Pathogenic CANVAS-causing but nonpathogenic *RFC1* DNA/RNA repeat motifs form quadruplex or triplex structures

Received for publication, July 20, 2023, and in revised form, August 22, 2023 Published, Papers in Press, September 1, 2023,
https://doi.org/10.1016/j.jbc.2023.105202

Mohammad Hossein Abdi[1], **Bita Zamiri**[1,*], **Gholamreza Pazuki**[1], **Soroush Sardari**[2], and **Christopher E. Pearson**[3,4,*]

From the [1]*Department of Chemical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran;* [2]*Drug Design and Bioinformatics Unit, Department of Medical Biotechnology, Pasteur Institute of Iran, Tehran, Iran;* [3]*Program of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada;* [4]*Program of Genetics & Genome Biology, The Hospital for Sick Children, The Peter Gilgan Centre for Research and Learning, Toronto, Ontario, Canada*

Reviewed by members of the JBC Editorial Board. Edited by Brian Strahl

Biallelic expansions of various tandem repeat sequence motifs are possible in *RFC1* (replication factor C subunit 1), encoding the DNA replication/repair protein RFC1, yet only certain repeat motifs cause cerebellar ataxia, neuropathy, and vestibular areflexia syndrome (CANVAS). CANVAS presents enigmatic puzzles: The pathogenic path for CANVAS neither is known nor is it understood why some, but not all expanded, motifs are pathogenic. The most common pathogenic repeat is (AAGGG)n•(CCCTT)n, whereas (AAAAG)n•(CTTTT)n is the most common nonpathogenic motif. While both intronic motifs can be expanded and transcribed, only r(AAGGG)n is retained in the mutant *RFC1* transcript. We show that only the pathogenic forms unusual nucleic acid structures. Specifically, DNA and RNA of the pathogenic d(AAGGG)4 and r(AAGGG)4 form G-quadruplexes in potassium solution. Nonpathogenic repeats did not form G-quadruplexes. Triple-stranded structures are formed by the pathogenic motifs but not by the nonpathogenic motifs. G- and C-richness of the pathogenic strands favor formation of G•G•G•G-tetrads and protonated C+-G Hoogsteen base pairings, involved in quadruplex and triplex structures, respectively, stabilized by increased hydrogen bonds and pi-stacking interactions relative to A-T Hoogsteen pairs that could form by the nonpathogenic motif. The ligand, TMPyP4, binds the pathogenic quadruplexes. Formation of quadruplexes and triplexes by pathogenic repeats supports toxic-DNA and toxic-RNA modes of pathogenesis at the *RFC1* gene and the *RFC1* transcript. Our findings with short repeats provide insights into the disease specificity of pathogenic repeat motif sequences and reveal nucleic acid structural features that may be pathogenically involved and targeted therapeutically.

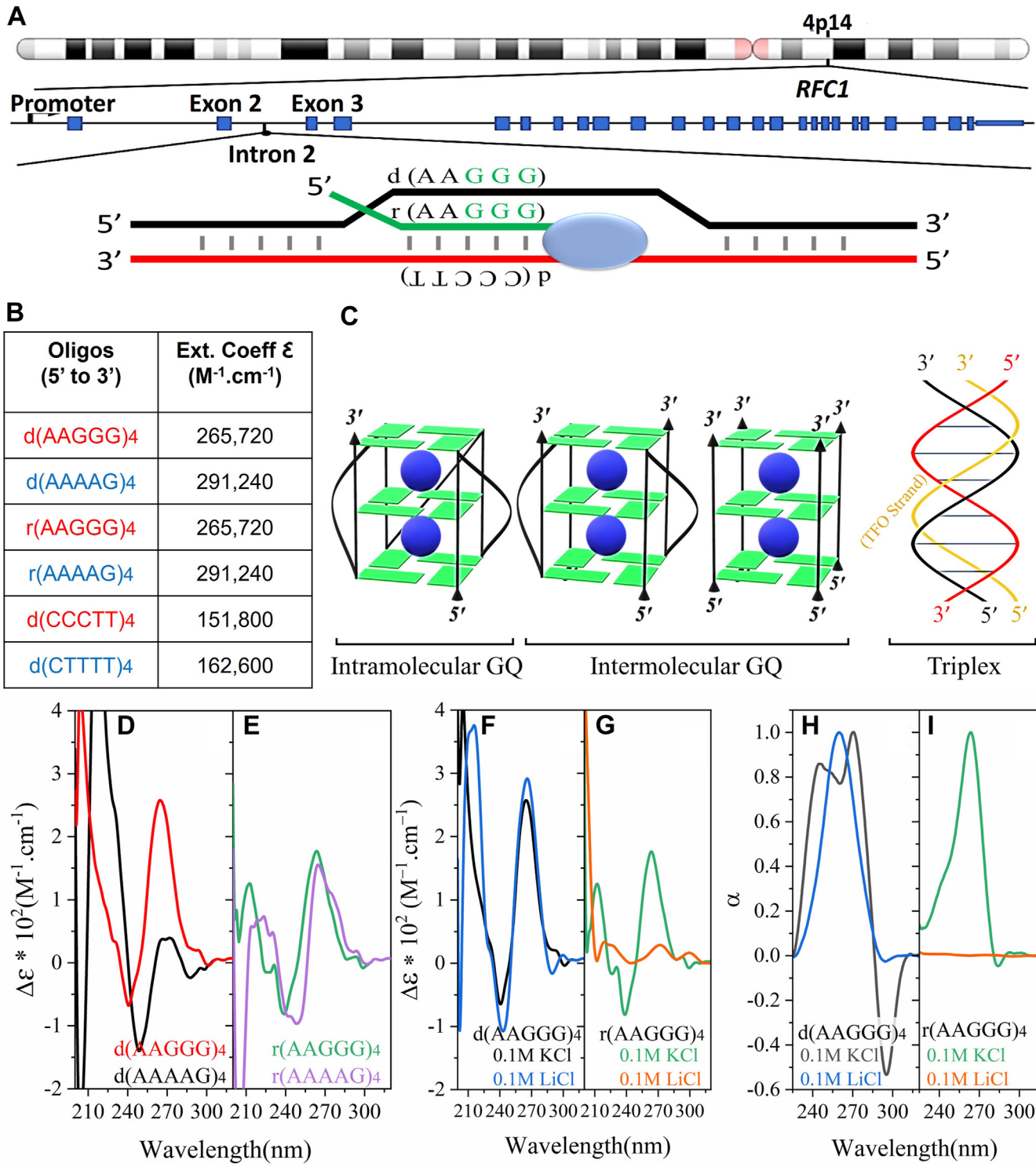
Cerebellar ataxia with neuropathy and vestibular areflexia syndrome (CANVAS) is a late-onset progressive neurodegenerative disease (1). Large biallelic repeat expansions of 250 to >2000 repeats in intron 2 of the replication factor C subunit 1 (*RFC1*) gene have been identified as the cause for disease (Fig. 1*A*) (2–4). A refractory chronic cough typically precedes neurological symptoms by >20 years and is present in monoallelic *RFC1*-expansion carriers (5, 6). Five different pathogenic expanded repeat motifs have been observed; with the AAGGG repeat having the highest frequency in patients (Fig. 1*A*) (2, 3). The nonpathogenic motifs contained AAAAG or AAAGG expanded repeats, with the former reported as the most common nonpathogenic allele (3, 4). Both nonpathogenic and pathogenic repeats can be massively expanded. It is not clear why some expanded motifs cause disease and other expanded motifs do not. The motif sequences suggest that higher-order nucleic acid structures might be involved and that pathogenic motifs are forming structures that nonpathogenic motifs are not.

The exact etiology and pathogenesis of CANVAS is puzzlingly unclear. Recent data support a loss-of-function mechanism caused by the homozygous *RFC1* pathogenic repeat expansion motifs (7). The RFC1 protein is part of the pentameric complex of RFC (replication factor C) that is involved in DNA replication and repair.

Homopurine sequences (guanines and adenines) and homopyrimidine (cytosines and thymines) DNA and RNA strands can self-associate or interassociate to form a wide variety of non-B-form structures. These structures can be biophysically polymorphic and involve different purine–purine base pairs, such as G•G, A•A, and G•A, which are the basis for multiple folding patterns (8). For instance, d(GA)n has been shown to self-associate into parallel and antiparallel homoduplexes, quadruplexes, and triplexes (9, 10). Expansions of d(GAA)n•(TTC)n, associated with Friedreich's ataxia, were shown to form higher-order structures revealed to be involved in disease (11–13). Environmental conditions including counterions, pH, and temperature can influence and determine the conformational polymorphism.

