Hairpin Structure-Forming Propensity of the (CCTG•CAGG) Tetranucleotide Repeats Contributes to the Genetic Instability Associated with Myotonic Dystrophy Type 2 *

Ruhee Dere§, Marek Napierala§, Laura P. W. Ranum‡, and Robert D. Wells§

From the
§ Institute of Biosciences and Technology
Center for Genome Research
Texas A&M University System Health Science Center
Texas Medical Center
2121 Holcombe Blvd.
Houston, TX 77030-3303
and
‡ University of Minnesota,
Department of Genetics, Cell Biology and Development,
Minneapolis, MN 55455
Instability of DM2 CCTG•CAGG Repeats
SUMMARY

The genetic instabilities of \((\text{CCTG}\bullet\text{CAGG})_n\) 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 \(\text{CTG}\bullet\text{CAG}\) repeats related to DM1, the instabilities were greater for the longer tetranucleotide repeat tracts. Also, the expansions and deletions predominated when 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 \(\text{d(CAGG)}_{26}\) and \(\text{d(CCTG)}_{26}\) as models of unpaired regions of the replication fork revealed that \(\text{d(CAGG)}_{26}\) has a marked propensity to adopt a defined base paired hairpin structure whereas the complementary \(\text{d(CCTG)}_{26}\) lacks this capacity. The effect of orientation described above differs from all previous results with three triplet repeat sequences (TRS) (including \(\text{CTG}\bullet\text{CAG}\)) 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.
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 (1). 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 (2-6). 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 (1). 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 (7-13). Additionally, two other non-triplet repeat neurological diseases, spinocerebellar ataxia type 10 (SCA10) (14) and progressive myoclonus epilepsy of Unverricht-Lundborg type (EPM1) (15), are caused by the expansions of pentanucleotide and dodecanucleotide repeating sequences, respectively.

Replication (9,10,16-23), recombination (24-28), and repair (10,21,29,30) were shown to be responsible for the instabilities of triplet repeat sequences (TRS). Slippage of the repeats (31-34) as promoted by non-B DNA structures (9-11,35-37) formed by these repeating sequences causes polymerases to pause during replication, as shown both
in vivo as well as in vitro (17,20,38-43), thereby generating instabilities. Furthermore, these structures are also recognized by mismatch repair (MMR) (29,44-48) and nucleotide excision repair (NER) (49,50); 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 (21,30,51-56). Triplet repeat sequences are hotspots for recombination, which may account for the massive expansions found in certain diseases (24-28,57,58).

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.
EXPERIMENTAL PROCEDURES

Construction of the (CCTG•CAGG)$_n$ Containing Shuttle Vector – The (CCTG•CAGG) tetranucleotide repeats (1) 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 (100µg/ml). Individual colonies were
grown in LB broth containing ampicillin (100µg/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 \((\text{CCTG}\cdot\text{CAGG})_n\) inserts were characterized using restriction mapping and DNA sequencing. An \(\text{EcoRI}\) digest was used for repeat tracts cloned proximal (\(\text{SmaI}\) site) to the SV40 origin of replication and an \(\text{AflIII/BglII}\) digest was used for inserts cloned into the site distal (\(\text{MfeI}\) site) to the same origin of replication. The restriction fragments were end-labeled with \([\alpha^{32}\text{P}]\) 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 \((\text{CCTG}\cdot\text{CAGG})_n\) 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 \((\text{CCTG}\cdot\text{CAGG})_{30}\) is a pure repeat (i.e. it contains no polymorphisms/interruptions) as determined by sequencing of the entire repeat containing tract. The \((\text{CCTG}\cdot\text{CAGG})_{114}\) and the \((\text{CCTG}\cdot\text{CAGG})_{200}\) carried a single bp interruption 11 repeats into the tract to give the sequence \((\text{CCTG})_{11}\text{CCTT(CTG)}_n\) where \(n = 102\) and 188, respectively. The \((\text{CCTG}\cdot\text{CAGG})_n\) (where \(n = 114\) or 200) repeats were sequenced using primers located at both ends of the repeating tract. In case of the \((\text{CCTG}\cdot\text{CAGG})_{114}\), 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 \((C\text{CTG}\cdot\text{CAGG})_{200}\) 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 \((C\text{CTG}\cdot\text{CAGG})_n\) repeat tracts in both orientations with respect to the SV40 origin of replication (Fig. 1). However, the SV40 origin is bi-directional (59) and is the operative origin in COS-7 cells (60). 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 \((C\text{CTG}\cdot\text{CAGG})_n\) Repeat Containing Plasmids into COS-7 Cells –*

