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

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

[1] Department of Genomic Neurology, Institute of Molecular Embryology and Genetics (IMEG), Kumamoto University; Kumamoto 860-0811, Japan.

[2] Graduate School of Pharmaceutical Sciences, Kumamoto University; Kumamoto 862-0973, Japan.

[3] Laboratory for Functional Non-coding Genomics, RIKEN Center for Integrative Medical Sciences (IMS); Yokohama, 230-0045, Japan

[a] These authors contributed equally to this work.

* Corresponding author: Norifumi Shioda

Address: 2-2-1 Honjo, Chuo-ku, Kumamoto 860-0811, Japan.

Phone number: +81-96-373-6633

Email: shioda@kumamoto-u.ac.jp

Running title: Nucleic acid structures of pathogenic repeats in CANVAS

Keywords: nucleic acid structure, biophysics, G-quadruplex, genetic disease, neurodegenerative disease

Abstract

Short tandem repeats are highly unstable, 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* repeat sequences 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 sequence formed double helical hairpin structures 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 non-pathogenic sequences. These findings provide novel insights into the structural polymorphism of the *RFC1* repeat sequences, which may be closely related to the disease mechanism of CANVAS.

Introduction

Short tandem repeats (STRs) are polymorphic repeat units of 1–6 base pairs that are abundant in eukaryotic genomes [1]. STRs are extremely unstable, depending on repeat length, and their expansion in the coding and non-coding 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 non-canonical DNA and its transcribed RNA, e.g., slipped hairpins, G-quadruplexes (G4s), and triplexes, that 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 non-pathogenic populations: the reference $(AAAAG)_{11}$, $(AAAAG)_{exp}$ ranging from 15 to 200 repeats, and $(AAAGG)_{exp}$ ranging from 40 to 1000 repeats (**Fig. 1A**) [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-Vis spectroscopy and thioflavin T (ThT) fluorescence and footprinting assays. Additionally, based on the results, the higher-order structures of nucleic acids in pathogenic repeat sequences 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 non-pathogenic repeat sequences form characteristic nucleic acid secondary structures. The disease-specific formation of non-canonical nucleic acid structures may be the underlying molecular cause of the pathology of CANVAS.

Results

CD of the DNAs in *RFC1* pentanucleotide repeats

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

Single-stranded guanine-rich sequences may form G4s, which are four-stranded nucleic acid structures comprising stacked G-quartets, i.e., 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-non-stabilizing 100 mM Li$^+$, the entire spectrum changes, and the typical features are not observed (**Fig. 1B**). To further confirm the G4 folding of d(AAGGG)$_4$, we assessed the cation-dependent thermal denaturing by measuring the melting temperature (Tm). d(AAGGG)$_4$ exhibits a much higher thermal stability under K$^+$ conditions than that under Li$^+$ conditions (ΔTm = 10.4 °C). No characteristic spectrum of the other pathological sequence, d(ACAGG)$_4$, is observed in the presence of K$^+$ or Li$^+$ (**Fig. 1C**).

Homopurine/-pyrimidine mirror repeats, such as AAGGG/CCCTT repeat sequences, 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. 1D, 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. 1D, 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. 1E**). No CD profile features of the i-motif structure are observed, with positive and negative peaks at 288 and 264 nm, respectively [19].

In terms of the non-pathological sequences, the CD spectra of 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. 1F**). 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 sequence 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. 1B**), a drastic increase in ThT fluorescence is detected for d(AAGGG)4 compared to those of other sequences (**Fig. 2A**).

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 sufficiently protected against DMS methylation and subsequent piperidine cleavage than in the presence of Li^+ (**Fig. 2B**). The results of CD (**Fig. 1B**) and the ThT assay (**Fig. 2A**), in addition to the DMS footprinting pattern, suggest G4 formation by d(AAGGG)4, where 4 tracts of 3 guanines form 4 strands with 3 two-base loops comprising AA (**Fig. 2C**).

CD of the RNAs in *RFC1* pentanucleotide repeats

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

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. 3A, 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. 3A, right**). In the thermal denaturation assay at the same concentration, the stabilizing effects of K^+ on r(AAGGG)$_{11}$ (ΔTm = 41.0 °C) and r(AAGGG)$_4$ (ΔTm = 44.0 °C) compared to those of Li^+ are observed. In the CD spectrum of ACAGG repeat RNA, no characteristic peaks representing 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. 3B**). These peaks are characteristic of the A-form duplex RNA conformation [22].