High guanine content in homopurine tracts under certain environmental conditions can assume quadruplex structures (Fig. 1*C*) (10). G-quartets are the building blocks for the four-stranded quadruplex structures. G-quartets are planar arrangements of four guanine molecules self-associated *via* Hoogsteen hydrogen bonds around a monovalent cation such as potassium

* For correspondence: Bita Zamiri, bita.zamiri@gmail.com; Christopher E. Pearson, cepearson.sickkids@gmail.com.



Oligos (5' to 3')	Ext. Coeff ε ($M^{-1}.cm^{-1}$)
d(AAGGG)4	265,720
d(AAAAG)4	291,240
r(AAGGG)4	265,720
r(AAAAG)4	291,240
d(CCCTT)4	151,800
d(CTTTT)4	162,600

Figure 1. Pathogenic but not nonpathogenic DNA and RNA repeats form quadruplex structures. *A*, pathogenic and nonpathogenic expanded repeat motifs identified in the transcribed intron 2 of the *RFC1* gene where the pathogenic repeat is retained, whereas the nonpathogenic repeat is excised by splicing mechanisms (3). Both are translated to the full-length RFC1 protein. *B*, sequences of the ribo-oligonucleotides and deoxyribo-oligonucleotides used here in *red*- and *blue*-fonted oligos are pathogenic and nonpathogenic, respectively. *C*, schematic of intramolecular and intermolecular G-quadruplex DNA structures consisting of three G-quartets form and schematic of a triplex DNA structure. *D*, CD spectra of 5 μM of DNA repeats. *E*, RNA repeats of pathogenic (AAGGG)4 and nonpathogenic (AAAAG)4 repeats. *F* and *G*, the effect of lithium ions on CD spectra. *H* and *I*, normalized thermal difference spectra (TDS) of 5 μM d(AAGGG)4 and *r(AAGGG)4* in 10 mM Tris and 0.1 mM EDTA (pH 7.4) at 25 °C. RFC1, replication factor C subunit 1.

or sodium. G-quartets stack on top of each other, stabilized by pi–pi interactions to form an overall helical arrangement. They can form from a single strand or multiple strands (14, 15). Quadruplexes have been linked to various biological processes, including genomic instability, gene regulation, transcript splicing, and RNA translation regulation (16–18).

The presence of a homopurine•homopyrimdine duplex and a single-stranded purine or pyrimidine strand can form triple-stranded DNAs (Fig. 1*C*). In triplexes, the third strand (*italicized*) is bound to the major groove of the complementary stranded double helix (indicated by "•") *via* alternative hydrogen bonding (Hoogsteen base pairing, indicated by a "-") resulting in C•G-C^+ and T•A-*T* triads where in most cases, cytosines require protonation (indicated by "+") to bind to the major groove. As previously reviewed, neurons under stressed conditions such as neurodegeneration can experience intranuclear imbalances of pH often toward acidic levels (pH < 7) (19, 20). Cytosine residues become protonated under slightly acidic pH (pH 4.5–6.2), favoring triplex formation (21).

Here, we report the formation of G-quadruplexes and triplexes in the pathogenic motif d(AAGGG)4 but not in the nonpathogenic d(AAAAG)4. Quadruplex formation is also confirmed in the pathogenic RNA r(AAAGGG)4 but not in the nonpathogenic repeat r(AAAAG)4. We also reveal triple-stranded structures formed by the pathogenic (AAGGG)4•(CCCTT)4-*(CCCTT)4* motifs but not by the nonpathogenic motif. Our findings support the concept that these unusual structures may participate in disease pathogenesis.

Results

Quadruplex structure formation by DNA/RNA of pathogenic repeats (CD)

Since the composition of the pathogenic repeats $(AAGGG)_n$•$(CCCTT)_n$ are both homopurine/homopyrimidine mirror repeats (22) and contain evenly spaced G3 runs (23), we hypothesized these repeats could assume triplex and/or quadruplex structures (Fig. 1, *A*–*C*). Previous studies revealed that a minimum of two to four contiguous repeat units was sufficient to detect triplex and quadruplex structures by GAA/TTC and CGG/CCG repeats (24–27). Therefore, we used CD spectroscopy to investigate the aforementioned and study the structural differences between pathogenic (AAGGG)4 and nonpathogenic (AAAAG)4 DNA and RNA repeats at physiological conditions. As seen in (Fig. 1*D*), the CD spectra of d(AAGGG)4 in KCl (pH 7.4) exhibits positive peaks around 265 and 210 nm with a negative peak around 240 nm, the signature spectra of parallel stranded G-quadruplexes (28). The nonpathogenic d(AAAAG)4 has a positive low-intensity peak around 270 nm and a high-intensity peak around 215 nm with negative peaks around 250 and 205 nm. For RNA (Fig. 1*E*), the pathogenic r(AAGGG)4 exhibits positive peaks around 265 nm and a negative peak around 240 nm; a signature of parallel-stranded RNA G-quadruplexes, whereas the nonpathogenic variant r(AAAAG)4 has positive peaks around 265 nm and 225 nm alongside a shoulder around 280 nm with negative peaks around 245 and 210 nm. A positive 260 nm peak and a negative 210 nm have been previously associated with A-form duplex formation in RNA molecules (28). Accordingly, potassium-induced G-quadruplex formation in both pathogenic DNA and RNA repeats of (AAGGG)4 is seen. In order to further confirm this, the effect of changing environmental conditions such as cation type has been studied.

G-quadruplex formation is largely cation dependent, and Li^+ ions are known to destabilize them. To show this, CD spectra (Fig. 1, *F* and *G*) and thermal difference spectra (TDS) (Fig. 1, *H* and *I*) of (AAGGG)4 DNA and RNA repeats in 100 mM LiCl were recorded. In Figure 1*F*, d(AAGGG)4 exhibits a positive peak at 265 nm and a negative peak at 240 nm (signature G-quadruplex signals) in both K^+ (*black*) and Li^+ (*blue*). In Li^+, negative peaks at 215 and 290 nm are seen. CD signals around 265, 250, and 215 nm have been attributed to homoduplexes (8). In Figure 1*G*, the effect of Li^+ on the CD spectra of r(AAGGG)4 is more pronounced and the CD signals at 260 and 240 nm disappear in Li^+. The TDS is the difference in the UV spectra of folded *versus* unfolded structures and was shown to provide a distinctive signature for nucleic acid structures (29). In Figure 1, *H* and *I*, TDS of d(AAGGG)4 and r(AAGGG)4 in K^+ and Li^+ are not superimposed. For d(AAGGG)4 in Li^+, the TDS exhibits a 257 nm positive peak with a minor negative peak around 293 nm, associated with GA duplex formation (29). In K^+, positive peaks at 243 and 273 nm and an intense negative peak at 295 nm confirm G-quadruplex formation. For r(AAGGG)4, consistent with the CD results, lithium was shown to destabilize the RNA higher-order structures according to the TDS as well (Fig. 1*I*). Thus, potassium-induced G-quadruplex formation in pathogenic DNA and RNA repeats is further confirmed. Next, the thermal stability of the repeats in K^+ *versus* Li^+ is studied.

Structure formation by DNA/RNA of pathogenic repeats (melting)

UV thermal melting profiles are used to gain structural insight by studying thermal stability in K^+ *versus* Li^+ and specify molecularity (based on the concentration dependence of T_m) (30). In Figure 2*A*, upon doubling the concentration of d(AAGGG)4, the T_m has increased from 45.32 to 48.90 °C, showing multimolecularity. Also, T_m was shown to decrease from 45.32 °C in K^+ to 32.17 °C in Li^+ and was concentration independent (Fig. 2*B*). Also, in K^+, thermal meltings were performed at 295 nm, a signature wavelength for recording G-quadruplex melting transitions, and in Li^+, no transition was recorded at 295 nm (not shown). The RNA G-quadruplexes formed in K^+ were shown to have a T_m of 35.28 ± 0.7 and 43.08 ± 0.5 at 5 μM and 10 μM strand concentrations, respectively, also showing multimolecularity (Fig. 2*C*), whereas in Li^+, no melting phenomenon was observed (not shown). Therefore, potassium-induced multimolecular G-quadruplexes are shown for DNA and RNA of d(AAGGG)4 repeats. Consequently, the tendency of the pathogenic and nonpathogenic repeats to form triplex structures was studied using mixing curves.

