JBC RESEARCH ARTICLE

Structural polymorphism of the nucleic acids in pentanucleotide repeats associated with the neurological disorder CANVAS

Received for publication, August 17, 2023, and in revised form, February 15, 2024 Published, Papers in Press, March 5, 2024,
https://doi.org/10.1016/j.jbc.2024.107138

Kenta Kudo[1,2,‡], **Karin Hori**[1,‡], **Sefan Asamitsu**[3], **Kohei Maeda**[1,2], **Yukari Aida**[1,2], **Mei Hokimoto**[1,2], **Kazuya Matsuo**[1], **Yasushi Yabuki**[1,2], **and Norifumi Shioda**[1,2,*]

From the [1]*Department of Genomic Neurology, Institute of Molecular Embryology and Genetics (IMEG), and* [2]*Graduate School of Pharmaceutical Sciences, Kumamoto University, Kumamoto, Japan;* [3]*Laboratory for Functional Non-coding Genomics, RIKEN Center for Integrative Medical Sciences (IMS), Yokohama, Japan*

Reviewed by members of the JBC Editorial Board. Edited by Karin Musier-Forsyth

Short tandem repeats are inherently unstable during DNA replication depending on repeat length, and the expansion of the repeat length in the human genome is responsible for repeat expansion disorders. Pentanucleotide AAGGG and ACAGG repeat expansions in intron 2 of the gene encoding replication factor C subunit 1 (*RFC1*) cause cerebellar ataxia, neuropathy, vestibular areflexia syndrome (CANVAS) and other phenotypes of late-onset cerebellar ataxia. Herein, we reveal the structural polymorphism of the *RFC1* repeats associated with CANVAS *in vitro*. Single-stranded AAGGG repeat DNA formed a hybrid-type G-quadruplex, whereas its RNA formed a parallel-type G-quadruplex with three layers. The RNA of the ACAGG repeat formed hairpin structure comprising C-G and G-C base pairs with A:A and GA:AG mismatched repeats. Furthermore, both pathogenic repeat RNAs formed more rigid structures than those of the nonpathogenic repeat RNAs. These findings provide novel insights into the structural polymorphism of the *RFC1* repeats, which may be closely related to the disease mechanism of CANVAS.

Short tandem repeats (STRs) are polymorphic repeat units of 1 to 6 base pairs that are abundant in eukaryotic genomes (1). STRs are inherently unstable during DNA replication depending on repeat length, and their expansion in the coding and noncoding regions across generations is the cause of almost 50 repeat expansion disorders, most of which primarily affect the nervous system (2). Thirteen STRs are known to cause repeat expansion disorders, and the structures of noncanonical DNA and its transcribed RNA, for example, slipped hairpins, G-quadruplexes (G4s), and triplexes, which form within repeat regions may induce cytotoxicity *via* abnormal genomic stability, transcription, epigenetics, mRNA translation, and RNA transport and decay (3–5).

Cerebellar ataxia, neuropathy, vestibular areflexia syndrome (CANVAS) is an autosomal recessive, slowly progressive, late-onset neurological disorder clinically characterized by the triad of cerebellar ataxia, bilateral vestibular hypofunction, and a peripheral sensory deficit (6, 7). Recently, a biallelic AAGGG repeat expansion $(AAGGG)_{exp}$ ranging from 400 to 2000 repeats in the gene encoding replication factor C subunit 1 (*RFC1*) was reported to cause CANVAS and other phenotypes of late-onset cerebellar ataxia (8, 9). In addition to the pathogenic $(AAGGG)_{exp}$ allele, two other pathogenic alleles were identified in the same *RFC1* locus: $(ACAGG)_{exp}$ in the Asia-Pacific and Japanese cohorts (10, 11) and $(AAAAG)_{10–25}(AAGGG)_{exp}(AAAGG)_{4–6}$ in the Māori-specific cohort (12). Remarkably, three different repeat conformations are observed in nonpathogenic populations: the reference $(AAAAG)_{11}$, $(AAAAG)_{exp}$ ranging from 15 to 200 repeats, and $(AAAGG)_{exp}$ ranging from 40 to 1000 repeats (Fig. 1*A*) (8). The cumulative allele frequency of $(AAAAG)_{exp}$ and $(AAAGG)_{exp}$ of approximately 20% in the general population is relatively high (8). A positive correlation between repeat length and G/C content is observed, whereas no correlation between repeat length and age of onset exists (10, 13). Thus, the onset of CANVAS may be due to the secondary structures of nucleic acids formed by the G/C-rich composition, independent of the repeat length. However, there are no studies on the secondary structures of nucleic acids associated with CANVAS.

Herein, we report the secondary structures of nucleic acids related to CANVAS using *in vitro* CD and UV-visible spectroscopy and thioflavin T (ThT) fluorescence and footprinting assays. Additionally, based on the results, the higher-order structures of nucleic acids in pathogenic repeats are simulated *via* molecular modeling. Single-stranded AAGGG repeat DNA and its RNA, respectively, form hybrid- and parallel-type G4s, and furthermore, ACAGG repeat RNA forms a unique slipped-hairpin structure. Conversely, none of the nonpathogenic repeats form characteristic nucleic acid secondary structures. The disease-specific formation of noncanonical nucleic acid structures may be the underlying molecular cause of the pathology of CANVAS.

‡ These authors contributed equally to this work.
* For correspondence: Norifumi Shioda, shioda@kumamoto-u.ac.jp.

© 2024 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Results

CD of the DNAs in RFC1 pentanucleotide repeats

We first investigated the topological unimolecular DNA structures of 20-mer single-stranded nonpathological $d(AAAAG)_4$ and $d(AAAGG)_4$, pathological $d(AAGGG)_4$ and $d(ACAGG)_4$, and their complementary $d(CTTTT)_4$, $d(CCTTT)_4$, $d(CCCTT)_4$, and $d(CCTGT)_4$ oligonucleotides using CD. It was performed at physiological pH, that is, pH 7.5.

Single-stranded guanine-rich sequences may form G4s, which are four-stranded nucleic acid structures comprising stacked G-quartets, that is, planar arrays of four guanines stabilized by Hoogsteen base pairing involving the N at position 7 of guanine (14). In the presence of 100 mM K^+, which stabilizes G4s, $d(AAGGG)_4$ forms a well-defined hybrid-type G4 structure (15), resulting in a CD spectrum containing positive peaks at approximately 295 and 260 nm. In contrast, in the presence of G4-nonstabilizing 100 mM Li^+, the entire spectrum changes, and the typical features are not observed (Fig. 1*B*). To further confirm the G4 folding of $d(AAGGG)_4$, we assessed the cation-dependent thermal denaturing by measuring the melting temperature (T_m). We measured CD spectra of $d(AAGGG)_4$ at different temperatures and confirmed that CD signals at 296 nm reflect the temperature-dependent unfolding of the G4 structure (Fig. S1*A*). In the thermal denaturation assay at 296 nm, $d(AAGGG)_4$ exhibited a much higher thermal stability under K^+ conditions than that under Li^+ conditions (ΔT_m = 10.4 °C) (Fig. S1*B*). No characteristic spectrum of the other pathological $d(ACAGG)_4$ is observed in the presence of K^+ or Li^+ (Fig. 1*C*).

Homopurine/homopyrimidine repeats, such as AAGGG/CCCTT repeat, may adopt intramolecular triplex structures (16). Triplexes are classified into two types according to the base composition and binding direction of the triplex-forming oligonucleotide (TFO) against the duplex DNA: pyrimidine and purine triplexes. Pyrimidine triplexes comprise TA-T and protonated CG-C^+ triplets, which require a weakly acidic environment for formation, whereas purine triplexes, which comprise CG-G and TA-A triplets, are stable at physiological pH in the presence of Mg^{2+} (17). Herein, we investigated whether purine triplexes form at physiological pH 7.5 with Mg^{2+} using the gel shift assay. Using single-stranded $d(AAGGG)_4$ as the TFO and fluorescein-labeled double-stranded $d(AAGGG/CCCTT)_4$, a triplex is detected in the presence of Li^+, depending on the TFO concentration (Fig. 1*D*, left). In the presence of K^+, where the TFO single-stranded $d(AAGGG)_4$ forms a G4, triplex formation is reduced compared to that under the Li^+ conditions (Fig. 1*D*, right). A G4 may be more likely to form than a triplex under near-physiological conditions when K^+ is present.

A cytosine-rich DNA sequence forms an i-motif, which is a structure comprising two intercalated parallel-stranded duplexes held together by hemi-protonated C:C^+ base pairs (18). However, the CD spectra of the complementary cytosine-rich single-stranded $d(CCCTT)_4$ and $d(CCTGT)_4$ display positive peaks at 275 nm (Fig. 1*E*). No CD profile features of the i-motif structure are observed, with positive and negative peaks at 288 and 264 nm, respectively (19).