None of the CD spectra of the complementary strands of the pathological sequences, r(CCCUU)$_4$ and r(CCUGU)$_4$, display characteristic peaks indicative of RNA secondary structures (**Fig. 3C**). In terms of the non-pathological sequences, 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. 3D, 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 d(CUUUU)$_4$ and d(CCUUU)$_4$ display small positive peaks at 275 nm (**Fig. 3D, right**), indicating the absence of complex structures, similar to poly(U) [23].

Analyses of the RNA structures of the pathological repeat sequences related to CANVAS

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

To identify more detailed RNA structures, we conducted RNase T1 footprinting of pathological AAGGG and ACAGG and non-pathological AAAAG repeats. As RNase T1 displays specificity for guanine residues in single-stranded RNA, 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 twelve guanosines (G3–G5, G8–G10, G13–G15, G18–G20) are protected against RNase T1 cleavage in the presence of K^+ (**Fig. 4B, left**). Based on the RNase footprinting pattern of r(AAGGG)$_{11}$ (**Fig. 4B,**

middle) and the results of CD (**Fig. 3A**), and the ThT assay (**Fig. 4A**), the pathological AAGGG repeat sequence displays a strong tendency to fold into an intramolecular parallel-form G4, which comprises 3 G-quartets with 3 two-base loops comprising AA (**Fig. 4B, 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. 4C, 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. 4C, middle**). RNA secondary structure prediction using deep learning based on MXfold2 (Keio University, Yokohama, Japan) [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. 4C, right**). This is consistent with the results obtained, including those of CD (**Fig. 3B**). The energy-minimized 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. 4C, right**). In the non-pathological AAAAG repeat RNA, the single guanosine in the repeat sequence 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. 4D**).

To determine the differences in the structural stabilities of these RNAs, we performed UV-monitored melting studies. The obtained results confirm that r(AAGGG)$_{11}$ and r(ACAGG)$_{11}$ form stable structures, with respective changes in folding free energy of $\Delta G37°C = -24.44 \pm 7.42$ and -1.72 ± 0.11 kcal/mol, whereas that of r(AAAAG)$_{11}$ is not detected (**Table 1**).

Discussion

We demonstrated the structural polymorphism of the *RFC1* repeat sequences 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 non-pathogenic repeat sequences. 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 repeat sequences in DNA and RNA could also form G4s based on the consensus motif ($G_{2-3} + N_{1-7}G_{2-3} + N_{1-7}G_{2-3} + N_{1-7}G_{2-3}$) [4,26]. 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 [27]. 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 extremely unstable [28]. 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 [29] predominate at high cytosine contents in the loops [28]. 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 [30,31], similar to that of the analogous base pairs in hammerhead ribozymes [32]. This study suggests that the combination of these factors results in ACAGG repeats with high energies, forming a rigid secondary structure not observed with non-pathogenic 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 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 [33,34]. Investigating the formation of these non-canonical DNA structures with various repeat sizes under different conditions is necessary.

The structures of non-canonical DNA and its transcribed RNA that form during repeat expansion disorders are central to the molecular mechanisms of toxicity via loss- or gain-of-function, RNA-binding protein (RBP) sequestration, and repeat-associated non-AUG (RAN) translation [5,35]. In C9orf72 amyotrophic lateral sclerosis and frontotemporal dementia (FTD) caused by GGGGCC repeat expansions (d(G$_4$C$_2$)$_{exp}$) in the *C9orf72* gene [36,37], its transcribed r(G$_4$C$_2$)$_{exp}$ forms G4s and disrupts neuronal function via various mechanisms by generating nuclear foci [38]. Additionally, sense r(G$_4$C$_2$)$_{exp}$ and antisense r(C$_4$G$_2$)$_{exp}$ transcribed RNAs are translated in all reading frames into dipeptide repeat proteins via RAN translation [39–41], resulting in these RAN proteins forming toxic aggregates [42–44]. In another pathological mechanism, G4-forming d(G$_4$C$_2$)$_{exp}$ inhibits transcription, causing loss of function by reducing C9orf72 protein levels [42]. Fragile X-associated tremor/ataxia syndrome (FXTAS) is a neurodegenerative disease caused by d(CGG)$_{exp}$ in the *FMR1* gene [45]. Transcribed

r(CGG)$_{exp}$ interacts with several RBPs that form RNA foci and induce cellular dysfunction [46]. Transcribed r(CGG)$_{exp}$ also produces the toxic protein FMRpolyG via RAN translation [47], and G4-forming r(CGG)$_{exp}$ induces the liquid-solid phase transition of FMRpolyG, causing neurological dysfunction [48]. 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 (polyQ) tracts, inducing polyQ toxicity [49]. 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 [39,50]. Additionally, r(CAG)$_{exp}$-derived RAN translation products co-aggregate with r(CAG)$_{exp}$ foci within the cytoplasm [51].

