

Antisense RNA C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms a triplex-like structure and binds small synthetic ligand

Leszek Błaszczuk^{1,†}, Marcin Ryczek^{1,†}, Bimolendu Das^{2,†}, Martyna Mateja-Pluta¹, Magdalena Bejger¹, Joanna Śliwiak¹, Kazuhiko Nakatani² and Agnieszka Kiliszek^{1,*}

¹Institute of Bioorganic Chemistry, Polish Academy of Sciences, Z. Noskowskiego 12/14, 61-704, Poland

²Department of Regulatory Bioorganic Chemistry, SANKEN (The Institute of Scientific and Industrial Research), Osaka University, 8-1 Mihogaoka, Ibaraki 567-0047, Japan

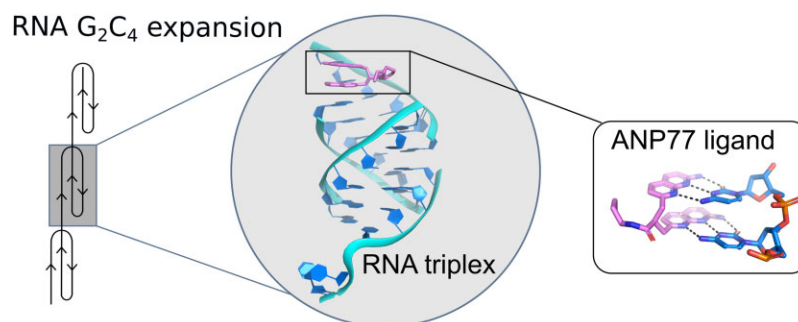
*To whom correspondence should be addressed. Tel: +48 61 852 8503 (Ext. 1278); Fax: +48 61 852 05 32; Email: kiliszek@ibch.poznan.pl

†The first three authors should be regarded as Joint First Authors.

Abstract

The abnormal expansion of GGGGCC/GGCCCC hexanucleotide repeats (HR) in *C9orf72* is associated with amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Structural polymorphisms of HR result in the multifactorial pathomechanism of ALS/FTD. Consequently, many ongoing studies are focused at developing therapies targeting pathogenic HR RNA. One of them involves small molecules blocking sequestration of important proteins, preventing formation of toxic nuclear foci. However, rational design of potential therapeutics is hindered by limited number of structural studies of RNA-ligand complexes. We determined the crystal structure of antisense HR RNA in complex with ANP77 ligand (1.1 Å resolution) and in the free form (0.92 and 1.5 Å resolution). HR RNA folds into a triplex structure composed of four RNA chains. ANP77 interacted with two neighboring single-stranded cytosines to form pseudo-canonical base pairs by adopting sandwich-like conformation and adjusting the position of its naphthyridine units to the helical twist of the RNA. In the unliganded structure, the cytosines formed a peculiar triplex i-motif, assembled by trans C•C⁺ pair and a third cytosine located at the Hoogsteen edge of the C•C⁺ pair. These results extend our knowledge of the structural polymorphisms of HR and can be used for rational design of small molecules targeting disease-related RNAs.

Graphical abstract



Introduction

Amyotrophic Lateral Sclerosis and Frontotemporal Dementia are fatal neurodegenerative repeat expansion disorders that affect the motor neurons in the brain and spinal cord. The progressive degeneration of nerve cells leads to changes in behavior, dysphagia, dysarthria, respiratory failure, and consequently, death (1,2). Currently, only symptomatic treatments are available to alleviate disease progression. The most frequent cause of ALS/FTD is a mutation in the *C9orf72* gene, which encodes a protein involved in the autophagy-lysosome pathway (3). The promoter region of *C9orf72* carries

microsatellite sequences consisting of hexanucleotide repeats (HR) 5'-GGGGCC-3'/5'-GGCCCC-3', which can undergo abnormal expansion. Healthy individuals possess up to 20 repeats, whereas the mutated form of the gene contains several thousand HR units (3–5). The presence of expanded repeats has a negative effect on DNA and RNA functions, resulting in the complex pathomechanism of ALS/FTD. One of these pathways involves transcriptional gene silencing and activation of the DNA damage response through the formation of DNA–RNA hybrids (R-loops) (3,6). On the RNA level, bidirectional transcription produces sense and

Received: October 24, 2023. Revised: April 5, 2024. Editorial Decision: April 18, 2024. Accepted: April 26, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

antisense transcripts containing complementary regions of repeated stretches of GGGGCC (G_4C_2) and GGCCCC (G_2C_4) that form stable secondary and tertiary structures (6,7). As a consequence, mutated transcripts gain the ability to sequester important proteins and form RNA foci that accumulate in the nucleus (8–10). Expanded repeats can also trigger repeat-associated non-ATG (RAN) translation, resulting in toxic cellular polydi-peptide aggregates (11–13).

The structural evaluation of the HR repeats has been performed only *in vitro* (in solution). These experiments show that sense and antisense transcripts containing HR fold into tertiary motifs and exhibit structural polymorphisms. Sense G_4C_2 RNA is rich in guanine residues. Thus, in the presence of potassium ions, inter- or intramolecular G-quadruplexes can be formed and assembled into G-wire or gel-like phases (6,7,14–17). G_4C_2 RNA repeats also fold into hairpin structures, which dominate in the absence of K^+ ions or at low refolding temperature (6,7,17). The antisense G_2C_4 repeats are cytosine rich and can fold into hairpin structures with long helical stems. Alternatively, cytosines can undergo protonation to form i-motifs or triplexes (6,7,18). So far, only one crystal structure of a G_2C_4 repeat has been reported (19). It folds into a duplex, representing part of the hairpin stem, consisting of G–C pairs interposed by non-canonical C–C pairs. The 3D structures of G_2C_4 repeats with protonated cytosine residues are currently unknown.

The structural polymorphism of the expanded HR has implications for the development of drugs against ALS/FTD. One proposed therapy involves small molecules that can interact with expanded HR and block its protein-binding sites, preventing the formation of toxic foci and inhibiting RAN translation (17). Most investigations have focused on targeting G_4C_2 repeats with small molecules using library screening, whereas the druggability of G_2C_4 repeats has not been explored (17,20–23). To be effective, targeted therapies require the development of ligands exhibiting high structural specificity against target RNA (23,24). The rational design of lead compounds and potential therapeutics can be greatly improved by three-dimensional structures of RNA in a ligand-bound form. However, this knowledge is limited with regards to disease-related repeat sequences (24,25).

Recently, we developed a small molecule ANP77 (2-Amino-1,8-naphthyridine dimer), which belongs to the family of mismatch-binding ligands (MBLs) (26). The MBL molecules were designed to recognize bulges, internal loops or mismatches present in RNA and DNA structures (27–31). They contain at least two heterocyclic moieties that can form hydrogen bonds with nucleobases of single-stranded residues (32,33). The aromatic components are connected by linker that can vary in length and atomic composition, affecting the conformational flexibility of the ligand (26,31). The ANP77 molecule is composed of two 2-amino-1,8-naphthyridine (NU) units connected by a short linker composed of three carbon atoms (Figure 1A). The 2-amino derivatives of the 1,8-naphthyridine were used to bind Watson-Crick interface of cytosines while the linker was restricting conformational diversity and promoting stacked structure of the molecule (26,34,35). Further research using systematic evolution of ligands by exponential enrichment (SELEX) and UV-melting confirmed that ANP77 recognized two consecutive cytosine residues located in single stranded parts of nucleic acids such as bulges or apical loops (26,36,37). Finally, ANP77 was

found to interact with cytosine rich sequence of G_2C_4 repeats (36).

