
Expanded CUG repeats in *DMPK* transcripts adopt diverse hairpin conformations without influencing the structure of the flanking sequences

REMCO T.P. VAN CRUCHTEN, BÉ WIERINGA, and DERICK G. WANSINK

Department of Cell Biology, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, 6525 GA, The Netherlands

ABSTRACT

Myotonic dystrophy type 1 (DM1) is a complex neuromuscular disorder caused by expansion of a CTG repeat in the 3'-untranslated region (UTR) of the *DMPK* gene. Mutant *DMPK* transcripts form aberrant structures and anomalously associate with RNA-binding proteins (RBPs). As a first step toward better understanding of the involvement of abnormal *DMPK* mRNA folding in DM1 manifestation, we used SHAPE, DMS, CMCT, and RNase T1 structure probing in vitro for modeling of the topology of the *DMPK* 3'-UTR with normal and pathogenic repeat lengths of up to 197 CUG triplets. The resulting structural information was validated by disruption of base-pairing with LNA antisense oligonucleotides (AONs) and used for prediction of therapeutic AON accessibility and verification of *DMPK* knockdown efficacy in cells. Our model for *DMPK* RNA structure demonstrates that the hairpin formed by the CUG repeat has length-dependent conformational plasticity, with a structure that is guided by and embedded in an otherwise rigid architecture of flanking regions in the *DMPK* 3'-UTR. Evidence is provided that long CUG repeats may form not only single asymmetrical hairpins but also exist as branched structures. These newly identified structures have implications for DM1 pathogenic mechanisms, like sequestration of RBPs and repeat-associated non-AUG (RAN) translation.

Keywords: myotonic dystrophy; RNA folding; antisense oligonucleotide; *DMPK*; trinucleotide repeat; hairpin

INTRODUCTION

RNA can adopt complex structures due to its flexible nature and ability to base pair intramolecularly. This folding of RNA plays an important role in processes like pre-mRNA splicing, intracellular mRNA transport, miRNA interaction, translation, and decay (Bevilacqua et al. 2016). Consequently, perturbation of normal RNA structure can lead to disease (Bernat and Disney 2015). Altered RNA structure caused by mutations may result in loss- or gain-of-function properties of the transcripts concerned—for example, splice alteration or sequestration of cellular factors, respectively—usually followed by a series of detrimental downstream effects in the cell.

Abnormal RNA structure plays a role in many repeat expansion diseases. Examples are myotonic dystrophy type 1 and 2 (DM1/2), Huntington's disease (HD), *C9orf72* amyotrophic lateral sclerosis/frontotemporal dementia (ALS/FTD), and fragile X-associated tremor-ataxia syndrome (FXTAS). In these disorders, mutant RNAs with an

expanded repeat tract are expressed, which can form stem-loop structures, commonly referred to as hairpins. In specific cases, like ALS/FTD and FXTAS, transcripts with GGGGCC and CGG repeats can also form anomalous G-quadruplex structures, stacked planar units of four G nucleotides associated via Hoogsteen interactions (Simone et al. 2015; Cammas and Millevoi 2017). G-quadruplexes and repeat hairpins are thought to facilitate anomalous interactions with RNA-binding proteins (RBPs) or have other *cis* and *trans* effects (Su et al. 2014; Zamiri et al. 2014; Bernat and Disney 2015; Błaszczuk et al. 2017; Cammas and Millevoi 2017) at the RNA level. Notably, expanded repeats have been found to participate in repeat-associated non-AUG (RAN) translation (Zu et al. 2010), although the relationship between this noncanonically initiated type of protein synthesis and repeat RNA folding structure is not yet well understood.

Here, to deepen our understanding of the pathological significance of abnormal topology of mutant RNA

Corresponding author: rick.wansink@radboudumc.nl

Article is online at <http://www.rnajournal.org/cgi/doi/10.1261/rna.068940.118>. Freely available online through the RNA Open Access option.

© 2019 van Cruchten et al. This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

transcripts with expanded repeats, we have focused on the multisystem disorder DM1. We have determined the secondary structure of differentially expanded CUG repeats, contained in the natural sequence context of the 3'-untranslated region (UTR) of *DMPK* mRNA. As a rule of thumb, healthy individuals carry repeat tracts with <37 triplets, whereas people with >50 triplets will likely suffer from DM1 symptoms at some point in life. Premutation individuals with 37–50 triplets are generally asymptomatic, but have a high risk of transmitting a repeat with pathogenic length to their offspring, due to genetic instability (Udd and Krahe 2012; De Antonio et al. 2016). CTG repeat expansion across generations in DM family members generally correlates with an earlier age of onset and more severe symptoms. In the most severe form of DM1, pathogenic repeats can consist of thousands of CTG triplets.

DM1 pathology is thought to occur mainly via aberrant protein interactions with the CUG repeat (Pettersson et al. 2015). These interactions may involve the sequestration of RNA-binding proteins (RBPs) such as splicing factors, for example, members of the muscleblind-like (MBNL) family, and the recruitment of ribosomes that engage the RNA in RAN translation. Other proteins such as CELF1 (CUGBP1), Staufin1 and DDX helicases are thought to associate also with, or be dysregulated by, expanded *DMPK* transcripts (Michalowski et al. 1999; Paul et al. 2011; Ravel-Chapuis et al. 2012). A classical hallmark of DM1 is the occurrence of ribonuclear foci in the cell nucleus, visualized by fluorescence in situ hybridization (FISH). These foci are thought to be complexes of one or at most a few expanded *DMPK* RNA molecules and associated RBPs (Gudde et al. 2016; Wojciechowska et al. 2018).

Earlier work on repeat RNA structure in DM1 has shown that CUG hairpins are “slippery” in nature and consist of G•C base pairs intermitted by relatively stable U•U mismatches (Napierała and Krzyzosiak 1997; Michalowski et al. 1999; Tian et al. 2000; Yuan et al. 2007). This conformation explains the binding preference of MBNL1 protein for the hairpin stem and the association of CELF1 with the hairpin base (Michalowski et al. 1999; Yuan et al. 2007). However, not yet all the effects of differential repeat lengths on the spectrum of pathology in DM1 can be easily explained by only this hairpin formation. Until now, the potential contribution of repeat-flanking sequences has been underappreciated in most studies.

RNA structure can be interrogated by chemical and enzymatic probes, combined with prediction of folding based on the thermodynamic properties of base-pairing. A well-known limit of secondary structure prediction solely based on thermodynamics is the generation of various possible structures with similar thermodynamic likelihood, especially in the case of long RNA molecules. However, by using constraints from probing studies the accuracy of these structural calculations can be greatly improved. Probes for RNA structure determination bind or cleave

either paired or unpaired nucleotides, often in a base-specific fashion, the locations of which can then be determined by reverse transcription (RT) primer extension (Weeks 2010; Sloma and Mathews 2015; Lorenz et al. 2016). In the last decade, selective 2' hydroxyl acylation analyzed by primer extension (SHAPE) has rapidly advanced RNA structure studies by providing a base-specific approach (Merino et al. 2005; Spitale et al. 2013).

Here, by combining SHAPE and dimethyl sulfate (DMS) probing analyzed using capillary electrophoresis, we determined the secondary structure of repeat tracts with 5–147 CUG triplets including almost 900 nucleotides (nt) of flanking sequence, as contained in the full length *DMPK* 3'-UTR encoded by exon 15. As a first step, we used in vitro approaches, enabling us to do structure determination of *DMPK* RNAs with different expansion lengths in isolation and avoid problems associated with mixing of normal and pathogenic allelic transcripts and the low abundance of *DMPK* transcripts in patient cells (Gudde et al. 2016). The inferred structure was validated by competitive binding with locked nucleic acid (LNA) antisense oligonucleotides (AONs). This allowed us to pinpoint how the CUG repeat interacts with the rest of the transcript. We next focused specifically on the folding of the CUG repeat itself by probing *DMPK* exon 15 containing up to 197 CUG triplets with RNase T1 and N-cyclohexyl-N'-(2-morpholinoethyl) carbodiimide methyl-p-toluenesulfonate (CMCT). Finally, the therapeutic AON targeting potential across *DMPK* exon 15 was calculated based on the structural information obtained, which correlated with previously determined AON efficacy in human muscle cells. Integrating this evidence, we present a model in which long CUG repeats can form branched hairpin structures in pathogenic *DMPK* transcripts, without altering the overall folding of the rest of the RNA molecule.

RESULTS

In vitro transcription generating *DMPK* exon 15 RNA with 5–197 CUG triplets

Our first step to find out whether there is a sequence-dominated effect of repeat expansion on *DMPK* RNA structure was to generate model molecules with a wide range of repeat lengths. To this end, we in vitro produced *DMPK* exon 15 RNA with a short poly(A) tail of 20 adenosine nucleotides carrying either 5, 21, 38, 69, 147, or 197 CUG triplets (Fig. 1A). With 378 nt 5' to the repeat and 505 nt 3' to the repeat, the size of these products ranged from 898 to 1474 nt. These transcripts are referred to as *DMPK*_{ex15} CUG_N RNAs. Capillary agarose gel electrophoresis analysis showed the formation of discrete products for the short-repeat transcripts (Fig. 1B). The RNA bands were broader for the two longest expansions with 147 and 197 triplets. This likely reflects the difficulty to fully

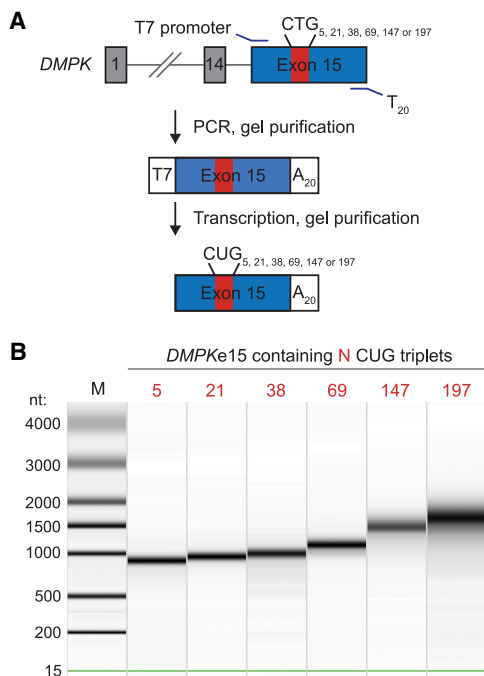


FIGURE 1. *DMPKe15* RNA production. (A) Scheme for the production of *DMPKe15* RNA containing various repeat lengths and a 3' A_{20} tail. To create templates for in vitro transcription, PCR using primers with overhang was performed on different plasmids containing the *DMPK* gene with different repeat lengths. RNA was generated by T7 polymerase-mediated transcription using the PCR products as templates. (B) Capillary gel electrophoresis of the produced *DMPKe15* RNAs to verify integrity and size. The range of expected product lengths is between 898 (CUG₅) and 1474 nt (CUG₁₉₇).

denature these expanded repeat RNAs for electrophoresis but may also indicate slight repeat-length heterogeneity. Sanger sequencing of the transcription templates confirmed the purity of the repeat and the correct composition of the flanking sequence.