Strand proportions affect structure formation

Mixing curves, resulting from recording the CD spectra of a mixture of predetermined proportions of d(AAGGG)4 and d(AAAAG)4 with their complementary strands, are used to demonstrate triplex formation in the pathogenic and nonpathogenic repeats (Fig. 3). The homopyrimidine d(CCCTT)4 exhibit a prominent positive peak around 285 nm

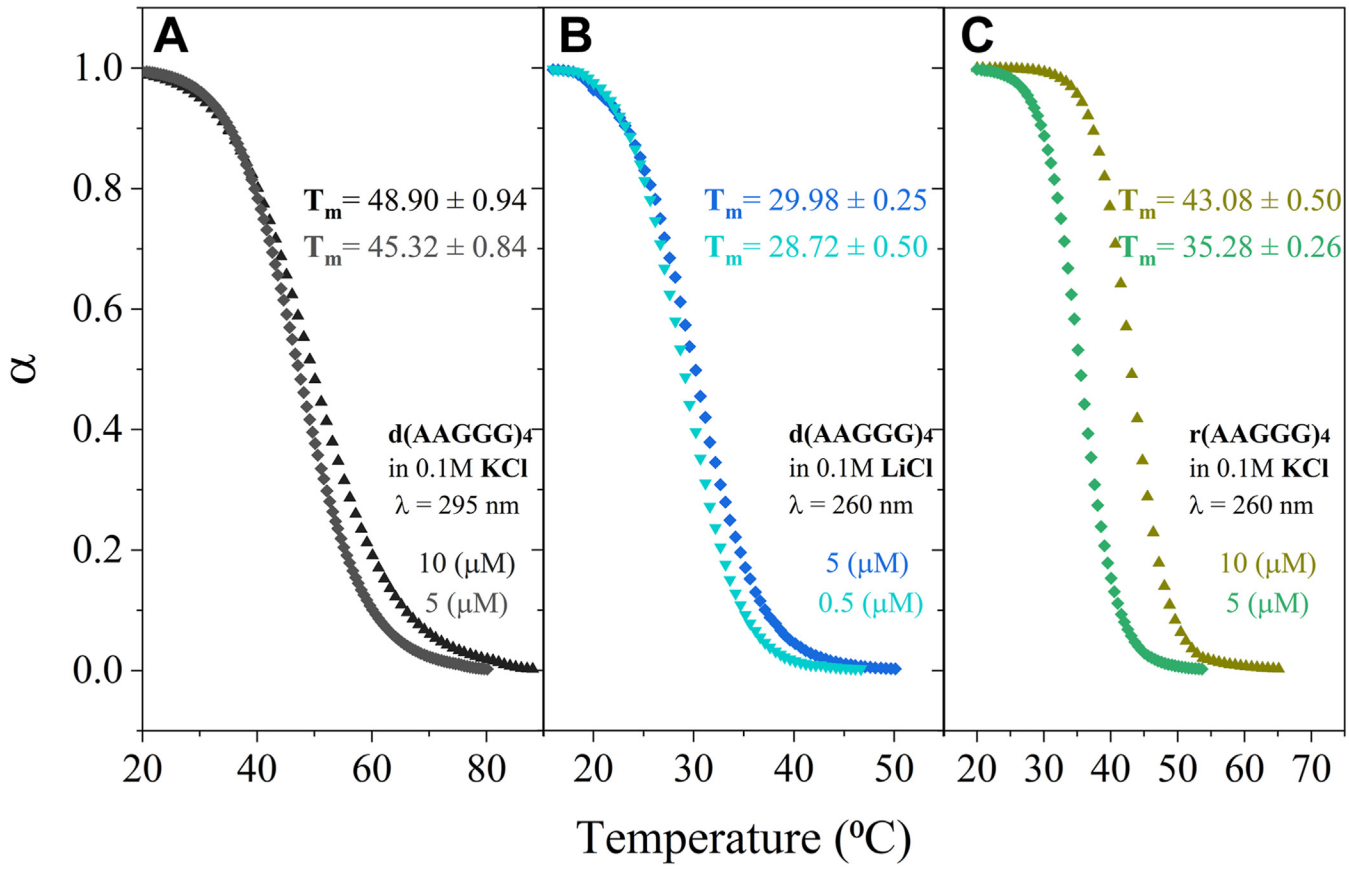


Figure 2. Effect of LiCl and strand concentration on the normalized UV thermal melting profiles. *A*, d(AAGGG)4 in 0.1 M KCl. *B*, d(AAGGG)4 in 0.1 M LiCl. *C*, r(AAGGG)4 in 0.1 M KCl in 10 mM Tris–HCl and 0.1 mM EDTA (pH 7.4) at different strand concentrations.

and a negative peak at 260 nm; and d(CTTTT)4 a positive peak around 280 nm and a negative peak around 250 nm at pH 5.6, confirming *i*-motif tetraplex formation (21). In Figure 3*A* for pathogenic d(AAGGG)4, upon increasing the proportion of d(CCCTT)4 in the d(CCCTT)4:d(AAGGG)4, mixture, the maximum intensity positive peak of the CD spectra red shifts to higher wavelengths initially accompanied by a reduction in intensity followed by an increase in intensity. The change in this peak is shown in Figure 3*C* (wavelength *versus* CT proportion, *blue points*). The breakpoint in the mixing curve (Δε *versus* CT proportion, *blue points*) indicates d(CCCTT)4•d(AAGGG)4-*d(CCCTT)4* triplex formation at 0.67:0.33 M proportion of d(CCCTT)4:d(AAGGG)4. This was confirmed by plotting the deviation of the weighted average of the spectra of the individual strands at different molar proportions of d(CCCTT)4, with the observed spectra of the mixing curves at the same molar proportions; which is maximal at 0.67 M proportion. At this molar proportion, isodichroic points are seen at 272 and 239 nm (Fig. S1). In contrast, the mixing curve for the nonpathogenic d(AAAAG)4 shown in Figure 3*B*, show that upon increasing the proportion of d(CTTTT)4, the maximum positive peak of the CD spectra red shifts to higher wavelengths with an intensity increase (also seen in Figure 3*C*, wavelength *versus* CT proportion, *red points*). Unlike the pathogenic repeats, the weighted average of the spectra of the individual strands of d(CTTTT)4 and d(AAAAG)4 at 0.67 M proportion of d(CTTTT)4 is superimposed with the observed mixing curve spectra (Fig. S1). This along with the absence of a breakpoint in the Δε *versus* CT proportion in Figure 3*C* (*red points*) is indicative of the absence of triplex DNA formation for the nonpathogenic repeats at the same environmental conditions and points toward the d(AAAAG)4 and d(CTTTT)4 strands forming self-complexes separately. Also, no isodichroic points are observed.

Both pH and counterions affect structure formation

Adenine-rich homopurine sequences such as d(AAAAG)4 can be biophysically polymorphic and largely dependent on environmental factors (8). Therefore, the effect of pH, temperature, and counterions has been studied (Fig. 4). In Figure 4, *A–C*, the effect of increasing Li^+ at two different pHs and increasing Mg^{2+} at neutral pH is shown. At neutral pH (Fig. 4*B*), in the absence of Li^+, the CD spectra exhibit a low-intensity positive peak around 270 nm, a negative peak around 250 nm, and a sharp positive peak around 220 nm. At acidic pH (Fig. 4*A*) and at 0 M LiCl, the 270 and 250 nm peaks are more pronounced. By increasing the ionic strength (Fig. 4, *A–C*), the 270 and 250 nm increase in intensity, showing secondary structure formation (21); consistent with reports in the literature that $MgCl_2$ (31) and LiCl (8) promote homoduplex formation in homopurine tracts. In Figure 4, *D–F*, the effect of temperature is studied. At 0.1 M K^+ (Fig. 4*D*), the 5 °C and 25

