non-B DNA structures, could lead to the development of various strategies for therapeutic intervention.

REFERENCES

1. Oberle, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M. F., and Mandel, J. L. (1991) *Science* **252**, 1097-1102
2. Verkerk, A. J., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F. P., Eussen, B. E., van Ommen, G. J. B., Blonden, L. A. J., Riggins, G. J., Chastain, J. L., Kunst, C. B., Galjaard, H., Caskey, C. T., Nelson, D. L., Oostra, B. A., and Warren, S. T. (1991) *Cell* **65**, 905-914
3. Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, J. C., Warren, S. T., Schlessinger, D., Sutherland, G. R., and Richards, R. I. (1991) *Science* **252**, 1179-1181
4. Cleary, J. D., and Pearson, C. E. (2003) *Cytogenet. Genome Res.* **100**, 25-55
5. Wells, R. D., and Warren, S. T. (eds) (1998) *Genetic Instabilities and Hereditary Neurological Diseases*, Academic Press, San Diego, CA
6. Lenzmeier, B. A., and Freudenreich, C. H. (2003) *Cytogenet. Genome Res.* **100**, 7-24
7. Wells, R. D., Dere, R., Hebert, M. L., Napierala, M., and Son, L. S. (2005) *Nucleic Acids Res.* **33**, 3785-3798
8. Gatchel, J. R., and Zoghbi, H. Y. (2005) *Nat. Rev. Genet.* **6**, 743-755
9. Charlesworth, B., Sniegowski, P., and Stephan, W. (1994) *Nature* **371**, 215-220
10. Moxon, E. R., and Wills, C. (1999) *Sci. Am.* **280**, 94-99

11. Toth, G., Gaspari, Z., and Jurka, J. (2000) *Genome Res.* **10**, 967-981

12. Bennett, P. (2000) *Mol. Pathol.* **53**, 177-183

13. De Bolle, X., Bayliss, C. D., Field, D., van de Ven, T., Saunders, N. J., Hood, D. W., and Moxon, E. R. (2000) *Mol. Microbiol.* **35**, 211-222

14. Modrich, P., and Lahue, R. (1996) *Annu. Rev. Biochem.* **65**, 101-133

15. Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998) *Nature* **396**, 643-649

16. Atkin, N. B. (2001) *Cytogenet. Cell Genet.* **92**, 177-181

17. Kolodner, R. D., Putnam, C. D., and Myung, K. (2002) *Science* **297**, 552-557

18. Marra, G., and Boland, C. R. (1995) *J. Natl. Cancer Inst.* **87**, 1114-1125

19. Cama, A., Genuardi, M., Guanti, G., Radice, P., and Varesco, L. (1996) *Tumori* **82**, 122-135

20. Craanen, M. E., Blok, P., Offerhaus, G. J., and Tytgat, G. N. (1996) *Scand. J. Gastroenterol. Suppl.* **218**, 92-97

21. Boland, C. R., Thibodeau, S. N., Hamilton, S. R., Sidransky, D., Eshleman, J. R., Burt, R. W., Meltzer, S. J., Rodriguez-Bigas, M. A., Fodde, R., Ranzani, G. N., and Srivastava, S. (1998) *Cancer Res.* **58**, 5248-5257

22. Chung, D. C., and Rustgi, A. K. (2003) *Ann. Intern. Med.* **138**, 560-570

23. Gurin, C. C., Federici, M. G., Kang, L., and Boyd, J. (1999) *Cancer Res.* **59**, 462-466

24. Basil, J. B., Goodfellow, P. J., Rader, J. S., Mutch, D. G., and Herzog, T. J. (2000) *Cancer* **89**, 1758-1764

25. Caldes, T., Perez-Segura, P., Tosar, A., de La Hoya, M., and Diaz-Rubio, E. (2000) *Teratog. Carcinog. Mutagen.* **20**, 283-291

26. Lax, S. F., Kendall, B., Tashiro, H., Slebos, R. J., and Hedrick, L. (2000) *Cancer* **88**, 814-824

27. Leach, F. S. (2002) *Curr. Opin. Urol.* **12**, 407-411

28. El-Rifai, W., and Powell, S. M. (2002) *Semin. Radiat. Oncol.* **12**, 128-140

29. Pestsova, E., Ganal, M. W., and Roder, M. S. (2000) *Genome* **43**, 689-697

30. Jouquand, S., Priat, C., Hitte, C., Lachaume, P., Andre, C., and Galibert, F. (2000) *Anim. Genet.* **31**, 266-272

31. Richman, M., Mellersh, C. S., Andre, C., Galibert, F., and Ostrander, E. A. (2001) *J. Biochem. Biophys. Methods* **47**, 137-149

32. Ikari, K., Onda, H., Furushima, K., Maeda, S., Harata, S., and Takeda, J. (2001) *J. Hum. Genet.* **46**, 207-210

33. Atwood, L. D., Samollow, P. B., Hixson, J. E., Stern, M. P., and MacCluer, J. W. (2001) *Hypertension* **37**, 425-428

34. Helminen, P., Ehnholm, C., Lokki, M. L., Jeffreys, A., and Peltonen, L. (1988) *Lancet* **1**, 574-576

35. Hagelberg, E., Gray, I. C., and Jeffreys, A. J. (1991) *Nature* **352**, 427-429

36. Jeffreys, A. J., Allen, M. J., Hagelberg, E., and Sonnberg, A. (1992) *Forensic Sci. Int.* **56**, 65-76

37. Cummings, C. J., and Zoghbi, H. Y. (2000) *Annu. Rev. Genomics Hum. Genet.* **1**, 281-328

38. Bowater, R. P., and Wells, R. D. (2001) *Prog. Nucleic Acid Res. Mol. Biol.* **66**, 159-202

39. Siianova, E., and Mirkin, S. M. (2001) *Mol. Biol.* **35**, 208-223

40. Lalioti, M. D., Antonarakis, S. E., and Scott, H. S. (2003) *Cytogenet. Genome Res.* **100**, 213-223

41. Lin, X., and Ashizawa, T. (2003) *Cytogenet. Genome Res.* **100**, 184-188

42. Ranum, L. P. W., and Day, J. W. (2004) *Am. J. Hum. Genet.* **74**, 793-804

43. Sutherland, G. R., and Richards, R. I. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 3636-3641

44. La Spada, A. R., Roling, D. B., Harding, A. E., Warner, C. L., Spiegel, R., Hausmanowa-Petrusewicz, I., Yee, W. C., and Fischbeck, K. H. (1992) *Nat. Genet.* **2**, 301-304

45. Leeflang, E. P., Zhang, L., Tavare, S., Hubert, R., Srinidhi, J., MacDonald, M. E., Myers, R. H., de Young, M., Wexler, N. S., Gusella, J. F., and et al. (1995) *Hum. Mol. Genet.* **4**, 1519-1526

46. Richards, R. I., Crawford, J., Narahara, K., Mangelsdorf, M., Friend, K., Staples, A., Denton, M., Easteal, S., Hori, T. A., Kondo, I., Jenkins, T., Goldman, A., Panich, V., Ferakova, E., and Sutherland, G. R. (1996) *Ann. Hum. Genet.* **60 (Pt 5)**, 391-400

47. Giuliano, P., De Cristofaro, T., Affaitati, A., Pizzulo, G. M., Feliciello, A., Criscuolo, C., De Michele, G., Filla, A., Avvedimento, E. V., and Varrone, S. (2003) *Hum. Mol. Genet.* **12**, 2301-2309