Chemical probing shows independent folding of the CUG repeat and flanking *DMPK* sequences

To assess the folding of normal and repeat-expanded *DMPKe15* RNA, we performed SHAPE on transcripts with 5, 38, 69, and 147 CUG triplets. SHAPE methodology interrogates the structure of all 4 nt, providing a detailed overview of the folding of a particular RNA (Merino et al. 2005). Because of the high GC-content and the repetitive nature of the CUG repeat, RT across and within long repeats and the region 5' thereof is challenging (Carrell et al. 2018). To overcome this issue, the RT reaction was primed just 3' of the repeat (D16) and with a primer that was anchored in the repeat (D14). Still, for transcripts containing ≥ 38 CUG triplets we were unable to accurately call the region directly 5' of the repeat. Further primers were positioned at the ultimate 3' end (D18), and between the transcript ends and the repeat (D1 and D2) (Fig. 2).

For transcripts containing ≥ 38 CUGs (premutation and expanded repeat lengths), prominent reactivity around the middle of the repeat clearly indicated the single-stranded loop of the slippery hairpin that these sequences form (Fig. 2; Supplemental Table S1). The peak reactivity resided just 5' of the exact middle of the repeat. Moreover, the longer the repeat the more reactivity was observed at other sites across the repeat tract. This phenomenon could be indicative for structures alternative to a single hairpin and was investigated further with additional chemical and enzymatic probing (see below).

Interestingly, we found no obvious differences in SHAPE reactivity between normal and expanded *DMPK* RNA in the sequences that these transcripts have in common. To complement these observations, DMS probing, which identifies single-stranded A and C nucleotides, was applied. We found a pattern of reactivity very similar to SHAPE, including the robust signal around the middle of the repeat, supporting the above findings (Supplemental Fig. S1; Supplemental Table S1).

When applying the SHAPE reactivities for *DMPKe15* CUG₅ and CUG₁₄₇ RNAs as constraints for RNA secondary structure prediction, various structures were generated. We specifically identified one pair of structures that could explain the uniform folding of these transcripts outside the CUG repeat, which also matched the DMS reactivities well (Supplemental Fig. S2).

Challenging *DMPK* RNA structure with LNA oligonucleotides

To verify the accuracy of the pair of RNA structures with 5 and 147 CUG triplets described above, we performed SHAPE in the presence of short LNA AONs that are expected to disrupt the predicted base-pairing. One LNA, LNA tr445569, was designed as a truncated variant of a therapeutic AON (Wheeler et al. 2012), and one LNA was complementary to the repeat. It is expected that LNA binding displaces the binding partner of its target sequence, leading to elevated SHAPE sensitivity in the now unpaired strand (Sztuba-Solinska and Le Grice 2014; Low et al. 2015). Binding of LNA tr445569 to *DMPKe15* CUG₅ RNA was indeed confirmed by complete inhibition of SHAPE reactivity at the target site (C240–C250), resulting from perfect base-pairing with the LNA (Fig. 3A). LNA tr445569 induced a rise in mean SHAPE reactivity from 0.17 to 0.63 at G190 and 0.02 to 0.41 at U197; both sites were predicted in our proposed structure. Reactivity of nucleotide A258, just downstream from the LNA's binding site was also elevated upon LNA binding and increased from 0.515 to 1.83, potentially due to structural rearrangement. Altogether, these observations provide strong supportive evidence for the proposed pair of RNA structures for this part of the *DMPK* transcript. Also, they indicate that the therapeutic AON 445569

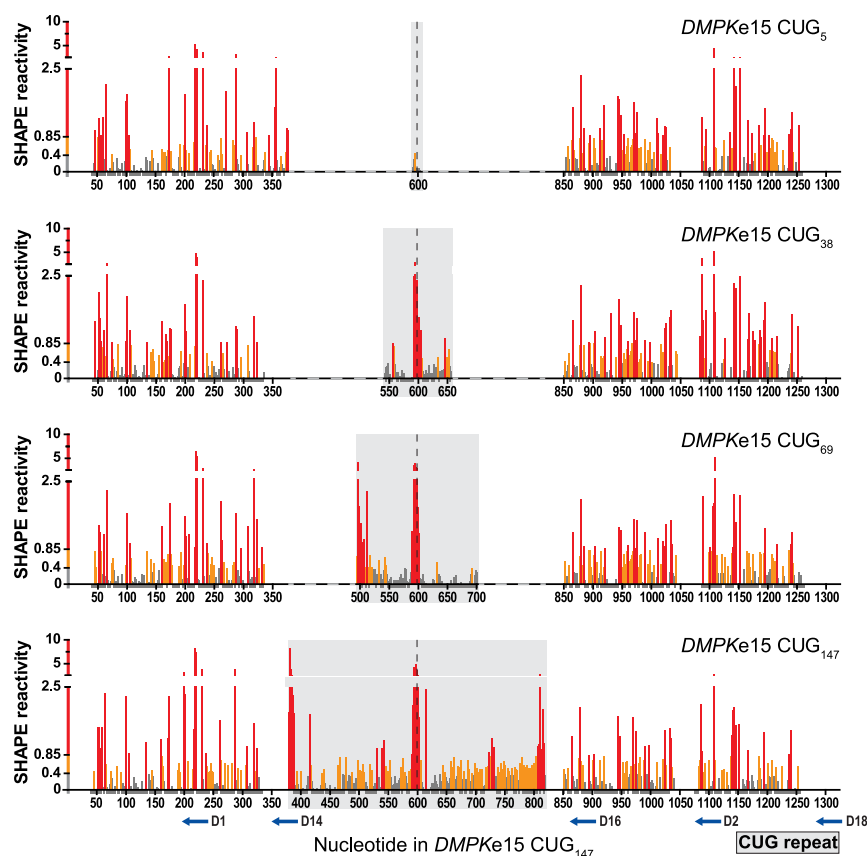


FIGURE 2. SHAPE reactivity of *DMPKe15* RNAs with either 5, 38, 69, or 147 CUG triplets. SHAPE experiments were performed in triplicate (*DMPKe15 CUG*₅) or duplicate (*CUG*₃₈, *CUG*₆₉, and *CUG*₁₄₇), after which data were averaged for bar plots. RT-primer locations are indicated with blue arrows below the x-axis. Numbering of the x-axis corresponds to the *DMPKe15 CUG*₁₄₇ RNA sequence. Values <−0.1, occurring from natural RT stops, were plotted as −0.1. Reactivity of 0.4–0.85 was considered moderate (in orange); >0.85 as high (in red). The dashed vertical line indicates the middle of the repeat. For the sequence directly 5' to the repeat, nucleotides 334–378, only data from *DMPKe15 CUG*₅ transcripts could be obtained. SHAPE reactivity profiles were clearly indicative of hairpin formation for *DMPKe15* transcripts with repeats containing ≥38 CUGs, with a main loop just 5' from the middle of the repeat and reactivity at both ends of the repeat. For transcripts containing 69 and 147 CUGs, reactivity was also observed along the rest of the repeat tract. Repeat expansion had minor or no effect on the reactivity of nonrepeat regions. All data points are listed in Supplemental Table S1.

targets a region that is moderately structured (Supplemental Fig. S3).

Pinpointing the interactions of repeat sequences with their flanking regions is of ongoing interest in the field (Napierała and Krzyzosiak 1997; Michlewski and Krzyzosiak 2004; Sobczak and Krzyzosiak 2005; De Mezer et al. 2011; Busan and Weeks 2013). We therefore considered it important to obtain independent evidence for the correctness of this part of our proposed structure for all different *DMPKe15 CUG*_N RNAs. The structure predicts that the 5' end of the repeat binds around 40 nt upstream and the 3' end binds around 110 nt upstream of the repeat's 5' end (Supplemental Fig. S2). By probing the structure after targeting the repeat with LNA (CAG)₃C, it was examined if and where the repeat bound back in *DMPKe15* transcripts

with 5 and 38 CUG triplets, as representative RNAs. In *DMPKe15 CUG*₅ RNA the CUG repeat itself featured low reactivity, which was maintained after LNA (CAG)₃C binding (Fig. 3B). For *DMPKe15 CUG*₃₈ RNA, the characteristic reactivity near the middle of the repeat was strongly decreased upon LNA binding and was spread along the entire repeat (Fig. 3C). This observation can be explained by the LNA consisting of three triplets and one additional C nucleotide, thus leaving a CpU dinucleotide unpaired for each bound LNA. Also, the 38 CUG triplets lower the effective RNA:LNA ratio due to the nine binding sites for LNA (CAG)₃C, which could leave a part of the repeat in every RNA molecule unbound at the 1:10 ratio that was used.

Increased reactivity upon LNA (CAG)₃C binding was indeed detected in the region ~110 nt upstream (274–283), where mean SHAPE reactivity at U283 rose from 0.96 to 1.27 for *DMPKe15 CUG*₅ RNA and from 0.60 to 0.90 for *DMPKe15 CUG*₃₈ RNA. The region ~40 nt upstream (335–339) showed a more profound increase: from 0.29 to 0.62 for G337 and 0.82 to 1.60 for G338 in *DMPKe15 CUG*₅ RNA, and from 0.16 to 0.39 for G337 and 0.40 to 1.13 for G338 in *DMPKe15 CUG*₃₈ RNA. Of note, in alternative RNA structures predicted by the folding software, the CUG repeat interacts downstream with nucleotides 420–424 and 470–479 in *DMPKe15 CUG*₅ RNA (data

not shown). We consider these structural conformations unlikely as no substantial change in SHAPE reactivity was observed at these sites upon LNA (CAG)₃C binding (Supplemental Fig. S4).