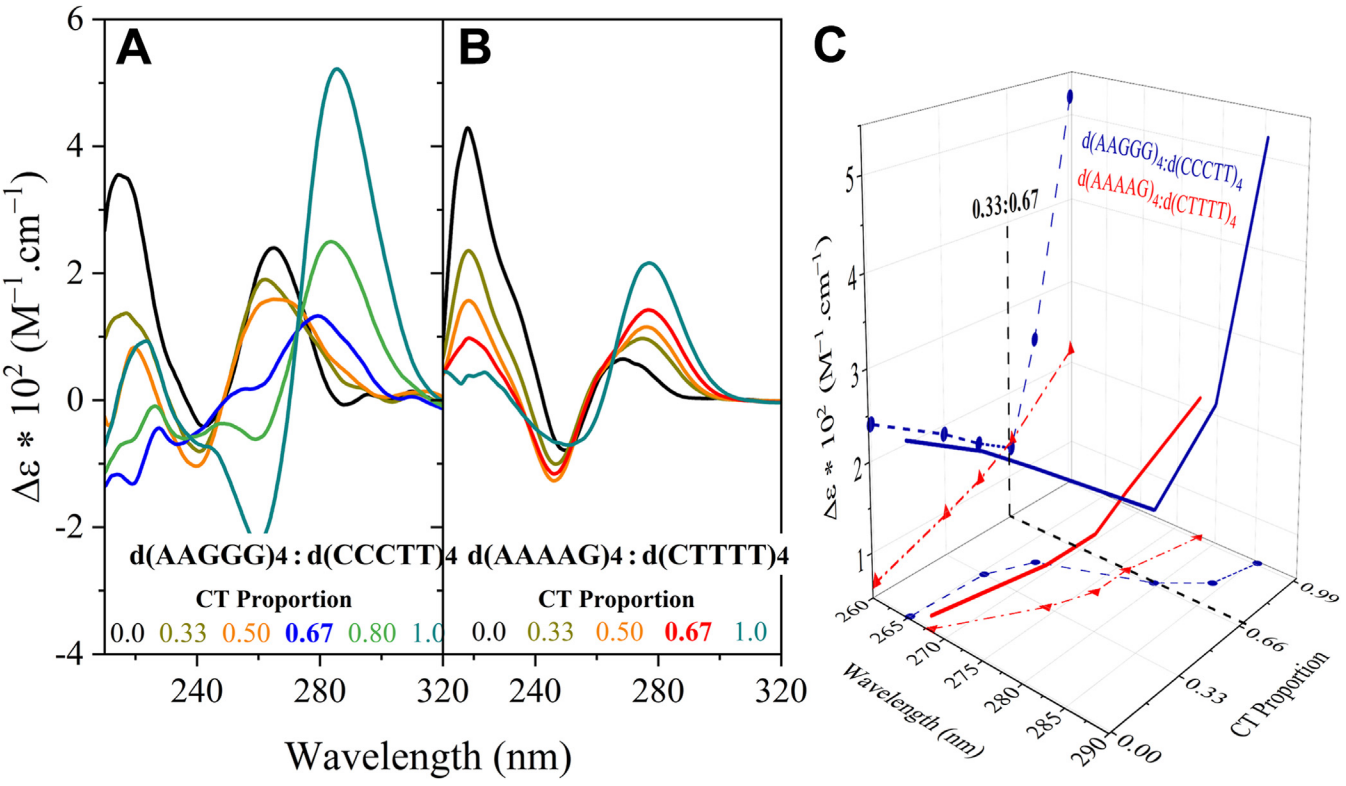


Figure 3. Triplex formation by pathogenic but not nonpathogenic DNA repeats. CD spectra of mixtures of (*A*) d(AAGGG)4:d(CCCTT)4 and (*B*) d(AAAAG)4:d(CTTTT)4 at the indicated molar portions; in 6 mM sodium phosphate buffer and 150 mM NaCl (pH 5.6) at 25 °C. *C*, effect of increasing the proportion of d(CCCTT)4 on the signature positive CD ellipticity of d(AAGGG)4•d(CCCTT)4 mixtures compared with the effect of parallel reactions of increasing the proportion of d(CTTTT)4 on the signature positive CD ellipticity of d(AAAAG)4•d(CTTTT)4 mixtures. Increased molar levels of d(CCCTT)4 relative to duplexes d(CCCTT)4•d(AAGGG)4 to 0.67:0.33 marks the maximum deviation in the CD spectrum of the d(TTCCC)4•d(AAGGG)4-*days(TTCCC)4* triplex from the weighted average of the spectra of the individual strands. A shift to a triplex signal was not observed for d(AAAAG)4•d(CTTTT)4) mixtures with increased proportions of d(CTTTT)4.

°C CD spectra are superimposed consistent with the absence of a melting transition (not shown), confirming a very low T_m. In 1.2 M Li^+ (Fig. 4*E*) and 0.5 M Mg^{2+}, this was not the case, and lower temperatures increased the 270 nm peak. At lower concentrations of Li^+, the CD spectra at two temperatures were superimposed (not shown). Furthermore, the TDS and melting transitions were recorded at 0.5 M $MgCl_2$ and 1.2 M LiCl (Fig. S2) confirming secondary structure formation induced by increasing the ionic strength.

Quadruplex ligand binds pathogenic repeats

The interaction between TMPyP4 and G-quadruplexes formed by d(AAGGG)4 was studied by measuring the visible absorption spectra (Fig. 5). The d(AAGGG)4 oligonucleotides were titrated into a solution of TMPyP4, and the Soret band was monitored as a function of DNA concentration (Fig. 5*A*). The hypochromicity seen upon increasing concentrations of d(AAGGG)4 was 68% with a bathochromic shift of 16 nm indicative of the binding of TMPyP4 to quadruplexes formed by d(AAGGG)4. The data also reveal an apparent isosbestic point at 435 nm indicating the presence of two species. The effect of TMPyP4 on the folding pattern of d(AAGGG)4 is seen in Figure 5*B*, where at the molar proportion of TMPyP4:d(AAGGG)4 equal to 0.9 (where binding is complete), we see a 260 nm peak with lower intensity.

Discussion

Homopurine sequences can adopt a variety of different folding patterns depending on the environmental conditions (8, 32). Here, we demonstrate that DNA and RNA of the pathogenic AAGGG but not the nonpathogenic AAAAG repeat motifs of *RFC1* form unusual structures whose formation is sensitive to pH, Li^+, K^+, and temperature. Quadruplexes form in DNA and RNA of the pathogenic repeat (AAGGG)4 in the presence of K^+, exhibiting the signature CD signal (Fig. 1, *D* and *E*) and TDS (Fig. 1, *H* and *F*) of G-quadruplexes. In the presence of Li^+, the DNA of the pathogenic repeats forms intramolecular foldbacks with lower stability through G•A and G•G pairs while preventing G-quadruplex formation (10). The similarity observed between the CD spectra of d(AAGGG)4 in K^+ *versus* Li^+ can been attributed to the fact that quadruplex formation can be the result of addition of two homoduplexes and that purine–purine stacking is conserved in the quadruplexes and intramolecular foldbacks (8). The DNA and RNA G-quadruplexes formed in K^+ by the pathogenic (AAGGG)4 are multimolecular as their melting temperatures are concentration-dependent (33). As for the nonpathogenic d(AAAAG)4 repeats, the CD spectra at physiological conditions (Fig. 1*D*) with low amplitudes from 240 to 300 nm point toward a single-stranded conformer consistent with the literature (8). Lowering pH and increasing ionic strength promoted secondary conformers for d(AAAAG)4. The