48. La Spada, A. R., and Taylor, J. P. (2003) *Neuron* **38**, 681-684

49. Rimoin, D. L., Rasmussen, I. M., Briggs, M. D., Roughley, P. J., Gruber, H. E., Warman, M. L., Olsen, B. R., Hsia, Y. E., Yuen, J., Reinker, K., and et al. (1994) *Hum. Genet.* **93**, 236-242

50. Briggs, M. D., and Chapman, K. L. (2002) *Hum. Mutat.* **19**, 465-478

51. Kennedy, J., Jackson, G., Ramsden, S., Taylor, J., Newman, W., Wright, M. J., Donnai, D., Elles, R., and Briggs, M. D. (2005) *Eur. J. Hum. Genet.* **13**, 547-555

52. Brais, B., Bouchard, J. P., Xie, Y. G., Rochefort, D. L., Chretien, N., Tome, F. M., Lafreniere, R. G., Rommens, J. M., Uyama, E., Nohira, O., Blumen, S., Korczyn, A. D., Heutink, P., Mathieu, J., Duranceau, A., Codere, F., Fardeau, M., and Rouleau, G. A. (1998) *Nat. Genet.* **18**, 164-167

53. Goodman, F. R. (2002) *Am. J. Med. Genet.* **112**, 256-265

54. Brais, B. (2003) *Cytogenet. Genome Res.* **100**, 252-260

55. Brown, L. Y., and Brown, S. A. (2004) *Trends Genet.* **20**, 51-58

56. Amiel, J., Trochet, D., Clement-Ziza, M., Munnich, A., and Lyonnet, S. (2004) *Hum. Mol. Genet.* **13 Spec No 2**, R235-243

57. Cummings, C. J., Mancini, M. A., Antalffy, B., DeFranco, D. B., Orr, H. T., and Zoghbi, H. Y. (1998) *Nat. Genet.* **19**, 148-154

58. Chai, Y., Koppenhafer, S. L., Bonini, N. M., and Paulson, H. L. (1999) *J. Neurosci.* **19**, 10338-10347

59. Zoghbi, H. Y., and Orr, H. T. (2000) *Annu. Rev. Neurosci.* **23**, 217-247

60. Opal, P., and Zoghbi, H. Y. (2002) *Trends Mol. Med.* **8**, 232-236

61. Nance, M. A. (1997) *Brain Pathol.* **7**, 881-900

62. Koshy, B. T., and Zoghbi, H. Y. (1997) *Brain Pathol.* **7**, 927-942

63. Sandberg, G., and Schalling, M. (1997) *Nucleic Acids Res.* **25**, 2883-2887

64. Tassone, F., Hagerman, R. J., Chamberlain, W. D., and Hagerman, P. J. (2000) *Am. J. Med. Genet.* **97**, 195-203

65. Greene, E., Handa, V., Kumari, D., and Usdin, K. (2003) *Cytogenet. Genome Res.* **100**, 65-76

66. Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J. P., Hudson, T., Sohn, R., Zemelman, B., Snell, R. G., Rundle, S. A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P. S., Shaw, D. J., and Housman, D. E. (1992) *Cell* **68**, 799-808.

67. Buxton, J., Shelbourne, P., Davies, J., Jones, C., Van Tongeren, T., Aslanidis, C., de Jong, P., Jansen, G., Anvret, M., Riley, B., Williamson, R., and Johnson, K. (1992) *Nature* **355**, 547-548

68. 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. J., Wieringa, B., Korneluk, R. G., Perryman, M. B., Epstein, H. F., and Caskey, C. T. (1992) *Science* **255**, 1256-1258.

69. Harley, H. G., Brook, J. D., Rundle, S. A., Crow, S., Reardon, W., Buckler, A. J., Harper, P. S., Housman, D. E., and Shaw, D. J. (1992) *Nature* **355**, 545-546

70. Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelö, J., O'Hoy, K., Leblond, S., Earle-MacDonald, J., de Jong, P. J., Wieringa, B., and Korneluk, R. G. (1992) *Science* **255**, 1253-1255

71. Koob, M. D., Moseley, M. L., Schut, L. J., Benzow, K. A., Bird, T. D., Day, J. W., and Ranum, L. P. (1999) *Nat. Genet.* **21**, 379-384

72. Timchenko, L. T., Miller, J. W., Timchenko, N. A., DeVore, D. R., Datar, K. V., Lin, L., Roberts, R., Caskey, C. T., and Swanson, M. S. (1996) *Nucleic Acids Res.* **24**, 4407-4414

73. Nemes, J. P., Benzow, K. A., Moseley, M. L., Ranum, L. P., and Koob, M. D. (2000) *Hum. Mol. Genet.* **9**, 1543-1551

74. Day, J. W., Schut, L. J., Moseley, M. L., Durand, A. C., and Ranum, L. P. (2000) *Neurology* **55**, 649-657

75. Ebralidze, A., Wang, Y., Petkova, V., Ebralidse, K., and Junghans, R. P. (2004) *Science* **303**, 383-387

76. 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, P. I., De Donato, S., Mandel, J. L., Cocozza, S., Koenig, M., and Pandolfo, M. (1996) *Science* **271**, 1423-1427

77. Montermini, L., Andermann, E., Labuda, M., Richter, A., Pandolfo, M., Cavalcanti, F., Pianese, L., Iodice, L., Farina, G., Monticelli, A., Turano, M., Filla, A., De Michele, G., and Cocozza, S. (1997) *Hum. Mol. Genet.* **6**, 1261-1266

78. Montermini, L., Richter, A., Morgan, K., Justice, C. M., Julien, D., Castellotti, B., Mercier, J., Poirier, J., Capozzoli, F., Bouchard, J. P., Lemieux, B., Mathieu, J., Vanasse, M., Seni, M. H., Graham, G., Andermann, F., Andermann, E., Melancon, S. B., Keats, B. J., Di Donato, S., and Pandolfo, M. (1997) *Annals of Neurology* **41**, 675-682

79. Campuzano, V., Montermini, L., Lutz, Y., Cova, L., Hindelang, C., Jiralerspong, S., Trottier, Y., Kish, S. J., Faucheux, B., Trouillas, P., Authier, F. J., Durr, A., Mandel, J. L., Vescovi, A., Pandolfo, M., and Koenig, M. (1997) *Hum. Mol. Genet.* **6**, 1771-1780

80. Bidichandani, S. I., Ashizawa, T., and Patel, P. I. (1998) *Am. J. Hum. Genet.* **62**, 111-121

81. Ohshima, K., Montermini, L., Wells, R. D., and Pandolfo, M. (1998) *J. Biol. Chem.* **275**, 14588-14595

82. Shelbourne, P., Winqvist, R., Kunert, E., Davies, J., Leisti, J., Thiele, H., Bachmann, H., Buxton, J., Williamson, B., and Johnson, K. (1992) *Hum. Mol. Genet.* **1**, 467-473

83. Brunner, H. G., Jansen, G., Nillesen, W., Nelen, M. R., de Die, C. E., Howeler, C. J., van Oost, B. A., Wieringa, B., Ropers, H. H., and Smeets, H. J. (1993) *N. Engl. J. Med.* **328**, 476-480

84. Ashizawa, T., Anvret, M., Baiget, M., Barcelo, J. M., Brunner, H., Cobo, A. M., Dallapiccola, B., Fenwick, R. G., Jr., Grandell, U., Harley, H., and et al. (1994) *Am. J. Hum. Genet.* **54**, 414-423

85. Krahe, R., Eckhart, M., Ogunniyi, A. O., Osuntokun, B. O., Siciliano, M. J., and Ashizawa, T. (1995) *Am. J. Hum. Genet.* **56**, 1067-1074

86. Tishkoff, S. A., Goldman, A., Calafell, F., Speed, W. C., Deinard, A. S., Bonne-Tamir, B., Kidd, J. R., Pakstis, A. J., Jenkins, T., and Kidd, K. K. (1998) *Am. J. Hum. Genet.* **62**, 1389-1402.

