Variant CCG and GGC repeats within the CTG expansion dramatically modify mutational dynamics and likely contribute toward unusual symptoms in some myotonic dystrophy type 1 patients

Claudia Braida¹, Rhoda K.A. Stefanatos¹, Berit Adam¹, Navdeep Mahajan¹, Hubert J.M. Smeets², Florence Niel³, Cyril Goizet³, Benoit Arveiler³, Michel Koenig⁴, Clotilde Lagier-Tourenne⁴, Jean-Louis Mandel⁴, Catharina G. Faber⁵, Christine E.M. de Die-Smulders⁶, Frank Spaans⁷ and Darren G. Monckton¹,*

¹Molecular Genetics, Faculty of Biomedical and Life Sciences, University of Glasgow, University Avenue, Glasgow G12 8QQ, UK, ²Department of Genetics and Cell Biology, Maastricht University, UNS 50, Postvak 16, Postbus 616, 6200 Maastricht, The Netherlands, ³Service de Genetique Medicale, CHU de Bordeaux, France, ⁴Laboratoire de Diagnostic Genetique, CHRU de Strasbourg, Faculte de Medecine, 11 rue Human, 67085 Strasbourg Cedex, France, ⁵Department of Neurology, ⁶Department of Clinical Genetics and ⁷Department of Clinical Neurophysiology, University Hospital Maastricht, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands

Received October 20, 2009; Revised and Accepted January 12, 2010

Myotonic dystrophy type 1 (DM1) is one of the most variable inherited human disorders. It is characterized by the involvement of multiple tissues and is caused by the expansion of a highly unstable CTG repeat. Variation in disease severity is partially accounted for by the number of CTG repeats inherited. However, the basis of the variable tissue-specific symptoms is unknown. We have determined that an unusual Dutch family co-segregating DM1, Charcot-Marie-Tooth neuropathy, encephalopathic attacks and early hearing loss, carries a complex variant repeat at the DM1 locus. The mutation comprises an expanded CTG tract at the 5'-end and a complex array of CTG repeats interspersed with multiple GGC and CCG repeats at the 3'-end. The complex variant repeat tract at the 3'-end of the array is relatively stable in both blood DNA and the maternal germ line, although the 5'-CTG tract remains genetically unstable and prone to expansion. Surprisingly though, even the pure 5'-CTG tract is more stable in blood DNA and the maternal germ line than archetypal DM1 alleles of a similar size. Complex variant repeats were also identified at the 3'-end of the CTG array of ~3–4% of unrelated DM1 patients. The observed polarity and the stabilizing effect of the variant repeats implicate a cis-acting modifier of mutational dynamics in the 3'-flanking DNA. The presence of such variant repeats very likely contributes toward the unusual symptoms in the Dutch family and additional symptomatic variation in DM1 via affects on both RNA toxicity and somatic instability.

Myotonic dystrophy type 1 (DM1) patients display a highly variable phenotype including myotonia, myopathy, cataracts and abnormalities in the heart, brain and endocrine systems (1). The distribution, severity and age of onset of symptoms vary dramatically both within and between families. Indeed, DM1 has been described as the most variable inherited
human disorder (1). The DM1 mutation is the expansion of a CTG repeat in the 3′-untranslated region of the DM protein kinase (DMPK) gene (2–7). Normal alleles at the DM1 locus contain 5 to 35 CTG repeats, whereas pathogenic alleles range from 50 up to more than 1000 repeats (5–7). There is a broad inverse correlation between the number of CTG repeats and the age at onset (8–10). Expanded alleles (>50 repeats) are genetically highly unstable in both the soma and germ line and are highly biased toward further expansion (5–9,11–14). Intergenerational expansions account for the observed anticipation (15) and age-dependent, tissue-specific somatic expansions are assumed to contribute toward the tissue-specificity and progressive nature of the symptoms (16). However, nothing is known about the genetic basis of the considerable inter-individual variability in the range and relative severity of the tissue-specific symptoms.

We previously described a unique Dutch family in which 14 patients over three generations co-segregated DM1 with a fully penetrant intermediate type Charcot-Marie-Tooth (CMT) neuropathy (Fig. 1A; 17–19). A subset of these patients also presented with early hearing loss and recurrent encephalopathic attacks characterized by confusion, decreased consciousness and fever (19). All patients from the DM1+CMT++(DM1+CMT+[early hearing loss]+[encephalopathic attacks]) family presented with classic DM1 features such as cataracts, distal muscle weakness, myotonia and ptosis. The patients also presented with a dominant intermediate CMT with reduced motor conduction velocity (<60% of normal), slight sensory loss in the lower limbs, absence of tendon reflexes and/or pes cavus (17,18). Although a mild and

Figure 1. An atypical DM1 expansion in the DM1+CMT++ family. (A) Pedigree of part of the DM1+CMT++ family. Further details of the clinical phenotypes and a more extended pedigree have been previously presented (17–19). The cross hatches represent obligate carrier status. DNA samples were available from III-9, III-16, III-17, IV-11, IV-12, IV-19, IV-20, IV-21 and IV-22. (B) PCR amplification of an expanded DM1 allele in the DM1+CMT++ family. The DM1 repeat was amplified using primers DM-C and DM-DR (11) in the presence of 10% DMSO, resolved on a 1.5% agarose gel and detected by Southern blot hybridization with a CTG repeat probe. The sizes of the molecular weight standard are indicated on the left and the number of triplet repeats on the right. Individuals IV-11 and IV-12 are both phenotypically normal and have inherited the large ‘38’ repeat normal allele from their affected father. Please note the faint third bands observed above the ‘38’-repeat allele in IV-11 and IV-12, and between the normal and mutant allele in III-16 and IV-22 are heteroduplex products derived from the cross-annealing of complementary strands from two different sized alleles. They are only observed in the latter stages of PCRs with relatively high input DNAs and are not observed in PCRs with lower amounts of input DNA. They are not visible on ethidium stained gels, but hybridize more efficiently due to their partially single stranded nature. (C) The 3′-complex variant repeat tract is stable in the DM1+CMT++ family. The 3′-complex variant repeat tract was amplified from a 1 in 100 dilution of the DM-C/DR reaction (described above) using a hemi-nested PCR with primers DM-GGC (5′-TG CTG CTG CTG GGC GGC G-3′) and DM-DR. The products were resolved on a 2% NuSieve, 1% agarose gel and the DNA stained with ethidium bromide (the image has been inverted to aid visualization). In addition to members of the DM1+CMT++ family, this analysis was also performed on unaffected and classic DM1 patient controls. The sizes of the molecular weight standard are indicated on the left and the number of triplet repeats on the right.
usually sub-clinical polyneuropathy occurs in many DM1 patients (20–22), CMT is not a recognized feature of DM1 (1). However, occasional coincidence of DM with a more severe neuropathy, including in some cases CMT, has been observed (23–25). In contrast to typical DM1 families (8,9,14,15), no congenital cases were detected in the fourth generation, despite four transmissions from two females with adult-onset disease. Interestingly, though, an increase in the severity of symptoms and decrease in the age of onset of CMT was observed between the third and fourth generations (19).