Expanded CUG repeats can form multi-stem-loop structures

In our chemical probing experiments, we observed reactivity in what would be the hairpin stem of an expanded CUG repeat (Fig. 2). Also, many of the predicted structures containing expanded repeats featured multibranching loops in the repeat or presence of multiple hairpins, with nearly identical predicted thermodynamic stability (data not shown). To investigate whether these structural

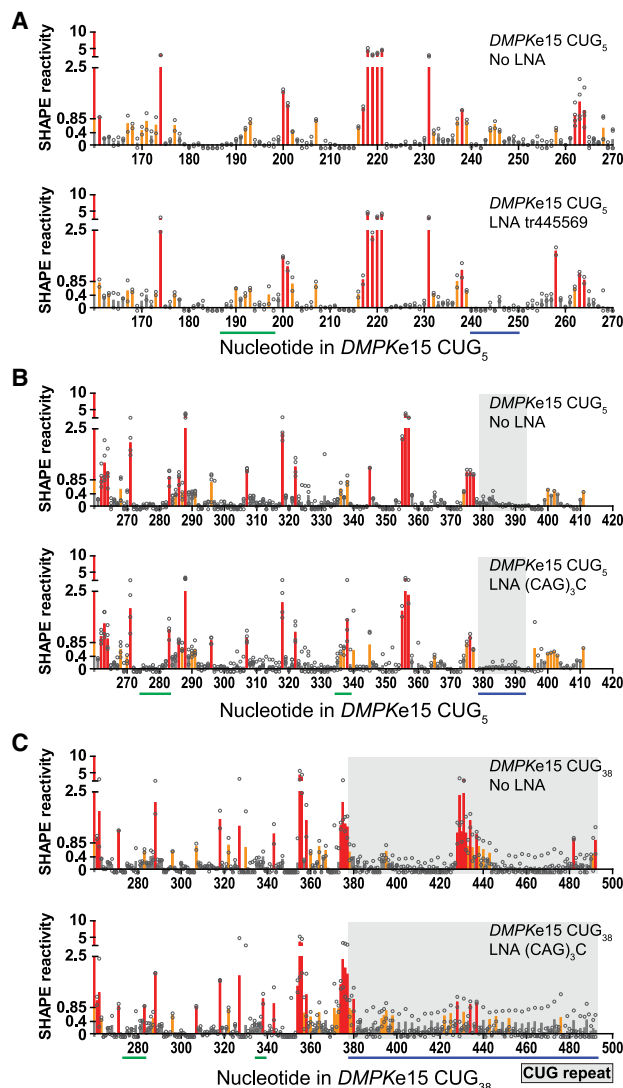


FIGURE 3. LNA-induced structure perturbation for validation of the proposed *DMPK*₁₅ RNA structure. *DMPK*₁₅ RNA containing 5 (A and B) or 38 CUGs (C) was probed using SHAPE in the presence of LNA tr445569 (A) or LNA (CAG)₃C (B and C). RT was performed from one (A) or two (B and C) primer locations. Each circle represents data from one out of two to four measurements, which were averaged for bar plots. Underlined in blue are the binding sites of the LNAs, and underlined in green are the expected interaction sites of these LNAs based on the structure in Supplemental Figure S2. Except for the CUG repeat and the region 3' thereof, the x-axis numbering corresponds to Figure 2; further representation of the data is as in Figure 2. See Supplemental Figure S3 for details on the binding site of LNA tr445569.

conformations were indeed differentially represented in *DMPK*₁₅ with different repeat lengths, we applied RNase T1 probing on *DMPK*₁₅ RNA with either 21, 38, 69, 147, or 197 CUG triplets and analyzed specifically the repeat region. RNase T1 cleaves RNA 3' of single-stranded guanines only, leading to less signal decay when probing long RNA stretches compared to NAI (SHAPE) and DMS,

which react with either all or two nucleotide types, respectively. As our experimental analyses implied a trade-off between signal:noise ratio and read length, we were generally only able to obtain reliable signals from the middle and 3' side of the repeat. The much weaker signal that was obtained from the 5' side showed a more or less symmetrical cleavage pattern when comparing the regions 5' and 3' from the main hairpin loop (Supplemental Fig. S5A). This finding served as an internal control and suggests that the two halves of the CUG repeat are structurally equivalent. We therefore focused on data obtained from the middle and 3' side of the repeat only.

The RNase T1 cleavage profile resembled the chemical probing results in that the most prominent reactivity was found just 5' of the exact repeat middle (Fig. 4). Reactivity at the ultimate 3' end of the repeat was less profound, potentially limited by the larger size of the T1 enzyme as compared to the small chemical probes. Interestingly, results obtained with transcripts containing 21 or 38 CUG triplets deviated markedly from those of expanded transcripts with 69, 147, or 197 CUGs. Transcripts with 21 or 38 CUG triplets featured only prominent cleavage just 5' of the middle of the repeat, indicating a single hairpin. In the expanded transcripts, however, RNase T1 cleavage was observed essentially all along the entire CUG repeat tract. Cleavage intensity was unevenly distributed, in a wavy pattern. Remarkably, the ~25 nt immediately adjacent to the middle of the repeat were relatively shielded from RNase T1 cleavage, which indicates that this region is double stranded in most conformations. Based on these observations, we consider it likely that long repeats do not only fold as one single hairpin, but also undergo dynamic changes, forming structures with multiple hairpins and multibranch loops.

Even at high doses of T1 enzyme (10- to 100-fold of that in Fig. 4), cleavage at sites other than the middle and the ends of the repeat was not observed in transcripts with 38 CUGs, supporting the notion that such reactivity is unique to transcripts with long repeats (Supplemental Fig. S5B). To rule out that these observations were caused by secondary cleavage events of refolded RNA after a first cleavage by RNase T1, a range of enzyme concentrations (0.001–0.1 U/μL) and incubation times (0.5–10 min) was tested on *DMPK*₁₅ CUG₁₄₇ RNA, but no differential effect on the cleavage pattern was ever seen (Supplemental Fig. S5C).

We also probed the repeat of *DMPK*₁₅ CUG₁₄₇ and CUG₁₉₇ RNA with CMCT, which binds at unpaired G and U bases and is thus not sensitive to secondary cleavage events. These experiments produced a less pronounced but similar pattern to DMS, SHAPE and RNase T1 experiments, with reactivity arising from the main hairpin loop in the middle of the repeat and in clusters in the remainder of the repeat (Supplemental Fig. S6).

Besides monovalent ions, also Mg²⁺ ions influence RNA folding stability, having a greater effect on tertiary structure

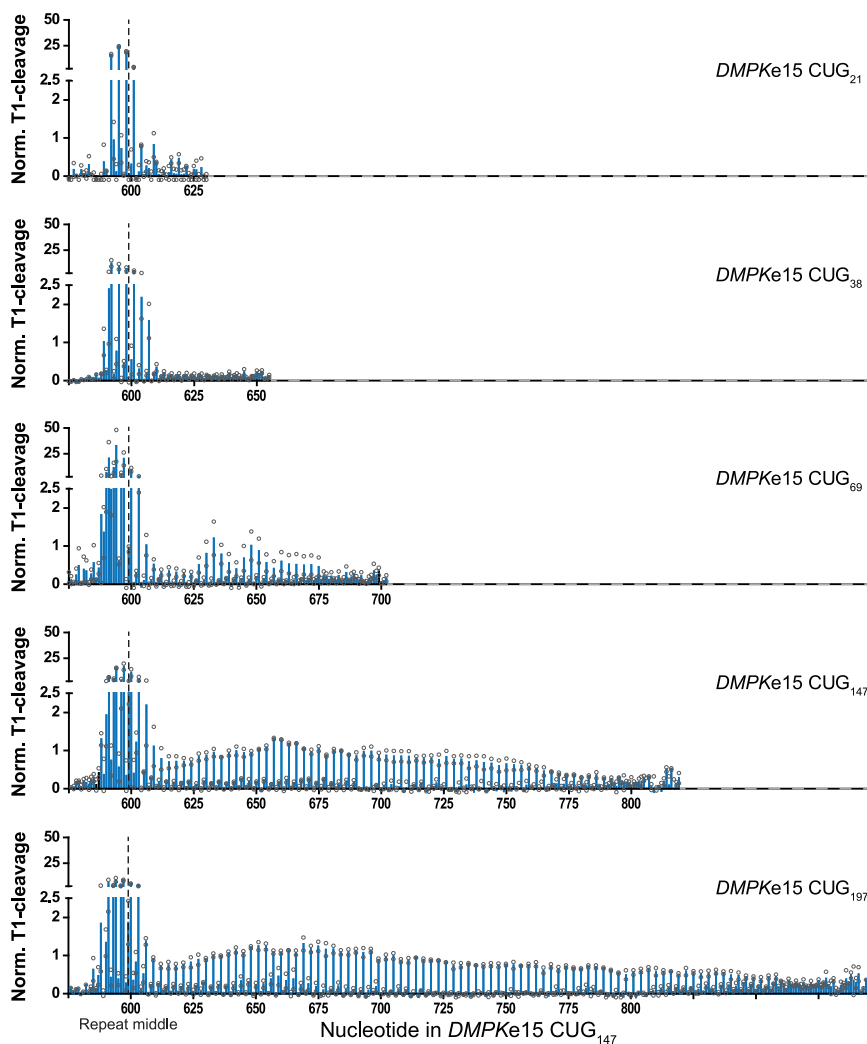


FIGURE 4. RNase T1 probing of *DMPKe15* RNA with either 21, 38, 69, 147, or 197 CUG triplets. *DMPKe15* RNA was incubated with 0.01 U/ μ L RNase T1 for 10 min at 37°C. Normalized T1 cleavage intensity per nucleotide for each replicate is represented as a circle and replicates were averaged for bar plots. Shown are only the region surrounding the middle of the repeat and the 3' side of the repeat, since reliable analysis of the 5' side was usually not possible due to signal decay in that area. Numbering corresponds to nucleotides in *DMPKe15 CUG*₁₄₇ RNA to facilitate comparison with the chemical probing experiments. The dashed line indicates the middle of the repeat. See Supplemental Figure S5 for additional RNase T1 probing conditions and analysis of the complete repeat in *DMPKe15 CUG*₁₄₇ RNA.

than on secondary structure of RNA, and with the ability to induce nonphysiological conformations at high concentration (Draper 2004). To exclude that the observed alternative folds arose from Mg^{2+} effects at the standard concentration used (6 mM), we performed RNase T1 probing on *DMPKe15 CUG*₁₄₇ RNA after folding in the presence of a range of $MgCl_2$ concentrations (0–10 mM). This probing revealed highly similar cleavage patterns for each condition, indicating that folding of the CUG repeat is not a function of Mg^{2+} concentration (Supplemental Fig. S7).

In an attempt to obtain independent evidence for existence of branched repeat-RNA conformations, we also separated normal and repeat-expanded *DMPKe15* RNAs on a

nondenaturing agarose gel (Supplemental Fig. S8). For all RNAs analyzed, gel staining revealed only one single intense band. Since in our experiments these alternative RNA conformations were only present in a subpopulation of molecules and these RNAs may occur in distinct conformations with different migratory properties, we assume that distinct RNA folds within this subpopulation are too rare to be detected by gel electrophoresis. Still, given the time-, dose-, probe-, and Mg^{2+} independency of signals in our probing experiments, we consider it unlikely that our findings arise from false-positive signals or in vitro artifacts. Based on the ratio of the intensity of T1 cleavage near the middle of the repeat and the wavy pattern in the 3' side, we estimate that 3% (*CUG*₆₉) to 15% (*CUG*₁₉₇) of the RNA molecules adopted a conformation other than a single hairpin.

The CUG repeat is structurally isolated via a clamp and GC-rich helices

By integrating the different types of experimental evidence, we propose the secondary structures for *DMPKe15* RNAs depicted in Figure 5 and Supplemental Figure S2. These models show that the CUG repeat base pairs with two distinct sites in the 5'-flanking sequence, in between of which an expanded repeat can bulge out. This explains why the most intense chemical modifications were not situated in the exact middle

of the repeat: 10 nt of the 3' end of the CUG repeat bind back on the flanking sequence versus only 4 nt of the 5' end. Thus, this binding forces the formation of a hairpin loop to a location just 5' of the middle of the repeat tract. The ultimate 5' end of an expanded repeat was initially predicted single stranded. However, after extrapolating the findings with LNA (CAG)₃C in *DMPKe15 CUG*₃₈ transcripts, we consider it most likely that an interaction, likely of dynamic nature, occurs between the repeat and nucleotides at positions 335–339 in the proposed structure.

