Quadruplex formation by both G-rich and C-rich DNA strands of the C9orf72 (GGGGCC)\(_8\)·(GGCCCC)\(_8\) repeat: effect of CpG methylation

Bita Zamiri\(^1,\dagger\), Mila Mirceta\(^2,3,\dagger\), Karol Bomsztyk\(^4\), Robert B. Macgregor, Jr.\(^1\) and Christopher E. Pearson\(^2,3,\star\)

\(^1\)Graduate Department of Pharmaceutical Sciences, Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, Ontario M5S 3M2, Canada, \(^2\)Program of Genetics & Genome Biology, The Hospital for Sick Children, Toronto, Ontario M5G 1L7, Canada, \(^3\)Program of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A1, Canada and \(^4\)UW Medicine South Lake Union, University of Washington, Seattle WA 98109, USA

Received July 20, 2015; Revised September 05, 2015; Accepted September 22, 2015

ABSTRACT

Unusual DNA/RNA structures of the C9orf72 repeat may participate in repeat expansions or pathogenesis of amyotrophic lateral sclerosis and frontotemporal dementia. Expanded repeats are CpG methylated with unknown consequences. Typically, quadruplex structures form by G-rich but not complementary C-rich strands. Using CD, UV and electrophoresis, we characterized the structures formed by (GGGGCC)\(_8\) and (GGCCCC)\(_8\) strands with and without 5-methylcytosine (5mCpG) or 5-hydroxymethylcytosine (5hmCpG) methylation. All strands formed heterogenous mixtures of structures, with features of quadruplexes (at pH 7.5, in K\(^+\), Na\(^+\) or Li\(^+\)), but no feature typical of i-motifs. C-rich strands formed quadruplexes, likely stabilized by G•C•G•C-tetrads and C•C•C•C-tetrads. Unlike G•G•G•G-tetrads, some G•C•G•C-tetrad conformations do not require the N7-Guanine position, hence C9orf72 quadruplexes still formed when N7-deazaGuanine replace all Guanines. 5mCpG and 5hmCpG increased and decreased the thermal stability of these structures. hnRNPK, through band-shift analysis, bound C-rich but not G-rich strands, with a binding preference of unmethylated > 5hmCpG > 5mCpG, where methylated DNA-protein complexes were retained in the wells, distinct from unmethylated complexes. Our findings suggest that for C-rich sequences interspersed with G-residues, one must consider quadruplex formation and that methylation of quadruplexes may affect epigenetic processes.

INTRODUCTION

The most common cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) is a (GGGGCC)n·(GGCCCC)n repeat expansion in the C9orf72 gene (1,2). In unaffected individuals, the most common repeat size is 2–19 units, while the pathogenic repeat size is 250–1600 repeats. ALS-FTD is one of more than 40 neurological, neurodegenerative, and neuromuscular diseases caused by gene-specific repeat expansions, such as myotonic dystrophy type 1 (DM1) and fragile X mental retardation syndrome (FXS). C9orf72-expansions show somatic repeat instability, aberrant CpG methylation, and lead to toxic-RNAs, toxic, RAN-peptides, and loss-of-function, pathological mechanisms, many of which have been reported in DM1 and FXS (3,4).

Epigenetic modifications occur at the C9orf72 gene with increased cytosine DNA methylation at CpG sites both in the promoter and within the repeat of many but not all C9orf72 expansion carriers (5–10). The CpG island and promoter upstream, but not downstream of the GGGGCC repeat is methylated in 36% of expansion carriers. However, methylation within the C9orf72 repeat is present in all individuals with greater than 90 repeats (10). Reports have not distinguished whether the C9orf72 methylation is the canonical 5-methylcytosine (5mCpG) or the more recently recognized 5-hydroxymethylcytosine (5hmCpG). C9orf72 methylation correlated with decreased levels of C9orf72 mRNA, RNA foci and RAN-dipeptide aggregates (6,7,9). The expanded and methylated C9orf72 repeat may differentially bind specific proteins. The expanded C9orf72 repeats, but not non-pathogenic repeats, have been reported to associate with a specific set of modified histones, which coincides with reduced C9orf72 mRNA expression (5). However, no study has assessed both C9orf72 DNA methylation state and protein interaction. Although the role of epige-
ethics in C9orf72-associated ALD/FTD is not yet understood, characterizing the structural and biological effects of C9orf72 repeat methylation may shed light on the mutagenic and pathogenic nature of the repeat.