Prior to the identification of the DM1 mutation, linkage analysis in this family showed that the DM and CMT phenotypes co-segregated with the APOC2 locus on chromosome 19 (~1 cm from the DMPK gene) with a maximum LOD score of 7.03 and with zero recombination (18). These data suggested that either a single or two closely linked mutations located near the APOC2 gene caused the DM and CMT. Indeed, subsequent analyses excluded mutations at the common CMT loci, and Southern blot analysis of restriction digested genomic DNA revealed a fragment equivalent to a small expansion (~200–400 CTG repeats) at the DM1 locus (19). However, despite the apparently small size of the expansion, the mutant allele was refractory to PCR amplification even using sensitive single molecule approaches. Base substitutions in control DM1 patients, the DM1 allele in all patients and across expanded CTG repeat alleles were present on the mutant chromosome, suggesting that rather than a deletion, the patients carried an inversion with a breakpoint within the array or an insertion into the array. These data also revealed that the mutant chromosome carried the classic DM1 A haplotype shared by all non-African DM1 alleles (31), suggesting a common origin for the DM1 mutation in this family.

Identification of a complex variant repeat expansion in the DM1 + CMT + + family

To identify the nature of the mutation at the 3'-end of the array, we performed vectorette PCR (32) to ‘walk’ across the putative breakpoints of either an inversion or insertion. We used a variety of restriction enzymes (TaqI, Saf3A1, HinfI, BamHI, EcoRI, PstI, Msel, AciI and HhaI) to generate vectorette libraries and attempted to recover the putative breakpoints using primers based in the 5'- and 3'-flanking DNA. Unexpectedly, despite ‘walking’ across the normal allele in all patients and across expanded CMT repeat alleles in control DM1 patients, the DM1 + CMT + + mutant allele was only visualized in the AciI (recognition sequence CCGC) and HhaI (recognition sequence CCGG) vectorette libraries (Supplementary Material, Fig. S1). Escherichia coli plasmid clones obtained from the HhaI vectorette library 5'-PCR products were highly unstable and comprised of fragments with a wide variety of sizes. Although only partial sequences could be obtained from these clones, the sequence (CTG)X(GGC)3GC was obtained. These data suggested that our previous failure to PCR amplify the expanded allele might have been mediated by an increase of the GC content within the repeat array. Indeed, using a modified PCR including 10% dimethyl sulfoxide (DMSO) we were then able to amplify the mutant allele as a heterogeneous smear of fragments with an estimated size between 200 and 400 triplets (Fig. 1B), consistent with the sizes obtained by Southern blot analysis of restriction digested genomic DNA (19). Escherichia coli plasmid clones of these products were also grossly unstable, but partial sequences obtained from collapsed subclones revealed the presence of CCG and CCGCTG variant repeats within the expanded alleles.

In order to determine the precise structure of the interrupted arrays, we used small pool-PCR (11) to separate out single mutant alleles, removing the confounding effects of somatic mosaicism. Restriction analysis with PvuII (recognition sequence GCNGC) which digests CNG repeats; AciI which cleaves GGC, CCG repeats and CCGCTG hexamers; MspA1I (recognition sequence CMGCKG) which cleaves only CCGCTG hexamers; and HhaI which cleaves the sequence GCGC (Supplementary Material, Fig. S2), yielded insights into the approximate positions of the various variant repeats within the array. These structures were further dissected using variant-repeat primed PCR (33) with tagged primers matching either (CCG)₃ or (CCGCTG)₃

RESULTS

Analysis of a putative deletion in the DM1 + CMT + + family

In order to investigate the possibility of a downstream deletion in the DM1 + CMT + + family, we performed a loss of heterozygosity analysis by genotyping multiple polymorphisms spanning ~40 kb across the DM1 locus. These analyses revealed that five patients were heterozygous at multiple polymorphisms (Supplementary Material, Table S1), arguing against a major deletion. Indeed, five patients were genotyped as heterozygous at the polymorphic site closest to the 3'-end of the CTG array (rs3745802), using a PCR primer located at the junction of the 3'-end of the CTG array (BAB452) and another primer located in intron 1 of SIX5 (SIX-DR; Fig. 2A, Supplementary Material, Tables S1 and S2). The cloning and sequencing of this entire ~2.5 kb region for both alleles from III-16 to IV-19 revealed no additional variants. These data indicated that the sequence downstream of the CTG repeat were present on the mutant chromosome, suggesting that either a single or two closely linked mutations located near the CTG array (rs3745802), using a PCR primer located at the DMPK 5'-end. These data suggested that either a single or two closely linked mutations located near the APOC2 gene caused the DM and CMT. Indeed, subsequent analyses excluded mutations at the common CMT loci, and Southern blot analysis of restriction digested genomic DNA revealed a fragment equivalent to a small expansion (~200–400 CTG repeats) at the DM1 locus (19). However, despite the apparently small size of the expansion, the mutant allele was refractory to PCR amplification even using sensitive single molecule approaches. Base substitutions in PCR primer sites were ruled out using independent pairs of flanking primers. Repeat primed-PCR (26) revealed an expanded CTG repeat allele at the 5'-end of the repeat array, but was negative at the 3'-end. These data suggested that the patients carried an atypical DM1 mutation with an additional lesion at the 3'-end of the repeat (19). In fragile X (27), Jacobian syndrome (28), Escherichia coli (29) and yeast (30), expanded repeats predispose to additional DNA instability leading to deletions and rearrangements encompassing the repeat and the flanking DNA. We therefore postulated that the CTG repeat expansion in the DM1 + CMT + + family was most likely accompanied by an additional lesion such as a deletion, insertion and/or rearrangement at the 3'-end. Such a mutation could modify the expression of DMPK and/or other nearby genes explaining the presence of CMT, acute encephalopathy and hearing loss (19). As a step to providing a molecular explanation for the complex phenotype observed in the DM1 + CMT + + family and to provide insight into symptomatic variation in DM1, we sought to characterize the precise nature of the mutation in this family.
(Supplementary Material, Fig. S3), and separately with the forward primer DM-GGC (5'-TG CTG CTG CTG GGC GGC G-3') which spans the single HhaI site identified (Figs 1C and 2A). The structure of the mutant allele was confirmed by cloning and sequencing of PCR fragments revealing a variable number of CTG repeats at the 5'-end followed by (GGC)₃G(CCG)₂₀(CCGCTG)₁₄(CTG)₃₅ at the 3'-end (Fig. 2B).