The plasmids containing the \((C\text{CTG}\cdot\text{CAGG})_n\) 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 *Dpn*I at 37°C for 2 h. *Dpn*I 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 *Dpn*I cleavage was assessed by digesting the parental plasmids with *Dpn*I, 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 *Dpn*I. The DNA after the *Dpn*I 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 *Eco*RI digest for plasmids carrying the repeats cloned into the *Sma*I site whereas an *Afl*III/*Bgl*II digest was used to excise the inserts cloned into the *Mfe*I 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-32P] 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 (OsO₄) (Aldrich), potassium permanganate (KMnO₄) (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 x 10⁵ cpm/reaction) in 10mM Tris, 40mM NaCl and 10mM MgCl₂ 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 (61). 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 MgCl₂ (pH 7.0) at 25°C for 60 min (40,62-67). The KMnO₄ reaction was performed at 20°C for 30 min in a buffer containing 50mM sodium cacodylate and 1mM EDTA (pH 7.0) (65,68-71). A final concentration of 8% DEPC and 0.1, 0.5 and 1mM KMnO₄ was used. The chemical modification with 1mM OsO₄ 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 (64,65,67,72-74). The modified oligonucleotides were then ethanol precipitated and cleaved at the
phosphodiester bonds using 1M piperidine (67,72,73,75,76). A series of ethanol precipitations and lyophilization steps then were used to ensure removal of the piperidine.

The enzymatic probes used included S1 nuclease (77,78) (Gibco BRL), P1 nuclease (65,71,79-81) (Gibco BRL) and mung bean nuclease (82-84) (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 MgCl₂ (pH 7.2). 1mM ZnCl₂ was included in the buffer used for the S1 nuclease reactions (75,78). 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 details of the actual concentrations are in the legends to Figs. 5 and 6. The reactions were stopped by addition of a urea-EDTA-dye solution (61) followed by quick freezing on dry ice.

The products of the chemical and enzymatic reactions were then analyzed on 10% denaturing polyacrylamide 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 (40,63,75). The gels were analyzed using a Molecular Dynamics – Storm 820 (Amersham Biosciences) phosphor-imager and quantitative analyses were performed using ImageQuant version 5.1.
RESULTS

Strategy of Study – The (CCTG•CAGG)<sub>n</sub> repeats were cloned either proximal (Smal site) or distal (MfeI site) to the SV40 origin of replication in the pcDNA3.1 shuttle vector (Fig. 1). An African green monkey kidney cell culture system was used to determine the genetic instability of the (CCTG•CAGG)<sub>n</sub> repeats associated with myotonic dystrophy type 2 (DM2). COS-7 cells were transfected with the tetranucleotide repeat containing plasmids (Fig. 1) 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.

To evaluate the effect of the length of the (CCTG•CAGG)<sub>n</sub> 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 (10,16,18,19,21,39,85,86). Therefore, we studied plasmids carrying the (CCTG•CAGG) repeats in both orientations. The distance of the (CCTG•CAGG)<sub>n</sub> repeats from the origin of replication may also play an important role in determining the genetic instabilities of these repeats (Iyer and Wells, unpublished work) (22). Thus, the repeats were cloned either proximal [~ 74 bp from the center of the SV40 core origin (64 bp in length) (87) at the Smal 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. 1 and 2) 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.