The structure surrounding the CUG repeat consists of numerous GC base pairs and is thus expected to be highly stable. Especially a series of six G nucleotides right next to

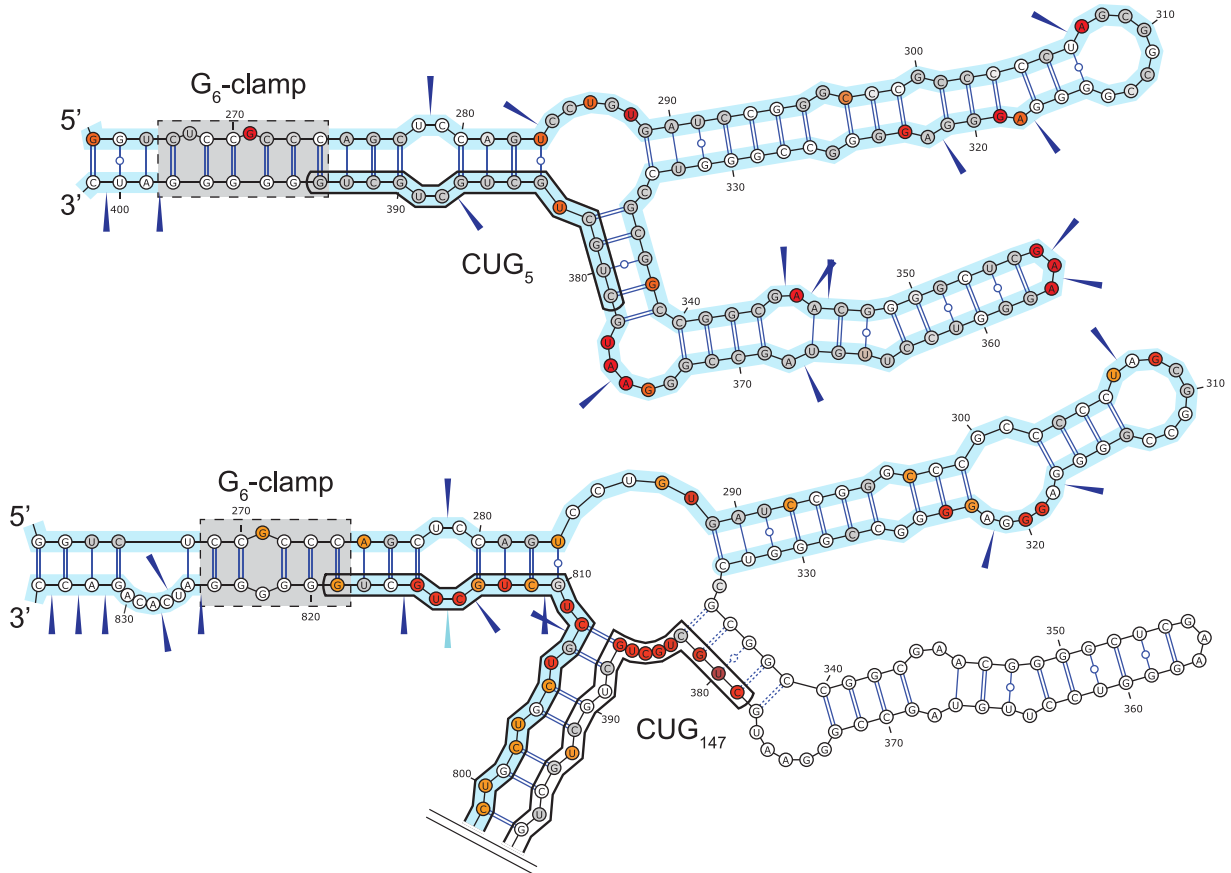


FIGURE 5. Proposed RNA structure of the region surrounding the CUG repeat in *DMPK*₁₅ RNA containing 5 and 147 CUG triplets. The CUG repeat (outlined in black) is likely kept in place by six G nucleotides at the 3' end of the repeat, which can base pair with a C-rich motif (~125 nt upstream in the case of a CUG₅ repeat) forming the G₆-clamp. The repeat is flanked by the G₆-clamp and two imperfect helices with a high GC-content, in between of which an expanded repeat bulges out. We expect that the two variants of the G₆-clamp (compare both structures) and the folding of the other helices shown here are not exclusive for these repeat lengths, but are more likely in equilibrium. Base-pairing of the four most 5' nucleotides of an expanded repeat with the flanking region is likely highly dynamic, and is therefore indicated with dotted lines. SHAPE reactivity is color coded in white (no data), gray (no or low reactivity, <0.4), orange (medium reactivity, between 0.4 and 0.85), and red (high reactivity, >0.85). Blue highlighting indicates availability of DMS data, dark blue triangles point upstream of nucleotides with high normalized DMS reactivity (>0.85). Light blue triangles indicate high DMS reactivity for U and G nucleotides, originating from misalignment or overlapping electrophoresis peaks. For complete structures of *DMPK*₁₅ RNA, see Supplemental Figure S2.

the 3' end of the repeat, which we coin the G₆-clamp, seems important. The G₆-clamp is predicted to bind a C-rich motif (CUCCGCC) ~125 nt upstream, in between of which two GC-rich double-stranded helices are formed. Supporting evidence for the interaction of the G₆-clamp with the C-rich motif is that nucleotide U283, 8 nt downstream from that motif, was found to be more reactive upon CAG LNA binding and is thus likely in close contact with the repeat. Of note, nucleotide A275, just downstream from the C-rich motif and binding to the repeat, was found to be more reactive upon repeat expansion (mean SHAPE reactivity of 0.00, 0.19, 0.11, and 0.66 for *DMPK*₁₅ CUG₅, -CUG₃₈, -CUG₆₉, and -CUG₁₄₇ transcripts respectively; DMS reactivity of 0.41 vs. 0.73 for CUG₅ and CUG₁₄₇ RNAs). This could point to a dynamic interaction with the expanded repeat, but it cannot be definitively ruled out that this particular observation

resulted from experimental noise because of multiple testing, since hundreds of nucleotides were measured per transcript.

To further study the influence of the G₆-clamp, we mutated the GGGGGG motif to CTTAAC in *DMPK*₁₅ CUG₅ RNA, and mutated the C-rich motif (CTCCGCC) to ATAAGAAA in transcripts with 5 and 69 triplets. Mutation of the six Gs in *DMPK*₁₅ CUG₅ RNA resulted in a moderately higher SHAPE reactivity in the repeat, suggesting that the stem formed by the G₆-clamp and the adjacent repeat sequence was indeed disrupted (Supplemental Fig. S9A). The typically high SHAPE reactivity of nucleotide G271 in the C-rich motif, which is in a constrained single-stranded conformation in the wild type transcript (see Fig. 5), was completely absent after mutating the six G nucleotides (Supplemental Fig. S9A). Mutation of the C-rich motif resulted in a slight increase

in SHAPE reactivity of the repeat in *DMPKe15* CUG₅ and CUG₆₉ transcripts (Supplemental Fig. S9A,B). This disruption of the repeat structure in C-mutant *DMPKe15* CUG₆₉ transcripts was confirmed by RNase T1 probing (Supplemental Fig. S9C). More strikingly, where in the wild type situation the peak SHAPE and RNase T1 reactivity resided 5' of the middle of the repeat, upon mutation of the C-rich motif this reactivity was situated almost exactly in the repeat middle. Both mutations induced no other overt effects in the folding structure of the ~150 nt upstream and downstream from the repeat that were measured. This indicates that it is not the G₆-clamp alone, but likely also the GC-rich helices enclosed by the repeat and the G₆-clamp that result in the limited effect of the repeat length on the overall folding of *DMPK* RNA. We attempted as well to disrupt the G₆-clamp by targeting it with an LNA oligo (ATCCCCCAGC, LNAs in bold), but found that this LNA did not bind efficiently and specifically to its intended target (data not shown). Taken combined, our observations strongly suggest that internal elements in *DMPK* exon 15 structurally isolate the repeat from the rest of the RNA molecule and contribute to the repeat structure.

SHAPE-determined *DMPK* RNA structure correlates with AON knockdown efficacy in cells

Knockdown of expanded *DMPK* transcripts by AONs is a promising therapeutic strategy for DM1, and structural insight into target accessibility for AONs would facilitate further development of these potential medicines (Mulders et al. 2009; Wheeler et al. 2009, 2012; Lee et al. 2012; González-Barriga et al. 2013). As proof of principle, we applied the OligoWalk algorithm in RNAstructure 5.8 to predict target sequences in *DMPK* exon 15 RNA which, based on their secondary structure, would be most accessible to AONs (Lu and Mathews 2008; Reuter and Mathews

2010). Figure 6 shows the free energy cost of disrupting the local structure in our *DMPKe15* RNA molecule by binding of every possible AON of 20 nt. Comparison of this energy plot with the *DMPK* knockdown efficacy in human skeletal muscle cells by >450 AONs targeted at 258 sites as reported previously, revealed that free energy cost correlated modestly but highly significantly ($P < 0.0001$) with the efficacy of the AONs tested (Swayze et al. 2014; Bennett et al. 2015). Around 6% of the efficacy was accounted for by the energy cost of structure disruption (Pearson's $r^2 = 0.055$). It is of note that this correlation was not found when the OligoWalk calculation was based on the folding prediction without our SHAPE constraints ($P = 0.23$, $r^2 = 0.0056$). The free energy cost-AON efficacy correlation was independent of AON chemistry, length, and concentration used (data not shown). Thus, prediction of RNA structure based on chemical probing in vitro is a reliable predictor for one of the many parameters that determine AON efficacy in cells.

DISCUSSION

In this study, we propose models for the secondary structure of normal and repeat-expanded *DMPK* transcripts and shed light on the behavior of long CUG repeats in their natural sequence context (summarized in Fig. 7). By in-depth chemical and enzymatic interrogation of RNA structure of significant CUG repeat lengths and flanking sequences, we found that the CUG repeat interacts with nucleotides upstream in the *DMPK* transcript and that repeat expansion does not influence overall folding of the rest of the RNA molecule. An expanded CUG repeat protrudes from the *DMPK* transcript and is able to form branched or multiple hairpin structures. These findings were confirmed using LNA oligonucleotides and placed into the context of targeting *DMPK* mRNA by AONs in the living cell.