During vectorette PCR we noticed that the normal allele from III-9 was also digested with AciI (Supplementary Material, Fig. S1) suggesting it too contained variant repeats. Indeed, cloning and sequencing of PCR fragments revealing a variable number of CTG repeats at the 5'-end followed by (GGC)₃G(CCG)₂₀(CCGCTG)₁₄(CTG)₃₅ at the 3'-end (Fig. 2B).

During vectorette PCR we noticed that the normal allele from III-9 was also digested with AciI (Supplementary Material, Fig. S1) suggesting it too contained variant repeats. Indeed, cloning and sequencing of PCR fragments revealing a variable number of CTG repeats at the 5'-end followed by (GGC)₃G(CCG)₂₀(CCGCTG)₁₄(CTG)₃₅ at the 3'-end (Fig. 2B). Variants repeat alleles at the DM1 locus are not uncommon. The schematic shows the structure of a series of atypical DM1 alleles identified in a cohort of French cases with unusual molecular diagnoses. Insufficient DNA was available from DM-UC1 to reconstruct the full allele structure and the length of the pure CTG tract at the 5'-end (Y) and the length (Z) and structure of the variant repeat array within the middle of the allele remain unknown. The length of the pure CTG repeat array at the 5'-end of the array in the progenitor allele of each patient was estimated by small pool-PCR analyses. The expanded DM1 allele from DM-UC9 does not contain any variant repeats, but has a 10 bp deletion in the immediate 3'-flanking DNA (ACAGACCATT). The relationships of DM1-UC1 and -UC2 and DM1-UC10 and -UC11 are also shown.

Mutational dynamics of the complex variant repeat expansion in the DM1+CMT++ family

Given the stabilizing effect of variant repeats at other expanded repeat loci (34–37), the presence of interruptions might be expected to increase the stability of the DM1+CMT+++ allele. PCR amplification of the interrupted 3'-end of the array revealed a single discrete band in all available family members (Fig. 1C) indicating that this end is relatively stable in both the maternal germ line and soma. In contrast, the pure CTG repeat tract at the 5'-end was highly variable and varied both within and between patients. To investigate the

Figure 2. Variant repeats at the DM1 locus. (A) Genomic organization around the DM1 expansion. The schematic shows the location of exons 14 and 15 of DMPK and exon A of SIX5. Coding regions are pictured as black boxes, untranslated regions as narrow white boxes, intronic sequences as a thick black line and intergenic sequences as a thin black line. The position of the closest flanking polymorphisms (rs635299 and rs3745802), and some of the PCR primers used in this study are also indicated. Using PCR primers BAB452 and SIX-DR it was shown that five of the eight affected subjects tested were heterozygous for rs3745802 (Supplementary Material, Table S1). (B) A complex variant repeat in the DM1+CMT+++ family. The schematic shows the structure of the complex variant repeat array at the 3'-end of the array. Each patient carries a different number (X) of pure CTG repeats at the 5'-end. The location of a 9 bp perfect inverted repeat within the array is indicated with white arrows and the presence of higher order repeats is indicated with black arrows. The structure of the ‘38’ repeat allele identified in III-9 and stably transmitted to IV-11 and IV-12 is also shown. (C) Variant repeat alleles at the DM1 locus are not uncommon. The schematic shows the structure of a series of atypical DM1 alleles identified in a cohort of French cases with unusual molecular diagnoses. Insufficient DNA was available from DM-UC1 to reconstruct the full allele structure and the length of the pure CTG tract at the 5'-end (Y) and the length (Z) and structure of the variant repeat array within the middle of the allele remain unknown. The length of the pure CTG repeat array at the 5'-end of the array in the progenitor allele of each patient was estimated by small pool-PCR analyses. The expanded DM1 allele from DM1-UC9 does not contain any variant repeats, but has a 10 bp deletion in the immediate 3'-flanking DNA (ACAGACCATT). The relationships of DM1-UC1 and -UC2 and DM1-UC10 and -UC11 are also shown.

(A) Genomic organization around the DM1 expansion. The schematic shows the location of exons 14 and 15 of DMPK and exon A of SIX5. Coding regions are pictured as black boxes, untranslated regions as narrow white boxes, intronic sequences as a thick black line and intergenic sequences as a thin black line. The position of the closest flanking polymorphisms (rs635299 and rs3745802), and some of the PCR primers used in this study are also indicated. Using PCR primers BAB452 and SIX-DR it was shown that five of the eight affected subjects tested were heterozygous for rs3745802 (Supplementary Material, Table S1). (B) A complex variant repeat in the DM1+CMT+++ family. The schematic shows the structure of the complex variant repeat array at the 3'-end of the array. Each patient carries a different number (X) of pure CTG repeats at the 5'-end. The location of a 9 bp perfect inverted repeat within the array is indicated with white arrows and the presence of higher order repeats is indicated with black arrows. The structure of the ‘38’ repeat allele identified in III-9 and stably transmitted to IV-11 and IV-12 is also shown. (C) Variant repeat alleles at the DM1 locus are not uncommon. The schematic shows the structure of a series of atypical DM1 alleles identified in a cohort of French cases with unusual molecular diagnoses. Insufficient DNA was available from DM-UC1 to reconstruct the full allele structure and the length of the pure CTG tract at the 5'-end (Y) and the length (Z) and structure of the variant repeat array within the middle of the allele remain unknown. The length of the pure CTG repeat array at the 5'-end of the array in the progenitor allele of each patient was estimated by small pool-PCR analyses. The expanded DM1 allele from DM1-UC9 does not contain any variant repeats, but has a 10 bp deletion in the immediate 3'-flanking DNA (ACAGACCATT). The relationships of DM1-UC1 and -UC2 and DM1-UC10 and -UC11 are also shown.
the pure 5'-interrupted array was subtracted from the total length of the array, Fig. 4A). Most interestingly, even when the length of the interage at sampling and progenitor allele length (\(P^0\) mutant alleles was lower than expected for the corresponding variation observed between individuals (Morales et al., manuscript in preparation). When compared quantitatively the degree of somatic mosaicism in the DM1+CTM++ mutant alleles was lower than expected for the corresponding age at sampling and progenitor allele length (\(P < 0.001\), Fig. 4A). Most interestingly, even when the length of the interrupted array was subtracted from the total length of the array, the pure 5'-end of the array was less somatically unstable than expected when compared with alleles from classic DM1 patients (\(P = 0.001\), Fig. 4B). These data indicate that not only does the presence of interruptions at the 3'-end of the CTG repeat decrease the somatic instability of the whole array, but they also decrease the relative somatic instability of the pure CTG tract at the 5'-end of the array.