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. 3). 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. 2, 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. 2). Thus, the percentages of both expansions and deletions as depicted in the bar graphs in Fig. 4 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. 3): the unchanged starting length tetracyclicidate 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. 3B, lane 5). Previous studies (88-90) 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 property of the vector rather than a result of the presence of the (CCTG•CAGG)\textsubscript{n} inserts, the pcDNA3.1 vector (lacking the (CCTG•CAGG)\textsubscript{n} repeats) was transfected into COS-7 cells. The cells were cultured for 2 weeks and the episomal DNA isolated, \textit{DpnI} cleaved and transformed into \textit{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)\textsubscript{n} repeats. Furthermore, no rearrangements were observed on analyzing individual colonies obtained from transforming the parental plasmids into \textit{E. coli} HB101. Likewise, rearrangements were not observed previously in other prior instability studies with triplet repeat inserts in \textit{E. coli} HB101 (10,16,18,19,39,86,91-93). This confirms that the rearrangements were due to an intrinsic property of the pcDNA3.1 shuttle vector that carries the SV40 origin of replication.

\textit{Length-dependent Instability of the (CCTG•CAGG)\textsubscript{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 (10,18,21,39,93,94). 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. 4. 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 both 48 h and 2 weeks (Fig. 3A 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 (10,16,18,22,52,85,95,96).

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 bidirectional 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. 4) 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 48h, 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. 4), even for 48h 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. 4); 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 48h 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 1. For the (CCTG•CAGG)n repeats cloned proximal (SmaI site) to the SV40 origin of replication (Table 1A), 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 1A). 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 1). 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\(\cdot\)CAGG)\(_n\) repeats cloned distal (\(MfeI\) site) to the SV40 origin (Table 1B) was 56 repeats or 1.5 fold for pRW5119. The largest increase for (CCTG\(\cdot\)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 1A). Approximately 20 clones containing expansions were chosen at random and the inserts sequenced; all inserts contained the (CCTG\(\cdot\)CAGG) repeat tracts within the limits of our detection.

The plasmids cultured in COS-7 cells for a period of 48h (data not shown) showed expansions ranging between 118 to 144 repeats for (CCTG\(\cdot\)CAGG)\(_{114}\) and between 202 to 370 repeats (up to 1.85 fold) for (CCTG\(\cdot\)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.

\textbf{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\(\cdot\)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) (65,71,77,78,80-82,84,97). 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 (65,71,78,79,97).

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 (77,78,98). On probing d(CAGG)$_{26}$ with S1 nuclease (Fig. 5), the greatest reactivity was observed between the cytosine and adenine residues of the 14$^{th}$ CAGG repeat, followed by weaker cleavage between the adenine and guanine of the 14$^{th}$ 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. 5) but the intensity of cleavage in the proposed stem is weaker than that found in the putative loop (Fig. 5A). Alternatively, on probing the labeled d(CCTG)$_{26}$ (Fig. 6), 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 (65,71,79-81,98). 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 13\textsuperscript{th} CAGG repeat and between the adenine and guanine of the 14\textsuperscript{th} CAGG repeat. Furthermore, cleavage was also observed between the cytosine and adenine residues of the 14\textsuperscript{th} CAGG repeat (Fig. 5). 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 in the stem) (Fig. 5). d(CCTG)\textsubscript{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. 6), again indicating that the d(CCTG)\textsubscript{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 (82-84,98). The probing of d(CAGG)\textsubscript{26} with mung bean nuclease showed predominant cleavage between the first and second guanines of the 13\textsuperscript{th} CAGG repeat, the cytosine and adenine, as well as the first and second guanine residues of the 14\textsuperscript{th} CAGG repeat. Less cleavage was observed between the adenine and guanine residues of the 14\textsuperscript{th} CAGG repeat (Fig. 5). These residues form the proposed terminal hairpin loop. The bonds between the G’s of each of the CAGG repeats (Fig. 5) 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)\textsubscript{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)\textsubscript{26} as opposed to d(CCTG)\textsubscript{26}.