Here, we present the crystal structure of G_2C_4 RNA in complex with the synthetic molecule ANP77. RNA folds into a triplex-like structure involving protonated cytosine residues, whereas ANP77 interacts directly with RNA *via* pseudo-canonical base pairs. These crystallographic models are rare examples demonstrating unexplored potential of cytosine-rich sequences to form complex RNA structures and provide sophisticated crystallographic templates for structure-guided approaches in the development of lead compounds for drug discovery against ALS/FTD.

Materials and methods

Synthesis, purification and crystallization of RNA oligomers and ligand

The RNA oligomers were synthesized by the solid phase method using Applied Biosystems DNA/RNA synthesizer and TOM protected phosphoramidites. The synthesis was carried out in DMT-ON mode and RNA was purified according to the protocol suitable for Glen-Pak cartridges (Glen Research). The purified RNA was lyophilized under vacuum using a Speed-Vac and stored at -20°C . The ANP77 was chemically synthesized according to the protocol described before (26). The lyophilized ligand was dissolved in water to the final concentration of 30 mM and used for further research. Before crystallization the RNA was dissolved in buffer containing 10 mM sodium cacodylate pH 7.0 and 100 mM NaCl. The final concentration of RNA was 0.5 mM. The oligomer was denatured for 5 min at 95°C and snap-cooled on ice for 10 min following incubation for 20 min in RT with or without 0.75 mM ANP77. A minimum amount of ligand was added to saturate potential binding sites in the RNA since higher concentration of ligand caused RNA precipitation. Crystals were grown by the sitting drop method at 19°C . The initial screening (covering pH range 5.6–8.5) was performed to determine optimal crystallization conditions. The best hit was observed for buffers with pH 6.0. We have performed crystallization of RNA with and without ANP77 also in other pH values (pH 5.5 and pH 7.0) but usually the sample was precipitating or giving small micro crystals not suitable for X-ray measurements. The crystals of RNA-ligand complex grew in 80 mM NaCl, 40 mM sodium cacodylate trihydrate pH 6.0, 45% v/v MPD, 12 mM spermine tetrahydrochloride. Unliganded RNA oligomer was crystallized in 80 mM NaCl, 20 mM magnesium chloride hexahydrate, 40 mM sodium cacodylate trihydrate pH 6.0, 35% v/v MPD and 12 mM spermine tetrahydrochloride. The crystals of unliganded RNA were also obtained in 0.08M magnesium acetate tetrahydrate, 0.05M sodium cacodylate trihydrate pH 6.5, 30% w/v polyethylene glycol 4000. The crystallization drop contained 2:1 or 1:1 RNA-to-crystallization solution ratio. Crystals appeared within several weeks. The size of the crystals was approx. 0.05 mm with cuboid shape.

X-ray data collection, structure solution and refinement

X-ray diffraction data were collected on in house diffractometer (XtaLab Synergy-R, Rigaku) and on P13 beam line of the EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany) (38). The data were integrated using

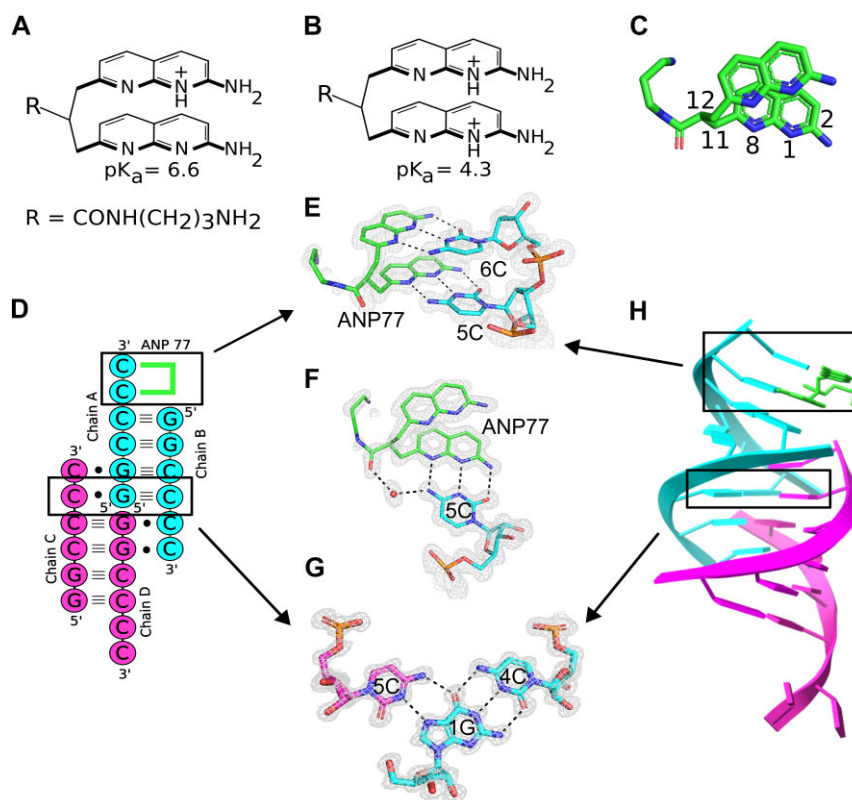


Figure 1. Interactions of ANP77 with G_2C_4 RNA. (A, B) chemical structure of the ANP77 molecule: (A) in single and (B) double-protonated state. (C) Stacked conformation of ANP77 observed in the crystal structure. Selected atoms are indicated (see text for details). (D) secondary structure of G_2C_4 -ANP77 complex. Green lines represent ANP77 ligand. (E, F) Pseudo-canonical base-pairs formed between ANP77 (green) and cytosine residues from chain A (cyan). (G) Base triple between cytosine from chain C (purple), guanosine from chain A (cyan), and cytosine from chain B (cyan). (H) Crystal model of RNA tetramer (chain A and B in cyan; chain C and D in pink) with bound ligand molecule (green sticks). The $2F_o - F_c$ electron density map (gray) is contoured at the 1σ level. The H-bonds are represented by black dashed lines.