Intergenerational transmission from females with adult-onset DM1 is associated with a very high risk of congenital DM1 in the next generation (8,9,13–15). Comparison of the estimated inherited progenitor allele length in each patient for whom DNA was available allowed us to measure the observed intergenerational differences in four maternal transmissions in the DM1+CTM++ family (no paternal transmissions of the mutant allele were observed in this family and DNA was unavailable from the deceased individuals in generation II). Although in all four cases an increase in the allele length transmitted was observed (Table 1), the sizes of the increases were relatively small (\(\sim 41–46\) CTG repeats at the 5'-end). Comparison of the maternal intergenerational length changes observed in this family with those in classic DM1 families (8) revealed that as with somatic instability, the presence of the variant repeats has a stabilizing effect on maternal transmissions of the DM1+CTM++ allele (\(P = 0.002\), Fig. 4C). This effect also extended to the pure 5'-end of the DM1+CTM++ allele and was apparent when comparing to all maternal transmissions (\(P = 0.02\), Fig. 4C). Closer examination of the length dependency of maternal transmissions suggested that as with paternal transmissions (38), the smaller maternal alleles show a different length dependency than longer alleles. Consequently, we repeated the analyses of the pure CTG 5'-end DM1+CTM++ alleles in comparison with other small expanded maternal alleles of \(\leq 100\) repeats (Fig. 4D), revealing an even greater stabilizing effect on maternal transmission (\(P < 0.001\)). The mean length change in maternal transmissions from classic DM1 patients with alleles \(\leq 100\) repeats was +793 repeats, compared with only +43 repeats in the DM1+CTM++ family. This dramatic alteration in intergenerational repeat dynamics accounts for the unexpected absence of congenital DM1 in the fourth generation of the DM1+CTM++ family.

**Variant repeat alleles are present in a subset of DM1 patients**

Having identified the presence of a complex variant repeat within the DM1+CTM++ family, we asked if variant repeats might be present within the expanded arrays of other DM1 patients. Routine diagnostic testing for DM1 in France is usually performed using a combination of traditional Southern blot analysis of restricted digested genomic DNA,
Figure 4. The DM1-CMT++ complex variant repeat allele is more stable than expected. (A) The DM1-CMT++ complex variant repeat allele is more stable than expected in the soma. The graph shows the relationship between the estimated progenitor allele length, the age at sampling and the degree of somatic variation (the range between the 10th and 90th percentiles) for classic DM1 alleles without GGC/CCG variants (diamonds; Morales et al., Couto et al., manuscripts in preparation) and the DM1-CMT++ alleles (triangles). Also shown is the linear regression plane with multiple variables \( r = 0.74, P < 0.0001 \) derived from the total data set (classic DM1 alleles plus the DM1-CMT++ alleles). Under the prior hypothesis that, as in all other studies (34–37), the variant repeats would stabilize the DM1-CMT++ allele, comparison of the mean residual variation in somatic variation not accounted for by age at sampling and progenitor allele length was significantly less in the DM1-CMT++ alleles than in the DM1 controls \( (P < 0.001, \text{one tailed } t\text{-test}) \). (B) The pure CTG repeat 5'-end of the DM1-CMT++ allele is more stable than expected in the soma. The graph shows the relationship between the estimated progenitor allele length, the age at sampling and the degree of somatic variation in classic DM1 alleles (as above) and the pure CTG repeat 5'-end of the DM1-CMT++ allele (filled triangles). Also shown is the linear regression plane with multiple variables \( r = 0.77, P < 0.0001 \) derived from the total data set (classic alleles plus the pure CTG repeat 5'-end of the DM1-CMT++ alleles). Comparison of the mean residual variation in somatic variation not accounted for by age at sampling and progenitor allele length was significantly less for the pure CTG repeat 5'-end of the DM1-CMT++ alleles than in the DM1 controls \( (P = 0.001, \text{one tailed } t\text{-test}) \). (C) The DM1-CMT++ complex variant repeat allele is more stable than expected in the maternal germ line. The graph shows the relationship between maternal allele length and transmitted allele length for classic DM1 families (circles) (8), and for the four transmissions in the DM1-CMT++ family counting all triplet repeats (open triangles) and counting only the pure CTG array at the 5'-end (filled triangles). Please note that two of the DM1-CMT++ transmissions were of identical lengths and the other two very close to each other (Table 1) and hence only two points can be clearly discerned on the graph. The points have been plotted on a log/log scale to aid visualization. The line shows the linear regression analysis between the log of the maternal allele and the transmitted allele for classic DM1 patients and the DM1-CMT++ family combined \( \text{transmitted allele length} = 1294.2 \times \log(\text{maternal allele length}) - 1674.9, r = 0.61, P < 0.0001 \). Mean residual variation in the DM1-CMT++ transmissions not accounted for by maternal allele length was significantly less than for classic DM1 patients \( (P = 0.002, \text{one tailed } t\text{-test}) \). Linear regression analysis between the log of the maternal allele and the transmitted allele for classic DM1 patients and the DM1-CMT++ family considering only the length of the pure 5'-end of the CTG array revealed a similar relationship \( \text{transmitted allele length} = 1349.1 \times \log(\text{maternal allele length}) - 1785, r = 0.64, P < 0.0001 \) \( (\text{regression line not shown}) \) and mean residual variation in the DM1-CMT++ transmissions not accounted for by maternal allele length remained significantly less than for classic DM1 patients \( (P = 0.02, \text{one tailed } t\text{-test}) \). (D) The pure CTG 5'-end of the DM1-CMT++ allele is more stable than expected in the maternal germ line. The graph shows the relationship between maternal allele length and transmitted allele length for classic DM1 families (circles) (8) and in the DM1-CMT++ family counting only the pure CTG array at the 5'-end (filled triangles) for transmissions with maternal allele lengths <100 repeats. The points have been plotted on a log/log scale to aid visualization. The line shows the linear regression analysis between the log of the maternal allele and the log of the transmitted allele for classic DM1 patients and the DM1-CMT++ family combined \( r = 0.35, P = 0.05 \). Mean residual variation in the pure 5'-end CTG DM1-CMT++ transmissions not accounted for by maternal allele length was significantly less than for classic DM1 patients \( (P < 0.001, \text{one tailed } t\text{-test}) \).
PCR amplification across the repeat and repeat primed-PCR at the 3'-end of the array. Over the last few years we have identified 11 French cases (~3–4%) in whom the routine molecular diagnosis was atypical (presence of gaps in the repeat primed-PCR ladder) or internally ambiguous (positive by Southern blot analysis of genomic DNA, but negative by repeat primed PCR and/or amplification across the array). These latter cases were initially assumed to be associated with deletions in the flanking DNA. Indeed, DM1-UC9 presents with an apparently pure CTG repeat expansion (~396 repeats) with a small deletion of 10 bp in the 3'-flanking DNA (Fig. 2C) that spans one of the commonly used PCR primer sites. In the rest of these cases, however, we identified an imperfect CTG repeat allele containing multiple CCG repeats all located at the 3'-end of the array (Fig. 2C). All alleles included short runs of pure CCG or CCGCTG hexamers. DM1-UC2, -UC3, -UC5, -UC6 and -UC8 also show clear alleles included short runs of pure CCG or CCGCTG hexamers. DM1-UC2, -UC3, -UC5, -UC6 and -UC8 also show clear.