In the CD spectra of nonpathological adenine-rich single-stranded $d(AAAAG)_4$ and $d(AAAGG)_4$ display strong positive peaks at 220 nm, and those of thymine-rich single-stranded $d(CTTTT)_4$ and $d(CCTTT)_4$ display positive peaks at 275 nm (Fig. 1*F*). Consistent with these results, previous studies report that the CD spectra of single-stranded poly(dA) and poly(dT) exhibit corresponding peaks at 220 and 275 nm, and correlations between these signal intensities and the total contents of adenine and thymine are observed, independent of the presence of secondary structures (19, 20).

G4 conformation of AAGGG repeat DNA

ThT binds preferentially to G4s and emits a strong fluorescence at a low background (21). To confirm the possibility that the AAGGG repeat forms a G4, we used DNA oligonucleotides to measure the intensity of ThT fluorescence in the presence of K^+. Consistent with the result of CD (Fig. 1*B*), a drastic increase in ThT fluorescence is detected for $d(AAGGG)_4$ compared to those of other oligonucleotides (Fig. 2*A*).

To elucidate the G4 formation of $d(AAGGG)_4$ in detail, we performed dimethyl sulfate (DMS) footprinting and identified the guanines involved in the G-quartets based on their resistances to methylation at position N^7 due to Hoogsteen hydrogen bonding. In the presence of K^+, all guanines (G3–G5, G8–G10, G13–G15, and G18–G20) are more protected against DMS methylation and subsequent piperidine cleavage than in the presence of Li^+ (Fig. 2*B*). The results of CD (Fig. 1*B*) and the ThT assay (Fig. 2*A*), in addition to the DMS footprinting pattern, suggest G4 formation by $d(AAGGG)_4$, where four tracts of three guanines form four strands with three two-base loops comprising adenine-adenine (AA) (Fig. 2*C*).

CD of the RNAs in RFC1 pentanucleotide repeats

Similar to CD of the DNAs, the topological unimolecular RNA structures were identified in CD using the nonpathological, pathological, and their complementary oligonucleotides under physiological conditions at pH 7.5.

In the presence of K^+, $r(AAGGG)_4$ forms a typical parallel-stranded G4, exhibiting positive and small negative peaks at 262 and 240 nm, respectively, and a positive band at 210 nm (15), which is not observed under Li^+ conditions (Fig. 3*A*, left). To investigate the higher-order RNA structures with more pathological repeat lengths, we performed CD using $r(AAGGG)_{11}$, and the spectrum displays a peak more characteristic of a parallel-type G4 compared to that in the spectrum of $r(AAGGG)_4$ (Fig. 3*A*, right). CD melting curves were measured as a function of temperature and identified 265 nm as the wavelength of their maximum Cotton effects (Fig. S2*A*). The stabilizing effects of K^+ on $r(AAGGG)_{11}$ (ΔT_m = 41.0 °C) and $r(AAGGG)_4$ (ΔT_m = 44.0 °C) compared to those of Li^+ were observed at 265 nm (Fig. S2*B*). In the CD spectrum of ACAGG repeat RNA, no characteristic peaks representing

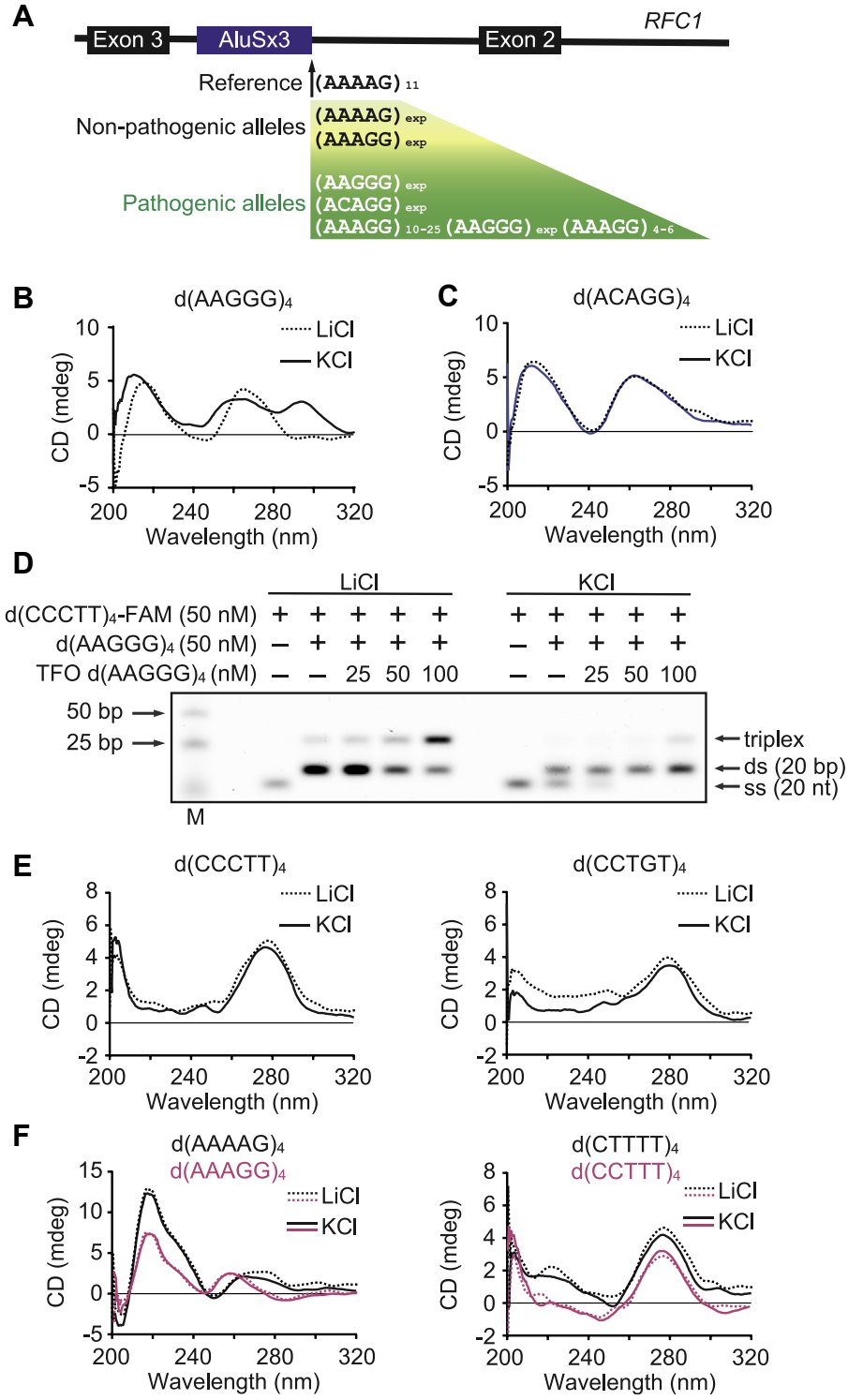

Figure 1. CD spectra of the DNAs in *RFC1* pentanucleotide repeats. *A*, schematic diagram showing the polymorphism at the *RFC1* repeat locus, consisting of nonpathogenic and pathogenic alleles. See also (7). *B* and *C*, representative CD spectra of the pathogenic $d(AAGGG)_4$ and $d(ACAGG)_4$ in the presence of 100 mM KCl (*solid line*) or LiCl (*dotted line*) at 25 °C and pH 7.5. *D*, gel mobility shift assays of a purine triplex at physiological pH 7.5 in the presence of Mg^{2+} and 100 mM LiCl (*left*) or KCl (*right*). *E*, representative CD spectra of the pathogenic complementary $d(CCCTT)_4$ and $d(CCTGT)_4$ in the presence of 100 mM KCl (*solid line*) or LiCl (*dotted line*) at 25 °C and pH 7.5. *F*, representative CD spectra of nonpathogenic $d(AAAAG)_4$ and $d(CTTTT)_4$ (*black lines*) and $d(AAAGG)_4$ and $d(CCTTT)_4$ (*magenta lines*) in the presence of 100 mM KCl (*solid line*) or LiCl (*dotted line*) at 25 °C and pH 7.5. All experiments were repeated at least twice, and the results were similar. ds, double strand; M; marker; *RFC1*, replication factor C subunit 1; ss; single strand.