CrysAlisPro software (Rigaku). The data were scaled using SCALA from CCP4 program suite (39,40). The structure of RNA-ligand complex was solved by molecular replacement with PHASER using part of RNA model containing C_4G_2 repeats (PDB code: 5ew7) (41). The phases of unliganded RNA model were taken from RNA-ligand structure. Early stages of the refinement were done using the Refmac5 from the CCP4 program suite. Further refinement was carried out with PHENIX (42,43). The manual model building was done using Coot (44). Restraints for the ANP77 ligand were generated by the grade web server (Global Phasing, <http://grade.globalphasing.org>). All figures were drawn using PyMOL (<https://pymol.org>). Atomic coordinates of the crystallographic models have been deposited in the Protein Data Bank (accession codes: 8QMH, 8QMI and 9EN6). The statistics of X-ray data collection and refinement are summarized in Supplementary Table S1. X-ray diffraction images are deposited in the MX-RDR database (Macromolecular Xtallography Raw Data Repository) (<https://mxrdr.icm.edu.pl/>) (doi:10.18150/AJNJBE for RNA-ANP77 structure and doi:10.18150/FESTPM, doi:10.18150/FRXAN8 for native RNA models).

CSI-TOF-MS measurements

Time-of-flight mass spectrometry analysis by cold spray ionization (CSI-TOF MS) data were obtained in negative mode using JEOL JMS-T100LP AccuTOF LC-plus 4G mass spec-

trometer. Samples containing the G_2C_4 RNA (20 μM) with or without ANP77 (50 μM) in an 8:2 mixture of water and methanol containing ammonium acetate (250 mM) were sprayed at a flow rate of 14 $\mu\text{l min}^{-1}$. During the injection, the spray temperature was set at -10°C .

Circular dichroism measurements

Circular dichroism (CD) experiments were carried out on a J-725 CD spectropolarimeter (JASCO) using a 10 mm path length cell at room temperature. Prior to measurements RNA (2.5 μM) was refolded in sodium cacodylate buffer (10 mM, pH 7.0) containing 100 mM NaCl for 5 min at 95°C and snap-cooled on ice for 10 min following incubation for 20 min at room temperature in the presence or absence of ANP77 ligand (25 μM). For each sample, five spectral scans were accumulated in the range from 220–400 nm.

CD in different temperature and pH values was carried out on a Jasco J-815 spectropolarimeter (JASCO) using a 5 mm path length. The pH-dependent CD spectra of RNA G_2C_4 (30 μM), $(G_2C_4)_2$ (14.6 μM) and $(G_2C_4)_3$ (10 μM) (refolded in the same way as above) were measured in a buffer containing 40 mM sodium cacodylate pH 5.5–7.0 or Tris-HCl pH 7.5–8.0 and 100 mM NaCl. The temperature-dependent CD spectra of G_2C_4 oligomer (30 μM) were obtained in sodium cacodylate buffer (40 mM, pH 6.0) containing 10 mM NaCl at 20–75 $^\circ\text{C}$. For each sample, five spectral scans were accumulated in the range from 205 to 340 nm.

DSC measurements

The RNA oligomers were dissolved in 100 mM sodium chloride and 10 mM sodium cacodylate buffer adjusted to pH 7.0, 6.0 or 5.3. The final concentration of RNA was 300 μ M. Samples were dialyzed against buffer overnight at 4°C to equilibrate RNA and the reference solutions. Before DSC measurements, RNA samples were denatured at 95°C for 5 min and snap-cooled on ice for 10 min. Next, the RNA was renatured for 20 min in RT with or without 300 μ M ANP77 ligand. DSC experiments were performed on a MicroCal PEAQ-DSC calorimeter (Malvern Instruments Ltd) Each measurement was carried out in five cycles of heating and cooling in the range of 2–110°C and a scan rate of 1°C/min. First, reference scans of the buffer were performed to establish the instrument thermal history and to reach a near perfect baseline repeatability. The results were analyzed using dedicated software implemented by Malvern Instruments. The melting temperature (T_m) was calculated by applying a two-state model fitting.

ITC measurements

ITC (Isothermal Titration Calorimetry) measurements were performed with a PEAQ-ITC calorimeter (Malvern) at 25°C in 10 mM sodium cacodylate pH 6.0 and 100 mM NaCl. 50 μ M RNA: G_2C_4 , $(G_2C_4)_2$ and $(G_2C_4)_3$ was titrated with 500 μ M ANP77 ligand until the saturation was observed. G_2C_4 RNA was titrated with 19 injections of 2 μ l volume, whereas $(G_2C_4)_2$ RNA with 38 injections separated with 160 s intervals. Data analysis was performed using Origin 7.0 software. The binding parameters were obtained by fitting the binding isotherms to a ‘One set of sites’ model for G_2C_4 RNA and ‘Sequential binding sites’ model (assuming two binding sites) for $(G_2C_4)_2$ RNA. The data obtained for $(G_2C_4)_3$ RNA could not be fitted and were unrepeatable. All measurements were performed in duplicate.

Native PAGE

For monitoring migration of RNA oligomers in native conditions 500–1000 pmol of G_2C_4 , $(G_2C_4)_2$, $(G_2C_4)_3$, 6U (6 uridines), 12U (12 uridines) and 18U (18 uridines) was renatured for 5 min at 95°C and slowly cooled (0.1°C/s) to 4°C in buffer containing 40 mM sodium cacodylate pH 6.0 and 100 mM NaCl. Next, glycerol was added to the final concentration of 1%. Samples were analyzed by native gel electrophoresis using 15% gel (19:1 acrylamide/bisacrylamide ratio) in 0.5 \times TB at 4°C (DNAspinter, Biovectis). RNA was visualized by UV shadowing.

Results and discussion

The overall structure of the G_2C_4 –ANP77 complex

The ANP77 ligand was designed to selectively bind to the internal loop of the Y/CC motif in RNA (Y denotes pyrimidine) (36,37). The ligand is composed of two 2-amino-1,8-naphthyridine units connected by a short aliphatic linker having attached aminopropyl carboxamide side chain to increase water solubility and interactions with negatively charged nucleic acids (Figure 1A–C). Each 2-amino naphthyridine unit served as a hydrogen bonding partner for interactions with cytosine residues. At a weakly acidic or neutral pH, protonation occurred at one of the NUs, whereas protonation of the second unit required a lower pH (Figure 1A and B) (36). The

preliminary experiments using UV-melting method confirmed the binding of ANP77 to the G_2C_4 sequence embedded in duplex structure (data not shown). Based on that finding we performed co-crystallization of G_2C_4 RNA with ligand molecule.