87. Pianese, L., Cavalcanti, F., De Michele, G., Filla, A., Campanella, G., Calabrese, O., Castaldo, I., Monticelli, A., and Cocozza, S. (1997) *Am. J. Hum. Genet.* **60**, 460-463

88. Zhang, L., Leeflang, E. P., Yu, J., and Arnheim, N. (1994) *Nat. Genet.* **7**, 531-535

89. Gecz, J., Gedeon, A. K., Sutherland, G. R., and Mulley, J. C. (1996) *Nat. Genet.* **13**, 105-108

90. Hammond, L. S., Macias, M. M., Tarleton, J. C., and Shashidhar Pai, G. (1997) *Am. J. Med. Genet.* **72**, 430-434

91. Wells, R. D., Parniewski, P., Pluciennik, A., Bacolla, A., Gellibolian, R., and Jaworski, A. (1998) *J. Biol. Chem.* **273**, 19532-19541

92. Pluciennik, A., Iyer, R. R., Parniewski, P., and Wells, R. D. (2000) *J. Biol. Chem.* **275**, 28386-28397.

93. Wojciechowska, M., Bacolla, A., Larson, J. E., and Wells, R. D. (2005) *J. Biol. Chem.* **280**, 941-952

94. Bacolla, A., Jaworski, A., Connors, T. D., and Wells, R. D. (2001) *J. Biol. Chem.* **276**, 18597-18604

95. Bacolla, A., Jaworski, A., Larson, J. E., Jakupciak, J. P., Chuzhanova, N., Abeysinghe, S. S., O'Connell, C. D., Cooper, D. N., and Wells, R. D. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**, 14162-14167

96. Bacolla, A., and Wells, R. D. (2004) *J. Biol. Chem.* **279**, 47411-47414

97. Meservy, J. L., Sargent, R. G., Iyer, R. R., Chan, F., McKenzie, G. J., Wells, R. D., and Wilson, J. H. (2003) *Mol. Cell. Biol.* **23**, 3152-3162

98. Abeliovich, D., Lerer, I., Pashut-Lavon, I., Shmueli, E., Raas-Rothschild, A., and Frydman, M. (1993) *Am. J. Hum. Genet.* **52**, 1175-1181

99. 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 et al. (1994) *Nat. Genet.* **6**, 409-414

100. Wohrle, D., Kennerknecht, I., Wolf, M., Enders, H., Schwemmle, S., and Steinbach, P. (1995) *Hum. Mol. Genet.* **4**, 1147-1153

101. Aoki, M., Abe, K., Tobita, M., Kameya, T., Watanabe, M., and Itoyama, Y. (1996) *Clin. Genet.* **50**, 199-201

102. Tanaka, F., Doyu, M., Ito, Y., Matsumoto, M., Mitsuma, T., Abe, K., Aoki, M., Itoyama, Y., Fischbeck, K. H., and Sobue, G. (1996) *Hum. Mol. Genet.* **5**, 1253-1257

103. Cancel, G., Gourfinkel-An, I., Stevanin, G., Didierjean, O., Abbas, N., Hirsch, E., Agid, Y., and Brice, A. (1998) *Hum. Mutat.* **11**, 23-27

104. Anvret, M., Ahlberg, G., Grandell, U., Hedberg, B., Johnson, K., and Edstrom, L. (1993) *Hum. Mol. Genet.* **2**, 1397-1400

105. Thornton, C. A., Johnson, K., and Moxley, R. T., III. (1994) *Ann. Neurol.* **35**, 104-107

106. Martorell, L., Martinez, J. M., Carey, N., Johnson, K., and Baiget, M. (1995) *J. Med. Genet.* **32**, 593-596

107. Chen, X., Mariappan, S. S. V., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoun, A., Smith, S. S., Bradbury, E. M., and Gupta, G. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 5199-5203

108. Gacy, A. M., Goellner, G., Juranic, N., Macura, S., and McMurray, C. T. (1995) *Cell* **81**, 533-540

109. Mitas, M., Yu, A., Dill, J., Kamp, T. J., Chambers, E. J., and Haworth, I. S. (1995) *Nucleic Acids Res.* **23**, 1050-1059

110. Yu, A., Dill, J., and Mitas, M. (1995) *Nucleic Acids Res.* **23**, 4055-4057

111. Mariappan, S. V. S., Catasti, P., Chen, X., Ratliff, R., Moyzis, R. K., Bradbury, E. M., and Gupta, G. (1996) *Nucleic Acids Res.* **24**, 784-792

112. Paiva, A. M., and Sheardy, R. D. (2004) *Biochemistry* **43**, 14218-14227

113. Petruska, J., Arnheim, N., and Goodman, M. F. (1996) *Nucleic Acids Res.* **24**, 1992-1998

114. Mitas, M. (1997) *Nucleic Acids Res.* **25**, 2245-2254

115. Ohshima, K., and Wells, R. D. (1997) *J. Biol. Chem.* **272**, 16798-16806

116. Yu, A., Barron, M. D., Romero, R. M., Christy, M., Gold, B., Dai, J., Gray, D. M., Haworth, I. S., and Mitas, M. (1997) *Biochemistry* **36**, 3687-3699

117. Gacy, A. M., Goellner, G. M., Spiro, C., Chen, X., Gupta, G., Bradbury, E. M., Dyer, R. B., Mikesell, M. J., Yao, J. Z., Johnson, A. J., Richter, A., Melancon, S. B., and McMurray, C. T. (1998) *Mol. Cell* **1**, 583-593

118. Mariappan, S. V. S., Chen, X., Catasti, P., Bradbury, E. M., and Gupta, G. (1998) in *Genetic Instabilities and Hereditary Neurological Diseases* (Wells, R. D., and Warren, S. T., eds), pp. 647-676, Academic Press, San Diego, CA