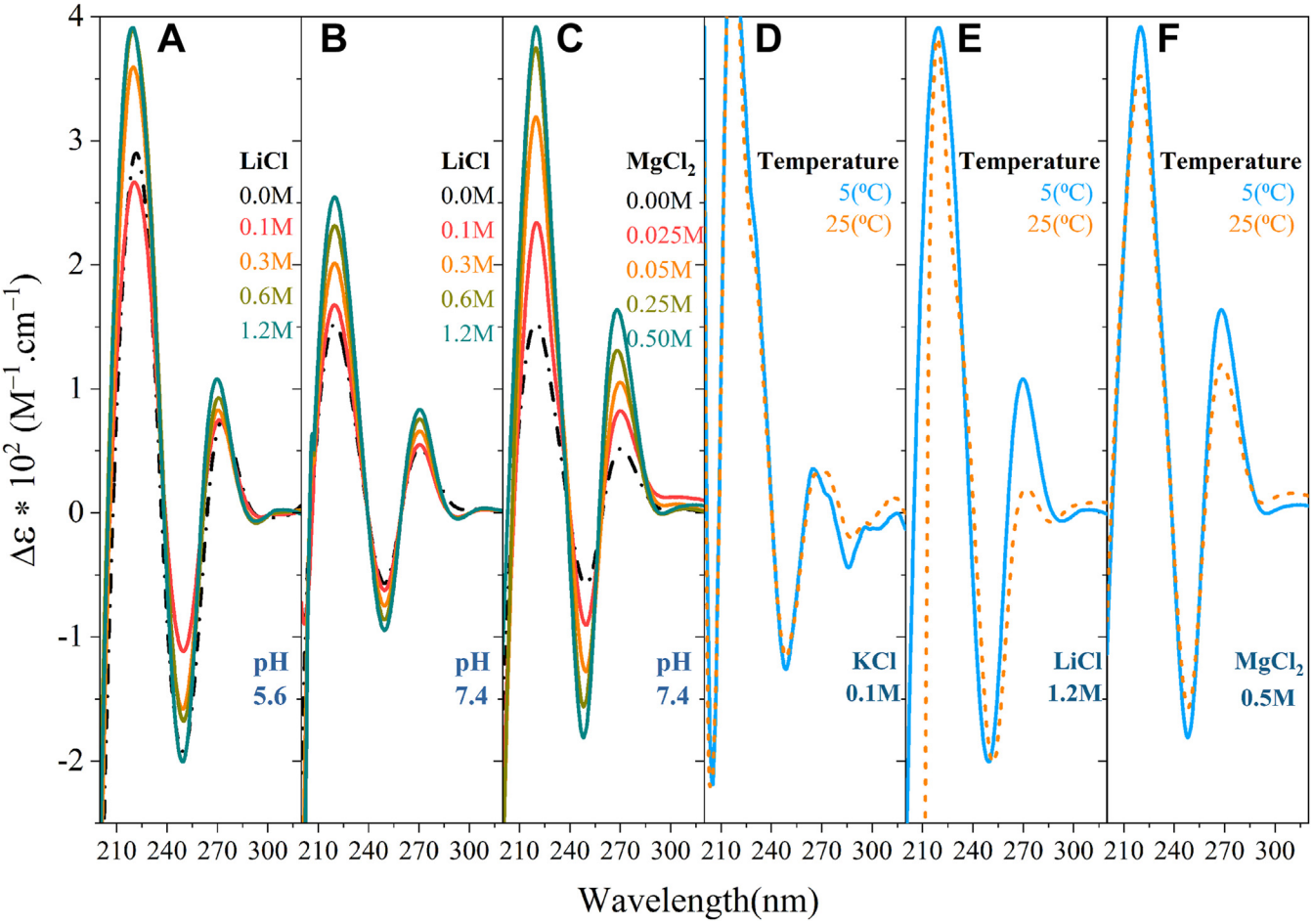


Figure 4. Assessment of homoduplex formation in the nonpathogenic DNA repeats using CD spectra of 5 μM d(AAAAG)4. *A*, indicated concentrations of LiCl in 6 mM potassium phosphate and 0.1 mM EDTA (acidic pH 5.6) at 5 °C. *B*, indicated concentrations of LiCl in 6 mM potassium phosphate and 0.1 mM EDTA (pH 7.4) at 5 °C. *C*, indicated concentrations of $MgCl_2$ in 10 mM Tris–HCl and 0.1 mM EDTA (pH 7.4) at 5 °C. *D*, 10 mM Tris–HCl, 0.1 mM EDTA, 100 mM KCl (pH 7.4) at 5 and 25 °C. *E*, 6 mM potassium phosphate, 0.1 mM EDTA, 1.2 M LiCl (pH 7.4) at 5 and 25 °C. *F*, 10 mM Tris–HCl, 0.1 mM EDTA, and 50 mM $MgCl_2$ (pH 7.4) at 5 and 25 °C.

arising CD spectra at acidic pH and high ionic strengths has been previously associated with bimolecular duplexes in homopurine tracts (34, 35).

We have also revealed that at slightly low pH levels that can arise in stressed cells (19, 20), (CCCTT)4•d(AAGGG)4-*d(CCCTT)4* triplex structures are formed, whereas the nonpathogenic strands did not form triplexes in the presence of their complementary strands. Several factors including base composition, base sequence, expansion size, and environmental factors determine triplex stability (21, 36–38). For instance, the more G-rich the strand is, the more biophysically stable the triplex is. At low pH, C•G-C^+ triads are more stable compared with T•A-*T* triads. While cytosine protonation permits G-C Hoogsteen pairs, protonation is not required for A-T Hoogsteen pairs in the T•A-*T* triads (39). Interrupting the G-richness with a single residue or multiple A residue amongst a run of G residues can diminish the propensity to form triplexes and their biophysical stability (37). Pi–pi stacking interactions of adjacent triads can dramatically affect stability of triplexes, G•G•G•G-quartets, or C•G-C^+ triads and can favor the biophysical stability of G-quadruplexes or triplexes, respectively. Moreover, the sequence organization of a G+A repeat, including the mirror repeat of the sequence, extrusion point, and loop location, can dramatically affect triplex structures (40). Moreover, Hoogsteen base pairings, involved in both quadruplex and triplex structures, where these pairings can be biophysically stabilized by increased hydrogen-bond and pi-stacking interactions relative to the A-T Hoogsteen base pairs formed by the nonpathogenic $(AAAAG)_n$•$(CTTTT)_n$ motif (41, 42).

The ability and propensity of the pathogenic but not nonpathogenic repeats to form unusual structures in the DNA and RNA suggests that such structures may participate in CANVAS pathogenesis. For example, the DNA structure formation may allow for larger expansion sizes, transmitted or somatic of the pathogenic *RFC1* repeats. CANVAS is caused by expansions of (AAGGG)250–2000 repeats. Our study has focused upon short tracts of RFC1 repeat motifs associated with CANVAS pathology and with neurotypically healthy individuals, we have not structurally assessed disease relevant lengths. While long tracts of the nonpathogenic motif, $(AAAAG)_n$•$(CTTTT)_n$, could form a triplex, its biophysical stability would likely be diminished relative to the expanded pathogenic motif, $(AAGGG)_n$•$(CCCTT)_n$, as A-tracts have been shown to disfavor triplexes (43). Similarly, the propensity to form a quadruplex by long tracts

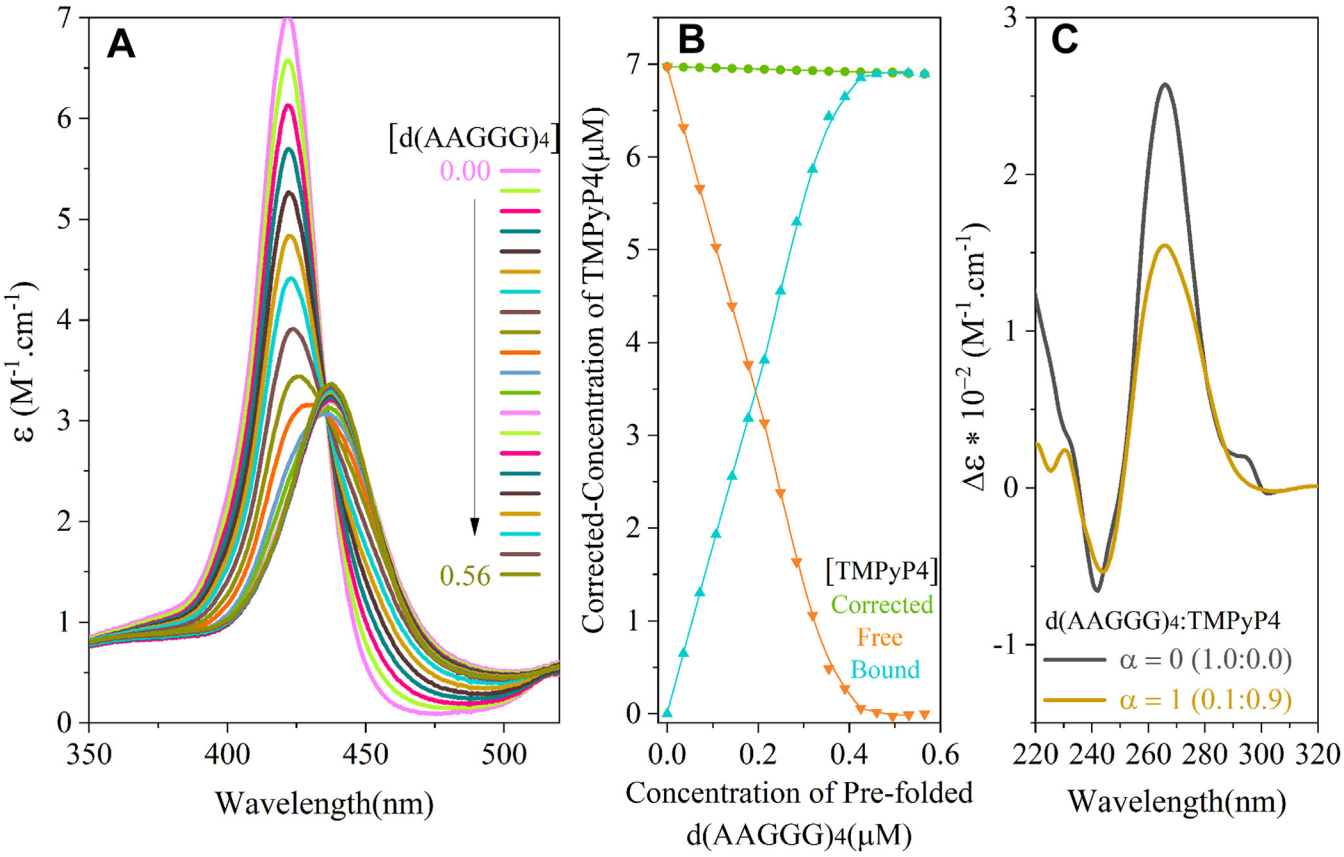


Figure 5. TMPyP4 binds to pathogenic DNA repeats. *A*, visible absorption spectra of TMPyP4 in the absence and presence of an increasing concentration of prefolded d(AAGGG)4. *B*, free and bound TMPyP4 *versus* the concentration of prefolded d(AAGGG)4. *C*, CD spectra of d(AAGGG)4 with and without TMPyP4.