Currently, how the repeat expansions in the gene encoding *RFC1* contribute to the pathogenesis of CANVAS is unknown. To date, no reductions in *RFC1* mRNA and protein levels have been observed in peripheral tissues and postmortem brain samples from patients with biallelic (AAGGG)$_{exp}$ in the *RFC1* gene. Additionally, no sense or antisense RNA foci were detected in the cerebellar vermes of patients. Conversely, *RFC1* pre-mRNA expression displayed increased intron 2 retention in the cerebella, frontal cortices, lymphoblasts, and muscles of patients compared to that of healthy controls [8]. GC-rich intron expansion, which is associated with intron retention, may inhibit splicing by altering RNA structure and/or splicing factor access to intron regulatory regions [52]. These results were obtained using a few samples, and thus, further studies are necessary to obtain accurate results regarding the transcription and protein translation of *RFC1*.

In conclusion, we identified the DNA and RNA secondary structures of pathological repeats associated with CANVAS *in vitro*. In the pathological sequences, 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 (Sapporo, Japan): 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 (Ibaraki, Japan): 5′ 6-FAM-r(AAAAG)$_{11}$, 5′ 6-FAM-r(AAGGG)$_{11}$, and 5′ 6-FAM-r(ACAGG)$_{11}$.

CD

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 200–400 nm using a J-1100 spectrometer (JASCO, Tokyo, Japan) and a quartz cuvette with a 3 mm path length. In the denaturing assays, the CD signal of each oligonucleotide at 265 (RNA) or 296 nm (DNA) was monitored. Temperature scans were performed continuously from 20 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, Tokyo, Japan). 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 420–600 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 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, Marlborough, MA, USA).

DMS footprinting

FAM-labeled DNA (2.5 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 1 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, Tokyo, Japan). The resulting residues were washed with RNase-free H_2O, re-dried, and dissolved in a loading solution (80 % formamide and 10 mM NaOH). 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 (1.5 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, Waltham, MA, USA) 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 8.0)) 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-vis spectral assays

RNA oligonucleotides (2.5 μM) were prepared in 10 mM Tris/HCl (pH 7.5) buffer containing 100 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 (hairpin) or 295 nm (G4) 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 to 95 °C at 1 °C/min, and the T_m values were determined using a 2-point method in the Melting Analysis program of the Spectra Manager software (JASCO) [53].

Thermodynamic analysis

The thermodynamic parameters of the hairpin and G4 structures were determined

according to a previously reported method [53,54]. The folding ($A_F(T)$) and unfolding ($A_U(T)$) baselines were manually selected in Excel (Microsoft, Redmond, WA, USA), where T represents the temperature in degrees Celsius.

Using the baselines, the equations representing the folded fractions (N_T) are as follows:

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

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

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

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

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

The Gibbs free enthalpy may be written as:

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

Where T, $\Delta H°$, $\Delta S°$, 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,

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

yields an Arrhenius plot [ln(Ka) as a function of 1/T] that displays a straight line, with a respective slope and y-axis intercept of $-\frac{\Delta H°}{R}$ and $\frac{\Delta S°}{R}$, which afford $\Delta H°$ and $\Delta S°$.

$\Delta G°$ is calculated at 37 °C.

Molecular modeling

Molecular modeling was performed using the Molecular Operating Environment software (Chemical Computing Group, Montreal, Canada) and Amber94/99 force field parameters and a distance-dependent dielectric constant of ε = 4r (where r is the distance between two atoms) [55,56]. 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) [57], respectively. Distance and dihedral-angle constraints were imposed on the guanine residues within the three G-tetrads during the process. The structure of the r(ACAGG)$_{11}$ hairpin conformation was created using MC-Sym pipeline (University of Montreal, Montreal, Canada) [58] with default parameters and then further refined to yield the initial structure for use in the final minimization. The resulting structures were solvated by adding H$_2$O molecules and Na$^+$ ions, and the systems were minimized without constraints to the points where the root-mean-square 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.

Corresponding author: Norifumi Shioda

Email: shioda@kumamoto-u.ac.jp

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

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**(4), 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**(5), 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**(1), 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**(12), 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**(13), 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**(22), 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**(5), 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**(4), 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**(1), 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**(10), 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**(12), 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**(9), 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**(2), 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**(6180), 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**(24), 7171–7175.

16. Moser, H. E. and Dervan, P. B. (1987) Sequence-specific cleavage of double helical DNA by triple helix formation. *Science.* **238**(4827), 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**(11), 3781–3785.

18. Gehring, K., Leroy, J.-L., and Guéron, M. (1993) A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature.* **363**(6429), 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**(18), 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**(7), 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**(1), 367–376.