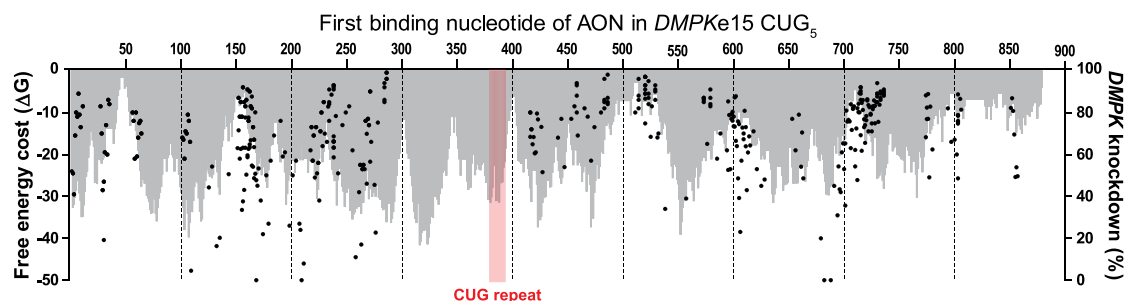


FIGURE 6. Calculated free energy cost for disruption of *DMPKe15* RNA structure by AON binding related to *DMPK* knockdown efficacy in cells. Gray bars indicate free energy cost (left y-axis) due to structure disruption by 18-nt AON binding as determined by OligoWalk (Lu and Mathews 2008), using 20 potential RNA structures generated by the *Fold* algorithm (Reuter and Mathews 2010) based on the SHAPE data presented in Figure 2 and Supplemental Table S1. Free energy cost is a parameter for AON accessibility, where a more negative value indicates a less accessible RNA sequence. Black dots show *DMPK* knockdown efficacy (right y-axis) by AONs in human skeletal muscle cells as previously reported (Swayze et al. 2014; Bennett et al. 2015). A bar or dot is plotted on the first, 5', nucleotide of *DMPKe15* CUG₅ RNA that the AON binds to. Note that in this figure, the numbering 3' of the repeat does not correspond to that in the other figures due to the different repeat length.

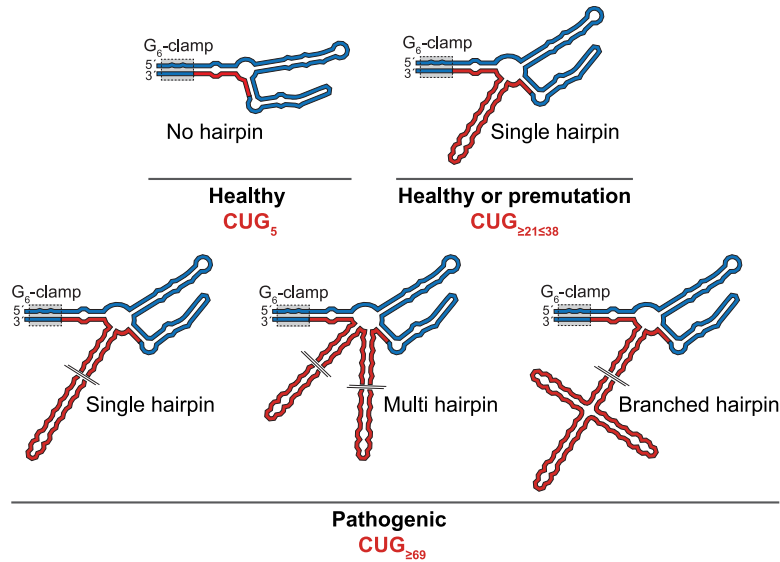


FIGURE 7. RNA secondary structure model of normal and expanded CUG repeats in the context of flanking *DMPK* sequences. Shown are RNA models of the CUG repeat (red) with ~ 100 5'-flanking nucleotides and ~ 10 3'-flanking nucleotides (blue) in *DMPK*. We propose that only a very short repeat of ~ 5 CUGs does not form a hairpin and that nondisease-causing repeats of up to 38 CUGs form exclusively small, single hairpins. The structure of pathogenic repeats of at least 69 CUG triplets forms an equilibrium between single, long slippery hairpins and branched or multihairpin structures. Particularly, these latter structures may contain protein binding sites unique to pathogenic transcripts in DM1 patients. Repeat expansion does not lead to changes in the RNA structure of repeat-flanking sequences. This is likely due to the repeat being isolated from the rest of the transcript via the G_6 -clamp and two stable GC-rich helices. The repeat is thus able to adopt alternate conformations without affecting the overall structure of the RNA.

The CUG repeat is structurally isolated from the rest of the *DMPK* transcript

We identified a highly structured environment especially 5' of the CUG repeat in *DMPK* RNA. This is composed of a G_6 -clamp along with two stable GC-rich helices that structurally isolate the repeat from the rest of the transcript, leading to the similar folding of sequences common to healthy and repeat-expanded transcripts. The interactions that make up this structure span further than the 30 repeat-flanking nucleotides tested earlier, explaining why these were found single stranded in that study (Napierała and Krzyzosiak 1997). Since the G_6 -clamp only fixates the 3' side of the CUG repeat, slippage can still take place. This is in contrast to the situation in CAG-repeat-expanded *CACNA1A* RNA (related to *SCA6*) for example, where the repeat is sandwiched between complementary GC-rich stretches that serve as clamp (Michlewski and Krzyzosiak 2004).

Expanded CUG repeats form branched hairpin structures in their natural sequence context

More than 20 years ago, pioneering work by Napierała and Krzyzosiak showed how CUG repeats of up to 49 triplets flanked by 30–35 *DMPK* nucleotides can form single slippery hairpins (Napierała and Krzyzosiak 1997). We also

observe that repeats of such length form slippery hairpins, but longer repeats in the context of *DMPK* exon 15 formed a more heterogeneous population. To our best knowledge, we are the first to provide experimental evidence that expanded repeat sequences actually can form branched or multiple hairpin structures. The formation of this type of RNA secondary structures in triplet repeat diseases has been hypothesized before, but was never experimentally demonstrated (Napierała and Krzyzosiak 1997; Koch and Leffert 1998; Broda et al. 2005). Michalowski et al. (1999) performed electron microscopy on RNAs consisting of 54–130 CUG triplets and ~ 200 –300 nt flanking *DMPK* sequence, where they only observed single hairpins. Tian et al. (2000) reported pure CUG repeats of up to 140 triplets to form only single hairpin structures based on chemical and enzymatic probing followed by PAGE. The sensitivity of these approaches is however low and rare species could be easily missed.

We found that the incidence of branched structures increased with re-

peat length, with around 15% of the population for *DMPK* exon 15 with 197 CUG triplets. This could have implications for the repeat lengths found in congenital DM1 patients, which exceed the lengths tested here by a factor of 10 (Udd and Krahe 2012). We expect that long branches of such repeats can form sub-branches as well, generating multibranch hairpins, but this warrants further study. In a subset of DM1 patients, the CTG repeat is interrupted with CAG, CCG, CGG, GGC, or CTC triplets (Leeflang and Arnheim 1995; Musova et al. 2009; Braida et al. 2010; Tomé et al. 2018). Except for the CAG interruptions, these interrupted repeats can be expected to form even more extensively branched (but also more rigid?) structures due to imperfect pairing, similar to CAA-interrupted CAG repeats in *SCA2* (Sobczak and Krzyzosiak 2005).

Our study adds to existing observations and knowledge on several points. Firstly, we have analyzed transcripts with unprecedented lengths of the CUG repeat as well as authentic *DMPK* flanking sequences and determined the structures of both in depth. Secondly, by using capillary electrophoresis to determine RT stop frequency, we have been able to identify subtle differences induced by long repeat expansions in vitro. Knowing the structure of *DMPK* transcripts in the living cell can be seen as the “holy grail,” but we expect studying the differences between normal and expanded *DMPK* RNA structure in a

cellular context to be highly challenging for several reasons. Firstly, *DMPK* RNAs in DM1 patient cells always occur as a mixture of normal and expanded variants, which are not readily distinguishable in next generation sequencing. Secondly, *DMPK* transcripts occur in low copy numbers of only five to 25 transcripts per cell (Gudde et al. 2016), leading to sensitivity problems. On top of that, normal and expanded *DMPK* RNAs show a different nuclear versus cytoplasmic distribution (Gudde et al. 2017), resulting in a differential decoration with RNPs that might influence both the RNA folding structure as well as the probe interaction (Lorenz et al. 2016).

Notably, when extrapolating our in vitro findings, we have to keep in mind that RNA is generally less structured in the living cell than in vitro and can be differentially and dynamically altered by protein interactions in a location-dependent manner, calling for careful interpretation of in vitro investigated RNA structures (Ding et al. 2014; Rouskin et al. 2014). It should be kept in mind also that we assumed the interaction of the repeat with sequences outside exon 15 to be minimal due to the large distance. From our in vitro system, we did learn that there are no intrinsic sequence-dominated differences in the structure of repeat-flanking regions between normal and repeat-expanded *DMPK* RNA, so our assumption may have value for future studies.

CUG repeat stem-loop structures as abnormal protein binding platforms

The branched hairpin model has implications for understanding aberrant protein interaction with expanded repeats, a central theme in the study of DM1 etiology. MBNL1 preferentially binds its epitope YGCY in an unstructured environment, which is abundant in the branched, dynamic structures that we propose, but rare in perfect hairpins (Taylor et al. 2018). This notion could answer the question raised before on how MBNL1 normally interacts with mostly single-stranded targets, but nonetheless binds CUG repeat RNA that was considered highly structured (Cass et al. 2011). This requirement regarding the dynamic nature of CUG repeats for MBNL1 binding was highlighted further by the finding that stabilization of the CUG hairpin structure by pseudouridylation lowers its affinity for MBNL1 (deLorimier et al. 2014). CELF1, on the other hand, has been shown to interact specifically with the base of CUG hairpins in single-stranded surroundings, occurring only once in a single hairpin (Michalowski et al. 1999). In multiple hairpins and branched loops, however, there are several locations that fulfill these requirements, resulting in multiple binding sites for CELF1 in one RNA molecule.

RNA structure also plays an important role in the initiation of RAN translation, a phenomenon most likely involved in the disease process of DM1 and many other repeat ex-

pansion disorders (Cleary and Ranum 2017). CGG- and GGGGCC-initiated RAN translation appears to be dependent on scanning ribosomes that stall in regions with a stable secondary structure and then start translation at non-AUG start codons (Kozak 1990; Kearse et al. 2016; Tabet et al. 2018). As in *FMR1*, the region upstream of the repeat in *DMPK* is high in GC (~75%) and forms a stable secondary structure. Scanning ribosomes, however, are not readily expected to occur in the 3'-UTR of *DMPK*, unless the stop codons are leaky and allow for read-through (at least three different stop codons exist in *DMPK* exon 15 as a result of alternative splicing [Wansink et al. 2003]). RAN translation of long, isolated CUG tracts can also be induced by the repeat itself, likely through an internal ribosome entry site (IRES)-like mechanism (Zu et al. 2010). For CUG as well as CAG repeats, it was shown that RAN translation is independent of a start codon, but rather depends on repeat length, with threshold repeat lengths of 35–50 triplets (Zu et al. 2010; Todd et al. 2013; Bañez-Coronel et al. 2015). Interestingly, this threshold corresponds to the length required for the formation of branched hairpin structures and thus could reflect structural features unique to expanded transcripts that facilitate RAN translation.