The three-dimensional model of the G_2C_4 –ANP77 complex exhibited a complicated triplex-like fold structure composed of four chains of G_2C_4 RNA (Protein Data Bank code 8QMH) (Supplementary Table S1). Chains A and B assembled into duplex AB, whereas chains C and D assembled into duplex CD (Figure 1D). In each duplex, four canonical G–C pairs were present as well as two single-stranded, overhanging cytosines at each 3′ end (Figure 1D). In the crystal lattice, duplexes AB and CD interacted with each other by docking 3′ overhanging cytosines from one end of the duplex into each other’s major grooves (Figure 1D). Cytosine from chain B interacted with duplex CD, whereas cytosine from chain C interacted with duplex AB. Consequently, chains A and D were arranged head-to-head, forming a pseudo-continuous strand. Overall, the RNA folded into a parallel triplex-like motif flanked by double-stranded duplexes and terminated by single-stranded cytosines (Figure 1D and H). The triplex motif was composed of four $C^+ \bullet G$ -C pairs representing standard major groove base triples (C^+ denotes protonated cytosine) (45,46). G and C formed canonical Watson–Crick pairs, whereas C^+ was located in the major groove and interacted with the Hoogsteen edge of guanosine from the G–C pair. The $C^+ \bullet G$ interaction included two H-bonds: one between the exo-amino group of C and the O6 carbonyl of G, and the second bond between the N3 amino group of C and the N7 imino atom of G (Figure 1G). The latter interaction required the protonation of cytosine, resulting in the conversion of the N3 imino atom of C^+ to an amino function. Although the crystal model was obtained at atomic resolution, the precise position of the hydrogen atoms could not be assigned. Since hydrogen atoms have a small contribution in scattering of X-rays they can be localized only in high quality and ultra-high resolution structures (47). On the other hand, the chemical knowledge and the distance (2.8–2.9 Å) between nitrogen atoms of cytosines and guanosines in the crystal model clearly indicated the presence of H-bond interactions and protonation of C residues in the observed triples.

In previous studies, G_2C_4 RNA was crystallized as a duplex (19). The cytosines were engaged in canonical pairing with guanosines and in the formation of non-canonical C–C pairs. Although cytosines can interact with each other to form H-bonds, crystallographic studies of C-rich sequences, such as CCG or CCUG repeats, have indicated that the system tends to minimize the number of C–C pairs while simultaneously maximizing the number of G–C pairs (19,48,49). This was achieved by strand slippage, resulting in 5′ or 3′ overhanging residues or the formation of other types of non-canonical pairs i.e. C–U instead of C–C (48,49). Our study confirmed this tendency. The G_2C_4 RNA formed slippery duplexes with two 3′ overhanging cytosines. In contrast to other studies, the terminal cytosines were ordered, and some were also engaged in the formation of a triplex structure.

Pseudocanonical ANP77–cytosine pairs

In the G_2C_4 –ANP77 structure, one ligand molecule was present. It interacted with the two overhanging cytosines of chain A, forming two pseudo-canonical base pairs (Figure 1D and H). The NUs of ANP77 mimicked the nitrogen bases

of nucleotides, forming three H-bonds with the functional groups of the overhanging cytosines from chain A at their Watson-Crick edges (Figure 1E and F). A hydrogen bond was observed between the 2-amino group of ANP77 and the O2 carbonyl atom of cytosine, a second bond between the N8 imino group of the ligand and the exo-amino group of C, and a third interaction between the N1 atom of naphthyridine and the N3 function of cytosine. This H-bonding implies the protonation of one of the nitrogen atoms. Despite the high quality of the electron density map, the localization of the protons within the two pseudo-pairs was ambiguous (Figure 1F). One proton was most likely located at one of the NUs because, at neutral pH, ANP77 is single-charged. The second proton can be attributed to either cytosine or second NU (Supplementary Figure S1). The second pair of overhanging cytosines from chain D was ordered and interacted with symmetry-related RNA and the 2-exo amino group of the ligand (Supplementary Figure S2A-B).

Conformation of ANP77 ligand

In the presented model, ANP77 exhibited a sandwich-like conformation, with the NUs shifting in relation to each other, resulting in limited stacking interactions between the aromatic moieties of the ligand (Figure 1C and E). One of the NUs overlapped extensively with the neighbouring 1G from chain B (Figure 1H), while the second unit did not form stacking interactions, despite its close proximity to the G-C pair of symmetry-related RNA molecules.

The aminopropyl carboxamide side chain of ANP77 was located in the major groove of RNA. Its amino-terminal region was disordered, indicating conformational freedom. Only the carbonyl group of the amide bond was modeled; however, it was in a double conformation. Depending on the position of the carbonyl group, it interacted with water molecules from the Hoogsteen edge of 5C involved in pseudo-pairing or with water molecules bridging the Hoogsteen edge of 4C from chain A (Figure 1F).

The crystallographic model of the G₂C₄-ANP77 complex confirmed that the ligand was capable of interacting directly with two consecutive cytosines. Through the formation of pseudo-base pairs, it merged into an RNA helix, extending the length of the double-stranded region (Figure 1H). This indicates that ANP77 fits into the structure not only in terms of H-bonding, but also in terms of the shape of the RNA helix. The geometry of the linker allowed the positioning of its aromatic rings in a similar manner as the nitrogen bases arranged in a nucleic acid helix. The NUs stacked one above another and were approximately 3 Å apart, reflecting the value of the rise parameter in A-RNA form. Moreover, they were twisted to each other by approximately 25°, and the degree of twist could easily be adjusted by rotation around the C11-C12 bond (Figure 1C). The length of the propane linker seemed to be properly selected, as it allowed the binding of only two consecutive cytosines. Adding more carbon atoms would increase the flexibility and conformational freedom of the ligand, which could have a negative impact on ANP77 selectivity.

In light of structural data, ANP77 seemed to be a well-designed small molecule for sensing single stranded neighbouring cytosine residues in RNA. In practice, the ANP77 could be considered as an independent ligand molecule targeting C-rich sequences in RNA or as a part of a larger multivalent compound (50). In the single or multimeric form, ANP77

could bind to the G₂C₄ repeats and block sequestration of proteins, preventing formation of nuclear foci. Alternatively, ANP77 could be attached to the other heterocycle locally activating target RNA degradation by RNase L (51). Another application could explore fluorescent properties of ANP77 *in vivo* as an indicator of the Y/CC motifs in the RNA structures as it was shown earlier (37).

Using ANP77 as a tool in medicine raises the question about possible off-target effects since the Y/CC motif can be present in different RNA molecules. However, the previous research using SELEX and UV-melting method showed that the ligand requires not only the presence of two consecutive cytosine residues in single stranded region but also the neighbouring sequences play important role in RNA-ligand interaction (26,36,37). Moreover, crystallographic data indicated that the structural arrangement of cytosine residues is important for ANP77 recognition. Upon ligand binding the cytosines had to be in a single stranded form but in a way to maintain helical character of double-stranded region. In case of the G₂C₄ repeats the off-target effect could be also minimized by the high abundance of mutated transcripts in relation to other RNA molecules. Nevertheless, the further improvement of selectivity against specific RNA could be obtained by modification of the ANP77 molecule or incorporation into the larger, multivalent compound.

How to improve the properties of the ANP77 ligand

The ANP77 molecule was designed to interact with two adjacent cytosines. Therefore, a number of features had to be taken into consideration during development of the ligand to ensure specific interaction with the target sequence (26,36). First, functional groups of NUs were selected as an interface for direct interactions with Watson-Crick edges of cytosines. Second, the aromatic rings were connected by short linker consisting of three carbon atoms in order to restrict conformational freedom of the molecule and facilitate formation of the stacked structure of the ligand. Third, the ANP77 possessed aminopropyl carboxamide side chain which easily undergo protonation increasing water solubility and interactions with negatively charged nucleic acids.