Unusual DNA, RNA and R-loop (RNA–DNA) structures have been suggested to be involved in disease-associated repeat instability, or pathogenesis (11–13). Typically, G-rich strands are thought to form four-stranded G-quadruplexes, and C-rich strands have been demonstrated to form intercalated i-motif structures at acidic pH (14,15). Formation of G-quadruplex structures has been reported in the RNA (GGGGCC)n repeat (16–18). The G-rich DNA strand of the C9orf72 repeat can form quadruplexes (13). Several proteins bind preferentially to quadruplex RNA and DNA (19). Numerous proteins that bind quadruplexes have been reported to bind the C9orf72 quadruplex RNAs or RNA foci, including ASF/SF2, nucleolin, and hnRNPK (13,17–20). There is little knowledge of protein interactions with the C9orf72 DNA, and of the structure of the C-rich strand.

Here, we examined the structures formed by the G-rich and C-rich strands of the C9orf72 repeat with various modifications, including 5mCpG, 5hmCpG and N7-deazaG. We show that both the C-rich and G-rich strands of the C9orf72 repeat can form four-stranded structures. We also demonstrate that 5mCpG, 5hmCpG and N7-deazaG modifications affect secondary structure. Furthermore, hnRNPK binds to the C-rich DNA repeat and binding is affected by methylation status.

MATERIALS AND METHODS

DNA oligonucleotide synthesis and Labeling

All the DNA oligodeoxyribonucleotides were purchased as purified from ACGT Corp., Toronto, Canada. The ODNs were dialyzed versus Milli-Q water and lyophilized. The concentrations were determined by recording their absorbance at 260 nm using molar extinction coefficients calculated using the nearest-neighbour method (21). Oligos were then dissolved in the desired buffer to give concentrations were determined by recording their absorbance at 260 nm using molar extinction coefficients calculated using the nearest-neighbour method (21). Oligos were then dissolved in the desired buffer to give concentrations ranging from 5 μM to 40 μM depending on the experiment. Oligonucleotides were end-labeled using [γ-32P]ATP. Equal sample amounts, based on Cerenkov counting (5000 cpm/sample), were electrophoresed on 4% or 8% polyacrylamide gels at 15 V/cm for 60 min or 9 V/cm for 120 min, respectively. We also carried out electrophoresis under denaturing conditions. In this case, the samples were heated in formamide, frozen on dry ice until loading and electrophoresed on pre-heated denaturing 6% 8M urea sequencing gels, run at 150 V/cm. Some degree of heterogeneity is evident as both faster and slower migrating species that were resistant to denaturation (Supplementary Figure S1). We, and others have reported heat- and denaturation-resistant structures formed by G- and C-rich sequences that aberrantly migrate faster or slower on denaturing gels (17,18,22–24).

CD spectroscopy

CD spectra were recorded at room temperature using an Aviv model 62 DS spectropolarimeter using 5- and 20-μM samples dissolved in the indicated buffers. The pH was adjusted to 7.5 with HCl. All the samples were heated to 95°C and allowed to cool to room temperature overnight. Three CD spectra, over the wavelength range of 320–220 nm were recorded in a 1-mm cuvette and averaged.

UV spectroscopy

Melting experiments were carried out on a Cary model 300 Bio spectrophotometer. Depending on the concentration, the samples were assessed in a 1-mm or 1-cm pathlength cuvette. Oligos were dissolved in 10 mM Tris containing either 100 mM NaCl, KCl or LiCl, pH 7.5. Melting was monitored at 280 nm (for all oligos but unmodified and G-rich 5mCpG in KCl) or 290 nm. The heating rate was 1°C/min. All experiments were done in triplicate.