119. Heidenfelder, B. L., Makhov, A. M., and Topal, M. D. (2003) *J. Biol. Chem.* **278**, 2425-2431

120. Dere, R., Napierala, M., Ranum, L. P., and Wells, R. D. (2004) *J. Biol. Chem.* **279**, 41715-41726

121. Aboul-ela, F., Koh, D., Tinoco, I., Jr., and Martin, F. H. (1985) *Nucleic Acids Res.* **13**, 4811-4824

122. Li, Y., Zon, G., and Wilson, W. D. (1991) *Biochemistry* **30**, 7566-7572

123. Huertas, D., Bellsolell, L., Casasnovas, J. M., Coll, M., and Azorin, F. (1993) *EMBO J.* **12**, 4029-4038

124. Li, Y., and Agrawal, S. (1995) *Biochemistry* **34**, 10056-10062

125. Allawi, H. T., and SantaLucia, J., Jr. (1998) *Nucleic Acids Res.* **26**, 2694-2701

126. Allawi, H. T., and SantaLucia, J., Jr. (1998) *Biochemistry* **37**, 2170-2179

127. Peyret, N., Seneviratne, P. A., Allawi, H. T., and SantaLucia, J., Jr. (1999) *Biochemistry* **38**, 3468-3477

128. Kang, S., Jaworski, A., Ohshima, K., and Wells, R. D. (1995) *Nat. Genet.* **10**, 213-218.

129. Ohshima, K., Kang, S., Larson, J. E., and Wells, R. D. (1996) *J. Biol. Chem.* **271**, 16773-16783

130. Shimizu, M., Gellibolian, R., Oostra, B. A., and Wells, R. D. (1996) *J. Mol. Biol.* **258**, 614-626

131. Maurer, D. J., O' Callaghan, B. L., and Livingston, D. M. (1996) *Mol. Cell. Biol.* **16**, 6617-6622

132. Freudenreich, C. H., Stavenhagen, J. B., and Zakian, V. A. (1997) *Mol. Cell. Biol.* **17**, 2090-2098

133. Miret, J. J., Pessoa-Brandao, L., and Lahue, R. S. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12438-12443

134. Iyer, R. R., and Wells, R. D. (1999) *J. Biol. Chem.* **274**, 3865-3877

135. Bichara, M., Pinet, I., Schumacher, S., and Fuchs, R. P. (2000) *Genetics* **154**, 533-542

136. Balakumaran, B. S., Freudenreich, C. H., and Zakian, V. A. (2000) *Hum. Mol. Genet.* **9**, 93-100

137. Cleary, J. D., Nichol, K., Wang, Y. H., and Pearson, C. E. (2002) *Nat. Genet.* **31**, 37-46

138. Pearson, C. E., Eichler, E. E., Lorenzetti, D., Acharya, S., Kramer, P. R., Kramer, S. F., Nelson, D. L., Zoghbi, H. Y., and Sinden, R. R. (1996) *Am. J. Hum. Genet.* **59(4, Suppl.)**, 244

139. Pearson, C. E., and Sinden, R. R. (1998) in *Genetic Instabilities and Hereditary Neurological Diseases* (Wells, R. D., and Warren, S. T., eds), pp. 585-621, Academic Press, San Diego, CA

140. Pearson, C. E., Tam, M., Wang, Y. H., Montgomery, S. E., Dar, A. C., Cleary, J. D., and Nichol, K. (2002) *Nucleic Acids Res.* **30**, 4534-4547

141. Pearson, C. E., Ewel, A., Acharya, S., Fishel, R. A., and Sinden, R. R. (1997) *Hum. Mol. Genet.* **6**, 1117-1123

142. Oussatcheva, E. A., Hashem, V. I., Zou, Y., Sinden, R. R., and Potaman, V. N. (2001) *J. Biol. Chem.* **276**, 30878-30884

143. Fry, M., and Loeb, L. A. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 4950-4954

144. Kettani, A., Kumar, R. A., and Patel, D. J. (1995) *J. Mol. Biol.* **254**, 638-656

145. Fojtik, P., and Vorlickova, M. (2001) *Nucleic Acids Res.* **29**, 4684-4690

146. Mills, M., Lacroix, L., Arimondo, P. B., Leroy, J. L., Francois, J. C., Klump, H., and Mergny, J. L. (2002) *Curr. Med. Chem. Anti-Canc. Agents* **2**, 627-644

147. Neidle, S., and Parkinson, G. N. (2003) *Curr. Opin. Struct. Biol.* **13**, 275-283

148. Usdin, K., and Woodford, K. J. (1995) *Nucleic Acids Res.* **23**, 4202-4209

149. Usdin, K. (1998) *Nucleic Acids Res.* **26**, 4078-4085

150. Fry, M., and Loeb, L. A. (1999) *J. Biol. Chem.* **274**, 12797-12802

151. Kamath-Loeb, A. S., Loeb, L. A., Johansson, E., Burgers, P. M., and Fry, M. (2001) *J. Biol. Chem.* **276**, 16439-16446

152. Fojtik, P., Kejnovska, I., and Vorlickova, M. (2004) *Nucleic Acids Res* **32**, 298-306

153. Sinden, R. R., Pearson, C. E., Potaman, V. N., and Ussery, D. W. (1998) in *Genes and Genomes* (Verma, R. S., ed) Vol. 5A, 1st Ed., pp. 1-141, 6 vols., JAI Press Inc., Greenwich, CT

154. Sakamoto, N., Chastain, P. D., Parniewski, P., Ohshima, K., Pandolfo, M., Griffith, J. D., and Wells, R. D. (1999) *Mol. Cell* **3**, 465-475

155. Mariappan, S. V., Catasti, P., Silks, L. A., III, Bradbury, E. M., and Gupta, G. (1999) *J. Mol. Biol.* **285**, 2035-2052

156. Tiner, W. J., Sr., Potaman, V. N., Sinden, R. R., and Lyubchenko, Y. L. (2001) *J. Mol. Biol.* **314**, 353-357

157. Potaman, V. N., Oussatcheva, E. A., Lyubchenko, Y. L., Shlyakhtenko, L. S., Bidichandani, S. I., Ashizawa, T., and Sinden, R. R. (2004) *Nucleic Acids Res.* **32**, 1224-1231

158. Kuryavyi, V. V., and Jovin, T. M. (1995) *Nat. Genet.* **9**, 339-341

159. Vetcher, A. A., Napierala, M., Iyer, R. R., Chastain, P. D., Griffith, J. D., and Wells, R. D. (2002) *J. Biol. Chem.*, (in press)

160. Vetcher, A. A., Napierala, M., and Wells, R. D. (2002) *J. Biol. Chem.* **277**, 39228-39234

161. Napierala, M., Dere, R., Vetcher, A., and Wells, R. D. (2004) *J. Biol. Chem.* **279**, 6444-6454

162. Potaman, V. N., Bissler, J. J., Hashem, V. I., Oussatcheva, E. A., Lu, L., Shlyakhtenko, L. S., Lyubchenko, Y. L., Matsuura, T., Ashizawa, T., Leffak, M., Benham, C. J., and Sinden, R. R. (2003) *J. Mol. Biol.* **326**, 1095-1111

163. Wong, L. J., Ashizawa, T., Monckton, D. G., Caskey, C. T., and Richards, C. S. (1995) *Am. J. Hum. Genet.* **56**, 114-122

164. Martorell, L., Monckton, D. G., Gamez, J., Johnson, K. J., Gich, I., Lopez de Munain, A., and Baiget, M. (1998) *Hum. Mol. Genet.* **7**, 307-312

165. Freudenreich, C. H., Kantrow, S. M., and Zakian, V. A. (1998) *Science* **279**, 853-856

166. Gordenin, D. A., Kunkel, T. A., and Resnick, M. A. (1997) *Nat. Genet.* **16**, 116-118

167. Kokoska, R. J., Stefanovic, L., Tran, H. T., Resnick, M. A., Gordenin, D. A., and Petes, T. D. (1998) *Mol. Cell. Biol.* **18**, 2779-2788

168. Schweitzer, J. K., and Livingston, D. M. (1998) *Hum. Mol. Genet.* **7**, 69-74

169. Spiro, C., Pelletier, R., Rolfsmeier, M. L., Dixon, M. J., Lahue, R. S., Gupta, G., Park, M. S., Chen, X., Mariappan, S. V., and McMurray, C. T. (1999) *Mol. Cell* **4**, 1079-1085

170. White, P. J., Borts, R. H., and Hirst, M. C. (1999) *Mol. Cell. Biol.* **19**, 5675-5684

171. Lee, S., and Park, M. S. (2002) *Exp. Mol. Med.* **34**, 313-317

172. Kucherlapati, M., Yang, K., Kuraguchi, M., Zhao, J., Lia, M., Heyer, J., Kane, M. F., Fan, K., Russell, R., Brown, A. M., Kneitz, B., Edelmann, W., Kolodner, R. D., Lipkin, M., and Kucherlapati, R. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 9924-9929