Our model provided the opportunity to rationally design modifications of ANP77 to improve its interaction with RNA in terms of the specificity, stability, and structural features of the target sequence. One modification would be to induce the protonation of both NUs because, at a neutral pH, only one of the nitrogen atoms (N1 or N1') possesses proton (36). In the crystal structure, it acted as a hydrogen-bond donor for cytosine's N3 imino group. In turn, a maximum of three H-bonds were formed between naphthyridine and cytosine. Securing the protonation of both units could facilitate tight binding and increase the ligand affinity for the target RNA. The second modification could involve redesigning the side chain of the ligand, which is located close to the major groove of the helix. The electrostatic potential generated by exo-amino groups of cytosines was predominantly positive, and most of the exo-amino groups interacted with water molecules, indicating their binding potential. Replacing the protonated amine located at the end of the side chain with a more negatively charged group could be more beneficial. One possible example is the hydroxyl group, which has an electronegative character and can serve as an acceptor of the H-bond.

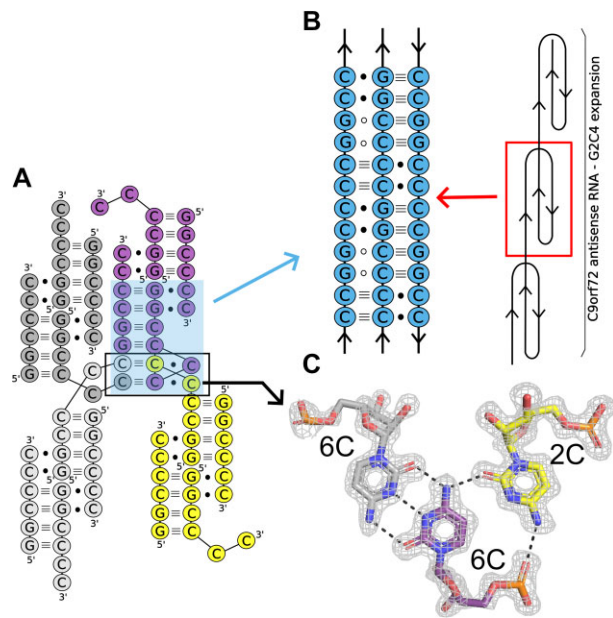


Figure 2. Higher-order interactions of unliganded RNA containing G_2C_4 repeats. **(A)** Cytosine residues of three symmetry-related molecules (grey, yellow and purple) form two unique $C\bullet C^+\bullet C$ base triples (black box). **(B)** Proposed model of triple helix structure formed by expanded G_2C_4 repeats (see text for details). Open circles denote close proximity of G residues and/or H-bond interactions with the C-G pairs. **(C)** Example of the $C\bullet C^+\bullet C$ triple assembled by the trans $C\bullet C^+$ pair (purple and gray) found in the *i*-motif structure and cytosine (yellow) located in the Hoogsteen edge moiety. The $2F_o - F_c$ electron density map (grey) is contoured at the 1σ level. The H-bonds are represented by black dashed lines.

Unliganded structure of G_2C_4 RNA

We determined the crystal structure of G_2C_4 in its free form at pH 6.0 (Protein Data Bank code: 8QMI) and at pH 6.5 (Protein Data Bank code 9EN6) (Supplementary Table S1). The superposition of both structures showed high similarity with r.m.s.d. of 0.38 Å. The overall assembly of the triple helix structure remained the same. The two terminal cytosines of chain A, which were bound to ANP77 in the liganded structure, engaged in higher-order interactions (Figure 2A). They formed peculiar $C\bullet C^+\bullet C$ base triples assembled by symmetry-related residues (5C of chain A, 6C of symmetry-related chain A, and 6C of symmetry-related chain D) (Figures 2A and C). The base triples consisted of a trans $C\bullet C^+$ pair commonly found in *i*-motif structures. The N-glycosidic bonds were oriented in trans relative to each other and one of the cytosines (the N3 atom) was protonated. The $C\bullet C^+$ pair formed three hydrogen bonds *via* Watson–Crick edges. The third C from the triples was located in the Hoogsteen edge of one of the cytosine residues from the *i*-motif pair. It interacted with two H-bonds between the wedged cytosine and the cytosine from the $C\bullet C^+$ pair (Figure 2C).

The unique triple helix-*i*-motif observed in ligand-free G_2C_4 RNA has not yet been described. This emphasizes the potential of cytosines to assist in the formation of higher-order RNA structures, presenting an example of diverse RNA folding pathways. In our opinion, the structural richness of G_2C_4 RNA could be implemented in structural biology, that is, in the folding of RNA nanoparticles, the assembly of RNA into multimers for CryoEM measurements, or the enhancement of crystal lattice formation during crystallization. In the con-

text of expanded G_2C_4 repeats, the obtained results indicated that the G_2C_4 RNA antisense strand can fold not only into simple hairpins, but also into diverse three-dimensional structures (triplexes or *i*-motifs). Based on our data, we propose a model in which the long tracts of G_2C_4 repeats form long triple helixes consisting of three different base triples: $C\bullet C^+\bullet C$, $C^+\bullet G-C$, and $G\bullet C-G$ (Figure 2B). This structural arrangement allows cytosines to engage in more favorable H-bond interactions than those observed in the crystal structure of the G_2C_4 duplex. Triple helixes are involved in the folding of RNAs into complex three-dimensional architectures, and they are crucial for the biological activity of RNA, such as telomere synthesis, ribosomal frame-shifting, regulation of gene expression through metabolite sensing, and the protection of RNA from degradation (52–61). The triple RNA helix formed by G_2C_4 repeats can be another example of a triple helix that plays an important role in cells, exemplifying the pathological role of HRs. In that case the triple helix of RNA G_2C_4 repeats, next to hairpin, tetraplex or *i*-motif structures, could be also involved in protein sequestration and nuclear foci formation or serving as a binding site of ribosomes during RAN-translation process. In case of the physiological range of HRs the triple helix structure could stabilize the mRNA or participate in binding of specific regulatory factors.

Interactions between RNA molecules and ions

In both structures interactions between RNA and ion species were observed. In the liganded structure three chloride ions were located: two in the major groove of the triple helix motif and the third ion between the end of the triple helix and beginning of the flanking duplex (Supplementary Figure S3A). All ions interacted with the exo-amino groups of cytosine residues. The distance between acceptor and H-bond donor was within the range of 3.2–3.4 Å, confirming the appropriate interpretation of electron density peaks. In the unliganded structure two magnesium cations were observed. Both of them directly interacted with oxygen atoms of the phosphate groups of RNA (Supplementary Figure S3B). The first one was bound to the 4C residue of chain A. Its coordination sphere consisted of the oxygen atom and five water molecules, typical for magnesium cations. The distances between acceptors and donors were in the range 2.0–2.3 Å. One of the water molecules from the magnesium coordination sphere interacted also with the oxygen atom of the phosphate group of the symmetry-related 6C residue of chain B (Supplementary Figure S3B). The second magnesium ion was located on a 2-fold axis between two phosphate groups of symmetry-related RNA molecules. Simultaneously, cation was directly bonded to two oxygen atoms (one O atom from phosphate of the 6C residue of chain C and second O atom of symmetry-related 6C residue of chain C) and four water molecules. The H-bond distances were also within the 2.0–2.2 Å range.