Electrophoretic mobility shift assays

Reactions were carried out in a 20-μl volume containing ~80 fmol of [γ-32P]ATP end-labeled DNA with the indicated amounts of hnRNPK (Abnova). Binding reactions contained a final buffer concentration of 10 mM Tris (pH 7.4), 50 mM KCl, 1 mM DTT, 2.5% glycerol, 5 mM MgCl2, 0.05% non-ideal P40. DNA was preincubated with the buffer described above for 30 min prior to the addition of protein and poly(dI-dC). Reactions were incubated at room temperature for 20 min, loaded onto 4% native polyacrylamide gels, electrophoresed at 200 V for 1 h in 1× Tris borate EDTA (0.089M Tris base, 0.089 M boric acid, 0.002 M EDTA), dried and autoradiographed. The extent of binding was empirically assessed, using ImageQuant software, by comparing the relative intensity of the band corresponding to the bound complex with that corresponding to the unbound material.

RESULTS

Quadruplex formation by both C9orf72 repeat strands and the effect of CpG methylation

We determined the structural effect of three different modifications, 5mCpG, 5hmCpG and N7-deazaG, on both the G-rich and C-rich strands of the C9orf72 repeat by CD spectroscopy and UV melting, in the presence of various ions (Figures 1 and 2). In KCl, the unmodified G-rich strand exhibits a CD spectrum with an intense positive peak at 290 nm, a negative peak around 260 nm a positive peak around 240 nm (Figure 1B). This spectrum is typical of antiparallel G-quadruplexes (25). In LiCl, the intensity at 290 nm decreases and the negative peak at 260 nm turns into a positive shoulder with a small negative peak around 240 nm. In LiCl, the population is shifted into a higher percentage of parallel quadruplex structures reflected by the 260-nm shoulder (Figure 1A). Methylation of the cytosines at CpG sites of the G-rich strand results in a more intense 260 nm peak. In KCl the spectra of the unmodified and 5mCpG G-rich strand are similar. When cytosines at CpG sites are 5-hydroxymethylated, in all salt conditions, the CD spectra exhibits a positive peak at 285 nm, a shoulder around 260 nm with a negative peak around 240 nm. Positive peaks
Figure 1. CD spectra of indicated C9orf72 DNAs. All experiments were done at 5 μM strand concentration, in 10 μM Tris (pH 7.5), and 100 μM of the indicated salt.

at 285 nm suggest duplex base stacking (25). The stability of the cyclic structure of G•G•G•G-tetrads, a critical component of G-quadruplexes, depends upon the ability of the N7 of guanine to accept a hydrogen bond (26). We directly tested the importance of this by replacing guanine residues with N7-deazaguanine, where the N7 position of guanine has been replaced by CH, which is not a hydrogen bond acceptor. Previous use of N7-deazaG revealed a strong dependence of G4-quadruplex (26). Here, modifying the N7 position of all guanines in the G-rich strands, in all salt conditions, results in a broad positive peak around 270 nm and negative peak around 250 nm showing complete structural transition. This spectrum has been associated with G•C•G•C-tetrad formation (27). These results support the suggestion that the peaks at 290 and 260 nm observed for the unmodified and 5mCpG G-rich strands arise from quadruplex formation. All G-rich sequences were also studied in the presence of LiCl or NaCl (Supplementary Figures S2 and S3). Although LiCl is generally considered too small to stabilize G-quadruplexes, interestingly the CD spectrum of the C9orf72 oligos in the presence of LiCl is very similar to that in NaCl (Supplementary Figures S2 and S3). Quadruplex formation in the presence of LiCl has previously been reported for the FRAXA and telomeric repeats (28,29).

The CD spectra of the C-rich strands are shown in Figure 1C and 1D. Spectra for the unmodified, 5mCpG, and 5hmCpG C-rich strands were qualitatively similar in LiCl, with a positive peak at 280 nm, a negative peak at 250 nm and positive shoulder at 260 nm. Also in LiCl, an isodichroic point is seen at 260 nm. A positive 260-nm shoulder was not evident for the unmodified strand in KCl, rather a negative 260-nm peak is evident which may due to the stabilizing effect of KCl, turning the shoulder into a more prominent peak. Methylation or hydroxymethylation of CpGs do not alter the CD spectra significantly. Both 5mCpG, and 5hmCpG of the C-rich strands decreased the intensity of the 280-nm peak, but retained the overall shape of the spectra. The CD spectrum of the N7-deazaG C-rich strands lost the 260-nm shoulder in LiCl, and retained the 280-nm and 250-nm peaks. This suggests that the 260-nm peak arises from quadruplex formation even in the C-rich
strands, which is probably due to the high fraction of gua-
nines present. Tetrabutylammonium (TBA) is too large to
stabilize G-quartet formation (30). Interestingly, in the pre-
ence of TBA, while the 280-nm peak of the unmodified,
5mCpG, 5hmCpG C-rich strands was retained, the 260-nm
shoulder disappeared. For N7-deazaG C-rich strands, the
CD spectra were the same in TBA, KCl and NaCl. Simi-
lar to the G-rich strands, all spectra were similar in LiCl
and NaCl, suggesting that the conformation does not de-
pend strongly on the nature of the cation (Supplementary
Figures S2 and S3).