$r(ACAGG)_4$ are observed. Conversely, the positive and negative CD absorption bands of $r(AAGGG)_{11}$ are detected as a maximum at ~260 nm, a minimum at ~210 nm, and a small negative peak between 290 and 300 nm under Li^+ and K^+ conditions (Fig. 3*B*). These peaks are characteristic of the A-form duplex RNA conformation (22).

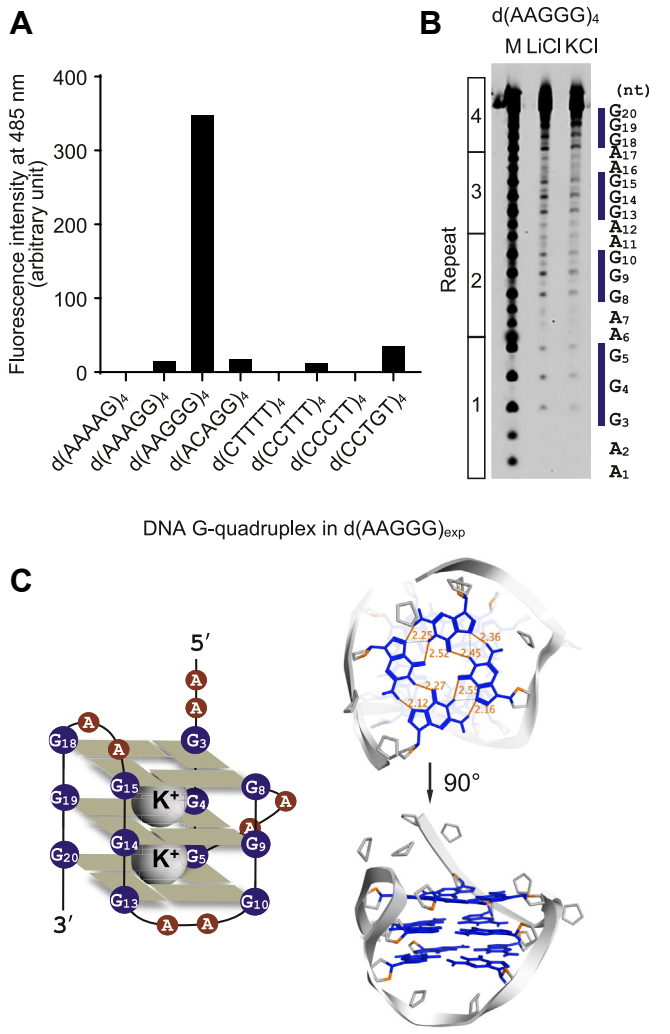

Figure 2. G4 conformation of AAGGG repeat DNA. *A*, *bar graph* of the fluorescence intensities of ThT at 485 nm in the presence of various DNA oligonucleotides. Samples were prepared at 2.5 μM in 10 mM Tris/HCl at pH 7.5 with 100 mM KCl, and ThT was added to yield a concentration of 0.2 μM. Experiments were repeated at least twice, and the results were similar. *B*, DMS footprinting of $d(AAGGG)_4$ oligonucleotides in the presence of 100 mM K^+ or Li^+. The sequence of $d(AAGGG)_4$ is shown to the *right* of the gel. Experiments were repeated at least twice, and the results were similar. *C*, schematic and 3D diagrams of the G4 conformation of $d(AAGGG)_4$, based on the results of the biophysical analyses (calculated lengths of hydrogen bonds of Hoogsteen base pairs range from 2.12 to 2.55 Å). dimethyl sulfate; G-quadruplex; M; marker; thioflavin T.

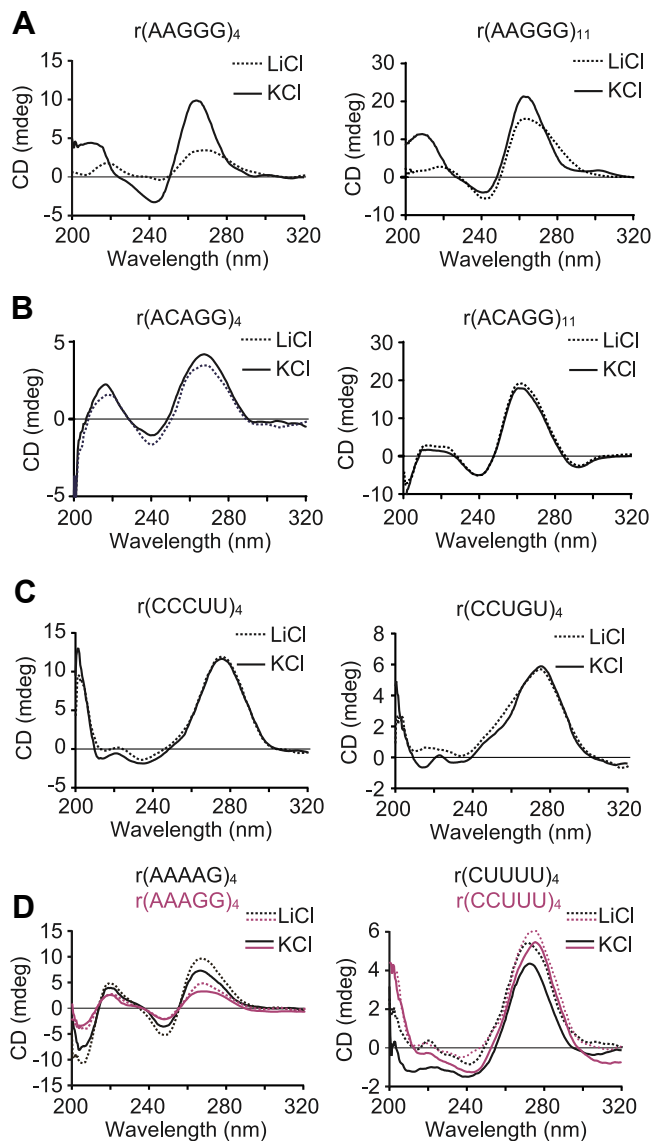

Figure 3. CD spectra of the RNAs in *RFC1*-derived pentanucleotide repeats. *A* and *B*, representative CD spectra of pathogenic $r(AAGGG)_4$, $r(AAGGG)_{11}$, $r(ACAGG)_4$, and $r(ACAGG)_{11}$ in the presence of 100 mM KCl (*solid lines*) or LiCl (*dotted lines*) at 25 °C and pH 7.5. *C*, representative CD spectra of the pathogenic complementary $r(CCCUU)_4$ and $r(CCUGU)_4$ in the presence of 100 mM KCl (*solid lines*) or LiCl (*dotted lines*) at 25 °C and pH 7.5. *D*, representative CD spectra of nonpathogenic $r(AAAAG)_4$ and $r(CUUUU)_4$ (*black lines*) and $r(AAAGG)_4$ and $r(CCUUU)_4$ (*magenta lines*) in the presence of 100 mM KCl (*solid line*) or LiCl (*dotted line*) at 25 °C and pH 7.5. All experiments were repeated at least twice, and the results were similar. *RFC1*, replication factor C subunit 1.

None of the CD spectra of the complementary $r(CCCUU)_4$ and $r(CCUGU)_4$ oligonucleotides display characteristic peaks indicative of RNA secondary structures (Fig. 3*C*). In terms of the nonpathological oligonucleotides, the CD spectra of the adenine-rich single strands $r(AAAAG)_4$ and $r(AAAGG)_4$ display strong positive and small negative peaks at 265 and 250 nm, respectively (Fig. 3*D*, left), which are consistent with the tendency of poly(A), based on its CD spectrum, to form unique single-stranded helical structures (23). The CD spectra of uridine-rich single-stranded $r(CUUUU)_4$ and $r(CCUUU)_4$ display small positive peaks at 275 nm (Fig. 3*D*, right), indicating the absence of complex structures, similar to poly(U) (23).

Analyses of the RNA structures of the pathological repeats related to CANVAS

The intensity of ThT fluorescence was then measured to investigate the possibility of G4 formation using 20-mer RNAs with nonpathological, pathological, and their complementary oligonucleotides. A drastic increase in ThT fluorescence is observed for $r(AAGGG)_4$ compared to those of the other repeat oligonucleotides (Fig. 4*A*).