The interactions between RNA molecules and cations can provide information about local charge generated by the functional groups of the nucleotides. In case of the G_2C_4 sequence the abundance of cytosine residues resulted in the positive charge of the major groove which was attracted by the chloride anions. The similar effect was observed for oligomers containing the UGG motif (62). The major groove of the helix was negatively charged by carbonyl groups of guanosine residues. This attracted cations, especially Ba^{2+} , whose presence increased thermal stability of RNA duplexes. For the

phosphate groups, being in close proximity, their high negative charge was usually neutralized by Mg^{2+} ions as it was observed in unliganded structure. In this case the Mg^{2+} participated in the crystal lattice formation but in many RNA crystal structures it was shown to be crucial for proper RNA folding and activity (63).

Biochemical analysis of G_2C_4 -ANP77 complex

Differential scanning calorimetry (DSC) was used to assess the thermodynamics of ligand binding and folding of G_2C_4 RNA. Similarly to crystallization experiments, we performed measurements at pH 6.0. Two peaks were observed for the unliganded RNA, indicating melting of the RNA structure: Tm1 (triplex melting point) at 38.1°C and Tm2 (duplex melting point) at 56.8°C (Figure 3A). In the presence of ANP77, Tm1 was shifted to 41.5°C, while Tm2 slightly changed to (57.4°C) (Figure 3B and Supplementary Table S2). Considering the importance of pH for the protonation of cytosines in terms of the folding of G_2C_4 RNA into a triplex structure and ligand binding, we performed DSC measurements at pH 7.0 and 5.3. At pH 7.0, for unliganded G_2C_4 , only a single peak (Tm2) was detected at 55.3°C (1.5°C lower than Tm2 at pH 6.0) (Figure 3A). The addition of the ligand did not alter the DSC profile (single peak at 54.4°C) (Figure 3B and Supplementary Table S2). At pH 5.3, one peak Tm1 + 2 (merged melting points of triplex and duplex) was also detected, but at higher temperatures, either for RNA alone (Tm1 + 2 = 59.6°C) and the RNA-ligand complex (Tm1 + 2 = 60.2°C). However, the height and profile of the peak at pH 5.3 was different from those at pH 6.0 and 7.0.

Mass spectra indicated the presence of double- and tetra-stranded form of unliganded G_2C_4 RNA and only a double-stranded form of the RNA-ligand sample. In the latter, two complexes could be distinguished, representing 1:1 and 1:2 RNA-to-ligand ratios, which are consistent with the crystallographic model showing two potential ANP77 binding sites (Figure 3C and D).

CD measurements were performed for RNA composed of one, two, and three G_2C_4 repeats at six different pH values (from pH 5.5 to 8.0). The CD spectra of all oligomers at pH 5.5 showed a similar profile: the high negative peak at 215 nm and the high broad positive peak at 272 nm (Figure 3E and Supplementary Figure S4). At higher pH values we observed a decrease in the peak height at 215 nm and a shift of the positive peak from 272 nm to 270 nm (Figure 3E and Supplementary Figure S4). The most significant changes in CD spectra were detected between pH 5.5 and 6.0. The obtained CD spectra, together with the observed reduction of the negative peak at 215 nm and shift of positive peak at higher pH suggests the presence of triplex structure which is in agreement with previous studies (64–66). The CD method was also used to identify the structural changes upon melting of the G_2C_4 oligomer at pH 6.0. Measurements were performed from 20 to 75°C (5°C step). Two sequential transitions could be distinguished: one from 20° to 35°C and second from 40° to 75°C. In the first transition, the 215 nm negative peak increased from –34.49 to –22.76 mdeg while the 272 nm peak decreased only slightly (Figure 3F). In the second transition, the negative peak disappeared at 55°C, while the 272 nm peak decreased significantly and shifted to 280 nm. Finally, CD measurements were also performed for the RNA oligomers having two and five repeats in the presence and absence of ANP77 ligand.

Binding of the ligand induced spectral changes in the range of 300–370 nm (Supplementary Figure S5) which may result from the fact that the achiral ANP77 became a part of the chiral RNA-ANP77 complex.

In order to obtain more information regarding the binding of ANP77 to investigated RNA we preformed ITC measurements for G_2C_4 RNA having one, two and three repeats. The G_2C_4 RNA oligomer interacted with ANP77 with K_d of $7.9 \pm 1.0 \mu M$ (Supplementary Figure S6A). In case of $(G_2C_4)_2$ the ITC showed the best fit to a sequential binding sites model assuming the presence of two binding sites (Supplementary Figure S6B). The average K_d , obtained from two separate experiments, for the first binding site was $976 \pm 377 nM$ while for the second binding site the K_d was $7.6 \pm 4.2 \mu M$. The data obtained for titration of the $(G_2C_4)_3$ could not be fitted and were unrepeatable (data not shown).

The affinity of ANP77 obtained for G_2C_4 RNA was rather low. This may be due to a fact that in solution two overhanging cytosines most likely presented structural freedom influencing the interaction with ANP77. In case of the $(G_2C_4)_2$ RNA two binding sites were identified with different affinity which suggested potential structural heterogeneity of the sample and coexistence of different structural folds of G_2C_4 RNA. The low affinity binding site (with similar K_d as for G_2C_4 RNA) may come from the triplex structure while high affinity binding site may represent single stranded cytosines located in constrained environment (inside the RNA duplex). The coexistence of multiple structural forms of longer G_2C_4 RNA has been observed by others (19). This potential heterogeneity could also explain why we were unable to obtain crystals of complexes having two and more G_2C_4 repeats.

The native PAGE electrophoresis of $(G_2C_4)_{1-3}$ oligomers showed that all oligomers migrated slower than the reference markers of the same length, suggesting formation of inter-molecular structures (Supplementary Figure S7). For the $(G_2C_4)_3$ oligomer additional fast migrating band was observed suggesting the presence of intra-molecular species.