We performed UV melting on the same samples, under
identical conditions as above. The unmodified and 5mCpG
G-rich strands exhibit the highest thermal stability in all
three ions (Figure 2A, Supplementary Figure S2). Interest-
ingly, the unmodified and 5mCpG, G-rich strands exhibited
lower stability in KCl than in NaCl or LiCl. This is likely due
to the fact that in KCl there is a single folding unit of an-
tiparallel quadruplexes and the melting curve reflects these
species (measured at 290 nm), while in NaCl and LiCl the
CD spectra suggests multiple folding forms and the melt-
ing profiles may reflect denaturation of other folding units
(recorded at 280 nm). The 5hmCpG G-rich strands all have
lower stability than their unmodified forms, but higher than
the N7-deazaG strands. For the C-rich strands the melting
temperatures are lower than for the G-rich strands (Fig-
ure 2C and D). The unmodified and 5mCpG oligos all show
concentration-dependent melting temperatures, suggesting
possible multimolecular structure formation. Both the G-
rich and the C-rich N7-deazaG-substituted oligos are more
stable in LiCl than in than NaCl or KCl.

Methylation status of the C9orf72 repeat DNA affects its
electrophoretic migration

To further evaluate the structural heterogeneity and com-
plexity of the various forms of the (GGGGCC)8 and
(GGCCCC)8 oligonucleotides, we assessed their elec-
trophoretic migration on native polyacrylamide gels follow-
The (GGCCCC)8 strand is preferentially bound by hnRNPK and affected by methylation

Toward addressing a biological effect of \( \text{C9orf72} \) repeat modifications, we assessed the binding of hnRNPK. To date, there is no evidence on the effect of DNA methylation at the \( \text{C9orf72} \) repeat and protein interactions. Using recombinant hnRNPK, we observed preferential binding of the protein to the C-rich repeat DNA (Figure 4). Unmethylated C-rich DNA was bound with complexes resolved as two distinct slower migrating species. The other modifications gave rise to broad smears of species with increasing amounts of hnRNPK, with most of the DNA-protein complex unable to enter the gel, likely due to increased structural complexity. Notably, the 5hmCpG form was preferentially bound (38%) over the 5mCpG form (14%), while the unmethylated form most efficiently bound (68%). Previously, protein interactions with \( \text{r(GGGGCC)} \) and other repeat sequences reported the retention of the nucleic acid-protein complex in the well (18,31,32). Treatment with proteinase K and SDS broke down these complexes, allowing the DNA to migrate similarly to the control without protein. The unmethylated G-rich DNA was also able to bind, although very minimally, to the protein, creating a complex that migrated faster than the DNA-protein complexes of C-rich DNA. We did not observe binding to the modified G-rich sequences. These results imply that methylation status and type can alter the biological recognition of the \( \text{C9orf72} \) repeats.

DISCUSSION

We demonstrate that both the C-rich and G-rich DNA strands of the \( \text{C9orf72} \) repeat can form a heterogenous population of structures, including those with features of four-stranded structures. It is accepted that a given G-rich and C-rich sequences can form numerous structural conformations (33). Previously, it was reported for shorter G-rich \( \text{C9orf72} \) repeats (≤4 repeat units) that these form high levels of structural variants (34) – our findings are consistent with their suggestion that such structural variation may be reflected in vivo and may be considerable for longer repeat tracts. That the C-rich strand can assume a quadruplex is distinct from many other studies, most of which have focused upon the G-rich strand, with little attention to the complementary C-rich strand (telomere repeats is one example). In cases where the complementary C-rich strand has been studied, it has often been reported that these form \( i \)-motifs under non-physiological pH. We propose that the four-stranded species formed by C-rich \( \text{C9orf72} \) strands are stabilized by \( \text{G} \text{C} \text{G} \text{G} \text{C} \text{-tetrads and C} \text{G} \text{C} \text{C} \text{G} \text{-tetrads, which would sandwich C} \text{C} \text{C} \text{C} \text{-tetrads (a predicted structure is shown in Figure 5A). Our findings suggest that for C-rich strands, which also contain G-residues, one must consider four-stranded quadruplex formation.}