Relevance of *DMPK* RNA structure to therapeutic AON design

Our newly gained insight into the in vitro RNA structure of *DMPK* exon 15 was applied for prediction of AON binding potential of all possible target sequences. SHAPE and DMS probing combined with LNA-mediated disruption of RNA structure brought us close to a single, most likely, in vitro RNA structure of *DMPK* exon 15, even though it cannot fully be ruled out that (parts of) other structures would fit our data as well. For the prediction of AON binding efficiency, we partly overcame this uncertainty using the OligoWalk algorithm, which takes into account suboptimal RNA structures and thus gives a less biased overview (Lu and Mathews 2008). Only when our SHAPE-based parameters were applied to the OligoWalk algorithm a significant correlation was found between the degree of RNA structure and the efficacy of AONs tested in cells. We recognize that the low r^2 value indicates that in vitro RNA structure has only limited predictive value for AON targetability as a whole. Still, since AON efficacy in living cells is dependent on so many factors that were not part of our prediction (e.g., RNA degradation mechanisms, base-pairing thermodynamics, endosomal release, nuclear transport, etc.), we find it remarkable that ~6% of the efficacy of AONs in cells can be predicted based on in vitro RNA structure.

The two most thoroughly investigated AONs for DM1, AON 445569 (IONIS-DMPKRx), and AON 486178, bind to nucleotides 238–257 and 803–818, respectively, in *DMPK* exon 15 CUG₅ RNA (Wheeler et al. 2012; Pandey et al. 2015; Jauvin et al. 2017). For AON 445569, we predict a

structure disruption cost ΔG of -24.9 , which is relatively unfavorable, and for AON 486178 we predict a ΔG of -11.6 , which is highly favorable. This difference in ΔG is reflected in the higher potency of the latter AON in a DM1 mouse model, but whether this represents a causal relationship cannot be proven based on experiments thus far (Jauvin et al. 2017). Based on the predicted free energy cost and excluding regions with low-complexity or high GC content (Supplemental Fig. S10), one can speculate that AONs starting around nucleotides 45 and 350, which have not yet been tested, could have high binding potential, assuming these nucleotides do not interact beyond exon 15.

Our findings further imply that since the repeat length does not influence the overall structure of the *DMPK* transcript, allele-specific targeting of the body of the transcript solely based on RNA structure is likely not a viable, discriminative strategy. Similar to what we observed for *DMPK*, sequence context of the repeat in the HD-causing *HTT* transcript has also been shown of importance for its structure and targetability (Busan and Weeks 2013). In this transcript, a healthy CAG repeat interacts extensively with flanking regions, which is linked to the specificity of repeat-binding AONs for expanded *HTT* transcripts (Hu et al. 2009). Since in *DMPK* the interaction of the CUG repeat with its flanking regions concerns only a few triplets, we do not expect a similar allele-selective effect there.

In summary, RNA structure determination can thus help in the development of AONs targeting the transcript outside the repeat, one of the main therapeutic approaches for DM and other trinucleotide expansion disorders today (Wheeler et al. 2012; Jauvin et al. 2017).

MATERIALS AND METHODS

Template preparation and T7 in vitro transcription

Plasmids containing the *DMPK* exon 15 (NCBI RefSeq NM_004409.4) sequence with either five, 21, 38, 69, 147, or 197 pure CTG triplets originated from a pKSS plasmid with a DM500 (Seznec et al. 2000) transgene fragment, in which during growth in *E. coli* contraction of the repeat had occurred naturally. T7 in vitro transcription templates were generated by PCR using these plasmids and forward primer 5'-TAATACGACTCACTATAGGGTCCCTAGGCCTGGCCTATC-3' consisting of the T7 promoter, the last two Gs of exon 14 (to obtain optimal T7 activity; underlined) and the first 20 nt of *DMPK* exon 15. The reverse primer, 5'-TTTTTTTTTTTTTTTTTTTTGGGCAGATGGAGGGCCTTTATTTCG-3', contained an overhang of 20 Ts to add a mimic of a poly(A) tail and was further designed to bind the final 25 nt of the *DMPK* gene. For templates with ≤ 69 CTGs, PCR was performed using Q5 High-Fidelity DNA Polymerase (NEB) as described by the manufacturer for GC-rich templates. For longer CTG repeats, the PCR protocol described by Meng et al. (2015) was applied in the presence of 200 μ M dNTPs. All PCR products were purified from agarose gel using the NucleoSpin kit

(Macherey-Nagel) and sequences were verified by Sanger sequencing (Radboud Genomics Technology Center). The only deviation from the NCBI reference gene was in the CTG₁₉₇ template, with a TG>CT substitution at positions 2408/2415 in NM_004409.4 (873/874 in *DMPKe15* CUG₁₉₇), which was used only for measurements of the repeat structure.

Side-directed mutagenesis by PCR was applied to mutate the G₆ and C-rich motif from the plasmid templates described above. To mutate the G₆-motif (G393–G398) to CTTAAC, first a PCR was carried out as described above with the T7 forward primer and 5'-CTGCTCTTAACATCACAGACCATTCTTCTTTTCGG-3' as reverse primer (mutation underlined). Secondly, a similar PCR was carried out with 5'-CTGTGATGTTAAGAGCAGCAGCAGCAGCATTCC-3' as forward primer and 5'-AGATGGAGGGCCTTTTATTTCG-3' as reverse primer. The purified products of these two PCRs were mixed in a 1:1 molar ratio, after which PCR and purification were performed to generate T7 templates as described above. The same approach was used to mutate the C-rich region (C267–C274) to TTTTCTTAT, with the primers 5'-CTGGAGC TTTTCTTATACCCACGCTCGG-3' and 5'-GTGGGTATAAGAA AAGCTCCAGTCCTGTGA-3' to produce the region 5' and 3' to the mutation, respectively.

In vitro transcription was performed using the HiScribe T7 High Yield RNA Synthesis Kit (NEB) according to manufacturer's instructions, except for transcripts with ≥ 147 CUGs, which were synthesized at 20°C. In vitro transcription resulted in transcripts composed of the last 154 nt of the *DMPK* open reading frame (splice isoforms A–D [Groenen et al. 2000]) and the complete 3'-UTR (1943–2863 in NM_004409.4) with 20 A nucleotides to mimic a poly(A) tail. We refer to these transcripts with N CUG triplets as *DMPK* exon 15 (*DMPKe15*) CUG_N RNA. To remove the DNA template after in vitro transcription, the reaction volume was supplemented to 100 μ L with DEPC-treated water and DNase I reaction buffer and was incubated for 15 min at 37°C in the presence of 2 U DNase I (Thermo). RNA was concentrated using a NucleoSpin column as described by the manufacturer, loaded on agarose gel under denaturing conditions (Masek et al. 2005) and purified using the NucleoSpin kit. An additional ethanol precipitation was performed to remove carryover of chaotropic salts, after which RNA purity was checked by UV-VIS spectrometry (Nanovue, GE). Integrity of the RNA was determined using a QIAxcel Advanced capillary gel electrophoresis machine (QIAGEN) according to manufacturer's instructions, using a double final concentration of denaturing agents to achieve more complete denaturation of the long CUG repeats. For non-denaturing gel electrophoresis, 200 ng *DMPKe15* RNA was first denatured and refolded as described below for the SHAPE procedure, mixed with Ficoll/bromophenol blue loading dye and then separated on a TAE-buffered 1.2% agarose gel. The gel was post-stained with ethidium bromide and visualized under UV light.

SHAPE

The SHAPE procedure was based on the protocol by Wilkinson et al. (2006): 0.35 pmol RNA was dissolved in 12 μ L 0.5 \times TE pH 7.6 and was denatured by heating at 65°C for 10 min. The RNA was placed directly on ice and 6 μ L folding buffer (333 mM HEPES pH 7.6, 20 mM MgCl₂, 433 mM NaCl) was added. The RNA was allowed to refold at 37°C for 30 min, after which a final

concentration of 20 mM 2-methylnicotinic acid imidazolide (NAI; Merck-Millipore) in DMSO or just DMSO (0.5 μ L) was added. After 15 min at 37°C, the reaction was quenched by ethanol precipitation. Modified RNA was recovered by centrifugation, washed twice with 70% ethanol, and dissolved in DEPC-treated water. RT was initiated at five sites in the transcript using SuperScript III (Invitrogen) as described by the manufacturer using the following HEX-labeled primers: D1, 5'-TCGGAGCGTTGTG AACTG-3'; D14, 5'-AGCAGCAGCAGCATTCCCCG-3'; D16, 5'-AG TTTGCCCATCCACGTCAGG-3'; D2, 5'-AGCAGCGCAAGTGAG GAG-3'; D18, 5'-TTTTTTTTTTTTTTTTTTGG-3' (IDT). To generate a reference ladder, the RT reaction was also performed supplemented with 0.5 mM ddGTP on nonmodified RNA using the same primer sequences with a 6-FAM label (IDT). After the RT reaction, the RNA was degraded with 200 mM NaOH and incubation at 95°C for 5 min, after which the solution was neutralized using an equimolar amount of HCl. cDNA was recovered by ethanol precipitation, washed twice with 70% ethanol, and dissolved in highly deionized (HiDi) formamide (Applied Biosystems) by incubation for 15 min at 65°C while shaking. Capillary electrophoresis was performed by the Radboud Genomics Technology Center using an ABI3730XL DNA Analyzer (Applied Biosystems). Electrophoresis traces were analyzed using QuSHAPE with standard settings (Karabiber et al. 2013). A fixed reactivity outlier percentage of 5 was applied to facilitate comparison between experiments. Duplicate or triplicate experiments were performed with RNA from different preparations. SHAPE reactivity data from overlapping reads and duplicate experiments were averaged. Two times three data points in the CUG₁₄₇ repeat were excluded due to different background signals (nucleotides 616–618 and 621–623). RNA secondary structures were generated using *Fold* in RNAstructure 5.8 (Reuter and Mathews 2010) with SHAPE reactivities as pseudo-energy (soft) constraints, where the SHAPE intercept and slope were set at -0.8 and 2.6 kcal/mole, respectively (Deigan et al. 2009). For visualization, SHAPE values <-0.1 were set at -0.1 . Graphical representations of RNA structure were generated using Varna (Darty et al. 2009) and formatted manually.

LNA-mediated structure perturbation

For structure verification experiments, a truncated variant of AON 445569 (Wheeler et al. 2012), LNA oligonucleotide tr445569, 5'-GTTGTGAAGT-3', or LNA (CAG)₃C, 5'-CAGCAGCAGC-3' (LNA nucleotides depicted in bold, Exiqon), was added before the denaturation step in a 1:10 RNA:AON molar ratio. SHAPE was performed as described above, only with an additional purification step after the modification because the bound LNAs interfered with the RT reaction (data not shown). Inspired by Busan et al. (Busan and Weeks 2013), after recovering the modified RNA, a 50-fold excess of DNA AON (5'-CAGTTCACAAC-3' or 5'-GCTGCTGCTG-3'; Exiqon) was added and then heated at 80°C for 5 min. After placing on ice, the RNA was directly purified from LNAs and oligos using the RNeasy MinElute RNA Cleanup kit (QIAGEN). RT reactions with primer D14 (in the case of LNA tr445569) or primers D14, D16, and D2 [for LNA (CAG)₃C] and data analysis were performed as described above. Because a slightly altered purification method was used compared to SHAPE without LNA perturbation, additional reactions without LNAs were performed in parallel.