22. Ranjbar, B. and Gill, P. (2009) Circular dichroism techniques: Biomolecular and nanostructural analyses- A review. *Chem Biol Drug Des.* **74**(2), 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**(6), 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**(22), 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. Huppert, J. L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**(9), 2908–2916.

27. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. (2006) Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res.* **34**(19), 5402–5415.

28. 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**(23), 6896–6905.

29. Watson, J. D. and Crick, F. H. C. (1953) Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature.* **171**(4356), 737–738.

30. Heus, H. A. and Pardi, A. (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science.* **253**(5016), 191–194.

31. 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**(23), 12851–12855.

32. Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature.* **372**(6501), 68–74.

33. 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**(1), 103–112.

34. 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**(16), 8038–8056.

35. 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**(9), 589–607.

36. 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**(2), 245–256.

37. 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**(2), 257–268.

38. Balendra, R. and Isaacs, A. M. (2018) *C9orf72*-mediated ALS and FTD: Multiple pathways to disease. *Nat Rev Neurol.* **14**(9), 544–558.

39. 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**(1), 260–265.

40. 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**(4), 639–646.

41. 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**(51), E4968–E4977.

42. 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**(7491), 195–200.

43. 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**(6), 1213–1225.

44. 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**(7567), 56–61.

45. 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**(1), 127–130.

46. 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**(1), 256–271.

47. 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**(3), 440–455.

48. 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**(3), eabd9440.

49. Lieberman, A. P., Shakkottai, V. G., and Albin, R. L. (2019) Polyglutamine repeats in neurodegenerative diseases. *Annual Rev Pathol: Mech Dis.* **14**, 1–27.

50. 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**(4), 667–677.

51. 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**(3), e2215071120.

52. 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**(16), 4234–4239.

53. Mergny, J.-L. and Lacroix, L. (2009) UV melting of G-quadruplexes. *Curr Protoc Nucleic Acid Chem.* **37**, 17.1.1–17.1.15.

54. 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**(13), 4065–4077.

55. Molecular Operating Environment (MOE), 2022.02 Chemical Computing Group, 910-1010 Sherbrooke St. W., Montreal, QC H3A 2R7, Canada, 2023.

56. 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**(11), 4070–4073.

57. 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 mesoporphyrin IX. *PLoS One.* **15**(11), e0241513.

58. Parisien, M. and Major, F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature.* **452**(7183), 51–55.

Footnotes

Abbreviations: CANVAS, cerebellar ataxia with neuropathy and vestibular areflexia syndrome; DMS, dimethyl sulfate; FAM, 6-carboxyfluorescein; FTD, frontotemporal dementia; FXTAS, fragile X-associated tremor/ataxia syndrome; IMEG, Institute of Molecular Embryology and Genetics; MOE, Molecular Operating Environment; RAN, repeat-associated non-AUG; RBP, RNA-binding protein; RFC1, replication factor C subunit 1; STR, short tandem repeat; TFO, triplex-forming oligonucleotide.

Figure legends

Figure 1. CD spectra of the DNAs in *RFC1* pentanucleotide repeats (A) Schematic diagram showing the polymorphism at the *RFC1* repeat locus, consisting of non-pathogenic and pathogenic alleles. *RFC1*: replication factor C subunit 1. See also Davies et al., (2022). (B, C) Representative CD spectra of the pathogenic sequences 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 sequences 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 non-pathogenic 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.

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.

Figure 3. CD spectra of the RNAs in *RFC1*-derived pentanucleotide repeats (A, 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 sequences 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 non-pathogenic 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.

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 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 (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)$_4$, based on the results of the biophysical analyses. The G-C and G:A base pairs are putatively formed (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.