To identify more detailed RNA structures, we conducted RNase T1 footprinting of pathological AAGGG and ACAGG and nonpathological AAAAG repeats. As RNase T1 displays

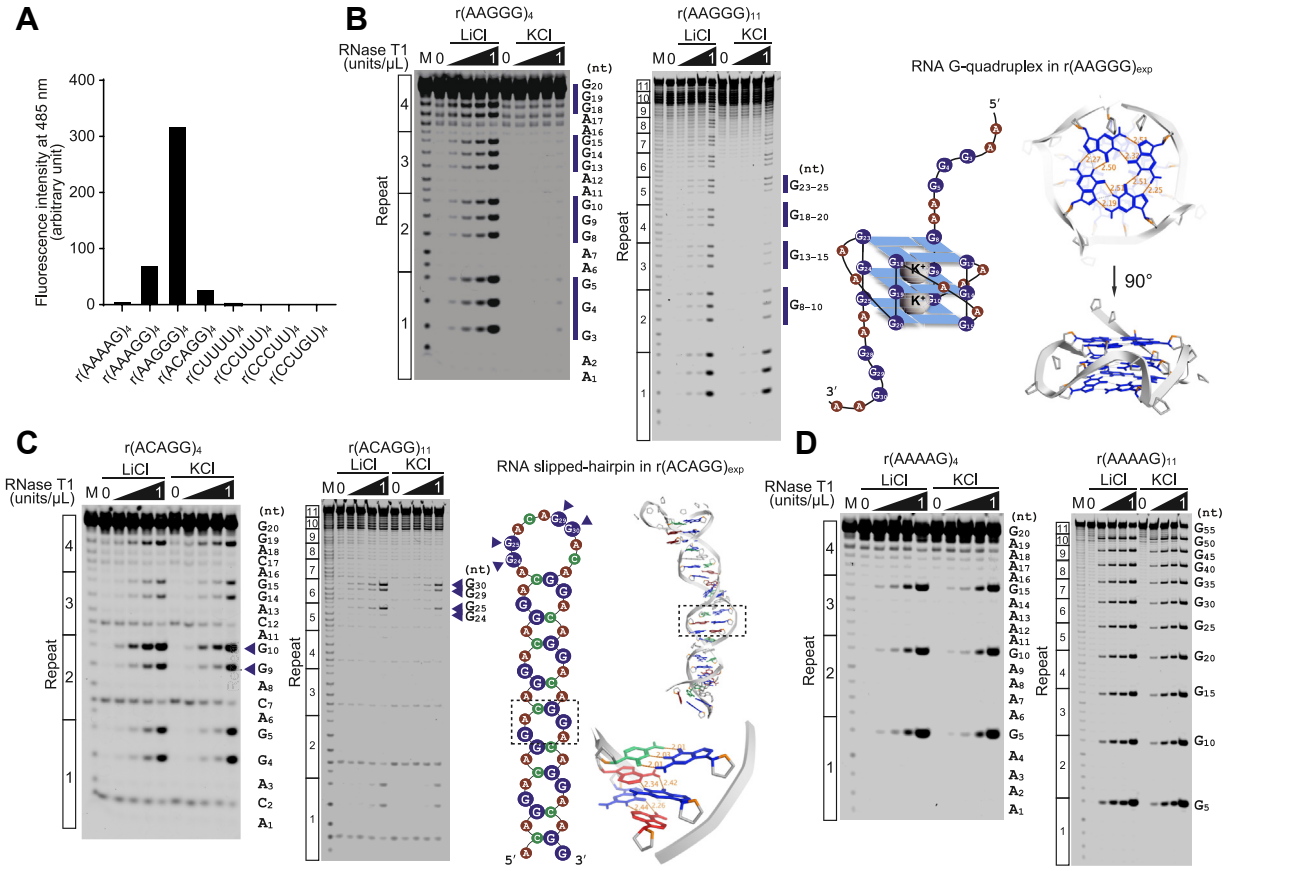

Figure 4. Secondary structures of the pathogenic repeat RNAs related to CANVAS. *A*, *bar graph* of the fluorescence intensities of ThT at 485 nm in the presence of various RNA oligonucleotides. Samples were prepared at 2.5 μM in 10 mM Tris/HCl at pH 7.5 with 100 mM KCl, and ThT was added to yield a concentration of 0.2 μM. *B*, patterns of RNase T1 cleavage of $r(AAGGG)_4$ (*left*) and $r(AAGGG)_{11}$ (*middle*). The concentrations of RNase T1 were 0, 0.25, 0.5, 0.75, and 1 units/μl. The *vertical blue lines* show the differences in the cleavage patterns obtained in the presence of LiCl and KCl. Schematic and 3D diagrams of the G4 conformation of $r(AAGGG)_4$, based on the results of the biophysical analyses (calculated lengths of hydrogen bonds of Hoogsteen base pairs range from 2.19 to 2.51 Å, *right*). *C*, patterns of RNase T1 cleavage of $r(ACAGG)_4$ (*left*) and $r(ACAGG)_{11}$ (*middle*). The *blue triangles* indicate strongly truncated guanosines that may be involved in loop formation. Schematic and 3D diagrams of the hairpin conformation of $r(ACAGG)_{11}$, based on the results of the biophysical analyses. The G-C and G:A base pairs are putatively formed (calculated lengths of hydrogen bonds of G-C or G:A base pairs range from 2.01 to 2.03 Å or 2.34 to 2.44 Å, respectively, *right*). *D*, patterns of RNase T1 cleavage of $r(AAAAG)_4$ (*left*) and $r(AAAAG)_{11}$ (*right*). All experiments were repeated at least twice, and the results were similar. cerebellar ataxia, neuropathy, vestibular areflexia syndrome; G-quadruplex; M; marker; thioflavin.

specificity for guanine residues in ssRNA, guanosines in loops and unstructured regions are more susceptible to cleavage than those directly involved in G4 and hairpin formation (24). In $r(AAGGG)_4$, all 12 guanosines (G3–G5, G8–G10, G13–G15, G18–G20) are protected against RNase T1 cleavage in the presence of K^+ (Fig. 4*B*, left). Based on the RNase footprinting pattern of $r(AAGGG)_{11}$ (Fig. 4*B*, middle) and the results of CD (Fig. 3*A*) and the ThT assay (Fig. 4*A*), the pathological AAGGG repeats displayed a strong tendency to fold into an intramolecular parallel-form G4, which comprises three G-quartets with three two-base loops comprising AA (Fig. 4*B*, right). Conversely, the guanosines in $r(ACAGG)_4$ are cleaved, regardless of the monovalent cation added, and the intensities of guanosines G9 and G10 are slightly increased in the cleavage pattern (Fig. 4*C*, left). In the RNase footprinting pattern of $r(ACAGG)_{11}$, four guanosines (G24, G25, G29, G30) predicted to occur in the single-stranded loop region are cleaved by RNase T1 (Fig. 4*C*, middle). RNA secondary structure prediction using deep learning based on MXfold2 (Keio Universit) (25) reveals that the stem structure formed by repeats of ACAGG on the 5′ and 3′ sides comprises C-G and G-C base pairs and A:A and GA:AG mismatch repeats (Fig. 4*C*, right). This is consistent with the results obtained, including those of CD (Fig. 3*B*). The energy-minimized

Table 1
Experimentally obtained thermodynamic parameters of the RNA oligonucleotides

Oligonucleotides	T_m (°C)	ΔH° (kcal/mol)	ΔS° (cal/mol*K)	ΔG° (kcal/mol)
$r(AAAAG)_{11}$	N.D.	N.D.	N.D.	N.D.
$r(ACAGG)_{11}$	38.53 ± 0.14	−90.13 ± 8.39	−289.26 ± 27.13	−0.46 ± 0.05
$r(AAGGG)_{11}$	65.07 ± 0.86	−50.09 ± 4.32	−148.40 ± 12.20	−4.08 ± 0.54

Experiments were repeated thrice. These measurements were performed under 10 mM KCl condition.
N.D., not detected.

structure of the r(ACAGG)$_{11}$ hairpin conformation suggests that the hydrogen bonds of the G-C and G:A base pairs contribute to the stability of the structure, whereas the A:A mismatch does not form hydrogen bonds (Fig. 4*C*, right). In the nonpathological AAAAG repeat RNA, the single guanosine is cleaved by RNase T1 to the same extent in 4 and 11 repeats, indicating that r(AAAAG)$_{exp}$ may not form a characteristic secondary structure (Fig. 4*D*).

To determine the differences in the structural stabilities of these RNAs, we performed UV-monitored melting studies. In the presence of 100 mM KCl, we were not able to draw the unfolding baseline of r(AAGGG)$_{11}$, which is required for calculating thermodynamic parameters (Fig. S3*A*). Therefore, we used 10 mM KCl conditions to be able to observe the entire melting behaviors of RNAs (Fig. S3*B*). The obtained results confirmed that r(AAGGG)$_{11}$ and r(ACAGG)$_{11}$ form stable structures, with respective changes in folding free energy of $\Delta G37\ ^{\circ}C = -4.08 \pm 0.54$ and -0.46 ± 0.05 kcal/mol, whereas that of r(AAAAG)$_{11}$ is not detected (Table 1). In addition, we calculated the thermal parameters corresponding to RNA hairpin structures using a MeltR method (26) to check the validity of our calculations and confirmed a similar tendency to values depicted in Table 1; T_m, ΔH°, ΔS°, and ΔG° for r(ACAGG)$_{11}$ were 38.45 ± 0.05 °C, −81.01 ± 10.52 kcal/mol, −260.27 ± 33.52 cal/mol*K, and −0.42 ± 0.05 kcal/mol, respectively.