We propose that both the C-rich and G-rich \( \text{C9orf72} \) quadruplexes involve \( \text{G} \text{C} \text{C} \text{G} \text{-tetrads and C} \text{G} \text{C} \text{C} \text{-tetrads (Figure 5A). Mixed G} \text{C} \text{C} \text{C} \text{-tetrads are structurally distinct from G} \text{G} \text{C} \text{G} \text{-tetrads. We propose that the distinct CD peak or a shoulder at 260 nm that we, and others observed for both \( \text{C9orf72} \) DNA strands (13) reflects the presence of G} \text{C} \text{C} \text{C} \text{-tetrads. Notably, G} \text{C} \text{C} \text{G} \text{-tetrads can assume at least four different inter-base interactions (27,35-46). Unlike G} \text{G} \text{C} \text{G} \text{-tetrads, which are stabilized by hydrogen bonding with the N7 position of the guanine residues (26), some, but not all G} \text{C} \text{C} \text{G} \text{-tetrads form independent of the N7-G, with hydrogen bonding of the cytosine amino protons either with the O6 oxygen of guanine (35,46). Other G} \text{C} \text{C} \text{G} \text{-tetrads conformations do utilize hydrogen bonding with the N7 nitrogen of guanine (35,46). The absence of an absolute requirement of the N7-G may explain the ability of the C-rich \( \text{C9orf72} \) strand to form four-stranded structures at physiological pH. The limited effect of N7-deazaG on the CD spectrum of the C-rich strand, which is not predicted to involve G} \text{G} \text{C} \text{G} \text{-tetrads, may be due to some of the GCGC-tetrads formed by the C-rich strand involve the N7 of G residues and hence are affected by the N7-deazaG. Moreover, unlike G} \text{G} \text{C} \text{G} \text{-tetrads, which require cation coordination for stability, G} \text{C} \text{C} \text{G} \text{-tetrads do not require cations for stability (35,44,46), which may explain the insensitivity of \( \text{C9orf72} \) quadruplex formation to LiCl or TBA. It is notable that some G4-quadruplexes, including that formed by the FRAXA CGG repeat, are extremely stable even in the presence of lithium ions (28,29). A recent study of an oligo with a repeat of \( \text{GGGGCC}3\text{GG(Br)}\text{GG} \) with an 8-bromoguanine substitution, did not detect G} \text{C} \text{C} \text{C} \text{-tetrads, which may be due to the preferential adoption of a syn-glycosidic conformation by 8-bromoguanines (47). Similarly, an NMR study of}
Figure 3. Methylation type and salt influence electrophoretic migration patterns. Migration of $[^{32}\text{P}]$ATP end-labeled DNA (frozen and thawed at room temperature) in native 8% polyacrylamide gel. $[^{32}\text{P}]$ATP end-labeled (GGGGCC)$_8$ and (CCCCCG)$_8$ containing the indicated modifications were dissolved in 50 mM Tris–HCl (pH 7.4) with the indicated salt, incubated for 20 min and then electrophoresed for 2 h at 120 V (9 V/cm).