### DISCUSSION

Unlike the loci associated with some of the other simple repeat expansion disorders (34–37), the DM1 locus has been assumed to be a pure CTG repeat. This assumption has been based on the sequencing of normal alleles and partial sequencing of small numbers of small expanded alleles (5–7). Until very recently, only one exception to this assumption had been described, a ‘37’ repeat allele was shown to have the structure (CTG)$_3$(CCGCTG)$_5$CTG (39). This allele was identified by Leeflang and Arnheim in an anonymous sperm donor and was shown to be more stable in the male germ line than expected. Such interrupted large normal alleles (>30 repeats) are rare in the general population, with only one example observed in over 3500 normal alleles (39,40). Thus, it was with considerable surprise that we identified the existence of a complex variant repeat mutation in the DM1+CMT++ family, and even more of a surprise that similar such variant repeats were found in ~3–4% of French DM1 diagnostic cases. In addition, we identified two more high end normal alleles with structures very similar to that reported by Leeflang and Arnheim. Very recently, and independently of our study, Musova et al. (41) reported three Czech DM1 families (~5%) with variant repeat containing alleles, and an additional two high end normal alleles, again with a structure similar to that reported by Leeflang and Arnheim. It is thus clear that expanded DM1 alleles with variant repeats are not uncommon.

Expanded simple sequence repeats are characterized by exceptionally high levels of somatic and germlinal instability that is assumed to be mediated by mutational pathways involving DNA mismatch repair proteins and/or DNA replication (16,42). As has been previously observed, the presence of variant repeats can result in the relative stabilization of simple sequence repeats in both the germ line (34–36) and soma (37). Here, we have established a similar stabilizing effect of variant repeats on the stability of the DM1+CMT++ allele in both the soma and female germ line. Although all four maternal transmissions of the DM1+CMT++ allele we observed resulted in expansion of the CTG tract, these expansions were all relatively small (<50 repeats) and did not result in the congenital DM1 phenotype often observed in the later generations of DM1 families. The one male germ line transmission we observed in the French cohort also resulted in a relatively small expansion (>65 triplet repeats). Similarly, Musova et al., noted that three of three paternal and two of three maternal transmissions observed in the Czech families with variant repeats, resulted in apparent contractions. These data thus suggest that the stabilizing effect we have quantified for the DM1+CMT++ allele in the female germ line reflects a more general stabilizing role for variant repeats in both the male and female germ line at the DM1 locus. It is generally assumed that the mutational pathway for length change mutations at expanded simple repeat loci involves the adoption of misaligned DNA structures such as slipped strand DNA (16,42). It has been previously shown that the presence of variant repeats reduces both the quantity and range of slipped strand structures that can be formed, presumably by the biophysical destabilization of slipped strand structures with misaligned variant repeats (43). Thus, the destabilization of slipped strand DNA mutational intermediates presents as a highly plausible explanation for the general stabilizing effect of variant repeats observed by us and others (34–37).

### Table 1. Progenitor allele length and intergenerational transmissions in the DM1+CMT++ family

<table>
<thead>
<tr>
<th>Estimated progenitor allele</th>
<th>Pure CTG repeats at 5'-end (CTG repeats)</th>
<th>Estimated intergenerational transmissions</th>
<th>Age of onset of DM1 (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total length (all triplet repeats)</td>
<td>Transmitted allele length (all triplet repeats)</td>
<td>Length change transmitted (CTG repeats)</td>
<td></td>
</tr>
<tr>
<td>III-9 225</td>
<td>139</td>
<td>–</td>
<td>25</td>
</tr>
<tr>
<td>III-16 170</td>
<td>84</td>
<td>213/213</td>
<td>43/43</td>
</tr>
<tr>
<td>IV-19 213</td>
<td>127</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>IV-20 213</td>
<td>127</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>III-17 179</td>
<td>93</td>
<td>220/225</td>
<td>41/46</td>
</tr>
<tr>
<td>IV-21 220</td>
<td>134</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>IV-22 225</td>
<td>139</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Downloaded from https://academic.oup.com/hmg/article-abstract/19/8/1399/569666 by guest on 08 April 2019
Very interestingly, the stabilization of the DM1+CMT++ allele we have observed here extends beyond the bounds of the variant repeat containing portion of the array. Specifically, we have shown that the remaining pure CTG repeat tract is more stable in both the soma and female germ line than similar sized alleles in classic DM1 families. Such a stabilizing effect of variant repeats on the pure portion of an expanded array has not been previously observed. Moreover, it is not obviously explainable by the destabilization of slipped strand mutational intermediates since the pure 5'-CTG array should be able to adopt the same range of slipped strand structures as a pure CTG array of the same length. There is considerable evidence that interlocus variability in the mutational dynamics of expanded CAG–CTG repeat loci extends beyond direct effects intrinsic to the repeat such as length and purity, and includes an important role for flanking DNA sequences as cis-acting modifiers (44–47). Little is known about the mode of action of such cis-acting modifiers, but the GC content of the flanking DNA (46) and the presence/absence/methylation status of nearby CTCF sites appears important (48). Of note, is that all of the variant repeats we and Musova et al. (41), have characterized at the DM1 locus are strongly clustered at the 3'-end of the array revealing considerable polarity in the generation and/or maintenance of variant repeats. Similar polarity has been previously observed at a number of tandem repeat loci (36,49–52) and suggests flanking cis-acting sequencing elements can modify mutational dynamics in a unidirectional manner. Thus, a plausible explanation for the observed stabilization of even the pure 5'-repeat tract, is that the variant repeat tract insulates the mutable pure CTG tract from the action of a critical modifier in the 3'-flanking DNA (Fig. 5A) or shifts the distance to the replication fork (42). Alternatively, the presence of the variant repeats may have a more direct inhibitory effect on the mutational dynamics of the adjacent pure CTG tract (Fig. 5A) by, for instance, altering local nucleosome phasing (53).