Discussion

We demonstrated the structural polymorphism of the *RFC1* repeats associated with CANVAS. The single-stranded AAGGG repeat DNA and its RNA, respectively, formed hybrid- and parallel-type G4s, and the RNA of the ACAGG repeat formed a unique slipped-hairpin structure. Furthermore, both pathogenic repeat RNAs formed rigid structures compared to those of the nonpathogenic repeats. These novel findings regarding structural polymorphism in *RFC1* repeat extensions may be relevant to the disease mechanism of CANVAS.

As G4s are formed by the stacking of at least two G-quartets, one possibility was that the ACAGG repeats in DNA and RNA could also form G4s based on the consensus motif ($G_{2-3} + N_{1-7}G_{2-3} + N_{1-7}G_{2-3} + N_{1-7}G_{2-3}$) (4, 27). However, ACAGG repeats in RNA could not form G4s, instead forming unique slipped hairpin conformations, with the C-G and G-C base pairs combined with the A:A and GA:AG mismatched pairs. This is likely because G4 stability depends on several factors, including the lengths of the loops, its base composition, and the number of G tracts (28). The stability of a G4 in RNA decreases with increasing loop length, particularly with two stacked G-quartets with the length of each loop longer than three residues, which is structurally unstable (29). Additionally, the presence of cytosine relative to adenine and uridine leads to instability in a G4 with two G-quartets, because G-C Watson-Crick base pairs (30) predominate at high cytosine contents in the loops (29). Furthermore, in ACAGG repeat RNA, the double helical structure of the A form may manifest, with sheared G:A and A:G base pairing in the loops (31, 32), similar to that of the analogous base pairs in hammerhead ribozymes (33). This study suggests that the combination of these factors results in ACAGG repeats with high energies, forming a rigid secondary structure not observed with nonpathogenic repeats.

We showed that neither single-stranded d(AAGGG)$_4$ with double-stranded d(AAGGG/TTCCC)$_4$ nor complementary pyrimidine-rich single-stranded d(CCCTT)$_4$ formed triplex and i-motif structures under physiological conditions in the presence of K^+. However, due to technical difficulties in preparing G-rich long single-stranded DNA, the repeat length-dependent structure could not be fully investigated. The formation of nucleic acid structures also varies depending on conditions such as molecular crowding and pH, and the stabilities of DNA in a triplex and i-motif vary with oligonucleotide length and acidic pH (34, 35). Investigating the formation of these noncanonical DNA structures with various repeat sizes under different conditions is necessary.

The structures of noncanonical DNA and its transcribed RNA that form during repeat expansion disorders are central to the molecular mechanisms of toxicity *via* gain of function (GOF) or loss of function (LOF) (5, 36). C9orf72 amyotrophic lateral sclerosis and frontotemporal dementia (C9ALS/FTD) is caused by GGGGCC repeat expansions (d(G_4C_2)$_{exp}$) in the *C9orf72* gene (37, 38). In GOF mechanism, transcribed r(G_4C_2)$_{exp}$ forms G4s and disrupts neuronal function *via* various mechanisms by RNA-binding protein (RBP) sequestration through RNA foci (39). Additionally, sense r(G_4C_2)$_{exp}$ and antisense r(C_4G_2)$_{exp}$ transcribed RNAs are translated in all reading frames into dipeptide repeat proteins *via* repeat-associated non-AUG (RAN) translation (40–42), resulting in these RAN proteins forming toxic aggregates (43–45). In LOF mechanism, G4-forming d(G_4C_2)$_{exp}$ inhibits transcription, reducing C9orf72 protein levels (43). Fragile X-associated tremor/ataxia syndrome is a neurodegenerative disease caused by d(CGG)$_{exp}$ in the *FMR1* gene (46). Transcribed r(CGG)$_{exp}$ interacts with several RBPs that form RNA foci and induce cellular dysfunction (47). Transcribed r(CGG)$_{exp}$ also produces the toxic protein FMRpolyG *via* RAN translation (48), and G4-forming r(CGG)$_{exp}$ induces the liquid-solid phase transition of FMRpolyG, causing neurological dysfunction (49). Slipped-hairpin–forming d(CAG/CTG)$_{exp}$ is a particularly common cause of repeat expansion disease (2). Transcribed r(CAG)$_{exp}$ within coding genes is translated into polyglutamine tracts, inducing polyglutamine toxicity (50). Transcribed r(CAG)$_{exp}$ and r(CUG)$_{exp}$ adopt highly stable hairpin structures that form nuclear foci and induce RNA toxicity (3), and the transcripts may also induce cell death *via* RAN translation (40, 51). Additionally, r(CAG)$_{exp}$-derived RAN translation products coaggregate with r(CAG)$_{exp}$ foci within the cytoplasm (52).

Although it is currently unclear how the repeat expansions in the gene encoding *RFC1* contribute to the pathogenesis of CANVAS, both GOF and LOF hypotheses are being considered. Three not mutually exclusive mechanisms have been proposed in CANVAS. First, GOF might result from repeat RNA foci sequestering functionally important RBPs. A second

ASBMB

GOF mechanism is based on the toxic peptides generated by RAN translation of the AAGGG, ACAGG, or their antisense repeat RNAs. Third, LOF mechanism might result from reduced RFC1 protein levels. Compound heterozygotic CANVAS patients in monoallelic *RFC1* expansions combined with LOF mutations of *RFC1* (truncation, nonsense, and splicing) in the nonexpanded allele (53–55) suggest that LOF mechanism might underlie the pathogenesis. *RFC1* encodes the largest subunit of replication factor C, the complex responsible for DNA replication and repair (56). A recent study reported that pathogenic AAGGG repeat forms a triplex *in vitro*, and stalls replication occurring only when AAGGG serves as the template strand (57), indicating non-B DNA-forming potential of the pathogenic repeat in *RFC1* during DNA replication results in fork stalling. However, no reductions in *RFC1* mRNA and protein levels have been observed in peripheral tissues, postmortem brain samples, and induced pluripotent stem cell–derived neurons from patients with biallelic $(AAGGG)_{exp}$ in the *RFC1* (8, 58). In the aspect of GOF hypotheses, RNA foci were not detected in patient brains in the initial analysis (8), but in a recent report, CCTGT- and CCCTT-containing RNA foci were identified in the neuronal nuclei of patients with CANVAS harboring biallelic $(ACAGG)_{exp}$ and $(AAGGG)_{exp}$, respectively (59). Both AAGGG repeat-RNA foci and CCCTT repeat-RNA foci were detected in control and CANVAS patient induced pluripotent stem cell–derived neurons, but the overall abundance of foci was low and the specificity was incomplete (58). In RAN translation, AAGGG repeat–derived KGREG peptide products were selectively detected within cerebellar granule cells in CANVAS patients, with no signal detected in controls (58). In the relationship between the nucleic acid conformation formed by pathogenic repeats and CANVAS pathology, lower motor neurons are more susceptible in patients exhibiting $(ACAGG)_{exp}$ than $(AAGGG)_{exp}$ (60), indicating that lower motor neurons may be more susceptible to ACAGG-forming hairpin structure than AAGGG-forming G4. These results were obtained using a few samples, and thus, further studies are necessary to obtain accurate results regarding the LOF and GOF mechanisms related to nucleic acid structures.

In conclusion, we identified the DNA and RNA secondary structures of pathological repeats associated with CANVAS *in vitro*. The pathological AAGGG repeat DNA and RNA formed G4s, and ACAGG repeat RNA formed a unique slipped-hairpin structure. These disease-specific rigid structures may aid in elucidating the pathological mechanism of CANVAS.