Figure 4. Methylation type influences DNA-protein interactions. Human recombinant hnRNPK band-shift with the $[^{32}\text{P}]$ATP end-labeled d(GGGGCC)$_8$ or d(GGCCCC)$_8$ DNAs (∼80 fmol) and 0, 50, 200 or 400 ng of purified recombinant hnRNPK. DNAs are either not CpG methylated, or contain 5-methylcytosine or 5-hydroxymethylcytosine bases at each CpG within the sequence. Samples were pre-incubated with the protein in a buffer solution containing 10 mM Tris (pH 7.4), 50 mM KCl, 1 mM DTT, 2.5% glycerol, 5 mM MgCl$_2$, and 0.05% Non-Idet P40. Samples were electrophoresed on a 4% native polyacrylamide gel for 1 hour at 200 volts (15V/cm). Lanes were densimetrically assessed using ImageQuant software to quantify the percent bound (%).
Figure 5. Proposed quadruplex structure in the C9orf72 8-repeat sequence and biological implications regarding R-loop formation. (A) Model of quadruplex structure forming in the (GGGGCC)₈ and (GGCCCC)₈ strands with various tetrad assemblies. Green colored boxes represent methylatable cytosines within the repeat sequence. (B) Intramolecular quadruplexes can form by either the G-rich or C-rich DNA strands of the C9orf72 repeat and may stabilized R-loops. They can also reflect the ‘pearls’ in a ‘pearls-on-a-string’ model of slipped-quadruplexes. See text.

short tracts of (GGGGCC)ₙ with either 1, 2 or 4 repeats, did not reveal G•C•G•C-tetrads (34). It is possible that the involvement of G•C•G•C-tetrads may require longer tracts of at least eight C9orf72 repeat units.

Our data imply that the C-rich C9orf72 quadruplex can involve C•C•C•C-tetrads (Figure 5A). These may be stabilized via stacking interactions that arise from sandwiching them between sets of G•C•G•C-tetrads or G•G•G•G-tetrads. Sandwiched G•C•G•C-tetrads in the FRAXA (CGG)ₙ repeat were proposed to be stabilized by adjacent G•G•G•G-tetrads (48) and elsewhere (49).

Most quadruplexes reported involve G•G•G•G-tetrads. However, some G-rich oligos with limited numbers of C residues can form quadruplexes involving G•C•G•C-tetrads and C•C•C•C-tetrads (27,35,36,42,46,50,51). For example, the disease-associated fragile X (CGG)ₙ repeat can form quadruplexes involving G•G•G•G-tetrads and C•C•C•C-tetrads (35,36,42,48). Notably, the FRAXA CD spectra also revealed a 260 nm peak/shoulder (48). It is widely accepted that quadruplex structures will involve G₄ tetrads composed of four interacting guanine residues. Such an expectation has led to the prediction that any sequence conforming to the sequence consensus G≥3N1–7G≥3N1–7G≥3N1–7G≥3 will form a quadruplex, the reverse is not necessarily true, in that all quadruplex-forming sequences may not conform to the consensus. Interestingly, G•C•G•C-tetrad-containing quadruplexes have long been proposed to act in various biological processes including the junction of replication forks (38) and homologous pairing of DNAs by the intercoiling of two Watson-Crick paired helices, which may facilitate recombination (52).

Methylation of DNA is an important epigenetic mark associated with altered gene expression, replication and genetic instability. Both methylcytosine and hydroxymethylcytosine at CpG sites of each repeat unit could alter the structures of both C9orf72 strands. The increased and decreased melting temperatures of the 5mCpG and 5hmCpG forms may serve a regulatory role when DNA strands must separate, as during DNA transcription, replication, repair, or recombination. Increased and decreased thermal stability have been reported for 5mCpG and 5hmCpG sites at non-repeated sequences (53–57). The altered thermal stability of the methylated repeats may affect the propensity of the C9orf72 repeats to form unusual DNA structures. The faster and slower electrophoretic migration of the 5mCpG and 5hmCpG forms we observed (Figure 3) finding here that the C-rich C9orf72 strand presents as a quadruplex-forming sequence does not conform to the expected consensus. Thus, while it is reasonable to expect that any sequence that conforms to the G≥3N1–7G≥3N1–7G≥3N1–7G≥3 consensus will form a quadruplex, the reverse is not necessarily true, in that all quadruplex-forming sequences may not conform to the consensus.
may reflect increased and decreased compaction of the DNA. The methylation state of non-repetitive sequences has been shown to enhance or impair the ability to form unwound DNA, bent-DNA, cruciforms, Z-DNA, quadruplexes and triple-DNA (58–60). We propose that 5mCpG of the C9orf72 repeats stabilizes the quadruplex, similar to its effect upon FRAXA repeats (29), while 5hmCpG destabilizes the quadruplexes. Unusual structure formation by the C9orf72 repeats may contribute to pathogenesis and their genetic length instability, as demonstrated for R-loops (13,17). In fragile X patients, an absence of CpG methylation of the CGG/CCG repeats has been linked to enhanced somatic contractions of the expanded repeats in patient tissues (61–63). It will be interesting to assess the possible link of methylation to C9orf72 repeat instability. Our findings suggest that methylation state, as well as the form of methylation, may alter C9orf72 repeat structure. Distinguishing the kind of CpG methylation at the C9orf72 promoter and repeat, in specific tissues may reveal functional roles of this epigenetic mark.