An obvious question that arises is how do variant repeat containing expanded alleles at the DM1 locus arise? In the DM1+CMT++ family the complex variant repeat structure at the 3'-end of the array must have been present in one of the two great-grandparents. Given that both of these individuals were reported as unaffected, it seems likely that the pure CTG tract at the 5'-end of the array was too small to cause significant symptoms, but must have been large enough to be unstable and expand in the subsequent generation. This still leaves unresolved the ultimate origin of the variant repeat containing allele. At the FRAXA, SCA1 and SCA2 loci, variant repeats are present in most normal alleles and absent in expanded alleles (35,36,50,51). The existence of the high end normal alleles containing variant repeats at the DM1 locus might represent a source for such expansions, but their observed stability [this study and (39,41)] suggests this is unlikely. An alternative proposition is that variant repeat containing alleles are secondary derivatives of pure

![Figure 5. Variant repeats and length change mutations. (A) Variant repeats increase the stability of the adjacent stretch of pure CTG repeats. For pure expanded alleles the mutational dynamics of the CTG tract (white squares) are driven by the action of a cis-acting modifier in the 3'-flanking sequence (i). In the presence of variant repeats (grey squares) at the 3'-end of the array, the distance between the mutable pure CTG tract and the cis-acting modifier in the 3'-flanking sequence will be increased and its effect may be reduced (ii). Alternatively, the presence of variant repeats may directly inhibit the mutability of the pure CTG tract (iii). (B) The spread of variant repeats within the array. A single variant repeat is assumed to arise as the product of a base substitution mutation (i). The most probable secondary mutation appears to involve small length change mutations of either one (ii) or two repeats (vi). Subsequent small length change mutations can result in longer runs of the variant repeat in either a pure array (iii) or in the context of the hexamer (vii). Rarer, longer, length change mutations can result in the generation (iv, viii) and spread (v, ix) of higher order repeats (black arrows).](https://academic.oup.com/hmg/article-abstract/19/8/1399/569666/1406-Human-Molecular-Gene)
expanded CTG repeat alleles and that variant repeats arise as de novo base substitutions within the array, as has been suggested for the SCAl8 locus at which the normal alleles are pure and interruptions are restricted to a subset of expanded alleles (52). All of the variant repeat containing alleles we observed (Supplementary Material, Table S1) were present on the classic DM1 A haplotype shared by all non-African DM1 alleles (31), consistent with a shared common origin for both pure and variant repeat-containing expanded DM1 alleles. The identification of more recently derived variant repeat-containing sub-lineages, if they exist, awaits a more detailed haplotype analysis extending beyond the conserved ±40 kb region analyzed here.

Although both intergenerational and acquired somatic length changes at the expanded simple repeat loci may be many tens, hundreds or even thousands of repeats in size, it is probable that such large length changes are in fact the products of many much smaller length changes (16). In this regard the structure of the variant repeat containing alleles we have determined is particularly informative. Notably, all of the alleles contain the CCG variant repeat in the context of either a short pure CCG array or in a CCGCTG hexamer (Fig. 2B and C). Assuming that these more complex arrays arose from a rare primary base substitutional variant, these data suggest that the most common secondary mutation resulting in the spread of the variant repeat involves very small misalignments of only one or two repeats (Fig. 5B). Such small mutations are not obviously compatible with the large hairpin structures frequently postulated as mutation intermediates (54). Nonetheless, many alleles contain clear evidence of higher order structures up to 11 repeats in length (Fig. 2C), indicating that some individual length change mutations do involve relatively large numbers of repeats (Fig. 5B). Of course, it remains probable that in addition to modifying absolute mutation frequency, the variant repeats also strongly bias the formation (43) and resolution of specific length change mutational intermediates. Nevertheless, these data provide insights into the range of mutational length mutations that are likely to occur in pure arrays, and direct insights into the types of process that likely govern the mutational dynamics of the many microsatellites spread throughout the genome, a high proportion of which contain variant repeats.