Table 1.
Experimentally obtained thermodynamic parameters of the RNA sequences.
Experiments were repeated thrice.
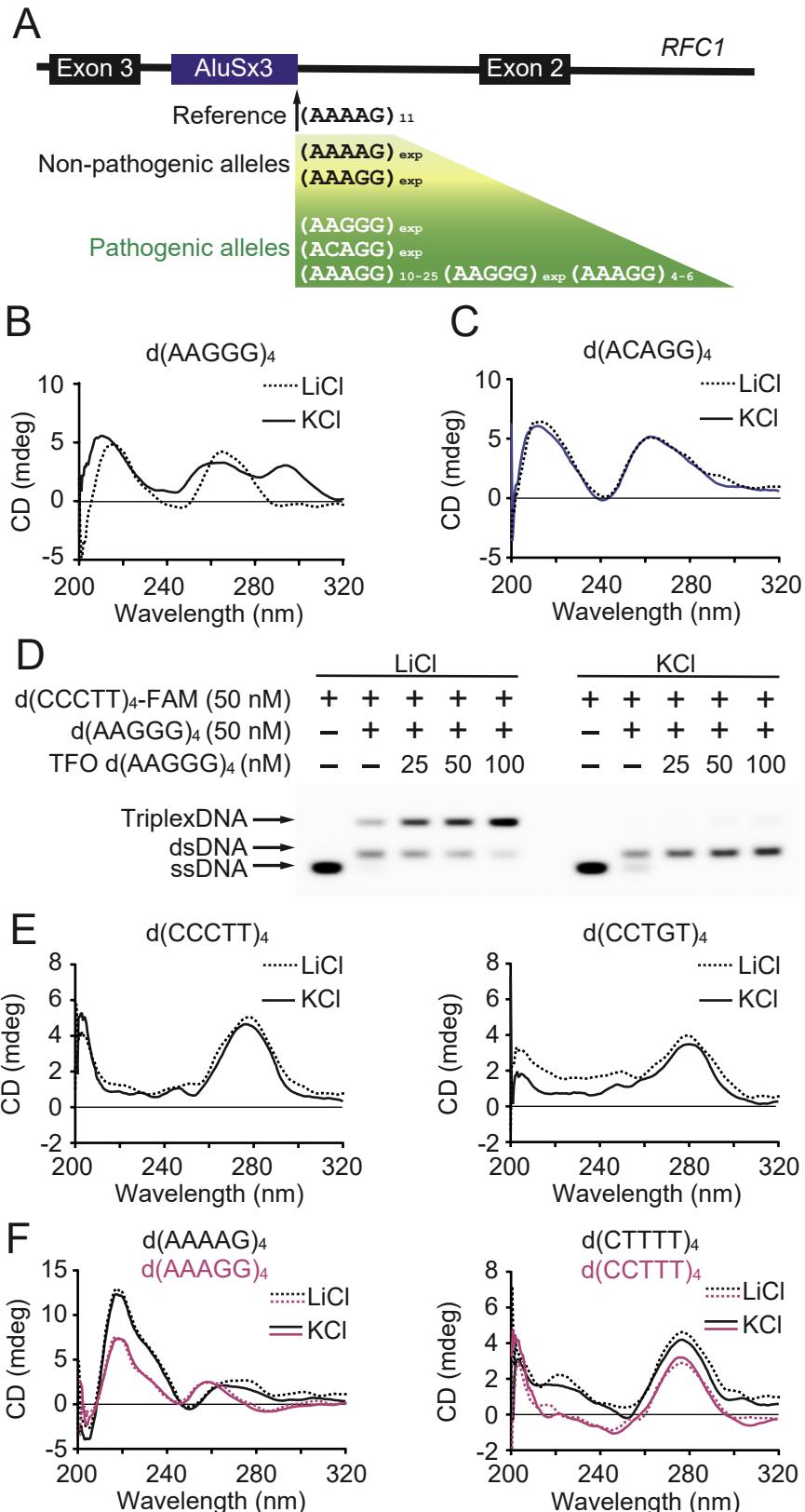
Figure 1
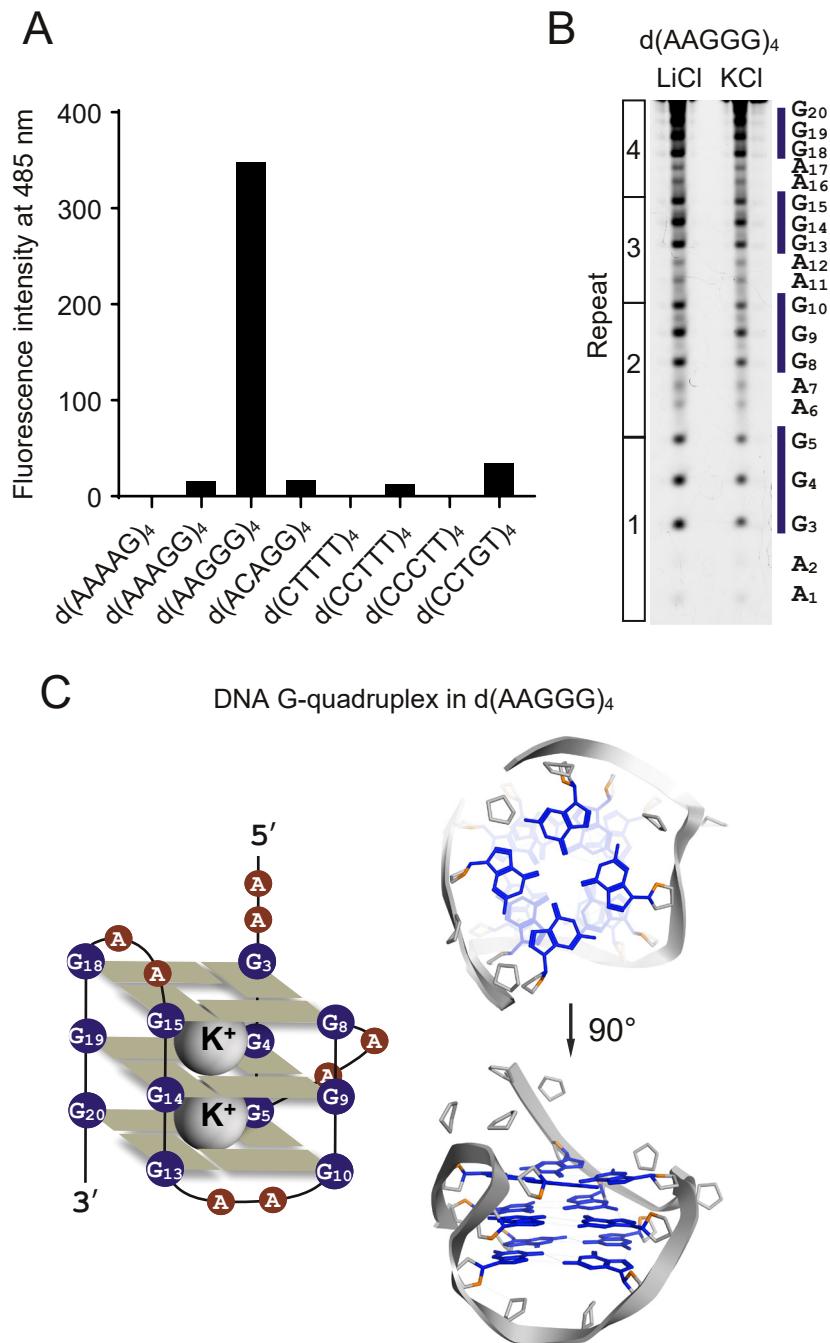