X-ray structure versus biochemical data

The results of the biophysical evaluation can be interpreted in terms of structural data. In G_2C_4 RNA, ANP77 interacted with cytosine residues located in the single-stranded region. Although six H-bonds were formed between RNA and the ligand, the stacking interactions were limited, and the entropy of the cytosines was reduced, resulting in an elusive thermal effect. DSC measurements, CD and mass spectra indicated that in solution, G_2C_4 existed as a tetramer folded into a triplex structure (Figure 3). The best conditions for triplex formation were slightly acidic (pH 5.5–6.0). At pH 6.0 the DSC and CD-melting spectra showed two peaks and two structural transitions, respectively (Figure 3A and F). The first peak and the first transition likely corresponded to the melting of the Hoogsteen interactions in the base triples, whereas the second peak and transition represented the melting of the slippery duplex. The DSC spectra at pH 5.3 also suggested the presence of one stable structure, either triplex or i-motif. Under acidic conditions the triplex motif could be stabilized, resulting in the merging of the duplex and tetramer signals into one peak (46). Alternatively, i-motif could be present since it can be easily formed at pH < 5.5, although the CD spectra did not support this hypothesis (67,68). At pH 7.0, the protonation of cytosines required for base triple formation was more

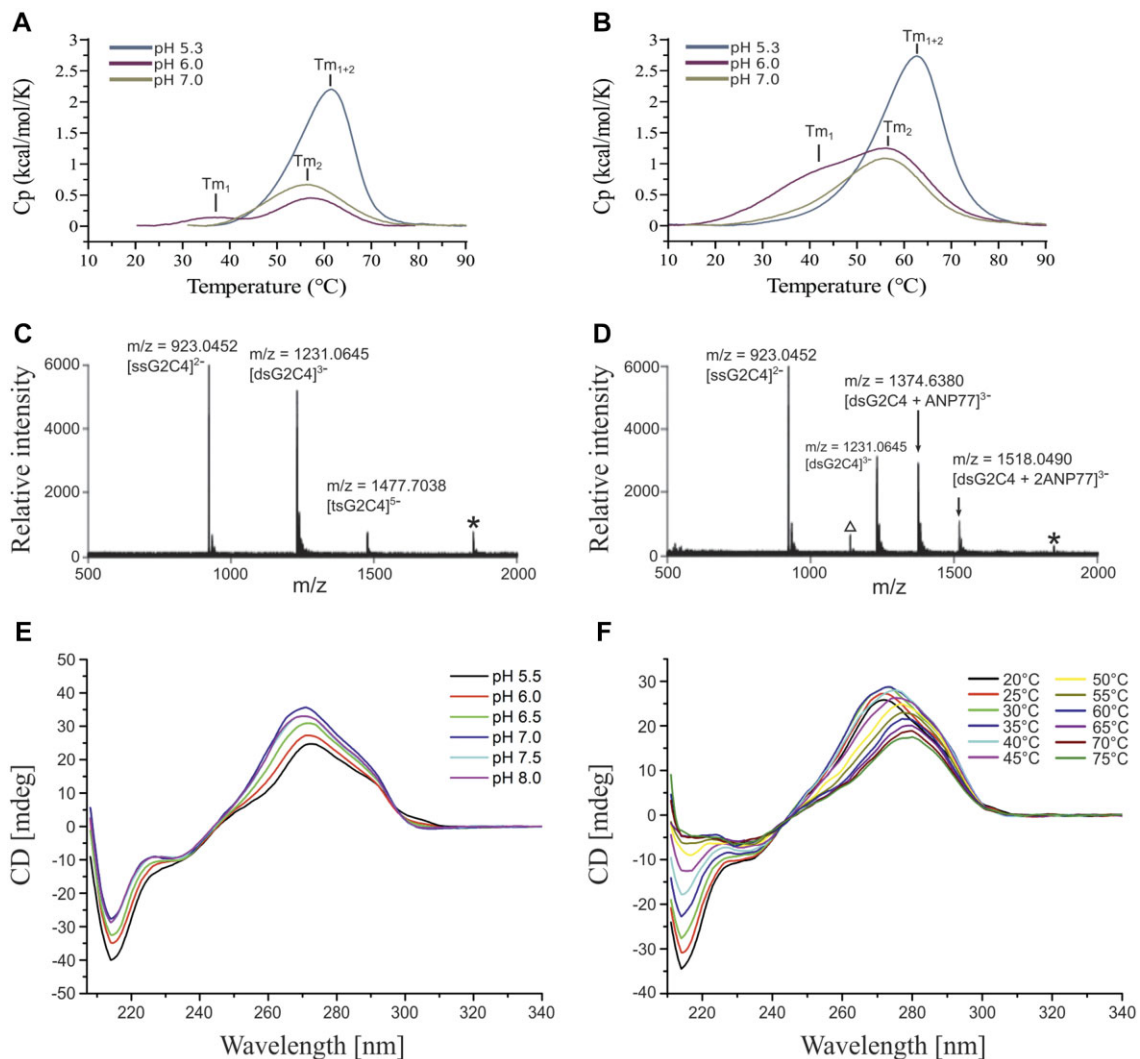


Figure 3. Bio-physical evaluation of interactions between RNA and ANP77 ligand. (A, B) Differential scanning calorimetry (DSC) spectra measured for (A) RNA oligomer and (B) for RNA-ligand complex in different pH conditions. (C-D) Cold-spray ionization time-of-flight mass spectrometry (CSI-TOF-MS) of G₂C₄ RNA (20 mM) in the (C) absence and (D) presence of ANP77 (50 mM). Asterisk (*) represents an ion corresponding to the 2⁻ ion of dsG₂C₄ overlapping with the 1⁻ ion of ssG₂C₄. The open triangle represents ion that could not be assigned. ssG₂C₄, dsG₂C₄, and tsG₂C₄ denote single-stranded, double-stranded, and tetra-stranded G₂C₄ RNA, respectively. (E, F) Circular dichroism spectra of G₂C₄ RNA (30 μM) measured in different pH (E) and temperatures (F).

difficult, explaining the sole presence of duplex species. Similar results of DSC and CD measurements were obtained by Dodd et al. (19). They investigated RNA composed of different numbers of G₂C₄ repeats. The authors suggested the presence of alternative structures including slipped intermolecular and intramolecular states or non-canonical structures.

pH-dependent stabilization of the pyrimidine major groove triplexes was demonstrated by thermodynamic and NMR studies. Under acidic conditions, a higher efficiency of cytosine residue protonation results in the tightening of Hoogsteen interactions in triple base pairs (65,69,70). However, this requirement for protonation to form tertiary contacts does not imply a low stability of this motif in the cellular environment (71). Although the calculated pK_a of the isolated cytosine was below pH 5.0, the structural context and solvent content indicate the protonation of C, even under basic conditions (72,73). The pK_a shift in cytosine has been observed in several crystal structures of pseudoknot from Beet Western Yellow Virus, Pea Enation Mosaic Virus and HDV ribozyme (57,74–78). These

studies demonstrated that formation of the C⁺GCA motif was crucial for proper folding and RNA activity. The protonated cytosine residue was engaged in formation of junction of two helical stems by forming H-bonds with Hoogsteen edges of G–C pair and neighboring adenosine residue. The assembly of G₂C₄ into the triple helix also requires the protonation of cytosines. Our crystallographic and biophysical results, together with previous reports showed that C⁺ species can be observed under near-physiological conditions, suggesting that they can also be stable in the cellular matrix and/or inside insoluble nuclear foci (18,79).

Structural polymorphisms of the sense and antisense transcripts of HR result in complex pathological pathways in ALS/FTD. Our study demonstrates the potential of C-rich G₂C₄ repeats to form higher-order structures. The ‘driving force’ for RNA folding are cytosine residues. The limited nucleotide composition of G₂C₄ RNA results in a pK_a shift of cytosines, extending their capabilities for H-bonding under near-physiological conditions. Consequently, a triplex struc-

ture, containing C⁺•G–C and unique C•C⁺•C i-motif base triples, is formed. In the cellular environment (particularly inside the nuclear foci) some local pH and ionic gradients as well as molecular crowding effects could stabilize HR repeats in the triplex form. Thus, the presented fold of G₂C₄ repeats is another example of RNA structural diversity and can be considered as a platform for RNA drug development against ALS/FTD.