of r(AAAAG)n would be severely diminished relative to long tracts of r(AAGGG)n (23). While it is likely that longer tracts will have at least the same, if not greater propensity to form the unusual structures we demonstrate herein, we make no claim as to a length-dependent effect. We note that numerous studies have used only short nondiseased lengths of repeat tracts to biophysically reveal formation of higher order structures. Those initial studies revealed key structural features in subsequent studies of disease-relevant lengths. For example, intrastrand hairpins are critical features of slipped-DNAs assumed by diseased-length repeat tracts in patient tissues (44–47). Similarly, triplexes and quadruplexes initially observed in short tracts were evident in diseased tracts in patient cells (24, 27, 48, 49), and in premutation/protomutation lengths, respectively (25, 26, 50–52). The (TTTCC)48•(GGAAA)48, of the quail *T64* gene, having a length shorter than *RFC1* CANVAS disease lengths, could form both PU-PY-PY and PU-PU-PY triplexes (53, 54). In the RNA, structure formation may contribute to the escape of repeat-containing intron excision, as many disease-associated G+C-rich repeats are retained in toxic RNAs (55). We use the term "toxic-DNA" much as the term "toxic-RNA," wherein an unusual DNA sequence can be the source of toxicity either as direct source of toxicity, in that it is mutagenic or indirect by being the template for toxic-RNAs. The involvement of RNA secondary structure of the actively or newly produced transcripts, including those with G/A-rich homopurine motifs on alternative splicing of pre-mRNA has been reviewed (56–60). Such structural changes can occur in the DNA and RNA. The subcellular localization and processing of the mutant *RFC1* transcript may be altered. A loss-of-function pathway may arise through the sequestration of proteins bound to the expanded repeat motif and/or to quadruplex or triplex structures formed by the pathogenic alleles. As in other repeat diseases, such sequestered proteins could include those involved in a broad range of functions, such as transcription, RNA metabolism, chromatin regulation, and DNA repair (61–63).

The possible involvement of higher-order nucleic acid structures in disease has led to chemical targeting of these using small molecules. Ligands that bind to G-quadruplexes or triplex structures are known. Ligands specific to the structures formed by disease-specific DNA and RNA repeat motifs have been shown to inhibit gene expression, alter gene splicing, or alter repeat instability (64–73). Understanding the disease-specific DNA and RNA structures formed by the CANVAS-causing repeat motifs in *RFC1*, relative to the nonpathogenic motifs, is a beginning to opening such therapeutic paths.

Experimental procedures

Sample preparation

Oligonucleotides listed in Figure 1*B* were purchased from IDT and ACGT Corp. Concentrations were determined by

recording the absorbance at 260 nm with molar extinction coefficients (Fig. 1*B*). Oligos were dissolved in 100 mM KCl, 10 mM Tris–HCl, and 0.1 mM EDTA (pH 7.4) unless otherwise stated. Samples were heated to 95 °C and allowed to cool to room temperature overnight.

CD spectroscopy

CD spectra were measured using a Jasco J-810 spectropolarimeter, using 5 to 20 μM sequence in the stated buffer. For the CD spectra, samples were in a 0.1 cm pathlength cuvette, and spectra were measured over the wavelength range of 190 to 330 nm at a scan rate of 200 nm/min. Spectra shown are the average of three such spectra and were corrected against the buffer. CD data were converted to molar circular-dichroic absorption (Δε) based on strand concentration (C) using the equation:

$$\Delta\varepsilon = \theta/(32980 * C * l) \quad (1)$$

where θ is the measured ellipticity value (mdeg), c is the molar strand concentration (M), and l is the optical path of a used cuvette (cm).

TDS

All experiments were performed on a Varian Cary 100 UV–Vis spectrophotometer using cuvettes with an optical path length of 1 cm in a Peltier multicell (6*6) holder. Absorbance spectra were recorded in the 220 to 340 nm range, with a scan speed of 200 nm/min and a data interval of 1 nm. In all cases, the sample concentration was set at 0.005 mM, and they were dissolved in 10 mM Tris–HCl, 0.1 mM EDTA (pH 7.4) and 100 mM KCl or LiCl, or 6 mM potassium phosphate, 0.1 mM EDTA (pH 5.6), and 1.5 M LiCl or 25 mM $MgCl_2$ when in need to lower the pH. TDS were normalized by dividing the raw data by its maximum positive value, where the highest positive peak gets a Y-value of +1.

UV thermal melting

UV thermal melting experiments were monitored on a Varian Cary 100 spectrometer by recording changes in the UV absorption at either 260 nm or 295 nm using a 1 cm path length cuvette. Heating rate was adjusted to 1 °C/min with data intervals of 0.5 °C. T_m values were determined from dual baseline-corrected 1 to 0 normalized curves (1-native and 0-denatured forms) as temperatures at which half of the molecules were folded.

Mixing curve

The single-stranded oligos of d(AAGGG)4 and d(TTCCC)4 were mixed together in weighed proportions to give d(AAGGG)4:d(TTCCC)4 mixtures with molar ratios of 80:20, 67:33, 60:40, 50:50, 40:60, 33:67, and 20:80. Together with samples of the individual strands, all were dissolved in 6 mM sodium phosphate buffer and 150 mM NaCl (pH 5.6) with a total DNA strand concentration of 5 μM. To overcome intrastrand pairing the samples were heated to 95 °C and allowed to cool overnight.

Soret band shift

A prefolded d(AAGGG)4 with 50 μM concentration was titrated into a cuvette containing 7 μM of TMPyP4. Absorption spectra were collected with a Varian Cary 100 UV–Vis spectrophotometer from 350 nm to 520 nm using cuvettes with an optical path length of 1 cm. The concentration of free TMPyP4 was determined using an extinction coefficient of $2.26 \times 10^5\ M^{-1}\ cm^{-1}$ and absorbance values at 424 nm. The titration was stopped when three successive additions of the sample resulted in no further shift of the Soret band.

All values were corrected for the dilution effect. The fraction (α) of TmPyP4 bond was determined as follows:

$$\alpha = \left(\text{Abs TMPyP4}_{\text{free}} - Abs_{\text{mixture}}\right) \Big/ \left(\text{Abs TMPyP4}_{\text{free}} - Abs_{\text{bond}}\right) \quad (2)$$

where Abs $TMPyP4_{free}$ is the absorbance of the free TMPyP4 in the absence of any added DNA, $Abs_{mixture}$ is the absorbance at any point after the beginning of the addition of the DNA, and Abs_{bond} is the absorbance of fully bonded TmPyP4 measured at 424 nm (the Soret maxima for TMPyP4). Concentrations of free TMPyP4 ($[TMPyP4]_{free}$) and the concentration of bound TMPyP4 ($[TMPyP4]_{bond}$) were calculated as follows:

$$[\text{TmPyP4}]_{\text{free}} = [\text{TmPyP4}]_{\text{corrected}}(1-\alpha) \quad (3)$$

$$[TMPyP4]_{\text{bond}} = \left([TMPyP4]_{\text{corrected}} - [TMPyP4]_{\text{free}}\right) \quad (4)$$

where $[TMPyP4]_{corrected}$ represents the concentration of TMPyP4 corrected for the volume changes because of the titrated DNA.

The percentage of hypochromicity of the Soret band is calculated as follows:

$$\%\text{hypochromicity} = \left[\left(\varepsilon_{\text{free}} - \varepsilon_{\text{bond}}\right)\right] * 10, \quad (5)$$

$$\text{Where } \varepsilon_{\text{bond}} = \text{Abs TMPyP4}_{\text{bond}} \,/\, [\text{TMPyP4}]_{\text{bond}}$$

Data availability

The authors confirm that the data supporting the findings of this study are available within the article and its supporting information.

Supporting information—This article contains supporting information.

Acknowledgments—We thank Dr Arturo López-Castel and Dr Amit L. Deshmukh for comments on the article.

We thank Dr Vladimir Potaman for his helpful discussions and recommendations.
We also thank Dr Tina Saberi Safaei and Dr Terence Gall-Duncan for their valuable help.