Figure 2
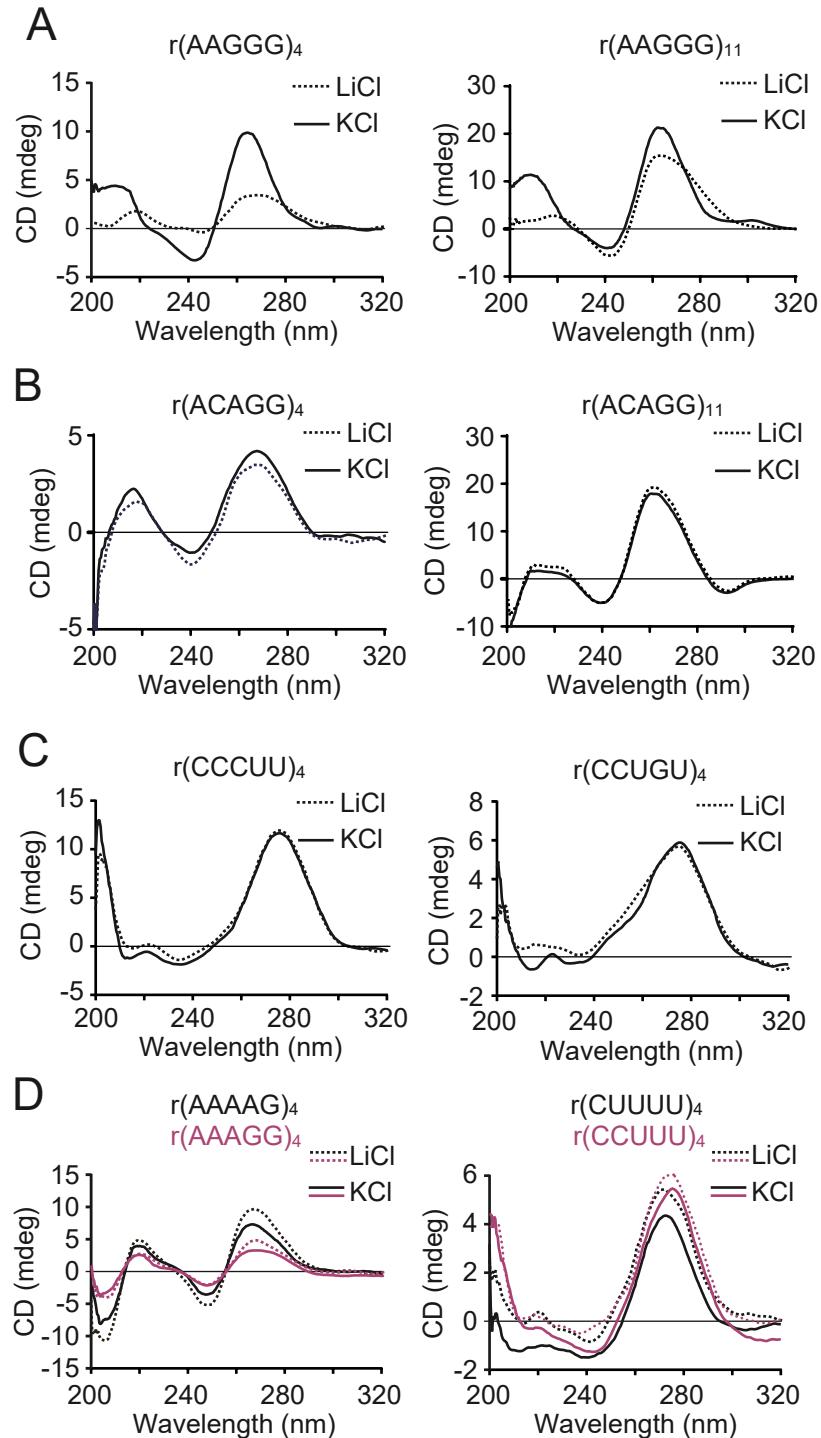
Figure 3
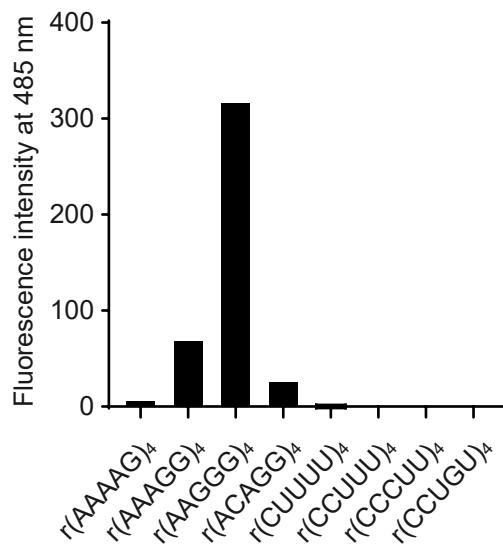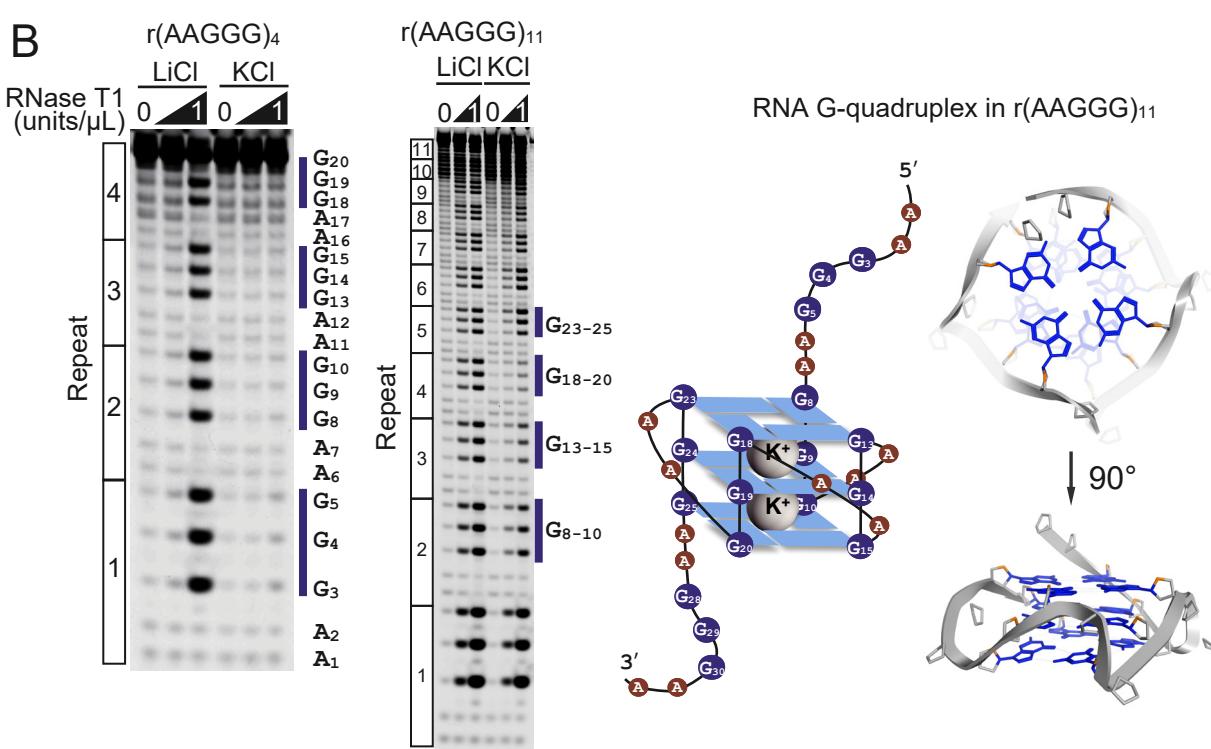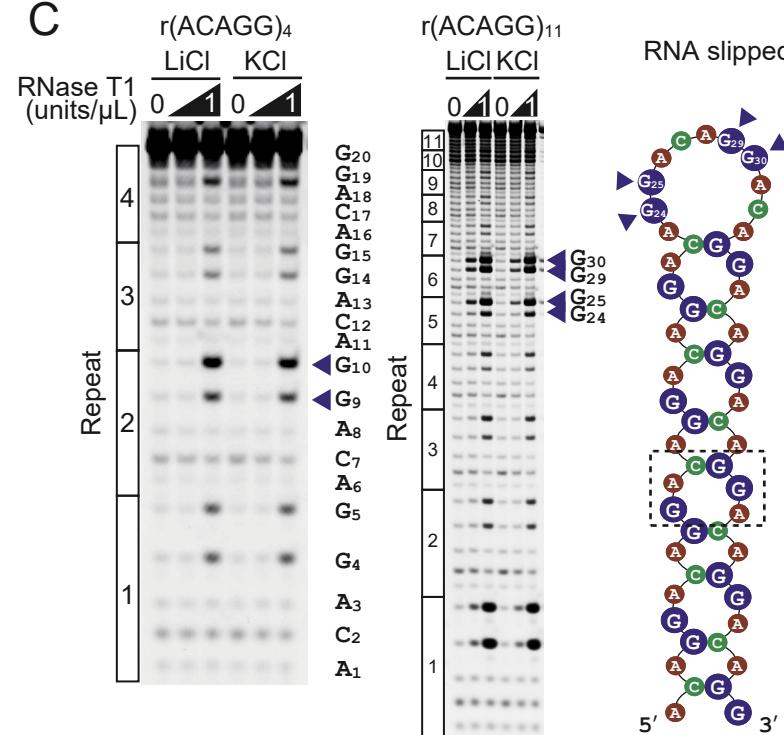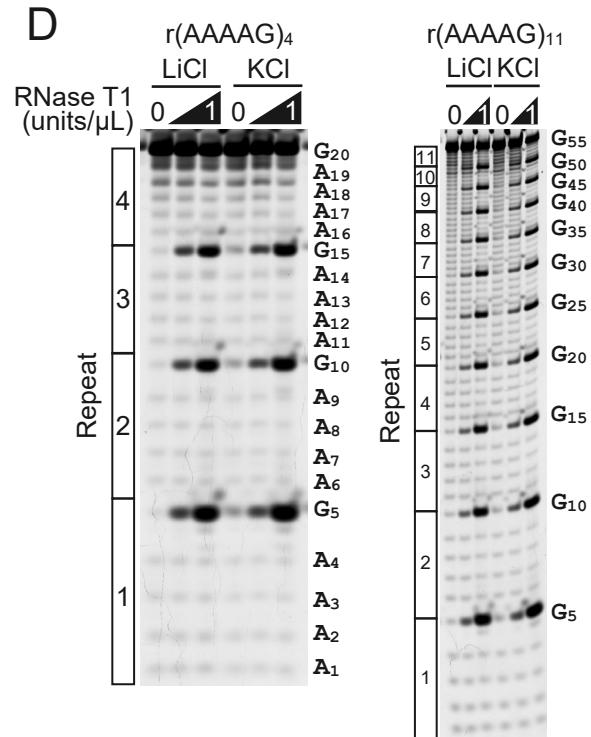
Figure 4

	T_{m} (°C)	$\Delta H°$ (kcal/mol)	$\Delta S°$ (cal/mol*K)	$\Delta G°$ (kcal/mol)
r(AAAAG)$_{11}$	N.D.	N.D.	N.D.	N.D.
r(AAGGG)$_{11}$	85.46±0.58	-178.52±49.78	-496.79±136.58	-24.44±7.42
r(ACAGG)$_{11}$	48.28±0.22	-55.11±1.49	-172.15±4.68	-1.72±0.11

N.D. = Not Detected

Table1

Experimentally obtained thermodynamic parameters of the RNA sequences. Experiments were repeated thrice.