During more than 25 years of follow-up it has been established that the DM1+CMT++ family presents with an increasingly unusual phenotype, including DM1, CMT, encephalopathic attacks and early hearing loss, accompanied by an atypical molecular lesion at the DM1 locus that provided ambiguous diagnostic testing results (17–19). We have now characterized the unusual molecular mutation at the DM1 locus in this family providing an explanation for the atypical diagnostic results. The presence of variant repeats at the 3′-end of the array increases the GC content of the array, making it highly resistant to PCR amplification and disrupting the binding of repeat primers targetted against a CTG array. All of the patients genotyped have a pure expanded (>84 repeats) CTG tract at the 5′-end that accounts for the relatively mild classic DM1 symptoms observed in this family such as cataracts and myotonia. These archetypal DM1 symptoms are presumably caused by the mis-regulation of alternative splicing that is observed in classic DM1 patients via expanded CUG repeat RNA sequestration of MBNL and effects on CUG-BP (55). It is our hypothesis that the CMT, early hearing loss and recurrent encephalopathic attacks observed in the DM1+CMT++ family are mediated by the complex variant repeat at the 3′-end of the DM1 array. The alternative explanation is that the CMT and other symptoms are caused by a third co-segregating dominant mutation near the DM1 locus. Without sequencing the entire co-segregating region (>3.5 Mb), the existence of a third co-segregating mutation cannot be formally excluded. However, a number of considerations argue against such an explanation. Firstly, if the additional symptoms were mediated by a third co-segregating mutation then the presence of the complex variant repeat would be entirely coincidental. We and Musova et al. (41), have established that variant repeats per se are not unique to the DM1+CMT++ family, but they are nonetheless relatively infrequent being present in <5% of families. Secondly, as both the great grandparents (I-6 and I-9) in the family had no reported signs of CMT (17), but transmitted the disorder to five of their offspring, any putative third mutation would have to have arisen as a de novo dominant mutation, present at very high frequency in the germ line of one of the unaffected grandparents. Thirdly, any third static mutation in another gene would provide no obvious explanation for the anticipation observed for the CMT aspect of the phenotype. In contrast, positing the complex variant repeat at the 3′-end of the DM1 array as the cause of the additional symptoms provides highly plausible explanations for the unusual inheritance patterns observed in the DM1+CMT++ family analogous to those mediated by the dynamic CTG repeat mutation in classic DM1 families. Moreover, the nature of the mutation also provides a plausible molecular pathology via extension of the toxic RNA gain of function underlying DM1 symptoms and by analogy to that observed in fragile X associated tremor ataxia syndrome (FXTAS) (56). In FXTAS, premutation (CGG)55–200 FMR1 alleles are associated with a tremor/ataxia phenotype, accompanied by a peripheral neuropathy in many patients, with a CMT-like phenotype in some of them (57). CGG RNA toxicity is associated with the presence of intranuclear protein inclusions (58) that contain numerous proteins (59), many previously associated with CMT, including HSP27, LMNA and NEFL (60) and the myelin sheath (MOG and MBP). The number of GGC repeats in the DM1+CMT++ transcript is very low (three), and it seems unlikely that these alone are sufficient to mediate significant toxicity. However, there are 34 CCG repeats in total (20 in the pure CCG array and 14 in the context of the CCGCTG hexamer). Although CCG transcripts have not previously been shown to be toxic to humans, CCG repeats produce a similar RNA mediated neurodegenerative phenotype to CGG repeats in Drosophila (61) consistent with a role for CCG repeats as important pathogenic determinants in the DM1+CMT++ allele. In DM1, the CUG RNA is thought to fold into a single stable hairpin (62) that sequesters MBNL (55). Mfold analysis (63) of the DM1+CMT++ transcript also predicted the adoption of a single hairpin at the 3′-end (Fig. 6A). The 3′ hairpin was characterized by a 9 bp perfect duplex formed at the apex of the hairpin stem by the sequence (GGC)3G(CCG)4 (Figs 2B and 6A). In this structure, the remainder of the CCG tract and the CCGCUG both paired
with CUG repeats with C–U and U–U mismatches. In other similarly stable structures, the (GGC)_3G sequence paired at the 3'–end of the CCG array generating a region of CCG•CCG near the apex of the hairpin incorporating C–C mismatches (not shown). Other structures with similar stability were also predicted, and although these had slightly more complicated branching structures, the overall pairing patterns were preserved. More detailed computational analyses of the relative stabilities of combinations of CUG, CCG and CGG repeats revealed that all combinations of CUG and CCG repeats were equally stable, but that structures with CGG repeats were much more stable (Fig. 6C). All of the structures with pyrimidine mismatches (U–U, C–U or C–C) are predicted to bind MBNL with similar affinities (64). However, MBNL would not be predicted to bind with high affinity to a CGG•CCG RNA duplex (64). The putative RNA toxicity of the DM1+CMT++ transcript might thus be associated with binding of other proteins to the CCG•CCG duplex, or by the binding of other proteins with greater affinity than MBNL to the other CCG structures.

Although the perfect hairpin and the total number of CCG repeats are relatively small, the relative toxicity of these sequences might be exacerbated by the expanded pure CUG tract that presumably traps these transcripts in ribonuclear foci (65) within the nucleus. These data highlight the probability that additional ‘pseudo-toxic’ sequences in the DMPK 3'–UTR (66) likely mediate the greater toxicity of the DM1 transcript relative to the DM2 transcript, despite the greater number (55) and higher affinity for MBNL (64) of DM2 CCUG repeats. Of course it is possible that the variant repeats might mediate an altered pathology via effects on the expression of the flanking genes. The introduction of multiple CpG dinucleotides into the expanded array might change the methylation status of the repeat and induce methylation changes in the flanking DNA.

Although a compelling case that the complex variant repeat allele observed in the DM1+CMT++ family causes the co-segregating CMT neuropathy can be made, it could be argued that this explanation is not definitive as the alternative scenario of a co-segregating third mutation has not been
formally excluded. The identification of additional DM1 patients carrying similar complex mutations and also presenting with a CMT phenotype would clearly make an even more persuasive case. Unfortunately, the French cases in which we have also identified variant repeats represent anonymous cases for whom detailed clinical evaluations have not been performed. Intriguingly, two of the patients with variant repeats described by Musova et al. (41), also present with a poly-neuropathy, although it is not known if this is greater in severity than the mild neuropathy observed in many DM1 patients (20–22). However, it remains possible that the CMT observed in the Dutch family is mediated by the specific sequence and combination of repeat variants present on the DM1 + CMT + + allele (the DM1 + CMT + + family remains the only cases in which GGC variant repeats and the perfect 9 bp CGG⇌CCG hairpin structure have been observed) and that other combinations of variant repeats may have different effects. Indeed, there are no other reports of large families clearly segregating DM and CMT, though it should be borne in mind that once a diagnosis of DM1 has been confirmed, the detailed clinical investigations necessary to reveal a co-presenting CMT may well not be performed. Additionally, although the complex variant repeat tract at the 3' end of the DM1 + CMT + + allele is stably inherited, the variant repeat tracts can in some instances change dramatically when transmitted [Fig. 2C and Musova et al. (41)]. Within family differences in the precise complement and structure of variant repeats could lead to a variable presentation of additional symptoms such as CMT within the same family. As noted, there appears to be a more frequent concurrence of DM1 [disease frequency ~1/8000 (1)] and a severe peripheral neuropathy [CMT frequency ~1/2500 (60)] than chance alone would predict (~1/20 000 000) and we expect that some of these patients (23–25) will present with similar complex variant repeats. Similar effects for variant repeats in modifying the range of symptoms observed have been proposed for SCA10 (67).

In addition to any new symptoms, the presence of variant repeats might also have an effect on the severity of the classic DM1 symptoms. As discussed above, the existence of CCG variant repeats would not be predicted to alter the structures formed by mutant DMPK transcripts or their affinity for MBNL. However, we expect that the dramatic stabilizing effect of the variant repeats will reduce the rate of expansion in muscle and other affected tissues leading directly to a delay in the onset and slowing of the progression of the DM1 symptoms. Indeed, Musova et al. (41) did suggest that the symptoms were generally less severe in their patients with variant repeats. Although it seems logical to assume that somatic expansions contributes toward disease pathology (16), there is actually very little direct evidence that this is the case (68).

The identification of variant repeats with extreme GC contents in a subset of DM1 patients has a number of important implications. This includes the potential for such alleles to yield false negatives in both repeat primed PCR and standard PCR based approaches to diagnostics. In addition, the probable association of variant repeats with atypical/less severe symptoms should shed light on the pathogenic processes involved in DM1 and provide new insights into symptomatic variation in this vastly variable disorder.