173. Spiro, C., and McMurray, C. T. (2003) *Mol. Cell. Biol.* **23**, 6063-6074

174. Kang, S., Ohshima, K., Shimizu, M., Amirhaeri, S., and Wells, R. D. (1995) *J. Biol. Chem.* **270**, 27014-27021

175. Viguera, E., Canceill, D., and Ehrlich, S. D. (2001) *EMBO J.* **20**, 2587-2595

176. Samadashwily, G. M., Raca, G., and Mirkin, S. M. (1997) *Nat. Genet.* **17**, 298-304

177. Pelletier, R., Krasilnikova, M. M., Samadashwily, G. M., Lahue, R., and Mirkin, S. M. (2003) *Mol. Cell. Biol.* **23**, 1349-1357

178. Krasilnikova, M. M., and Mirkin, S. M. (2004) *Mol. Cell. Biol.* **24**, 2286-2295

179. Schweitzer, J. K., and Livingston, D. M. (1999) *Genetics* **152**, 953-963

180. Karthikeyan, R., Vonarx, E. J., Straffon, A. F., Simon, M., Faye, G., and Kunz, B. A. (2000) *J. Mol. Biol.* **299**, 405-419

181. Iyer, R. R., Pluciennik, A., Rosche, W. A., Sinden, R. R., and Wells, R. D. (2000) *J. Biol. Chem.* **275**, 2174-2184

182. Jin, Y. H., Obert, R., Burgers, P. M., Kunkel, T. A., Resnick, M. A., and Gordenin, D. A. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 5122-5127

183. Callahan, J. L., Andrews, K. J., Zakian, V. A., and Freudenreich, C. H. (2003) *Mol. Cell. Biol.* **23**, 7849-7860

184. Ireland, M. J., Reinke, S. S., and Livingston, D. M. (2000) *Genetics* **155**, 1657-1665

185. Bae, S. H., Bae, K. H., Kim, J. A., and Seo, Y. S. (2001) *Nature* **412**, 456-461

186. Bhattacharyya, S., and Lahue, R. S. (2004) *Mol. Cell. Biol.* **24**, 7324-7330

187. Wang, W., and Bambara, R. A. (2005) *J. Biol. Chem.* **280**, 5391-5399

188. Rosche, W. A., Jaworski, A., Kang, S., Kramer, S. F., Larson, J. E., Geidroc, D. P., Wells, R. D., and Sinden, R. R. (1996) *J. Bacteriol.* **178**, 5042-5044.

189. Lahue, R. S., and Slater, D. L. (2003) *Front. Biosci.* **8**, 653-665

190. Richards, R. I., and Sutherland, G. R. (1992) *Cell* **70**, 709-712

191. Kremer, B., Almqvist, E., Theilmann, J., Spence, N., Telenius, H., Goldberg, Y. P., and Hayden, M. R. (1995) *Am. J. Hum. Genet.* **57**, 343-350

192. Kovtun, I. V., Goellner, G., and McMurray, C. T. (2001) *Biochem. Cell. Biol.* **79**, 325-336

193. Jakupciak, J. P., and Wells, R. D. (2000) *J. Biol. Chem.* **275**, 40003-40013

194. Arai, N., Akiyama, R., Niimi, N., Nakatsubo, H., and Inoue, T. (1999) *Genes Genet. Syst.* **74**, 159-167

195. Cohen, H., Sears, D. D., Zenvirth, D., Hieter, P., and Simchen, G. (1999) *Mol. Cell. Biol.* **19**, 4153-4158

196. Jankowski, C., Nasar, F., and Nag, D. K. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 2134-2139

197. Schweitzer, J. K., Reinke, S. S., and Livingston, D. M. (2001) *Genetics* **159**, 1861-1865

198. Jankowski, C., and Nag, D. K. (2002) *Mol. Genet. Genomics* **267**, 64-70

199. Richard, G. F., and Paques, F. (2000) *EMBO Rep.* **1**, 122-126

200. Richard, G. F., Cyncynatus, C., and Dujon, B. (2003) *J. Mol. Biol.* **326**, 769-782

201. Napierala, M., Parniewski, P. P., Pluciennik, A., and Wells, R. D. (2002) *J. Biol. Chem.* **277**, 34087-34100

202. Pluciennik, A., Iyer, R. R., Napierala, M., Larson, J. E., Filutowicz, M., and Wells, R. D. (2002) *J. Biol. Chem.* **277**, 34074-34086

203. Hebert, M. L., Spitz, L. A., and Wells, R. D. (2004) *J. Mol. Biol.* **336**, 655-672

204. Hashem, V. I., Rosche, W. A., and Sinden, R. R. (2004) *Mutat. Res.* **554**, 95-109

205. Marcadier, J. L., and Pearson, C. E. (2003) *J. Biol. Chem.* **278**, 33848-33856

206. Savouret, C., Brisson, E., Essers, J., Kanaar, R., Pastink, A., te Riele, H., Junien, C., and Gourdon, G. (2003) *EMBO J.* **22**, 2264-2273

207. Sarkar, P. S., Chang, H. C., Boudi, F. B., and Reddy, S. (1998) *Cell* **95**, 531-540

208. Richard, G. F., Goellner, G. M., McMurray, C. T., and Haber, J. E. (2000) *EMBO J.* **19**, 2381-2390

209. Sharples, G. J., and Leach, D. R. (1995) *Mol. Microbiol.* **17**, 1215-1217

210. Connelly, J. C., Kirkham, L. A., and and Leach, D. R. F. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7969-7974

211. D'Amours, D., and Jackson, S. P. (2002) *Nat. Rev. Mol. Cell. Biol.* **3**, 317-327

212. Lobachev, K. S., Gordenin, D. A., and Resnick, M. A. (2002) *Cell* **108**, 183-193

213. Jaworski, A., Rosche, W. A., Gellibolian, R., Kang, S., Shimizu, M., Bowater, R. P., Sinden, R. R., and Wells, R. D. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 11019-11023

214. Schumacher, S., Fuchs, R. P. P., and Bichara, M. (1998) *J. Mol. Biol.* **279**, 1101-1110

215. Parniewski, P., Jaworski, A., Wells, R. D., and Bowater, R. P. (2000) *J. Mol. Biol.* **299**, 865-874.

216. Schmidt, K. H., Abbott, C. M., and Leach, D. R. (2000) *Mol. Microbiol.* **35**, 463-471

217. Hashem, V. I., Rosche, W. A., and Sinden, R. R. (2002) *Mutat Res* **502**, 25-37

218. Miret, J. J., Pessoabrandao, L., and Lahue, R. S. (1997) *Mol. Cell. Biol.* **17**, 3382-3387

219. Schweitzer, J. K., and Livingston, D. M. (1997) *Hum. Mol. Genet.* **6**, 349-355

220. Maurer, D. J., O'Callaghan, B. L., and Livingston, D. M. (1998) *Mol. Cell. Biol.* **18**, 4597-4604

221. Manley, K., Shirley, T. L., Flaherty, L., and Messer, A. (1999) *Nat. Genet.* **23**, 471-473

222. van den Broek, W. J., Nelen, M. R., Wansink, D. G., Coerwinkel, M. M., te Riele, H., Groenen, P. J., and Wieringa, B. (2002) *Hum. Mol. Genet.* **11**, 191-198