The expanded and methylated C9orf72 repeat may differentially bind specific proteins. Specific histone modifications have been reported at the expanded C9orf72 repeats, but not non-pathogenic repeats (5,6); however, the DNA methylation status in these studies was not assessed. We show CpG methylation-sensitive binding of hnRNPK to C9orf72 repeat DNAs. hnRNPK bound preferentially to the C-rich DNA strand, consistent with previous studies showing an association of hnRNPK with the C-rich RNA foci in cells with C9orf72 expansions (13,64). However, direct binding of hnRNPK to the C-rich C9orf72 RNA was not observed (64). Our finding that hnRNPK bound to the C-rich DNA strand, is consistent with the reported preference of hnRNPK to bind DNA over RNA (65). Our data shows that hnRNPK binding is affected by CpG methylation of the C9orf72 repeats. Protein complexes with the r(GGGGCC)n and r(CUG)n repeat sequences have been reported to be retained in the wells of electrophoresis gels (18,31). We postulate that the methylated C9orf72 DNA–protein complexes are retained in the wells because they form multimeric DNA-protein structures incapable of entering the gels as opposed to a structural alteration of the DNA. This is supported by the observation that treatment of the complexes with proteinase K permits the protein-free DNA to co-migrate with the starting DNA. The implications of this differential binding is unknown, however, the observation that DNA methylation can alter both the biophysical traits and the interaction of proteins supports the concept that this epigenetic modification may be mediated by structural and protein selective avenues. hnRNPK functions in many genetic processes, including transcription, translation, chromatin remodeling and DNA repair (66).

Considerable evidence supports the formation of unusual DNA, RNA or RNA-DNA structures by disease associated repeats, structures that participate in mutagenic and pathogenic processes. It is very likely that our preparations contain mixtures of different folded forms, antiparallel, parallel quadruplexes along with hairpins, as previously suggested (13). The intramolecular structures formed by the G-rich and C-rich C9orf72 DNA strands that we report here, could reflect the displaced DNA strand in the RNA-DNA hybrids (R-loop) (Figure 5B), previously shown to form on either of the C9orf72 strands (13,17). R-loops can be biophysically stabilized by quadruplex formation of the displaced G-rich DNA strand at non-repetitive sequences (67). Notably, our observation herein, that the C-rich C9orf72 DNA can also assume a quadruplex structure suggests that it may stabilize R-loops having RNAs hybridized to the G-rich strand. Thus, R-loop formation on either DNA strand of the C9orf72 repeats could be stabilized by quadruplex formation by either of the displaced DNA strands. Similarly, the intramolecular quadruplexes formed by the G-rich and C-rich C9orf72 DNA strands that we report here, could reflect the ‘pearls’ in a ‘pearls-on-a-string’ type model of slipped-quadruplexes (Figure 5B). Alternatively, a ‘bubblie’ of separated strands, where each forms a series of quadruplexes may stabilize strand separation for some function, possibly the initiation of DNA replication, as several studies reveal that metazoan replication initiation sites coincide with quadruplex-forming sequences (68–72). Since transcription could potentially occur in the 64% of C9orf72 expansion carriers that do not have adjacent promoter CpG methylation but have a methylated repeat, the kind of methylation could modulate the formation of the above structures. Although the role of epigenetics in C9orf72-associated ALD/FTD is not understood, characterizing the structural and biological affects of C9orf72 repeat methylation may shed light on the mutagenic and pathogenic nature of the repeat.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

CEP conceived, designed, and coordinated the study. BZ, MM, RBM and CEP wrote the manuscript. BZ performed and analyzed the experiments in Figures 1 and 2. RBM designed and interpreted the results of Figure 1. MM performed and analyzed the experiments in Figures 3 and 4. KB provided intellectual input regarding hnRNPK. All authors reviewed the results and approved the final version of the manuscript.