*Chemical Probe Determinations* – d(CAGG)\textsubscript{26} and d(CCTG)\textsubscript{26} were individually probed with OsO\textsubscript{4} or DEPC or KMnO\textsubscript{4}. OsO\textsubscript{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 (39,64,65,72-74). OsO\textsubscript{4}/piperidine was used to probe d(CCTG)\textsubscript{26} (Fig. 6). 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 (39,40,62-64,66,67). DEPC/piperidine was used to probe both d(CAGG)\textsubscript{26} and d(CCTG)\textsubscript{26}. Similar to the enzymatic probing, the most effective modification occurred at the adenine and the first guanine residues of the 13\textsuperscript{th} CAGG repeat and at the adenine of the 14\textsuperscript{th} CAGG repeat (Fig. 5) 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. 5) as previously stated. In the case of d(CCTG)\textsubscript{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. 6).
KMnO$_4$ preferentially oxidizes unpaired or unstacked thymines (and to a much lesser extent cytosines) resulting in strand cleavage upon subsequent treatment with piperidine (65,68,69,71,76). 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)$_{26}$, KMnO$_4$ 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.
DISCUSSION

DM2 is caused by the expansion of a (CCTG•CAGG) tetranucleotide repeat (1) 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 (Fig. 8).

Prior genetic instability studies on (CTG•CAG), (CGG•CCG) and (GAA•TTC) triplet repeats (8-11,16-18,21,39,86,91-93,99) 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. 7). d(CAGG)$_{26}$ can form hairpin structures with either 6, 5, 4 or 3 unpaired residues in the terminal loop (Fig. 7A, 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. 7B and C, there would be no Watson-Crick pairing making them less stable than the structures shown in Fig. 7A and D. However, the structure formed with 6 unpaired residues in the terminal loop (Fig. 7A) is more stable than that with 3 residues in the loop (Fig. 7D) due to the two non-Watson-Crick A$\cdot$G pairs that flank the Watson-Crick G$\cdot$C pairs (Fig. 7A) compared to the A$\cdot$A and G$\cdot$G oppositions that flank the G$\cdot$C pairs in structure 7D. Although the stability of the mispairs is context dependent, an approximation of the pairing stabilities is G$\cdot$C > A$\cdot$T > G$\cdot$G > G$\cdot$T ≈ G$\cdot$A > A$\cdot$C$^+$ > T$\cdot$T ≈ A$\cdot$A ≈ C$\cdot$C$^+$ > T$\cdot$C ≥ A$\cdot$C ≥ C$\cdot$C (100-106).

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 (107,108) as well as the favored closing of the loop with a 5$'$ pyrimidine and a 3$'$ purine pair (107,108). Thus, our chemical and enzymatic probe results showed the formation of the structure (Fig. 7A) which the above stated rationale predicts to be most stable. Furthermore, DEPC modified both the adenines and guanines at the N7 positions showing that this position is not involved in hydrogen bonding and that the G$\cdot$A pairs are of the G(anti)$\cdot$A(anti) or A(anti)$\cdot$G(anti) types (102,104,109).

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. 7E – H. The structures with 4 and 3 unpaired residues in the terminal loops lack any Watson-Crick pairs (Fig. 7G and
H, respectively) making them less stable than the structures with 6 and 5 residues in the loops (Fig. 7E and F, respectively). Furthermore, the T•T mismatches as well as the C•T mispairs are extremely destabilizing (105,110) 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 SacI 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 (9,10,16,18,21,22,39,52,85,86,95,96). However, the CCTG sequences are genetically unstable in the orientation prone to expand (orientation II) (Fig. 8) as compared to the TRS that are unstable in the orientation prone to delete (orientation II).

Fig. 8 shows a mechanism for the orientation-dependant instability of the (CCTG•CAGG) sequences. In 1995, Kang et al. (16) 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. 8). This is the first case of a “reversed” orientation behavior (9,10). 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 (Iyer and Wells, unpublished work) (16,19,22,111). 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) (22,112), 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 instability observed (12,22,112). 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 (21,113-118).