DMS probing

DMS probing was performed as in SHAPE, only instead of NAI, the RNA was incubated with a final concentration of 0.25% DMS (Sigma-Aldrich) in ethanol or just ethanol (1 μ L) for 6 min. The reactions were quenched by adding 1/4th volume 2-mercaptoethanol and immediate ethanol precipitation. RT reactions, electrophoresis, and data analysis were performed as for SHAPE.

RNase T1 probing

Denaturation and refolding of the RNA were performed as described for SHAPE experiments. After these steps, RNase T1 enzyme (Thermo) dissolved in 10 mM Tris pH 7.5 was added to a final concentration of 0.01 units per μ l reaction mix, unless indicated otherwise. Reactions without enzyme served as control. Cleavage was performed for 10 min at 37°C, unless indicated otherwise. The reaction was stopped by phenol-chloroform extraction, followed by isopropanol precipitation. RT reactions, electrophoresis, and data analysis were performed as described for SHAPE using D16 primer.

CMCT probing

N-Cyclohexyl-N'-(2-morpholinoethyl) carbodiimide methyl-p-toluenesulfonate (CMCT) probing was performed as in SHAPE, only instead of NAI, the RNA was incubated with a final concentration of 50 mM CMCT (Sigma-Aldrich) in 0.5 \times TE pH 7.6 or pure buffer (2 μ L) for 5 min. Quenching, RT reactions, electrophoresis, and data analysis were performed as for SHAPE using the D16 primer.

OligoWalk

The OligoWalk (Lu and Mathews 2008) algorithm in RNAstructure 5.8 (Reuter and Mathews 2010) was applied in *Break Local Structure* mode for 1 mM 20-nt DNA oligos to a set of 20 suboptimal structures generated as described above with SHAPE constraints. For comparison, this was also performed for a set of 20 structures predicted in the same fashion but without SHAPE constraints. For comparison to the IONIS AON efficacy experiments, AON sequences and *DMPK* knockdown values were extracted from patent US9,765,338 (Bennett et al. 2015) (Tables 1, 5–7, 12, and 13 in this patent) and patent application US2016/0304877A1 (Swayze et al. 2014) (Tables 4–7 and 12–18 in this application). IONIS AONs were given an identifier based on the first binding nucleotide in *DMPK* exon 15, and for overlapping identifiers the effect was averaged. Linear regression was performed using GraphPad Prism version 5.01 for Windows (GraphPad Software) on % *DMPK* knockdown by IONIS AONs versus the predicted Δ G of structure disruption by AON binding. In our comparison, we pooled all AONs irrespective of length, chemistry and concentration used, since each set analyzed individually resulted in the same outcome (data not shown).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

COMPETING INTEREST STATEMENT

D.G.W. is an inventor on a patent related to AONs for treatment of DM1.

ACKNOWLEDGMENTS

We thank Dr. Colin Logie, Dr. Hans Heus, and the members of our department for helpful discussions. This work was supported by an internal grant from the Radboud University Medical Center (Internal PhD Round 2015).

Received October 1, 2018; accepted January 24, 2019.

REFERENCES

- Bañez-Coronel M, Ayhan F, Tarabochia AD, Zu T, Perez BA, Tusi SK, Pletnikova O, Borchelt DR, Ross CA, Margolis RL, et al. 2015. RAN translation in Huntington disease. *Neuron* **88**: 667–677. doi:10.1016/j.neuron.2015.10.038
- Bennett FC, Freier SM, MacLeod RA, Pandey SK, Thornton CA, Wheeler T, Cheng SH, Leger A, Wentworth BM. 2015. *Modulation of dystrophin myotonic-protein kinase (DMPK) expression*. U.S. patent no. US9,765,338B2.
- Bernat V, Disney MD. 2015. RNA structures as mediators of neurological diseases and as drug targets. *Neuron* **87**: 28–46. doi:10.1016/j.neuron.2015.06.012
- Bevilacqua PC, Ritchey LE, Su Z, Assmann SM. 2016. Genome-wide analysis of RNA secondary structure. *Annu Rev Genet* **50**: 235–266. doi:10.1146/annurev-genet-120215-035034
- Błaszczak L, Rypniewski W, Kiliszek A. 2017. Structures of RNA repeats associated with neurological diseases. *Wiley Interdiscip Rev RNA* **8**: e1412. doi:10.1002/wrna.1412
- Braida C, Stefanatos RKA, Adam B, Mahajan N, Smeets HJM, Niel F, Goizet C, Arveiler B, Koenig M, Lagier-Tourenne C, et al. 2010. 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. *Hum Mol Genet* **19**: 1399–1412. doi:10.1093/hmg/ddq015
- Broda M, Kierzek E, Gdaniec Z, Kulinski T, Kierzek R. 2005. Thermodynamic stability of RNA structures formed by CNG trinucleotide repeats. Implication for prediction of RNA structure. *Biochemistry* **44**: 10873–10882. doi:10.1021/bi0502339
- Busan S, Weeks KM. 2013. Role of context in RNA structure: flanking sequences reconfigure CAG motif folding in *Huntingtin* exon 1 transcripts. *Biochemistry* **52**: 8219–8225. doi:10.1021/bi401129r
- Cammis A, Millevoi S. 2017. RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Res* **45**: 1584–1595. doi:10.1093/nar/gkw1280
- Carrell ST, Tang Z, Mohr S, Lambowitz AM, Thornton CA. 2018. Detection of expanded RNA repeats using thermostable group II intron reverse transcriptase. *Nucleic Acids Res* **46**: e1. doi:10.1093/nar/gkx867
- Cass D, Hotchko R, Barber P, Jones K, Gates DP, Berglund JA. 2011. The four Zn fingers of MBNL1 provide a flexible platform for recognition of its RNA binding elements. *BMC Mol Biol* **12**: 20. doi:10.1186/1471-2199-12-20
- Cleary JD, Ranum LP. 2017. New developments in RAN translation: insights from multiple diseases. *Curr Opin Genet Dev* **44**: 125–134. doi:10.1016/j.gde.2017.03.006
- Darty K, Denise A, Ponty Y. 2009. VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**: 1974–1975. doi:10.1093/bioinformatics/btp250
- De Antonio M, Dogan C, Hamroun D, Mati M, Zerrouki S, Eymard B, Katsahian S, Bassez G, French Myotonic Dystrophy Clinical Network. 2016. Unravelling the myotonic dystrophy type 1 clinical spectrum: a systematic registry-based study with implications for disease classification. *Rev Neurol (Paris)* **172**: 572–580. doi:10.1016/j.neuro.2016.08.003
- Deigan KE, Li TW, Mathews DH, Weeks KM. 2009. Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci* **106**: 97–102. doi:10.1073/pnas.0806929106
- deLorimier E, Coonrod LA, Copperman J, Taber A, Reister EE, Sharma K, Todd PK, Guenza MG, Berglund JA. 2014. Modifications to toxic CUG RNAs induce structural stability, rescue missplicing in a myotonic dystrophy cell model and reduce toxicity in a myotonic dystrophy zebrafish model. *Nucleic Acids Res* **42**: 12768–12778. doi:10.1093/nar/gku941
- De Mezer M, Wojciechowska M, Napierala M, Sobczak K, Krzyzosiak WJ. 2011. Mutant CAG repeats of Huntingtin transcript fold into hairpins, form nuclear foci and are targets for RNA interference. *Nucleic Acids Res* **39**: 3852–3863. doi:10.1093/nar/gkq1323
- Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. 2014. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**: 696–700. doi:10.1038/nature12756
- Draper DE. 2004. A guide to ions and RNA structure. *RNA* **10**: 335–343. doi:10.1261/rna.5205404
- González-Barriga A, Mulders SA, van de Giessen J, Hooijer JD, Bijl S, van Kessel ID, van Beers J, van Deutekom JC, Franssen JA, Wieringa B, et al. 2013. Design and analysis of effects of triplet repeat oligonucleotides in cell models for myotonic dystrophy. *Mol Ther Nucleic Acids* **2**: e81. doi:10.1038/mtna.2013.9
- Groenen PJTA, Wansink DG, Coerwinkel M, van den Broek W, Jansen G, Wieringa B. 2000. Constitutive and regulated modes of splicing produce six major myotonic dystrophy protein kinase (DMPK) isoforms with distinct properties. *Hum Mol Genet* **9**: 605–616. doi:10.1093/hmg/9.4.605
- Gudde AEEG, González-Barriga A, van den Broek WJAA, Wieringa B, Wansink DG. 2016. A low absolute number of expanded transcripts is involved in myotonic dystrophy type 1 manifestation in muscle. *Hum Mol Genet* **25**: 1648–1662. doi:10.1093/hmg/ddw042
- Gudde AEEG, van Kessel IDG, André LM, Wieringa B, Wansink DG. 2017. Trinucleotide-repeat expanded and normal DMPK transcripts contain unusually long poly(A) tails despite differential nuclear residence. *Biochim Biophys Acta Gene Regul Mech* **1860**: 740–749. doi:10.1016/j.bbtagm.2017.04.002
- Hu J, Matsui M, Gagnon KT, Schwartz JC, Gabillet S, Arar K, Wu J, Bezprozvanny I, Corey DR. 2009. Allele-specific silencing of mutant huntingtin and ataxin-3 genes by targeting expanded CAG repeats in mRNAs. *Nat Biotechnol* **27**: 478–484. doi:10.1038/nbt.1539
- Jauvin D, Chrétien J, Pandey SK, Martineau L, Revillod L, Bassez G, Lachon A, MacLeod AR, Gourdon G, Wheeler TM, et al. 2017. Targeting DMPK with antisense oligonucleotide improves muscle strength in myotonic dystrophy type 1 mice. *Mol Ther Nucleic Acids* **7**: 465–474. doi:10.1016/j.omtn.2017.05.007
- Karabiber F, McGinnis JL, Favorov O V, Weeks KM. 2013. QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA* **19**: 63–73. doi:10.1261/ma.036327.112
- Kearse MG, Green KM, Krans A, Rodriguez CM, Linsalata AE, Goldstrohm AC, Todd PK. 2016. CGG repeat-associated non-AUG translation utilizes a cap-dependent scanning mechanism of initiation to produce toxic proteins. *Mol Cell* **62**: 314–322. doi:10.1016/j.molcel.2016.02.034