Author contributions—B. Z. and C. E. P. conceptualization; B. Z. methodology; B. Z., M. H. A., and G. P. formal analysis; M. H. A. investigation; S. S. resources; B. Z., C. E. P., and M. H. A. writing–original draft; C. E. P. and M. H. A. visualization; B. Z. and C. E. P. supervision; B. Z. and C. E. P. project administration.

Funding and additional information—C. E. P. is supported by the Canadian Institutes of Health Research (grant nos.: FRN-148910 and FRN-173282), the Natural Sciences and Engineering Research Council of Canada (grant nos.: RGPIN-2016-08355 and RGPIN-2016-06355/498835), The Petroff Family Foundation, Tribute Communities, The Marigold Foundation, and the Fox Family Foundation. C. E. P. holds a Tier 1 Canada Research Chair in Disease-Associated Genome Instability.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: CANVAS, cerebellar ataxia, neuropathy, and vestibular areflexia syndrome; RFC1, replication factor C subunit 1; TDS, thermal difference spectra.

References

1. Szmulewicz, D. J., McLean, C. A., MacDougall, H. G., Roberts, L., Storey, E., and Halmagyi, G. M. (2014) CANVAS an update: clinical presentation, investigation and management. *J. Vestib. Res.* **24**, 465–474
2. Akcimen, F., Ross, J. P., Bourassa, C. V., Liao, C., Rochefort, D., Gama, M. T. D., *et al.* (2019) Investigation of the RFC1 repeat expansion in a Canadian and a Brazilian ataxia cohort: identification of novel conformations. *Front. Genet.* **10**, 1219
3. Cortese, A., Simone, R., Sullivan, R., Vandrovcova, J., Tariq, H., Yau, W. Y., *et al.* (2019) Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nat. Genet.* **51**, 649–658
4. Rafehi, H., Szmulewicz, D. J., Bennett, M. F., Sobreira, N. L. M., Pope, K., Smith, K. R., *et al.* (2019) Bioinformatics-based identification of expanded repeats: a non-reference intronic pentamer expansion in RFC1 causes CANVAS. *Am. J. Hum. Genet.* **105**, 151–165
5. Guilleminault, L., Chazelas, P., Melloni, B., Magdelaine, C., Villeneuve, T., Brouquières, D., *et al.* (2023) Repeat expansions of RFC1 in refractory chronic cough: a missing piece of the puzzle? *Chest* **163**, 911–915
6. Yang, B. E., and Moss, J. (2023) Cough, cough: how much of the symptom resides in your genes? *Chest* **163**, 746–748
7. Ronco, R., Perini, C., Currò, R., Dominik, N., Facchini, S., Gennari, A., *et al.* (2022) Truncating variants in RFC1 in cerebellar ataxia, neuropathy, and vestibular areflexia syndrome. *Neurology*. https://doi.org/10.1212/WNL.0000000000201486
8. Kejnovska, I., Kypr, J., Vondrusková, J., and Vorlickova, M. (2007) Towards a better understanding of the unusual conformations of the alternating guanine - adenine repeat strands of DNA. *Biopolymers* **85**, 19–27
9. Kejnovská, I., Tůmová, M., and Vorlícková, M. (2001) (CGA)(4): parallel, anti-parallel, right-handed and left-handed homoduplexes of a trinucleotide repeat DNA. *Biochim. Biophys. Acta* **1527**, 73–80
10. Kejnovska, I., Kypr, J., and Vorlickova, M. (2003) Circular dichroism spectroscopy of conformers of (guanine + adenine) repeat strands of DNA. *Chirality* **15**, 584–592
11. Krasilnikova, M. M., and Mirkin, S. M. (2004) Replication stalling at Friedreich's ataxia (GAA)n repeats *in vivo*. *Mol. Cell Biol.* **24**, 2286–2295
12. Belotserkovskii, B. P., Liu, R., and Hanawalt, P. C. (2009) Peptide nucleic acid (PNA) binding and its effect on *in vitro* transcription in friedreich's ataxia triplet repeats. *Mol. Carcinog* **48**, 299–308
13. Rider, S. D., Jr., Gadgil, R. Y., Hitch, D. C., Damewood, F. J., Zavada, N., Shanahan, M., *et al.* (2022) Stable G-quadruplex DNA structures promote replication-dependent genome instability. *J. Biol. Chem.* **298**, 101947
14. Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **5**, 182–186
15. Bugaut, A., and Balasubramanian, S. (2012) 5′-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.* **40**, 4727–4741
16. Piazza, A., Boulé, J. B., Lopes, J., Mingo, K., Largy, E., Teulade-Fichou, M. P., and Nicolas, A. (2010) Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae. *Nucleic Acids Res.* **38**, 4337–4348
17. Wu, Y., and Brosh, R. M., Jr. (2010) G-quadruplex nucleic acids and human disease. *FEBS J.* **277**, 3470–3488
18. Lipps, H. J., and Rhodes, D. (2009) G-quadruplex structures: *in vivo* evidence and function. *Trends Cell Biol.* **19**, 414–422
19. Chesler, M. (2003) Regulation and modulation of pH in the brain. *Physiol. Rev.* **83**, 1183–1221
20. Zamiri, B., Mirceta, M., Abu-Ghazalah, R., Wold, M. S., Pearson, C. E., and Macgregor, R. B., Jr. (2018) Stress-induced acidification may contribute to formation of unusual structures in C9orf72-repeats. *Biochim. Biophys. Acta Gen. Subj.* **1862**, 1482–1491
21. Lee, J. S., Johnson, D. A., and Morgan, A. R. (1979) Complexes formed by (pyrimidine)n . (purine)n DNAs on lowering the pH are three-stranded. *Nucleic Acids Res.* **6**, 3073–3091
22. Mirkin, S. M., and Frank-Kamenetskii, M. D. (1994) H-DNA and related structures. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 541–576
23. Ding, Y., Fleming, A. M., and Burrows, C. J. (2018) Case studies on potential G-quadruplex-forming sequences from the bacterial orders Deinococcales and Thermales derived from a survey of published genomes. *Sci. Rep.* **8**, 15679
24. LeProust, E. M., Pearso, C. E., Sinden, R. R., Gao, X., and Pearso, C. E. (2000) Unexpected formation of parallel duplex in GAA and TTC trinucleotide repeats of Friedreich's ataxia11Edited by I. Tinoco. *J. Mol. Biol.* **302**, 1063–1080
25. Chen, X., Mariappan, S. V., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoun, A., *et al.* (1995) Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 5199–5203
26. Mariappan, S. V., Catasti, P., Chen, X., Ratliff, R., Moyzis, R. K., Bradbury, E. M., and Gupta, G. (1996) Solution structures of the individual single strands of the fragile X DNA triplets (GCC)n.(GGC)n. *Nucleic Acids Res.* **24**, 784–792
27. Mariappan, S. V., Catasti, P., Silks, L. A., 3rd, Bradbury, E. M., and Gupta, G. (1999) The high-resolution structure of the triplex formed by the GAA/TTC triplet repeat associated with Friedreich's ataxia. *J. Mol. Biol.* **285**, 2035–2052
28. Vorlickova, M., Kejnovska, I., Bednarova, K., Renciuk, D., and Kypr, J. (2012) Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes. *Chirality* **24**, 691–698
29. Mergny, J. L., Li, J., Lacroix, L., Amrane, S., and Chaires, J. B. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.* **33**, e138
30. Lane, A. N., Chaires, J. B., Gray, R. D., and Trent, J. O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **36**, 5482–5515
31. Rippe, K., Fritsch, V., Westhof, E., and Jovin, T. M. (1992) Alternating d(G-A) sequences form a parallel-stranded DNA homoduplex. *EMBO J.* **11**, 3777–3786
32. Dolinnaya, N. G., and Fresco, J. R. (1992) Single-stranded nucleic acid helical secondary structure stabilized by ionic bonds: d(A(+)-G)10. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 9242–9246
33. Chaires, J. B. (2010) Human telomeric G-quadruplex: thermodynamic and kinetic studies of telomeric quadruplex stability. *FEBS J.* **277**, 1098–1106
34. Ortiz-Lombardia, M., Eritja, R., Azorin, F., Kypr, J., Tejralova, I., and Vorlickova, M. (1995) Divalent zinc cations induce the formation of two