223. Gomes-Pereira, M., Fortune, M. T., Ingram, L., McAbney, J. P., and Monckton, D. G. (2004) *Hum. Mol. Genet.* **13**, 1815-1825

224. Parniewski, P., Bacolla, A., Jaworski, A., and Wells, R. D. (1999) *Nucleic Acids Res.* **27**, 616-623

225. Bowater, R. P., Jaworski, A., Larson, J. E., Parniewski, P., and Wells, R. D. (1997) *Nucleic Acids Res.* **25**, 2861-2868

226. Mochmann, L. H., and Wells, R. D. (2004) *Nucleic Acids Res.* **32**, 4469-4479

227. Parsons, M. A., Sinden, R. R., and Izban, M. G. (1998) *J. Biol. Chem.* **273**, 26998-27008

228. Lalioti, M. D., Scott, H. S., Buresi, C., Rossier, C., Bottani, A., Morris, M. A., Malafosse, A., and Antonarakis, S. E. (1997) *Nature* **386**, 847-851

229. Lalioti, M. D., Scott, H. S., Genton, P., Grid, D., Ouazzani, R., M'Rabet, A., Ibrahim, S., Gouider, R., Dravet, C., Chkili, T., Bottani, A., Buresi, C., Malafosse, A., and Antonarakis, S. E. (1998) *Am. J. Hum. Genet.* **62**, 842-847

230. Larson, G. P., Ding, S., Lafreniere, R. G., Rouleau, G. A., and Krontiris, T. G. (1999) *Hum. Mol. Genet.* **8**, 1985-1988

231. Jithesh, P. V., Singh, P., and Joshi, R. (2001) *J. Biomol. Struct. Dyn.* **19**, 479-495

232. Pataskar, S. S., Dash, D., and Brahmachari, S. K. (2001) *J. Biomol. Struct. Dyn.* **19**, 307-313

233. Pataskar, S. S., Dash, D., and Brahmachari, S. K. (2001) *J. Biomol. Struct. Dyn.* **19**, 293-305

234. Saha, T., and Usdin, K. (2001) *FEBS Lett.* **491**, 184-187

235. Matsuura, T., Yamagata, T., Burgess, D. L., Rasmussen, A., Grewal, R. P., Watase, K., Khajavi, M., McCall, A. E., Davis, C. F., Zu, L., Achari, M., Pulst,

S. M., Alonso, E., Noebels, J. L., Nelson, D. L., Zoghbi, H. Y., and Ashizawa, T. (2000) *Nat. Genet.* **26**, 191-194

236. Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., Day, J. W., and Ranum, L. P. W. (2001) *Science* **293**, 864-867

237. Meola, G., and Moxley, R. T., 3rd. (2004) *J. Neurol.* **251**, 1173-1182

238. Schoser, B. G., Kress, W., Walter, M. C., Halliger-Keller, B., Lochmuller, H., and Ricker, K. (2004) *Brain* **127**, 1868-1877

239. Thornton, C. A., Griggs, R. C., and Moxley, R. T., III. (1994) *Ann. Neurol.* **35**, 269-272

240. Ricker, K., Koch, M. C., Lehmann-Horn, F., Pongratz, D., Otto, M., Heine, R., and Moxley, R. T., 3rd. (1994) *Neurology* **44**, 1448-1452

241. Udd, B., Krahe, R., Wallgren-Pettersson, C., Falck, B., and Kalimo, H. (1997) *Neuromuscul. Disord.* **7**, 217-228

242. Ranum, L. P., Rasmussen, P. F., Benzow, K. A., Koob, M. D., and Day, J. W. (1998) *Nat. Genet.* **19**, 196-198

243. Meola, G., Sansone, V., Perani, D., Colleluori, A., Cappa, S., Cotelli, M., Fazio, F., Thornton, C. A., and Moxley, R. T. (1999) *Neurology* **53**, 1042-1050

244. Ricker, K., Grimm, T., Koch, M. C., Schneider, C., Kress, W., Reimers, C. D., Schulte-Mattler, W., Mueller-Myhsok, B., Toyka, K. V., and Mueller, C. R. (1999) *Neurology* **52**, 170-171

245. Miller, J. W., Urbinati, C. R., Teng-Umnuay, P., Stenberg, M. G., Byrne, B. J., Thornton, C. A., and Swanson, M. S. (2000) *EMBO J.* **19**, 4439-4448

246. Fardaei, M., Rogers, M. T., Thorpe, H. M., Larkin, K., Hamshere, M. G., Harper, P. S., and Brook, J. D. (2002) *Hum. Mol. Genet.* **11**, 805-814

247. Ranum, L. P., and Day, J. W. (2004) *Trends Genet.* **20**, 506-512

248. Timchenko, L. T., Timchenko, N. A., Caskey, C. T., and Roberts, R. (1996) *Hum. Mol. Genet.* **5**, 115-121

249. Michalowski, S., Miller, J. W., Urbinati, C. R., Paliouras, M., Swanson, M. S., and Griffith, J. (1999) *Nucleic Acids Res.* **27**, 3534-3542

250. Fardaei, M., Larkin, K., Brook, J. D., and Hamshere, M. G. (2001) *Nucleic Acids Res.* **29**, 2766-2771

251. Mankodi, A., Urbinati, C. R., Yuan, Q. P., Moxley, R. T., Sansone, V., Krym, M., Henderson, D., Schalling, M., Swanson, M. S., and Thornton, C. A. (2001) *Hum. Mol. Genet.* **10**, 2165-2170

252. Moxley, R. T., 3rd, and Ricker, K. (1995) *Muscle Nerve* **18**, 557-558

253. Finsterer, J. (2002) *Eur. J. Neurol.* **9**, 441-447

254. Mankodi, A., and Thornton, C. A. (2002) *Curr. Opin. Neurol.* **15**, 545-552

255. Day, J. W., Ricker, K., Jacobsen, J. F., Rasmussen, L. J., Dick, K. A., Kress, W., Schneider, C., Koch, M. C., Beilman, G. J., Harrison, A. R., Dalton, J. C., and Ranum, L. P. W. (2003) *Neurology* **60**, 657-664

256. Cummings, C. J., and Zoghbi, H. Y. (2000) *Hum. Mol. Genet.* **9**, 909-916

257. Sinden, R. R., Potaman, V. N., Oussatcheva, E. A., Pearson, C. E., Lyubchenko, Y. L., and Shlyakhtenko, L. S. (2002) *J. Biosci.* **27**, 53-65

258. Jakupciak, J. P., and Wells, R. D. (1999) *J. Biol. Chem.* **274**, 23468-23479

259. Schlotterer, C., and Tautz, D. (1992) *Nucleic Acids Res.* **20**, 211-215

260. Ji, J., Clegg, N. J., Peterson, K. R., Jackson, A. L., Laird, C. D., and Loeb, L. A. (1996) *Nucleic Acids Res.* **24**, 2835-2840

261. Petruska, J., Hartenstine, M. J., and Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 5204-5210

262. Sinden, R. R. (2001) *Nature* **411**, 757-758

263. Darlow, J. M., and Leach, D. R. F. (1998) *J. Mol. Biol.* **275**, 3-16

264. Wells, R. D., Bacolla, A., and Bowater, R. P. (1998) *Instabilities of Triplet Repeats: Factors and Mechanisms* (Oostra, B. A., Ed.), 21, Springer Press Inc., Berlin

265. Krasilnikov, A. S., Panyutin, I. G., Samadashwily, G. M., Cox, R., Lazurkin, Y. S., and Mirkin, S. M. (1997) *Nucleic Acids Res.* **25**, 1339-1346