Experimental procedures

Oligonucleotides

The following DNA and RNA oligonucleotides were synthesized by Hokkaido System Science: $d(AAAAG)_4$, $d(AAAGG)_4$, $d(AAGGG)_4$, $d(ACAGG)_4$, $d(CTTTT)_4$, $d(CCTTT)_4$, $d(CCCTT)_4$, $d(CCTGT)_4$, 5′ 6-carboxyfluorescein (FAM)-$d(CCCTT)_4$, $r(AAAAG)_4$, $r(AAAGG)_4$, $r(AAGGG)_4$, $r(ACAGG)_4$, $r(CTTTT)_4$, $r(CCTTT)_4$, $r(CCCTT)_4$, $r(CCTGT)_4$, 5′ 6-FAM-$r(AAAAG)_4$, 5′ 6-FAM-$r(AAGGG)_4$, and 5′ 6-FAM-$r(ACAGG)_4$; and Ajinomoto Bio-Pharma Services: 5′ 6-FAM-$r(AAAAG)_{11}$, 5′ 6-FAM-$r(AAGGG)_{11}$, and 5′ 6-FAM-$r(ACAGG)_{11}$.

Circular dichroism

Oligonucleotides (2.5 μM) were prepared in 10 mM Tris/HCl (pH 7.5) buffer containing 100 mM KCl or LiCl. The oligonucleotides were heated at 95 °C and gradually cooled to room temperature over 1.5 h prior to use. CD was conducted at 25 °C in the range 200 to 400 nm using a J-1100 spectrometer (JASCO) and a quartz cuvette with a 3-mm path length. In the denaturing assays, CD spectra were acquired in the 200 to 400 nm range at 5 °C intervals from 20 °C to 95 °C, and the CD signal of each oligonucleotide at 265 (RNA) or 296 nm (DNA) was monitored. Temperature scans were performed continuously from 20 °C to 95 °C at 1 °C/min, and the T_m values were determined as the temperature at half of the maximum decrease in the signal.

ThT assays

Oligonucleotides (2.5 μM) were prepared in 10 mM Tris/HCl (pH 7.5) containing 100 mM KCl, and they were heated at 95 °C and gradually cooled to room temperature over 1.5 h prior to use. ThT (0.2 μM) was added to each sample, which was then incubated for 2 h at room temperature using a Thermo-Shaker (Bio-Medical Science). The fluorescence spectra were recorded using a quartz cuvette with a 3-mm path length and an FP-8300 spectrometer (JASCO) at an excitation wavelength of 440 nm in the wavelength range 420 to 600 nm. The fluorescence signals of each oligonucleotide at 485 nm were used for comparison.

Gel shift assays

$d(AAGGG)_4$ and FAM-labeled $d(CCCTT)_4$ were prepared in 20 mM Tris/HCl (pH 7.5) containing 20 mM $MgCl_2$, 100 (or 0) mM LiCl, 100 (or 0) mM KCl, 2.5 mM spermidine, and $d(AAGGG)_4$ (TFO) in the same conditions. The oligonucleotides were heated at 95 °C and gradually cooled to room temperature over 1.5 h prior to use. FAM-labeled duplex DNA (50 nM) was incubated with increasing concentrations of $d(AAGGG)_4$ (TFO) in the buffer for 12 h at 37 °C. Electrophoresis was performed at 4 °C with tris-borate-EDTA (TBE)-urea polyacrylamide (29:1 acrylamide:bisacrylamide) (w/v) gel containing 20 mM $MgCl_2$ at 100 V for 3 h in 1 × TBE running buffer. DNAs labeled with FAM on the gel were visualized using an Amersham TYPHOON (Cytiva).

DMS footprinting

FAM-labeled DNA (0.5 μM) was prepared in 10 mM Tris/HCl (pH 7.5) containing 100 mM KCl or LiCl. The DNA oligomers were heated at 95 °C and gradually cooled to room temperature over 1.5 h prior to use. Each sample was incubated with 10% DMS in 50% ethanol for 10 min on ice. The reaction was stopped with a stop solution (0.1 M mercaptoethanol, 0.5 M sodium acetate, and 50 μg/ml calf thymus DNA). Samples purified *via* isopropanol precipitation were incubated

with 10% piperidine at 95 °C for 5 min and then dried using a CC-105 centrifugal concentrator (TOMY Digital Biology). The resulting residues were washed with RNase-free H_2O, redried, and dissolved in a loading solution (80% formamide and 10 mM NaOH). Hydroxyl radical reactions were used to generate a marker for DMS footprinting (61). DNA was adjusted to a final concentration of 1.5 μM, and the following three solutions were added in equal volumes (25 mM sodium ascorbate, 1.2% H_2O_2 and iron(II)EDTA (0.4 mM Fe(II)/0.8 mM EDTA). The reaction was stopped with 0.2 M thiourea and was then purified by phenol-chloroform and isopropanol reaction, followed by piperidine reaction as in the sample. Aliquots of the resultant solution were loaded onto a 15% TBE-urea polyacrylamide (19:1 acrylamide:bisacrylamide) (w/v) gel and electrophoresed at 1500 V for 2 h in 1 × TBE running buffer. DNAs labeled with FAM on the gels were visualized using the Amersham TYPHOON (Cytiva).

RNase T1 footprinting

FAM-labeled RNA oligomers (1 μM) were prepared in 10 mM Tris/HCl (pH 7.5) buffer containing 100 mM KCl or LiCl. The RNA oligomers were heated at 95 °C and gradually cooled to room temperature over 1.5 h prior to use. The RNA was treated with RNase T1 (Thermo Fisher Scientific) at room temperature for 3 min, and then a stop solution (8 M urea, 0.4 M 2-mercaptoethanol, and 0.5 M EDTA [pH 8]) was added to terminate the reaction. Also, FAM-labeled RNA oligomers (1 μM) were prepared in 50 mM $NaHCO_3/Na_2CO_3$ (pH 9.2) buffer containing 1 mM EDTA (pH 8), 0.25 mg/ml tRNA on ice. The RNA oligomers were heated at 90 °C for 2 min, and then a stop solution was added to terminate the reaction. Aliquots of the resultant solution were loaded onto a 15% TBE-urea polyacrylamide (19:1 acrylamide:bisacrylamide) (w/v) gel and electrophoresed at 1500 V for 2 h in 1 × TBE running buffer. The RNAs labeled with FAM on the gels were visualized using the Amersham TYPHOON (Cytiva).

UV-visible spectral assays

RNA oligonucleotides (1.25, 2.5, 5 μM) were prepared in 10 mM Hepes/NaOH (pH 7.5) buffer containing 100 mM or 10 mM KCl. The RNA oligonucleotides were heated at 95 °C and gradually cooled to room temperature over 1.5 h prior to use. The absorbance of each RNA oligonucleotide at 260 [$r(AAAAG)_{11}$, $r(ACAGG)_{11}$] or 295 nm [$r(AAGGG)_{11}$] was recorded using a V-750 spectrometer (JASCO) in a series of cell cuvettes with path lengths of 10 mm. Temperature scans were performed continuously from 20 °C to 95 °C at 1 °C/min, and the T_m values were determined using a two-point method in the melting analysis program of the Spectra Manager (https://jascoinc.com/products/spectroscopy/circular-dichroism/software/spectra-manager/) software (JASCO) (62).

Thermodynamic analysis

The thermodynamic parameters of the hairpin and G4 structures (2.5 μM) were determined according to a previously reported method (62, 63). The following methods all assume a two-state transition from single-stranded (unfolding) to folding. The folding ($A_F(T) = A_F(0) + S_F.T$) and unfolding ($A_U(T) = A_U(0) + S_U.T$) baselines were manually selected in Excel (Microsoft), where T represents the temperature in degrees Celsius (Fig. S4*A*).

Using the baselines, the equations representing the folded fractions (N_T) from absorbance (A) are as follows (Fig. S4*B*):

$$N_T = \left[A(T) - A_F(T)\right] / \left[A_U(T) - A_F(T)\right] \cdots \text{hairpin}$$

$$N_T = \left[A(T) - A_U(T)\right] / \left[A_F(T) - A_U(T)\right] \cdots \text{G4}$$

Under the two-transition state assumption, the equilibrium constants (K_a) are related to the folded fractions according to the following equations:

$$Ka = (1 - N_T)/N_T \cdots \text{hairpin}$$

$$Ka = N_T/(1 - N_T) \cdots \text{G4}$$

By definition, the Gibbs free enthalpy may be written as:

$$\Delta G^\circ = \Delta H^\circ - (T\Delta S^\circ) = -RTln(Ka)$$

where T, ΔH°, ΔS°, and R are the temperature in Kelvin, standard enthalpy of the reaction, standard entropy, and ideal gas constant, respectively.

The following van't Hoff equation:

$$\ln(Ka) = -\frac{\Delta H^\circ}{R}(1 / T) + \frac{\Delta S^\circ}{R}$$

yields an van't Hoff plot [ln(Ka) *versus* 1/T] that displays a straight line (Fig. S4*C*), with a respective slope and *y*-axis intercept of $-\frac{\Delta H^\circ}{R}$ and $\frac{\Delta S^\circ}{R}$, which afford ΔH° and ΔS°. ΔG° is calculated at 37 °C.