Furthermore, polymerase switching (PolI/PolIII) has been implicated in *E. coli* to contribute to the genetic instability of triplet repeats (16,47). 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 (87,119). 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 (L. Mochmann and R. D. Wells, unpublished work) (49,93,120-123). The (CCTG•CAGG)_n repeats in our experiments, cloned into the *Sma*I 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 (17,20,23,122). This in turn could lead to activation of repair (10,12,21,29,30,44,45,47,48,52,94,95,124-126) and recombination (24-28) 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, unpublished data). 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.
REFERENCES


81. Sakamoto, N., Chastain, P. D., Parniewski, P., Ohshima, K., Pandolfo, M.,
82. Chan, H. W., Dodgson, J. B., and Wells, R. D. (1977) Biochemistry 16, 2356-
   2366
84. Pearson, C. E., Tam, M., Wang, Y. H., Montgomery, S. E., Dar, A. C., Cleary, J.
   17, 2090-2098
   62, 3738-3746
   Biochem. 267, 5665-5678
   543-547
92. La Spada, A. R. (1997) Brain Pathology 7, 943-963
   Chem. 275, 14588-14595


* This work was supported by grants from the National Institutes of Health (NS37554 and ES11347), the Robert A. Welch Foundation, and the Muscular Dystrophy (Seek a Miracle) Association to R.D.W. and NIH (NS35870) and the Muscular Dystrophy Association to L.P.W.R.

To whom correspondence should be addressed: Center for Genome Research, Institute of Biosciences and Technology, Texas A&M University Health Science Center, 2121 W. Holcombe Blvd. Houston, TX 77030-3303. Tel.: 713-677-7651; Fax: 713-677-7689; E-mail: rwells@ibt.tamushsc.edu.
Legends to Figures

Fig. 1. Plasmids used in the study. The (CCTG•CAGG)<sub>n</sub> 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 bi-directional SV40 origin of replication. Orientations I and II are defined in Experimental Procedures.

Fig. 2. Experimental strategy – Mammalian cell culture assay. The (CCTG•CAGG)<sub>n</sub> 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. 1). 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).

Fig. 3. Genetic instability of the (CCTG•CAGG)<sub>n</sub> 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\textbullet CAGG)\textsubscript{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\textbullet CAGG)\textsubscript{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.

**Fig. 4. Relative amounts of expansions and deletions of (CCTG\textbullet CAGG)\textsubscript{n} repeats.**

The episomal DNA was isolated from COS-7 cells, which were cultured for 2 weeks, and the lengths of the tetranucleotidic 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\textbullet CAGG)\textsubscript{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 gray 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.

Fig. 5. 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 10$^{th}$ 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.

Fig. 6. 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].

Fig. 7. Theoretical hairpin structures formed by d(CAGG) and d(CCTG) oligonucleotides. A portion of the various folded-back structures that can be formed by the d(CAGG) and d(CCTG) oligomers are shown. 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 oligonucleotides of any length. All oligomers are numbered from their 5′-ends.

Fig. 8. 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).
TABLE I

Expansions and deletions of \((\text{CCTG} \cdot \text{CAGG})_n\) 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 \((\text{CCTG} \cdot \text{CAGG})_n\) repeats are shown; the estimated repeat length varied by \(\pm 5\) repeats for the plasmids carrying either 200 or 114 repeats. The \((\text{CCTG} \cdot \text{CAGG})_{30}\) however was sequenced to indicate the precise number of repeats. 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.
TABLE I