266. Kramer, P. R., Pearson, C. E., and Sinden, R. R. (1996) *Hum. Genet.* **98**, 151-157

267. Richard, G. F., Dujon, B., and Haber, J. E. (1999) *Mol. Gen. Genet.* **261**, 871-882

268. Nasar, F., Jankowski, C., and Nag, D. K. (2000) *Mol. Cell. Biol.* **20**, 3449-3458

269. Vispe, S., Ho, E. L., Yung, T. M., and Satoh, M. S. (2003) *J. Biol. Chem.* **278**, 35279-35285

270. Cemal, C. K., Huxley, C., and Chamberlain, S. (1999) *Gene* **236**, 53-61

271. Zhou, Z. H., Akgun, E., and Jasin, M. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8326-8333.

272. Deb, S., DeLucia, A. L., Baur, C. P., Koff, A., and Tegtmeyer, P. (1986) *Mol. Cell. Biol.* **6**, 1663-1670

273. Gluzman, Y. (1981) *Cell* **23**, 175-182

274. Napierala, M., and Krzyzosiak, W. J. (1997) *J. Biol. Chem.* **272**, 31079-31085

275. Hanvey, J. C., Shimizu, M., and Wells, R. D. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 6292-6296

276. Hanvey, J. C., Shimizu, M., and Wells, R. D. (1989) *J. Biol. Chem.* **264**, 5950-5956

277. Shimizu, M., Hanvey, J. C., and Wells, R. D. (1989) *J. Biol. Chem.* **264**, 5944-5949

278. Panyutin, I. G., and Wells, R. D. (1992) *J. Biol. Chem.* **267**, 5495-5501

279. Kohwi-Shigematsu, T., and Kohwi, Y. (1992) *Methods Enzymol.* **212**, 155-180

280. Ohshima, K., Kang, S., Larson, J. E., and Wells, R. D. (1996) *J. Biol. Chem.* **271**, 16784-16791

281. Sasse-Dwight, S., and Gralla, J. D. (1989) *J. Biol. Chem.* **264**, 8074-8081

282. Jiang, H., Zacharias, W., and Amirhaeri, S. (1991) *Nucleic Acids Res.* **19**, 6943-6948

283. Kang, S., Wohlrab, F., and Wells, R. D. (1992) *J. Biol. Chem.* **267**, 19435-19442

284. Rahmouni, A. R., and Wells, R. D. (1989) *Science* **246**, 358-363

285. Shimizu, M., Hanvey, J. C., and Wells, R. D. (1990) *Biochemistry* **29**, 4704-4713

286. Palecek, E. (1992) *Methods Enzymol.* **212**, 139-155

287. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd, Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

288. Kang, S., and Wells, R. D. (1992) *J. Biol. Chem.* **267**, 20887-20891

289. Chan, H. W., and Wells, R. D. (1974) *Nature* **252**, 205-209.

290. Amirhaeri, S., Wohlrab, F., Major, E. O., and Wells, R. D. (1988) *J. Virol.* **62**, 922-931

291. Ohshima, K., Sakamoto, N., Labuda, M., Poirier, J., Moseley, M. L., Montermini, L., Ranum, L. P. W., Wells, R. D., and Pandolfo, M. (1999) *Neurology* **53**, 1854-1857

292. Chan, H. W., Dodgson, J. B., and Wells, R. D. (1977) *Biochemistry* **16**, 2356-2366

293. Pearson, C. E., and Sinden, R. R. (1996) *Biochemistry* **35**, 5041-5053

294. Bullock, P. A. (1997) *Crit. Rev. Biochem. Mol. Biol.* **32**, 503-568

295. Heinzel, S. S., Krysan, P. J., Calos, M. P., and DuBridge, R. B. (1988) *J. Virol.* **62**, 3738-3746

296. Dutch, R. E., Bianchi, V., and Lehman, I. R. (1995) *J. Virol.* **69**, 3084-3089

297. Van Craenenbroeck, K., Vanhoenacker, P., and Haegeman, G. (2000) *Eur. J. Biochem.* **267**, 5665-5678

298. Kang, S., Ohshima, K., Jaworski, A., and Wells, R. D. (1996) *J. Mol. Biol.* **258**, 543-547

299. La Spada, A. R. (1997) *Brain Pathol.* **7**, 943-963

300. McMurray, C. T. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 1823-1825

301. Mitas, M., Yu, A., Dill, J., and Haworth, I. S. (1995) *Biochemistry* **34**, 12803-12811

302. Wohlrab, F. (1992) *Methods Enzymol.* **212**, 294-301

303. Davison, A., and Leach, D. R. (1994) *Nucleic Acids Res.* **22**, 4361-4363

304. Darlow, J. M., and Leach, D. R. F. (1995) *Genetics* **141**, 825-832

305. Huertas, D., and Azorin, F. (1996) *Biochemistry* **35**, 13125-13135

306. Tibanyenda, N., De Bruin, S. H., Haasnoot, C. A., van der Marel, G. A., van Boom, J. H., and Hilbers, C. W. (1984) *Eur. J. Biochem.* **139**, 19-27

307. Nenguke, T., Aladjem, M. I., Gusella, J. F., Wexler, N. S., and Arnheim, N. (2003) *Hum. Mol. Genet.* **12**, 1021-1028

308. Anderson, S., and DePamphilis, M. L. (1979) *J. Biol. Chem.* **254**, 11495-11504

309. Henricksen, L. A., Veeraraghavan, J., Chafin, D. R., and Bambara, R. A. (2002) *J. Biol. Chem.* **277**, 22361-22369

310. Veeraraghavan, J., Rossi, M. L., and Bambara, R. A. (2003) *J. Biol. Chem.* **278**, 42854-42866

311. Waga, S., and Stillman, B. (1998) *Annu. Rev. Biochem.* **67**, 721-751

312. Beilina, A., Tassone, F., Schwartz, P. H., Sahota, P., and Hagerman, P. J. (2004) *Hum. Mol. Genet.* **13**, 543-549

313. La Spada, A. R., Paulson, H. L., and Fischbeck, K. H. (1994) *Ann. Neurol.* **36**, 814-822

314. Savouret, C., Garcia-Cordier, C., Megret, J., te Riele, H., Junien, C., and Gourdon, G. (2004) *Mol. Cell. Biol.* **24**, 629-637

315. Jakupciak, J. P., and Wells, R. D. (2000) *IUBMB Life* **50**, 355-359

316. Bierne, H., Ehrlich, S. D., and Michel, B. (1997) *EMBO J.* **16**, 3332-3340.

317. Saintigny, Y., Delacote, F., Vares, G., Petitot, F., Lambert, S., Averbeck, D., and Lopez, B. S. (2001) *EMBO J.* **20**, 3861-3870

318. Gonzalez-Barrera, S., Cortes-Ledesma, F., Wellinger, R. E., and Aguilera, A. (2003) *Mol. Cell* **11**, 1661-1671

319. Read, L. R., Raynard, S. J., Ruksc, A., and Baker, M. D. (2004) *Nucleic Acids Res.* **32**, 1184-1196

320. Saleh-Gohari, N., and Helleday, T. (2004) *Nucleic Acids Res.* **32**, 3683-3688

321. Dudas, A., and Chovanec, M. (2004) *Mutat. Res.* **566**, 131-167

322. Hebert, M. L., and Wells, R. D. (2005) *J. Mol. Biol.* **353**, 961-979

323. Kuzminov, A. (1995) *Mol. Microbiol.* **16**, 373-384

324. Kanaar, R., Hoeijmakers, J. H., and van Gent, D. C. (1998) *Trends Cell Biol.* **8**, 483-489

325. Jeggo, P. A. (1998) *Adv. Genet.* **38**, 185-218

326. Paques, F., and Haber, J. E. (1999) *Microbiol. Mol. Biol. Rev.* **63**, 349-404

327. Cox, M. M., Goodman, M. F., Kreuzer, K. N., Sherratt, D. J., Sandler, S. J., and Marians, K. J. (2000) *Nature* **404**, 37-41