FUNDING
Weston Brain Institute (to C.E.P.); Amyotrophic Lateral Sclerosis Association of America (to C.E.P.); Natural Science and Engineering Research Council of Canada (to C.E.P., R.B.M.); National Institutes of Health [R01 DK083310, R21 GM111439, R33 CA191135 to K.B.]. Funding for open access charge: Weston Brain Institute Grant to Christopher E. Pearson.

Conflict of interest statement. None declared.

REFERENCES

1. DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J. et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron, 72, 245–256.
57. Hardin, C.C., Corregan, M., Brown, B.A. and Frederick, L.N. (1993) Cytosine-cytosine+ base pairing stabilizes DNA quadruplexes and cytosine methylation greatly enhances the effect. *Biochemistry*, 32, 5870–5880.
58. Diekman,S. (1987) DNA methylation can enhance or induce DNA curvature. *EMBO J.*, 6, 4213–4217.
59. Hodges-Garcia,Y. and Hagerman,P.J. (1992) Cytosine methylation can induce local distortions in the structure of duplex DNA. *Biochemistry*, 31, 7595–7599.
60. Zacharias,W. (1993) Methylation of cytosine influences the DNA structure. *EYS*, 64, 27–38.
61. Nichol Edamura,K., Leonard,M.R. and Pearson,C.E. (2005) Role of replication and CpG methylation in fragile X syndrome CGG deletions in primate cells. *Am. J. Hum. Genet.*, 76, 302–311.
62. Wohrle,D., Salat,U., Glaser,D., Mucke,J., Meisel-Stosiek,M., Schindler,D., Vogel,W. and Steinbach,P. (1998) Unusual mutations in high functioning fragile X males: apparent instability of expanded unmethylated CGG repeats. *J. Med. Genet.*, 35, 103–111.
63. Taylor,A.K., Tassone,F., Dyer,P.N., Hersch,S.M., Harris,J.B., Greenough,W.T. and Hagerman,R.J. (1999) Tissue heterogeneity of the FMR1 mutation in a high-functioning male with fragile X syndrome. *Am. J. Med. Genet.*, 84, 233–239.
64. Cooper-Knock,J., Higginbottom,A., Stopford,M.J., Highley,J.R., Ince,P.G., Wharton,S.B., Pickering-Brown,S., Kirby,J., Hautbergue,G.M. and Shaw,P.J. (2015) Antisense RNA foci in the motor neurons of C9ORF72-ALS patients are associated with TDP-43 proteinopathy. *Acta Neuropathol.*, 130, 63–75.
65. Tomonaga,T. and Levene,D. (1995) Heterogeneous nuclear ribonucleoprotein K is a DNA-binding transactivator. *J. Biol. Chem.*, 270, 4875–4881.
66. Bomszyt,K., Denisenko,O. and Ostrowski,J. (2004) hnRNP K: one protein multiple processes. *Bioessays*, 26, 629–638.
67. Duquette,M.L., Handa,P., Vincent,J.A., Taylor,A.F. and Maizels,N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Nature Genet.*, 35, 1618–1629.
68. Walton,A.L., Hassan-Zadeh,V., Lema,I., Boggetto,N., Alberti,P., Santome,C., Riou,J.F. and Prioleau,M.N. (2014) G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J.*, 33, 732–746.
69. Hoshina,S., Yura,K., Teranishi,H., Kiyasu,N., Tominaga,A., Kadoma,H., Nakatsuka,A., Kunichika,T., Obuse,C. and Waga,S. (2013) Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. *J. Biol. Chem.*, 288, 30161–30171.
70. Cayrou,C., Coulombe,P., Puy,A., Riaille,S., Kaplan,N., Segal,E. and Meehan,M. (2012) New insights into replication origin characteristics in metazoans. *Cell Cycle*, 11, 658–667.
71. Lombrana,R., Almeida,R., Almeida,R. and Gomez,M. (2015) R-loops and initiation of DNA replication in human cells: a missing link? *Front. Genet.*, 6, 158.
72. Bartholdy,B., Mukhopadhyay,R., Lajugie,J., Aladjem,M.I. and Bouhissara,E.E. (2015) Allele-specific analysis of DNA replication origins in mammalian cells. *Nat. Commun.*, 6, 7051.