Expansions and deletions of \((\text{CCTG}\bullet\text{CAGG})_n\) repeats

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Number of repeats</th>
<th>Orientation</th>
<th>Expansions</th>
<th>Deletions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Inserts cloned proximal to the SV40 origin of replication (SmaI site)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pRW5115</td>
<td>200</td>
<td>I</td>
<td>284 (284)</td>
<td>0-197 (90)</td>
</tr>
<tr>
<td>pRW5114</td>
<td>200</td>
<td>II</td>
<td>202-300 (246)</td>
<td>29-197 (130)</td>
</tr>
<tr>
<td>pRW5113</td>
<td>114</td>
<td>I</td>
<td>117-190 (148)</td>
<td>13-112 (85)</td>
</tr>
<tr>
<td>pRW5112</td>
<td>114</td>
<td>II</td>
<td>116-210 (158)</td>
<td>0-112 (82)</td>
</tr>
<tr>
<td>pRW5111</td>
<td>30</td>
<td>I</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>pRW5110</td>
<td>30</td>
<td>II</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>B. Inserts cloned distal to the SV40 origin of replication (MfeI site)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pRW5121</td>
<td>200</td>
<td>I</td>
<td>206-214 (210)</td>
<td>71-197 (139)</td>
</tr>
<tr>
<td>pRW5120</td>
<td>200</td>
<td>II</td>
<td>200 (200)</td>
<td>9-198 (130)</td>
</tr>
<tr>
<td>pRW5119</td>
<td>114</td>
<td>I</td>
<td>116-170 (133)</td>
<td>0-111 (85)</td>
</tr>
<tr>
<td>pRW5118</td>
<td>114</td>
<td>II</td>
<td>116-124 (119)</td>
<td>16-108 (63)</td>
</tr>
<tr>
<td>pRW5117</td>
<td>30</td>
<td>I</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>pRW5116</td>
<td>30</td>
<td>II</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
(CCTG•CAGG)\textsubscript{n} repeats cloned into \textit{SmaI} site

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
Plasmid & Number of Repeats (n) & Orientation \\
\hline
pRW5110 & 30 & II \\
pRW5111 & 30 & I \\
pRW5112 & 114 & II \\
pRW5113 & 114 & I \\
pRW5114 & 200 & II \\
pRW5115 & 200 & I \\
\hline
\end{tabular}
\end{table}

(CCTG•CAGG)\textsubscript{n} repeats cloned into \textit{MfeI} site

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
Plasmid & Number of Repeats (n) & Orientation \\
\hline
pRW5116 & 30 & II \\
pRW5117 & 30 & I \\
pRW5118 & 114 & II \\
pRW5119 & 114 & I \\
pRW5120 & 200 & II \\
pRW5121 & 200 & I \\
\hline
\end{tabular}
\end{table}

\textbf{Fig. 1}
Plasmid harboring tetranucleotide repeats

Transformation of *E. coli*

Transfection of *COS-7* cells

Isolation of DNA, *DpnI* digestion

Transformation of *E. coli*

Biochemical analyses
- Restriction Analyses
- Southern Hybridization
- Sequencing
(CCTG•CAGG)$_n$ cloned distal to the SV40 origin of replication (*MfeI* site)

<table>
<thead>
<tr>
<th>Expansion Percentage</th>
<th>n = 114</th>
<th>n = 200</th>
<th>Deletion Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>10</td>
<td>10</td>
<td>I</td>
</tr>
<tr>
<td>II</td>
<td>20</td>
<td>20</td>
<td>II</td>
</tr>
</tbody>
</table>

(CCTG•CAGG)$_n$ cloned proximal to the SV40 origin of replication (*SmaI* site)

<table>
<thead>
<tr>
<th>Expansion Percentage</th>
<th>n = 114</th>
<th>n = 200</th>
<th>Deletion Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>30</td>
<td>30</td>
<td>I</td>
</tr>
<tr>
<td>II</td>
<td>40</td>
<td>40</td>
<td>II</td>
</tr>
</tbody>
</table>

Fig. 4
A. \((\text{CAGG})_{26}\)

B. 

Enzymatic Probing

Chemical Probing

Fig. 5
(CCTG)$_{26}$
Fig. 7
Fig. 8

A. Orientation II
EXPANSION

5'CAGGCAGGCAGGCA 3'

3'GTCCGTCCGTCCGT

Leading strand template

Lagging strand template

B. Orientation I
DELETION

5'CCTGCCTGCCTG 3'

3'GGACGGACGGAC 5'

Leading strand template

Lagging strand template
Hairpin structure-forming propensity of the \((\text{CCTG}^5\text{CAGG})\) tetranucleotide repeats contributes to the genetic instability associated with myotonic dystrophy type 2
Ruhee Dere, Marek Napierala, Laura P. W. Ranum and Robert D. Wells

\textit{J. Biol. Chem.} published online July 29, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M406415200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2004/07/29/jbc.M406415200.citation.full.html#ref-list-1