- Koch KS, Leffert HL. 1998. Giant hairpins formed by CUG repeats in myotonic dystrophy messenger RNAs might sterically block RNA export through nuclear pores. *J Theor Biol* **192**: 505–514. doi:10.1006/jtbi.1998.0679
- Kozak M. 1990. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci* **87**: 8301–8305. doi:10.1073/pnas.87.21.8301
- Lee JE, Bennett CF, Cooper TA. 2012. RNase H-mediated degradation of toxic RNA in myotonic dystrophy type 1. *Proc Natl Acad Sci* **109**: 4221–4226. doi:10.1073/pnas.1117019109
- Leeflang EP, Arnheim N. 1995. A novel repeat structure at the myotonic dystrophy locus in a 37 repeat allele with unexpectedly high stability. *Hum Mol Genet* **4**: 135–136. doi:10.1093/hmg/4.1.135
- Lorenz R, Wolfinger MT, Tanzer A, Hofacker IL. 2016. Predicting RNA secondary structures from sequence and probing data. *Methods* **103**: 86–98. doi:10.1016/j.jymeth.2016.04.004
- Low JT, Garcia-miranda P, Mouzakis KD, Gorelick RJ, Butcher SE, Weeks KM. 2015. Structure and dynamics of the HIV-1 frameshift element RNA. *Biochemistry* **53**: 4282–4291. doi:10.1021/bi5004926
- Lu ZJ, Mathews DH. 2008. OligoWalk: an online siRNA design tool utilizing hybridization thermodynamics. *Nucleic Acids Res* **36**: W104–W108. doi:10.1093/nar/gkn250
- Masek T, Vopalensky V, Suchomelova P, Pospisek M. 2005. Denaturing RNA electrophoresis in TAE agarose gels. *Anal Biochem* **336**: 46–50. doi:10.1016/j.ab.2004.09.010
- Meng YX, Shen HR, Zhao Z, Bing Q, Li N, Hu J. 2015. Optimization PCR for detection CTG/CCTG-repeat expansions in the diagnosis of myotonic dystrophies. *Ann Clin Lab Sci* **45**: 502–507.
- Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. 2005. RNA structure analysis at single nucleotide resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *J Am Chem Soc* **127**: 4223–4231. doi:10.1021/ja043822v
- Michlewski G, Krzyzosiak WJ. 2004. Molecular architecture of CAG repeats in human disease related transcripts. *J Mol Biol* **340**: 665–679. doi:10.1016/j.jmb.2004.05.021
- Michalowski S, Miller JW, Urbinati CR, Paliouras M, Swanson MS, Griffith J. 1999. Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein. *Nucleic Acids Res* **27**: 3534–3542. doi:10.1093/nar/27.17.3534
- Mulders SAM, van den Broek WJAA, Wheeler TM, Croes HJE, van Kuik-Romeijn P, de Kimpe SJ, Furling D, Platenburg GJ, Gourdon G, Thornton CA, et al. 2009. Triplet-repeat oligonucleotide-mediated reversal of RNA toxicity in myotonic dystrophy. *Proc Natl Acad Sci* **106**: 13915–13920. doi:10.1073/pnas.0905780106
- Musova Z, Mazanec R, Krepelova A, Ehler E, Vales J, Jankova R, Prochazka T, Koukal P, Marikova T, Kraus J, et al. 2009. Highly unstable sequence interruptions of the CTG repeat in the myotonic dystrophy gene. *Am J Med Genet Part A* **149A**: 1365–1374. doi:10.1002/ajmg.a.32987
- Napierała M, Krzyzosiak WJ. 1997. CUG repeats present in myotonin kinase RNA form metastable “slippery” hairpins. *J Biol Chem* **272**: 31079–31085. doi:10.1074/jbc.272.49.31079
- Pandey SK, Wheeler TM, Justice SL, Kim A, Younis HS, Gattis D, Jauvin D, Puymirat J, Swayze EE, Freier SM, et al. 2015. Identification and characterization of modified antisense oligonucleotides targeting DMPK in mice and nonhuman primates for the treatment of myotonic dystrophy type 1. *J Pharmacol Exp Ther* **355**: 310–321. doi:10.1124/jpet.115.226969
- Paul S, Dansithong W, Jog SP, Holt I, Mittal S, Brook JD, Morris GE, Comai L, Reddy S. 2011. Expanded CUG repeats dysregulate RNA splicing by altering the stoichiometry of the muscleblind 1 complex. *J Biol Chem* **286**: 38427–38438. doi:10.1074/jbc.M111.255224
- Petersson OJ, Aagaard L, Jensen TG, Damgaard CK. 2015. Molecular mechanisms in DM1—a focus on foci. *Nucleic Acids Res* **43**: 2433–2441. doi:10.1093/nar/gkv029
- Ravel-Chapuis A, Bélanger G, Yadava RS, Mahadevan MS, DesGroseillers L, Côté J, Jasmin BJ. 2012. The RNA-binding protein Staufen1 is increased in DM1 skeletal muscle and promotes alternative pre-mRNA splicing. *J Cell Biol* **196**: 699–712. doi:10.1083/jcb.201108113
- Reuter JS, Mathews DH. 2010. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**: 129. doi:10.1186/1471-2105-11-129
- Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. 2014. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**: 701–705. doi:10.1038/nature12894
- Seznec H, Lia-Baldini S, Duros C, Fouquet C, Lacroix C, Hofmann-Radvanyi H, Junien C, Gourdon G. 2000. Transgenic mice carrying large human genomic sequences with expanded CTG repeat mimic closely the DM CTG repeat intergenerational and somatic instability. *Hum Mol Genet* **9**: 1185–1194. doi:10.1093/hmg/9.8.1185
- Simone R, Fratta P, Neidle S, Parkinson GN, Isaacs AM. 2015. G-quadruplexes: emerging roles in neurodegenerative diseases and the non-coding transcriptome. *FEBS Lett* **589**: 1653–1668. doi:10.1016/j.febslet.2015.05.003
- Sлома MF, Mathews DH. 2015. Improving RNA secondary structure prediction with structure mapping data. *Methods Enzymol* **553**: 91–114. doi:10.1016/bs.mie.2014.10.053
- Sobczak K, Krzyzosiak WJ. 2005. CAG repeats containing CAA interruptions form branched hairpin structures in spinocerebellar ataxia type 2 transcripts. *J Biol Chem* **280**: 3898–3910. doi:10.1074/jbc.M409984200
- Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, Chang HY. 2013. RNA SHAPE analysis in living cells. *Nat Chem Biol* **9**: 18–20. doi:10.1038/nchembio.1131
- Su Z, Zhang Y, Gendron TF, Bauer PO, Chew J, Yang WY, Fostvedt E, Jansen-West K, Belzil VV, Desaro P, et al. 2014. Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. *Neuron* **83**: 1043–1050. doi:10.1016/j.neuron.2014.07.041
- Swayze EE, Pandey SK, MacLeod RA, Bennett CF, Freier SM. 2014. *Compounds and methods for modulation of dystrophin myotonic-protein kinase (DMPK) expression*. U.S. patent no. US2016/0304877A1.
- Sztuba-Solinska J, Le Grice SFJ. 2014. Insights into secondary and tertiary interactions of dengue virus RNA by SHAPE. *Methods Mol Biol* **1138**: 225–239. doi:10.1007/978-1-4939-0348-1_14
- Tabet R, Schaeffer L, Freyermuth F, Jambéau M, Workman M, Lee CZ, Lin CC, Jiang J, Jansen-West K, Abou-Hamdan H, et al. 2018. CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. *Nat Commun* **9**: 152. doi:10.1038/s41467-017-02643-5
- Taylor K, Sznajder ŁJ, Cywoniuk P, Thomas JD, Swanson MS, Sobczak K. 2018. MBNL splicing activity depends on RNA binding site structural context. *Nucleic Acids Res* **46**: 9119–9133. doi:10.1093/nar/gky565
- Tian B, White RJ, Xia T, Welle S, Turner DH, Mathews MB, Thornton CA. 2000. Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. *RNA* **6**: 79–87. doi:10.1017/S1355838200991544
- Todd PK, Oh SY, Krans A, He F, Sellier C, Frazer M, Renoux AJ, Chen K, Scaglione KM, Basrur V, et al. 2013. CGG repeat-associated translation mediates neurodegeneration in fragile × tremor

- ataxia syndrome. *Neuron* **78**: 440–455. doi:10.1016/j.neuron.2013.03.026
- Tomé S, Dandelot E, Dogan C, Bertrand A, Geneviève D, Péréon Y, Simon M, Bonnefont J-P, Bassez G, Gourdon G. 2018. Unusual association of a unique CAG interruption in 5' of DM1 CTG repeats with intergenerational contractions and low somatic mosaicism. *Hum Mutat* **39**: 970–982. doi:10.1002/humu.23531
- Udd B, Krahe R. 2012. The myotonic dystrophies: molecular, clinical, and therapeutic challenges. *Lancet Neurol* **11**: 891–905. doi:10.1016/S1474-4422(12)70204-1
- Wansink DG, van herpen REMA, Coerwinkel-Driessen MM, Groenen PJTA, Hemmings BA, Wieringa B. 2003. Alternative splicing controls myotonic dystrophy protein kinase structure, enzymatic activity, and subcellular localization. *Mol Cell Biol* **23**: 5489–5501. doi:10.1128/MCB.23.16.5489-5501.2003
- Weeks KM. 2010. Advances in RNA structure analysis by chemical probing. *Curr Opin Struct Biol* **20**: 295–304. doi:10.1016/j.sbi.2010.04.001
- Wheeler TM, Sobczak K, Lueck JD, Osborne RJ, Lin X, Dirksen RT, Thornton CA. 2009. Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA. *Science* **325**: 336–339. doi:10.1126/science.1173110
- Wheeler TM, Leger AJ, Pandey SK, MacLeod AR, Cheng SH, Wentworth BM, Bennett CF, Thornton CA. 2012. Targeting nuclear RNA for in vivo correction of myotonic dystrophy. *Nature* **488**: 111–117. doi:10.1038/nature11362
- Wilkinson KA, Merino EJ, Weeks KM. 2006. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **1**: 1610–1616. doi:10.1038/nprot.2006.249
- Wojciechowska M, Sobczak K, Kozłowski P, Sedehizadeh S, Wojtkowiak-Szlachcic A, Czubak K, Markus R, Lusakowska A, Kaminska A, Brook JD. 2018. Quantitative methods to monitor RNA biomarkers in myotonic dystrophy. *Sci Rep* **8**: 5885. doi:10.1038/s41598-018-24156-x
- Yuan Y, Compton SA, Sobczak K, Stenberg MG, Thornton CA, Griffith JD, Swanson MS. 2007. Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res* **35**: 5474–5486. doi:10.1093/nar/gkm601
- Zamiri B, Reddy K, Macgregor RB, Pearson CE. 2014. TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)*n* repeat of the *C9orf72* gene and blocks interaction of RNA-binding proteins. *J Biol Chem* **289**: 4653–4659. doi:10.1074/jbc.C113.502336
- Zu T, Gibbens B, Doty NS, Gomes-pereira M, Huguet A, Stone MD, Margolis J, Peterson M, Markowski TW, Ingram MAC, et al. 2010. Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci* **108**: 260–265. doi:10.1073/pnas.1013343108