For hairpin structures, thermodynamic parameters were also calculated using MeltR to validate the above methods.

Molecular modeling

Molecular modeling was performed using the molecular operating environment (https://www.chemcomp.com/) software (Chemical Computing Group) and Amber94/99 force field parameters and a distance-dependent dielectric constant of ε = 4r (where r is the distance between two atoms) (64, 65). The initial structures of the d(AAGGG)4 and r(AAGGG)4 G4 conformations were constructed based on the crystal structure of the hybrid-type G4 (PDB ID: 7DJU) and d(TGGGT)4 G4 (PDB ID: 6p45) (66), respectively. The AA residues at the 5′ end of $d(AAGGG)_4$ and $r(AAGGG)_4$ were refined to be stacked with each other, in accordance with the prevalent stacking motif observed at the termini of G4 structures and recent literature elucidating the solution structure of an

AAGGG-derived G4 [ref]. These refinements were carried out using a minimization program, with tether deviation set to 1 and RMSD set to 0.1. Distance and dihedral-angle constraints were imposed on the guanine residues within the three G-tetrads during the process; specifically, O-H and N-H distances at the Hoogsteen hydrogen bonds were constrained to 2 to 2.5 Å, and C-N-N-C and C-O-N-C dihedral angles at the Hoogsteen hydrogen bonds were constrained to 0°.

For the hairpin conformations, 30 distinct hairpin conformations derived from the r(ACAGG)11 were produced using MC-Sym pipeline (University of Montreal) with default parameters (67). The most likely hairpin conformation was selected and then further minimized to obtain the refined structure for final minimization. Distance constraints were imposed on the guanine and cytosine residues involved in Watson-Crick GC base pairs during this process, with O-H and N-H hydrogen bond distances for GC base pairing constrained to 1.7 to 2 Å. The resulting structures were solvated by adding H_2O molecules and K^+ ions, and the systems were minimized without constraints to the points where the RMS gradients were <0.001 kcal/(mol Å).

Data availability

All data needed to evaluate the conclusions in the paper are present in the paper. Additional data related to this paper may be requested from the corresponding author upon reasonable request.

Supporting information—This article contains supporting information.

Acknowledgments—We thank editage for the English language review.

Author contributions—K. K., K. H., K. M., Y. A., and M. H. investigation; K. K., K. H., S. A., K. M., Y. A., and M. H. formal analysis; K. M. and Y. Y. visualization; N. S. supervision; N. S. writing-original draft.

Funding and additional information—This work was supported by the Japan Agency for Medical Research and Development (JP21ak0101131, JP23ek0109651, and JP23ek0109591 [to N. S.]), the Japan Society for the Promotion of Science KAKENHI (JP23H00373 and JP22K19297 to [N. S.]), the Japan Science and Technology Agency Fusion Oriented Research for disruptive Science and Technology Program (JPMJFR2043 [to N. S.]), program of the Joint Usage/Research Center for Developmental Medicine (Institute of Molecular Embryology and Genetics, Kumamoto University), program of the Inter-University Research Network for High Depth Omics (Institute of Molecular Embryology and Genetics, Kumamoto University), the Mochida Memorial Foundation for Medical and Pharmaceutical Research, and the Astellas Foundation for Research on Metabolic Disorders.

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, vestibular areflexia syndrome; DMS, dimethyl sulfate; G4, G-quadruplex; GOF, gain of function; LOF, loss of function; RAN, repeat-associated non-AUG; RBP, RNA-binding protein; RFC1, replication factor C subunit 1; STR, short tandem repeat; TBE, tris-borate-EDTA; TFO, triplex-forming oligonucleotide; ThT, thioflavin T.

References

1. Richard, G.-F., Kerrest, A., and Dujon, B. (2008) Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol. Mol. Biol. Rev.* **72**, 686–727
2. Depienne, C., and Mandel, J.-L. (2021) 30 years of repeat expansion disorders: what have we learned and what are the remaining challenges? *Am. J. Hum. Genet.* **108**, 764–785
3. Krzyzosiak, W. J., Sobczak, K., Wojciechowska, M., Fiszer, A., Mykowska, A., and Kozlowski, P. (2012) Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Res.* **40**, 11–26
4. Asamitsu, S., Takeuchi, M., Ikenoshita, S., Imai, Y., Kashiwagi, H., and Shioda, N. (2019) Perspectives for applying G-quadruplex structures in neurobiology and neuropharmacology. *Int. J. Mol. Sci.* **20**, 2884
5. Khristich, A. N., and Mirkin, S. M. (2020) On the wrong DNA track: molecular mechanisms of repeat-mediated genome instability. *J. Biol. Chem.* **295**, 4134–4170
6. Szmulewicz, D. J., Waterston, J. A., Halmagyi, G. M., Mossman, S., Chancellor, A. M., McLean, C. A., *et al.* (2011) Sensory neuropathy as part of the cerebellar ataxia neuropathy vestibular areflexia syndrome. *Neurology* **76**, 1903–1910
7. Davies, K., Szmulewicz, D. J., Corben, L. A., Delatycki, M., and Lockhart, P. J. (2022) *RFC1*-related disease: molecular and clinical insights. *Neurol. Genet.* **8**, e200016
8. 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
9. 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
10. Scriba, C. K., Beecroft, S. J., Clayton, J. S., Cortese, A., Sullivan, R., Yau, W. Y., *et al.* (2020) A novel RFC1 repeat motif (ACAGG) in two Asia-Pacific CANVAS families. *Brain* **143**, 2904–2910
11. Tsuchiya, M., Nan, H., Koh, K., Ichinose, Y., Gao, L., Shimozono, K., *et al.* (2020) *RFC1* repeat expansion in Japanese patients with late-onset cerebellar ataxia. *J. Hum. Genet.* **65**, 1143–1147
12. Beecroft, S. J., Cortese, A., Sullivan, R., Yau, W. Y., Dyer, Z., Wu, T. Y., *et al.* (2020) A Māori specific *RFC1* pathogenic repeat configuration in CANVAS, likely due to a founder allele. *Brain* **143**, 2673–2680
13. Cortese, A., Tozza, S., Yau, W. Y., Rossi, S., Beecroft, S. J., Jaunmuktane, Z., *et al.* (2020) Cerebellar ataxia, neuropathy, vestibular areflexia syndrome due to RFC1 repeat expansion. *Brain* **143**, 480–490
14. Sen, D., and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364–366
15. del Villar-Guerra, R., Trent, J. O., and Chaires, J. B. (2018) G-quadruplex secondary structure obtained from circular dichroism spectroscopy. *Angew. Chem. Int. Ed.* **57**, 7171–7175
16. Moser, H. E., and Dervan, P. B. (1987) Sequence-specific cleavage of double helical DNA by triple helix formation. *Science* **238**, 645–650
17. Kohwi, Y., and Kohwi-Shigematsu, T. (1988) Magnesium ion-dependent triple-helix structure formed by homopurine-homopyrimidine sequences in supercoiled plasmid DNA. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 3781–3785
18. Gehring, K., Leroy, J.-L., and Guéron, M. (1993) A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature* **363**, 561–565
19. Iaccarino, N., Cheng, M., Qiu, D., Pagano, B., Amato, J., Di Porzio, A., *et al.* (2021) Effects of sequence and base composition on the CD and TDS profiles of i-DNA. *Angew. Chem. Int. Ed.* **60**, 10295–10303