distinct homoduplexes of a d(GA)20 DNA sequence. *Biochemistry* **34**, 14408–14415
35. Kypr, J., Kejnovska, I., Renciuk, D., and Vorlickova, M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **37**, 1713–1725
36. Beck, A., Vijayanathan, V., Thomas, T., and Thomas, T. J. (2013) Ionic microenvironmental effects on triplex DNA stabilization: cationic counterion effects on poly(dT).poly(dA).poly(dT). *Biochimie* **95**, 1310–1318
37. Leitner, D., and Weisz, K. (2000) Sequence-dependent stability of intramolecular DNA triple helices. *J. Biomol. Struct. Dyn.* **17**, 993–1000
38. Debin, A., Laboulais, C., Ouali, M., Malvy, C., Le Bret, M., and Svinarchuk, F. (1999) Stability of G,A triple helices. *Nucleic Acids Res.* **27**, 2699–2707
39. Fox, K. R. (1990) Long (dA)n.(dT)n tracts can form intramolecular triplexes under superhelical stress. *Nucleic Acids Res.* **18**, 5387–5391
40. Shimizu, M., Kubo, K., Matsumoto, U., and Shindo, H. (1994) The loop sequence plays crucial roles for isomerization of intramolecular DNA triplexes in supercoiled plasmids. *J. Mol. Biol.* **235**, 185–197
41. Potaman, V. N., and Soyfer, V. N. (1994) Divalent metal cations upon coordination to the N7 of purines differentially stabilize the PyPuPu DNA triplex due to unequal Hoogsteen-type hydrogen bond enhancement. *J. Biomol. Struct. Dyn.* **11**, 1035–1040
42. Frank-Kamenetskii, M. D., and Mirkin, S. M. (1995) Triplex DNA structures. *Annu. Rev. Biochem.* **64**, 65–95
43. Sandström, K., Wärmländer, S., Gräslund, A., and Leijon, M. (2002) A-tract DNA disfavours triplex formation. *J. Mol. Biol.* **315**, 737–748
44. Mitas, M., Yu, A., Dill, J., Kamp, T. J., Chambers, E. J., and Haworth, I. S. (1995) Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: (CTG)15. *Nucleic Acids Res.* **23**, 1050–1059
45. Gacy, A. M., Goellner, G., Juranić, N., Macura, S., and McMurray, C. T. (1995) Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*. *Cell* **81**, 533–540
46. Pearson, C. E., and Sinden, R. R. (1996) Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry* **35**, 5041–5053
47. Axford, M. M., Wang, Y.-H., Nakamori, M., Zannis-Hadjopoulos, M., Thornton, C. A., and Pearson, C. E. (2013) Detection of slipped-DNAs at the trinucleotide repeats of the myotonic dystrophy type I disease locus in patient tissues. *PLoS Genet.* **9**, e1003866
48. Napierala, M., Bacolla, A., and Wells, R. D. (2005) Increased negative superhelical density *in vivo* enhances the genetic instability of triplet repeat sequences. *J. Biol. Chem.* **280**, 37366–37376
49. Matos-Rodrigues, G., van Wietmarschen, N., Wu, W., Tripathi, V., Koussa, N. C., Pavani, R., *et al.* (2022) S1-END-seq reveals DNA secondary structures in human cells. *Mol. Cell* **82**, 3538–3552.e5
50. Fratta, P., Mizielinska, S., Nicoll, A. J., Zloh, M., Fisher, E. M. C., Parkinson, G., and Isaacs, A. M. (2012) C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci. Rep.* **2**, 1016
51. Reddy, K., Zamiri, B., Stanley, S. Y. R., Macgregor, R. B., Jr., and Pearson, C. E. (2013) The disease-associated r(GGGGCC)n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. *J. Biol. Chem.* **288**, 9860–9866
52. Haeusler, A. R., Donnelly, C. J., Periz, G., Simko, E. A. J., Shaw, P. G., Kim, M. S., *et al.* (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**, 195–200
53. Michel, D., Chatelain, G., Herault, Y., and Brun, G. (1992) The long repetitive polypurine/polypyrimidine sequence (TTCCC)48 forms DNA triplex with PU-PU-PY base triplets *in vivo*. *Nucleic Acids Res.* **20**, 439–443
54. Michel, D., Chatelain, G., Herault, Y., Harper, F., and Brun, G. (1993) H-DNA can act as a transcriptional insulator. *Cell Mol. Biol. Res.* **39**, 131–140
55. Sznajder Ł, J., Thomas, J. D., Carrell, E. M., Reid, T., McFarland, K. N., Cleary, J. D., *et al.* (2018) Intron retention induced by microsatellite expansions as a disease biomarker. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 4234–4239
56. Nussinov, R. (1988) Conserved quartets near 5' intron junctions in primate nuclear pre-mRNA. *J. Theor. Biol.* **133**, 73–84
57. Nussinov, R. (1987) (A)GGG(A), (A)CCC(A) and other potential 3' splice signals in primate nuclear pre-mRNA sequences. *Biochim. Biophys. Acta* **910**, 261–270
58. Nussinov, R. (1989) Conserved signals around the 5' splice sites in eukaryotic nuclear precursor mRNAs: G-runs are frequent in the introns and C in the exons near both 5' and 3' splice sites. *J. Biomol. Struct. Dyn.* **6**, 985–1000
59. Buratti, E., and Baralle, F. E. (2004) Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol. Cell Biol.* **24**, 10505–10514
60. Eperon, L. P., Graham, I. R., Griffiths, A. D., and Eperon, I. C. (1988) Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? *Cell* **54**, 393–401
61. Baud, A., Derbis, M., Tutak, K., and Sobczak, K. (2022) Partners in crime: proteins implicated in RNA repeat expansion diseases. *Wiley Inter. Rev. RNA* **13**, e1709
62. Malik, I., Kelley, C. P., Wang, E. T., and Todd, P. K. (2021) Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat. Rev. Mol. Cell Biol.* **22**, 589–607
63. Wang, X., Goodrich, K. J., Conlon, E. G., Gao, J., Erbse, A. H., Manley, J. L., and Cech, T. R. (2019) C9orf72 and triplet repeat disorder RNAs: G-quadruplex formation, binding to PRC2 and implications for disease mechanisms. *RNA* **25**, 935–947
64. Zamiri, B., Reddy, K., Macgregor, R. B., Jr., and Pearson, C. E. (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
65. Mori, K., Gotoh, S., Yamashita, T., Uozumi, R., Kawabe, Y., Tagami, S., *et al.* (2021) The porphyrin TMPyP4 inhibits elongation during the noncanonical translation of the FTLD/ALS-associated GGGGCC repeat in the C9orf72 gene. *J. Biol. Chem.* **297**, 101120
66. Simone, R., Balendra, R., Moens, T. G., Preza, E., Wilson, K. M., Heslegrave, A., *et al.* (2018) G-quadruplex-binding small molecules ameliorate C9orf72 FTD/ALS pathology in vitro and in vivo. *EMBO Mol. Med.* **10**, 22–31
67. Hasuike, Y., Tanaka, H., Gall-Duncan, T., Mehkary, M., Nakatani, K., Pearson, C. E., *et al.* (2022) CAG repeat-binding small molecule improves motor coordination impairment in a mouse model of Dentatorubral-pallidoluysian atrophy. *Neurobiol. Dis.* **163**, 105604
68. Nakamori, M., Panigrahi, G. B., Lanni, S., Gall-Duncan, T., Hayakawa, H., Tanaka, H., *et al.* (2020) A slipped-CAG DNA-binding small molecule induces trinucleotide-repeat contractions in vivo. *Nat. Genet.* **52**, 146–159
69. Erwin, G. S., Grieshop, M. P., Ali, A., Qi, J., Lawlor, M., Kumar, D., *et al.* (2017) Synthetic transcription elongation factors license transcription across repressive chromatin. *Science* **358**, 1617–1622
70. Burnett, R., Melander, C., Puckett, J. W., Son, L. S., Wells, R. D., Dervan, P. B., and Gottesfeld, J. M. (2006) DNA sequence-specific polyamides alleviate transcription inhibition associated with long GAA.TTC repeats in Friedreich's ataxia. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11497–11502
71. Cheng, A., Liu, C., Ye, W., Huang, D., She, W., Liu, X., *et al.* (2022) Selective C9orf72 G-quadruplex-binding small molecules ameliorate pathological signatures of ALS/FTD models. *J. Med. Chem.* **65**, 12825–12837
72. Hirayanagi, K., Ozaki, H., Tsukagoshi, S., Furuta, N., and Ikeda, Y. (2021) Porphyrins ameliorate spinocerebellar ataxia type 36 GGCCTG repeat expansion-mediated cytotoxicity. *Neurosci. Res.* **171**, 92–102
73. Satange, R., Rode, A. B., and Hou, M. H. (2022) Revisiting recent unusual drug-DNA complex structures: implications for cancer and neurological disease diagnostics and therapeutics. *Bioorg. Med. Chem.* **76**, 117094