Such an understanding is important for providing patients and families with improved prognostic information and risk assessments to future generations, and will be crucial for identifying homogeneous cohorts for clinical trials. These data also add a critical proviso to the interpretation of animal and cell models in which interrupting linker sequences (5’-CTCGA-3’) are used to generate large expanded ‘CTG’ repeat arrays (69,70). Although it is clear that these models replicate many features of DM1 in humans, the data presented here suggest that caution should be used in inferring that the observed pathology is mediated solely by the CUG component of the arrays. All of these implications can be further extended to the other repeat expansion disorders, further highlighting (67) in particular, the potential for variant repeats to be present in a subset of patients in disorders for which the expanded arrays are assumed to be pure based on relatively little direct testing (e.g. Huntington disease). Additionally, the association of variant repeats with a CMT neuropathy raises the possibility that other currently idiopathic peripheral neuropathies and/or other neurological disorders are associated with the expansion of non-coding CGG/CCG arrays.

**MATERIALS AND METHODS**

**DNA extraction**

DNA was extracted from blood samples following standard procedures. QIAquick PCR purification kit and QIAquick gel extraction kit (Qiagen) were used to purify DNA from PCRs and from agarose gel slices, respectively.

**Polymerase chain reaction**

PCR was performed as previously described (11) using 20 ng input DNA and 1 μM primers in a thermal cycler (Biometra) cycled through 28–30 rounds of (96°C for 45 s, 68°C for 45 s, 70°C for 3 min), 68°C for 1 min and 70°C for 10 min. DMSO (10%) was added to reactions indicated and the annealing temperature was reduced from 68 to 63.5°C. Small pool-PCR was performed as described (11) using primers DM-A and DM-DR in the presence of 10% DMSO. Progenitor allele lengths were estimated as the lower boundary of the highly skewed distributions (11).

**Variant repeat primed-PCR**

Variant repeat PCR was performed as previously described (26,33) using 20 ng genomic DNA and 28 cycles. At the 5'-end of the DM1 array the DM-A (11) flanking primer was used in combination with tagged variant repeat primers to detect CTG repeats (TAG-AGC, 5’-TCA TGC GTC CAT GGT CCG GAA GCA GCA GCA GCA GC-3’), CCGCTGG hexamers (TAG-CAGCGG 5’-TCA TGC GTC CAT GGT CCG GAC AGC GCC AGC AGC GG-3’) or CCG repeats (TAG-GGC 5’-TCA TGC GTC CAT GGT CCG GAG CGG CGG CGG CGG CGG CGG CGG CGG CGG GG-3’) and the TAG primer (5’-TCA TGC GTC CAT GGT CCG GAA GCA GCA GCA GCA GC-3’). At the 3'-end of the DM1 array the DM-DR (11) flanking primer was used in combination with tagged variant repeat primers to detect CTG repeats (TAG-GTC 5’-TCA TGC
GTC CAT CCG GAT GCT GCT GCT GC-3', CCGCTG hexamers (TAG-CCGCTG 5'-TCA TGG GTC CAT GGT CCG GAC CGC TGC CGC TG-3') or CCG repeats (TAG-GCC 5'-TCA TGG GTC CAT GGT CCG GAC CGC CGC CGC CGC CGC CG-3') and TAG. Products were resolved on a 1.5% agarose gel and Southern blot hybridized with either the 5'-flanking or DM1 CTG repeat probe (Supplementary Material, Fig. S2) for analysis at the 5'- and 3'-ends, respectively.

**Vectorette PCR**

DNA vectorette libraries (32) were generated by digesting genomic DNA (1 μg) overnight with 5 U of restriction enzyme in the presence of 1 mM of spermidine in a final volume of 20 μl. Following digestion, the enzyme was heat inactivated for 20 min at 65 or 80°C. Oligonucleotide linkers were prepared using 2.5 μg of each oligonucleotide in 20 μl of annealing buffer [167 mM NaCl, 17 mM MgCl2, 17 mM Tris–HCl, (pH 7.5)] heated to 100°C and cooled slowly to room temperature. The digested DNA (20 μl) and 2 μg of linker were incubated at 16°C overnight with 200 U of T4 DNA ligase and 1X T4 ligase buffer (New England Biolabs, including ATP) in a total volume of 100 μl. Excess linker was removed using an NH4 acetate/isopropanol precipitation. Vectorette library DNA was dissolved in 100 μl water and 1 μl used in the first round PCR using flanking primers DM-C (11) or DM-CC (5'-GGC TCG TGT CCT TGT AG-3') at the 5'-end, and DM-DR at the 3'-end, and TAG4 (5'-TCA TGG GTC CAT GAG TGC GAT-3'). The sequence of the oligos used to prepare the linker were for: HhaI: V 5'-GG ACT TGC TAC GGT AAT CAG-3' and V-NL-TAG4-GC 5'-TCA TGG GTC CAT GAG TGC GAT-3'. The sequence of the oligos used to prepare the linker were for: HhaI: V 5'-GG ACT TGC TAC GGT AAT CAG-3' and V-NL-TAG4-GC 5'-TCA TGG GTC CAT GAG TGC GAT-3'. The sequence of the oligos used to prepare the linker were for: HhaI: V 5'-GG ACT TGC TAC GGT AAT CAG-3' and V-NL-TAG4-GC 5'-TCA TGG GTC CAT GAG TGC GAT-3'. The sequence of the oligos used to prepare the linker were for: HhaI: V 5'-GG ACT TGC TAC GGT AAT CAG-3' and V-NL-TAG4-GC 5'-TCA TGG GTC CAT GAG TGC GAT-3'.

**Cloning and sequencing**

Purified PCR products were cloned with the TOPO TA Cloning® Kit for Sequencing (Invitrogen) with the following modifications: 4 μl of PCR product was ligated to the vector, the ligation was incubated for 30 min at room temperature and 4 μl TOPO cloning reaction was added into the vial of one shot TOP10 chemically competent cells. Plasmid DNA was isolated either with Qiaprep spin miniprep kit or endofree plasmid maxp kit (Qiagen). Sequences were provided by Geneservice (www.geneservice.co.uk) and analyzed using 4Peaks software.

**Statistical analyses**

Statistical analyses were carried out using Microsoft Excel, Statistica and MiniTab.

**RNA folding predictions**

RNA folding patterns were predicted using Mfold (63) at: http://www.bioinfo.rpi.edu/applications/mfold. Comparative analyses of different CUG/CCG/CGG/CCGCUG duplexes was performed using an experimental sequence: G35 (60 nucleotide test sequence 1)G35 (60 nucleotide test sequence 2)C5 (Fig. 6B).

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG online.

**ACKNOWLEDGEMENTS**

The authors would like to thank colleagues in the DGM group and the Molecular Genetics Research Theme for productive discussions and helpful comments on the manuscript.

**FUNDING**

This work was supported by awards to D.G.M. from the Association Francaise contre les Myopathies, the Myotonic Dystrophy Support Group and the Muscular Dystrophy Campaign.

**REFERENCES**