328. Dronkert, M. L., Beverloo, H. B., Johnson, R. D., Hoeijmakers, J. H., Jasin, M., and Kanaar, R. (2000) *Mol. Cell Biol.* **20**, 3147-3156

329. Pastink, A., Eeken, J. C., and Lohman, P. H. (2001) *Mutat. Res.* **480-481**, 37-50

330. van den Bosch, M., Lohman, P. H., and Pastink, A. (2002) *Biol. Chem.* **383**, 873-892

331. Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D., and Rehrauer, W. M. (1994) *Microbiol. Rev.* **58**, 401-465

332. Kuzminov, A. (1999) *Microbiol. Mol. Biol. Rev.* **63**, 751-813

333. Pastink, A., and Lohman, P. H. (1999) *Mutat. Res.* **428**, 141-156

334. Cox, M. M. (1999) *Prog. Nucleic Acid Res. Mol. Biol.* **63**, 311-366

335. Symington, L. S. (2002) *Microbiol. Mol. Biol. Rev.* **66**, 630-670

336. Tsukamoto, Y., and Ikeda, H. (1998) *Genes Cells* **3**, 135-144

337. Critchlow, S. E., and Jackson, S. P. (1998) *Trends Biochem. Sci.* **23**, 394-398

338. Lambert, S., Saintigny, Y., Delacote, F., Amiot, F., Chaput, B., Lecomte, M., Huck, S., Bertrand, P., and Lopez, B. S. (1999) *Mutat. Res.* **433**, 159-168

339. Kuzminov, A. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8461-8468

340. Aravind, L., and Koonin, E. V. (2001) *Genome Res.* **11**, 1365-1374

341. Weller, G. R., Kysela, B., Roy, R., Tonkin, L. M., Scanlan, E., Della, M., Devine, S. K., Day, J. P., Wilkinson, A., d'Adda di Fagagna, F., Devine, K. M., Bowater, R. P., Jeggo, P. A., Jackson, S. P., and Doherty, A. J. (2002) *Science* **297**, 1686-1689

342. Kuzminov, A. (1995) *J. Theor. Biol.* **177**, 29-43

343. Morel, P., Reverdy, C., Michel, B., Ehrlich, S. D., and Cassuto, E. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10003-10008

344. Courcelle, J., Khodursky, A., Peter, B., Brown, P. O., and Hanawalt, P. C. (2001) *Genetics* **158**, 41-64

345. Courcelle, J., and Hanawalt, P. C. (2003) *Annu. Rev. Genet.* **37**, 611-646

346. Hishida, T., Han, Y. W., Shibata, T., Kubota, Y., Ishino, Y., Iwasaki, H., and Shinagawa, H. (2004) *Genes Dev.* **18**, 1886-1897

347. Bi, X., and Liu, L. F. (1994) *J. Mol. Biol.* **235**, 414-423

348. Bi, X., Lyu, Y. L., and Liu, L. F. (1995) *J. Mol. Biol.* **247**, 890-902

349. Bi, X., and Liu, L. F. (1996) *Prog. Nucleic Acid Res. Mol. Biol.* **54**, 253-292

350. Bi, X., and Liu, L. F. (1996) *J. Mol. Biol.* **256**, 849-858

351. Morag, A. S., Saveson, C. J., and Lovett, S. T. (1999) *J. Mol. Biol.* **289**, 21-27

352. Miller, J. H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

353. Bowater, R. P., Rosche, W. A., Jaworski, A., Sinden, R. R., and Wells, R. D. (1996) *J. Mol. Biol.* **264**, 82-96.

354. Yamamoto, K., Takahashi, N., Fujitani, Y., Yoshikura, H., and Kobayashi, I. (1996) *Genetics* **143**, 27-36

355. Lyu, Y. L., Lin, C. T., and Liu, L. F. (1999) *J. Mol. Biol.* **285**, 1485-1501

356. Napierala, M., Bacolla, A., and Wells, R. D. (2005) *J. Biol. Chem.* **280**, 37366-37376

APPENDIX

AUTHORSHIP AND COPYRIGHT INFORMATION

Chapter II

The studies were designed by the candidate and Drs. Marek Napierala and Robert D. Wells. The CCTG•CAGG repeat containing plasmids were kind gifts from Dr. Laura P. W. Ranum (University of Minnesota). The cloning of the plasmids used in the study, the culturing of the COS-7 cells and the chemical and enzymatic probe studies were conducted by the candidate and Dr. Marek Napierala. The single colony analyses were performed by the candidate. The manuscript was composed and edited primarily by the candidate and Dr. Wells, with inputs from Drs. Marek Napierala and Laura P. W. Ranum.

Chapter II was reprinted from "Hairpin Structure-forming Propensity of the CCTG•CAGG Tetranucleotide Repeats Contributes to the Genetic Instability Associated with Myotonic Dystrophy Type 2" by Ruhee Dere, Marek Napierala, Laura P. W. Ranum and Robert D. Wells (2004) Journal of Biological Chemistry, volume 279, pages 41715-41726. Copyright permission was obtained from the American Society for Biochemistry and Molecular Biology, Inc., Bethesda, MD.

Chapter III

The studies presented in chapter III were designed by the candidate and Dr. Robert D. Wells. All the experiments were conducted by the candidate. The manuscript

entitled “The Expansion Prone Recombination Hotspot Activity of the CCTG•CAGG Repeats in *Escherichia coli*” is being prepared for submission to the Journal of Molecular Biology for publication. The authors are Ruhee Dere and Robert D. Wells and the manuscript has being composed and edited by the candidate and Dr. Wells.

VITA

Name Ruhee J. Dere

Address Institute of Biosciences and Technology
2121 W. Holcombe Blvd.
Houston, TX 77030

Email address rdere@ibt.tamhsc.edu

Education

1995 – 1998	University of Mumbai, Mumbai, India Bachelor of Science Major subject: Life Sciences
1998 – 2000	University of Mumbai, Mumbai, India Master of Science Major subjects: Life Sciences and Biotechnology
2000 – 2006	Texas A&M University Doctor of Philosophy Major subject: Genetics

Previously published material

1. Marek Napierala, Ruhee Dere, Alexandre A. Vetcher and Robert D. Wells (2004) *J. Biol. Chem.* **279** (8): 6444 – 6454
2. Ruhee Dere, Marek Napierala, Laura P. W. Ranum and Robert D. Wells (2004) *J. Biol. Chem.* **279** (40): 41715 – 41726
3. Robert D. Wells, Ruhee Dere, Micheal Hebert, Marek Napierala and Leslie Spitz-Son (2005) *Nucleic Acids Res.* **33** (12): 3785 – 3798

Book Chapter

R. Dere, M. Hebert and M. Napierala (2006) Involvement of Genetic Recombination in Microsatellite Instability. In Wells, R. D. and Ashizawa, T. (eds.) *Genetic Instabilities and Hereditary Neurological Diseases*, Volume 2, Elsevier-Academic Press, Inc., San Diego (in press).