20. Kejnovská, I., Kypr, J., and Vorlíčková, M. (2003) Circular dichroism spectroscopy of conformers of (guanine + adenine) repeat strands of DNA. *Chirality* **15**, 584–592
21. Mohanty, J., Barooah, N., Dhamodharan, V., Harikrishna, S., Pradeepkumar, P. I., and Bhasikuttan, A. C. (2013) Thioflavin T as an efficient inducer and selective fluorescent sensor for the human telomeric G-quadruplex DNA. *J. Am. Chem. Soc.* **135**, 367–376
22. Ranjbar, B., and Gill, P. (2009) Circular dichroism techniques: biomolecular and nanostructural analyses- A review. *Chem. Biol. Drug Des.* **74**, 101–120
23. Tang, T. T. L., Stowell, J. A. W., Hill, C. H., and Passmore, L. A. (2019) The intrinsic structure of poly(A) RNA determines the specificity of Pan2 and Caf1 deadenylases. *Nat. Struct. Mol. Biol.* **26**, 433–442
24. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.-P., and Ehresmann, B. (1987) Probing the structure of RNAs in solution. *Nucleic Acids Res.* **15**, 9109–9128
25. Sato, K., Akiyama, M., and Sakakibara, Y. (2021) RNA secondary structure prediction using deep learning with thermodynamic integration. *Nat. Commun.* **12**, 941
26. Sieg, J. P., Arteaga, S. J., Znosko, B. M., and Bevilacqua, P. C. (2023) MeltR software provides facile determination of nucleic acid thermodynamics. *Biophys. Rep.* **3**, 100101
27. Huppert, J. L., and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908–2916
28. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **34**, 5402–5415
29. Pandey, S., Agarwala, P., and Maiti, S. (2013) Effect of loops and G-quartets on the stability of RNA G-quadruplexes. *J. Phys. Chem. B* **117**, 6896–6905
30. Watson, J. D., and Crick, F. H. C. (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738
31. Heus, H. A., and Pardi, A. (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **253**, 191–194
32. Baeyens, K. J., De Bondt, H. L., Pardi, A., and Holbrook, S. R. (1996) A curved RNA helix incorporating an internal loop with G·A and A·A non-Watson-Crick base pairing. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 12851–12855
33. Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68–74
34. Li, J., Begbie, A., Boehm, B. J., Button, A., Whidborne, C., Pouferis, Y., *et al.* (2019) Ion mobility-mass spectrometry reveals details of formation and structure for GAA·TCC DNA and RNA triplexes. *J. Am. Soc. Mass Spectrom.* **30**, 103–112
35. Abou Assi, H., Garavís, M., González, C., and Damha, M. J. (2018) i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.* **46**, 8038–8056
36. 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
37. DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., *et al.* (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of *C9ORF72* causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245–256
38. Renton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J. R., *et al.* (2011) A hexanucleotide repeat expansion in *C9ORF72* is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268
39. Balendra, R., and Isaacs, A. M. (2018) *C9orf72*-mediated ALS and FTD: multiple pathways to disease. *Nat. Rev. Neurol.* **14**, 544–558
40. Zu, T., Gibbens, B., Doty, N. S., Gomes-Pereira, M., Huguet, A., Stone, M. D., *et al.* (2011) Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 260–265
41. Ash, P. E. A., Bieniek, K. F., Gendron, T. F., Caulfield, T., Lin, W.-L., DeJesus-Hernandez, M., *et al.* (2013) Unconventional translation of *C9ORF72* GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* **77**, 639–646
42. Zu, T., Liu, Y., Bañez-Coronel, M., Reid, T., Pletnikova, O., Lewis, J., *et al.* (2013) RAN proteins and RNA foci from antisense transcripts in *C9ORF72* ALS and frontotemporal dementia. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E4968–E4977
43. 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
44. Wen, X., Tan, W., Westergard, T., Krishnamurthy, K., Markandaiah, S. S., Shi, Y., *et al.* (2014) Antisense proline-arginine RAN dipeptides linked to *C9ORF72*-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death. *Neuron* **84**, 1213–1225
45. Zhang, K., Donnelly, C. J., Haeusler, A. R., Grima, J. C., Machamer, J. B., Steinwald, P., *et al.* (2015) The *C9orf72* repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525**, 56–61
46. Hagerman, R. J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., *et al.* (2001) Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* **57**, 127–130
47. Iwahashi, C. K., Yasui, D. H., An, H.-J., Greco, C. M., Tassone, F., Nannen, K., *et al.* (2006) Protein composition of the intranuclear inclusions of FXTAS. *Brain* **129**, 256–271
48. Todd, P. K., Oh, S. Y., Krans, A., He, F., Sellier, C., Frazer, M., *et al.* (2013) CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* **78**, 440–455
49. Asamitsu, S., Yabuki, Y., Ikenoshita, S., Kawakubo, K., Kawasaki, M., Usuki, S., *et al.* (2021) CGG repeat RNA G-quadruplexes interact with FMRpolyG to cause neuronal dysfunction in fragile X-related tremor/ataxia syndrome. *Sci. Adv.* **7**, eabd9440
50. Lieberman, A. P., Shakkottai, V. G., and Albin, R. L. (2019) Polyglutamine repeats in neurodegenerative diseases. *Annu. Rev. Pathol.* **14**, 1–27
51. Bañez-Coronel, M., Ayhan, F., Tarabochia, A. D., Zu, T., Perez, B. A., Tusi, S. K., *et al.* (2015) RAN translation in Huntington disease. *Neuron* **88**, 667–677
52. Das, M. R., Chang, Y., Anderson, R., Saunders, R. A., Zhang, N., Tomberlin, C. P., *et al.* (2023) Repeat-associated non-AUG translation induces cytoplasmic aggregation of CAG repeat-containing RNAs. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2215071120
53. Ronco, R., Perini, C., Currò, R., Dominik, N., Facchini, S., Gennari, A., *et al.* (2023) Truncating variants in RFC1 in cerebellar ataxia, neuropathy, and vestibular areflexia syndrome. *Neurology* **100**, e543
54. Arteche-López, A., Avila-Fernandez, A., Damian, A., Soengas-Gonda, E., de la Fuente, R. P., Gómez, P. R., *et al.* (2023) New Cerebellar Ataxia, Neuropathy, Vestibular Areflexia Syndrome cases are caused by the presence of a nonsense variant in compound heterozygosity with the pathogenic repeat expansion in the RFC1 gene. *Clin. Genet.* **103**, 236–241
55. Weber, S., Coarelli, G., Heinzmann, A., Monin, M.-L., Richard, N., Gerard, M., *et al.* (2023) Two RFC1 splicing variants in CANVAS. *Brain* **146**, e14–e16
56. Ogi, T., Limsirichaikul, S., Overmeer, R. M., Volker, M., Takenaka, K., Cloney, R., *et al.* (2010) Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Mol. Cell* **37**, 714–727
57. [preprint] Hisey, J. A., Radchenko, E. A., Ceschi, S., Rastokina, A., Mandel, N. H., McGinty, R. J., *et al.* (2023) Pathogenic CANVAS (AAGGG)n repeats stall DNA replication due to the formation of alternative DNA structures. *bioRxiv*. https://doi.org/10.1101/2023.07.25.550509
58. [preprint] Maltby, C. J., Krans, A., Grudzien, S. J., Palacios, Y., Muiños, J., Suárez, A., *et al.* (2023) AAGGG repeat expansions trigger RFC1-independent synaptic dysregulation in human CANVAS Neurons. *bioRxiv*. https://doi.org/10.1101/2023.12.13.571345
59. Wada, T., Doi, H., Okubo, M., Tada, M., Ueda, N., Suzuki, H., *et al.* (2023) RNA foci in two bi-allelic RFC1 expansion carriers. *Ann. Neurol.* **95**, 607–613. Online ahead of print
60. Miyatake, S., Yoshida, K., Koshimizu, E., Doi, H., Yamada, M., Miyaji, Y., *et al.* (2022) Repeat conformation heterogeneity in cerebellar ataxia, neuropathy, vestibular areflexia syndrome. *Brain* **145**, 1139–1150

61. Prigodich, R. V., and Martin, C. T. (1990) Reaction of single-stranded DNA with hydroxyl radical generated by iron(II)-ethylenediaminetetraacetic acid. *Biochemistry* **29**, 8017–8019
62. Mergny, J.-L., and Lacroix, L. (2009) UV melting of G-quadruplexes. *Curr. Protoc. Nucleic Acid Chem.* **37**, 17.1.1–17.1.15
63. Amrane, S., Saccà, B., Mills, M., Chauhan, M., Klump, H. H., and Mergny, J.-L. (2005) Length-dependent energetics of $(CTG)_n$ and $(CAG)_n$ trinucleotide repeats. *Nucleic Acids Res.* **33**, 4065–4077
64. Molecular Operating Environment (MOE) (2023) *2022.02 Chemical Computing Group, 910-1010 Sherbrooke St. W., Montreal*, QC H3A 2R7, Canada
65. Park, S., Zheng, L., Kumakiri, S., Sakashita, S., Otomo, H., Ikehata, K., *et al.* (2014) Development of DNA-based hybrid catalysts through direct ligand incorporation: toward understanding of DNA-based asymmetric catalysis. *ACS Catal.* **4**, 4070–4073
66. Lin, L. Y., McCarthy, S., Powell, B. M., Manurung, Y., Xiang, I. M., Dean, W. L., *et al.* (2020) Biophysical and X-ray structural studies of the $(GGGTT)_3GGG$ G-quadruplex in complex with *N*-methyl meso-porphyrin IX. *PLoS One* **15**, e0241513
67. Parisien, M., and Major, F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51–55