ANP77 is a binding candidate for C-rich G₂C₄ RNA. It directly binds to adjacent cytosine residues and forms pseudo-canonical base pairs. This is our second study indicating that small molecules interact with the Watson–Crick edges of nucleotides. We previously showed that cyclic mismatch-binding ligand (CMBL) exhibited specificity toward adenosine residues in CAG repeats associated with polyglutamine disorders (33). We demonstrated that ANP77 and CMBL do not require the formation of ‘pockets’ for specific RNA recognition but rather the presence of particular structural motifs. Bioinformatic tools are one strategy for design and identification of lead compounds (24). The crystallographic data for ANP77 and CMBL ligands and its derivatives can be used for *in silico* predictions, providing detailed information about their structure and interactions with RNA which could accelerate drug development against RNA-driven disorders.

Data availability

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession numbers 8QMH and 8QMI. Diffraction images have been deposited in Macromolecular Xtallography Raw Data Repository at <https://doi.org/10.18150/AJNJB>; <https://doi.org/10.18150/FESTPM>; <https://doi.org/10.18150/FRXAN8>.

Supplementary data

Supplementary Data are available at NAR Online.

Funding

National Science Centre [UMO-2022/45/B/NZ7/03543, UMO-2017/26/E/NZ1/00950 to A.K.]; Japan Society for the Promotion of Science KAKENHI Grant-in-Aid for Scientific Research (A) [22H00351 to K.N.]. For Open Access publishing, the authors have applied a CC-BY public copyright license to any Author Accepted Manuscript (AAM) version arising from this submission. The synchrotron data was collected at P13 beamline operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany) [Proposal number MX-989]. We would like to thank Selina Storm for the assistance in using the beamline. Funding for open access charge: National Science Centre [UMO-2022/45/B/NZ7/03543].

Conflict of interest statement

None declared.

References

- Devenney,E.M., Ahmed,R.M. and Hodges,J.R. (2019) Frontotemporal dementia. *Handb. Clin. Neurol.*, **167**, 279–299.
- Geser,F., Lee,V.M. and Trojanowski,J.Q. (2010) Amyotrophic lateral sclerosis and frontotemporal lobar degeneration: a spectrum of TDP-43 proteinopathies. *Neuropathology*, **30**, 103–112.
- DeJesus-Hernandez,M., Mackenzie,I.R., Boeve,B.F., Boxer,A.L., Baker,M., Rutherford,N.J., Nicholson,A.M., Finch,N.A., Flynn,H., Adamson,J., *et al.* (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*, **72**, 245–256.
- Beck,J., Poulter,M., Hensman,D., Rohrer,J.D., Mahoney,C.J., Adamson,G., Campbell,T., Uphill,J., Borg,A., Fratta,P., *et al.* (2013) Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. *Am. J. Hum. Genet.*, **92**, 345–353.
- van Blitterswijk,M., DeJesus-Hernandez,M., Niemantsverdriet,E., Murray,M.E., Heckman,M.G., Diehl,N.N., Brown,P.H., Baker,M.C., Finch,N.A., Bauer,P.O., *et al.* (2013) Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study. *Lancet Neurol.*, **12**, 978–988.
- Haeusler,A.R., Donnelly,C.J., Periz,G., Simko,E.A., Shaw,P.G., Kim,M.S., Maragakis,N.J., Troncoso,J.C., Pandey,A., Sattler,R., *et al.* (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*, **507**, 195–200.
- Kumar,V., Kashav,T., Islam,A., Ahmad,F. and Hassan,M.I. (2016) Structural insight into C9orf72 hexanucleotide repeat expansions: towards new therapeutic targets in FTD-ALS. *Neurochem. Int.*, **100**, 11–20.
- Gendron,T.F., Bieniek,K.F., Zhang,Y.J., Jansen-West,K., Ash,P.E., Caulfield,T., Daugherty,L., Dunmore,J.H., Castanedes-Casey,M., Chew,J., *et al.* (2013) Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathol.*, **126**, 829–844.
- Mizielinska,S., Lashley,T., Norona,F.E., Clayton,E.L., Ridler,C.E., Fratta,P. and Isaacs,A.M. (2013) C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. *Acta Neuropathol.*, **126**, 845–857.
- Mori,K., Lammich,S., Mackenzie,I.R., Forne,I., Zilow,S., Kretschmar,H., Edbauer,D., Janssens,J., Kleinberger,G., Cruts,M., *et al.* (2013) hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. *Acta Neuropathol.*, **125**, 413–423.
- Mizielinska,S., Gronke,S., Niccoli,T., Ridler,C.E., Clayton,E.L., Devoy,A., Moens,T., Norona,F.E., Woollacott,I.O.C., Pietrzyk,J., *et al.* (2014) C9orf72 repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins. *Science*, **345**, 1192–1194.
- Mori,K., Arzberger,T., Grasser,F.A., Gijssels,I., May,S., Rentzsch,K., Weng,S.M., Schludi,M.H., van der Zee,J., Cruts,M., *et al.* (2013) Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathol.*, **126**, 881–893.
- Mori,K., Weng,S.M., Arzberger,T., May,S., Rentzsch,K., Kremmer,E., Schmid,B., Kretschmar,H.A., Cruts,M., Van Broeckhoven,C., *et al.* (2013) The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTD/ALS. *Science*, **339**, 1335–1338.
- Bose,K., Maity,A., Ngo,K.H., Vandana,J.J., Shneider,N.A. and Phan,A.T. (2022) Formation of RNA G-wires by G(4)C(2) repeats associated with ALS and FTD. *Biochem. Biophys. Res. Commun.*, **610**, 113–118.
- Fratta,P., Mizielinska,S., Nicoll,A.J., Zloh,M., Fisher,E.M., Parkinson,G. and Isaacs,A.M. (2012) C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci. Rep.*, **2**, 1016.
- Jain,A. and Vale,R.D. (2017) RNA phase transitions in repeat expansion disorders. *Nature*, **546**, 243–247.

17. Su,Z., Zhang,Y., Gendron,T.F., Bauer,P.O., Chew,J., Yang,W.Y., Fostvedt,E., Jansen-West,K., Belzil,V.V., Desaro,P., *et al.* (2014) Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. *Neuron*, **83**, 1043–1050.
18. Kovanda,A., Zalar,M., Sket,P., Plavec,J. and Rogelj,B. (2015) Anti-sense DNA d(GGCCCC)n expansions in C9ORF72 form i-motifs and protonated hairpins. *Sci. Rep.*, **5**, 17944.
19. Dodd,D.W., Tomchick,D.R., Corey,D.R. and Gagnon,K.T. (2016) Pathogenic C9ORF72 antisense repeat RNA forms a double helix with tandem C:c mismatches. *Biochemistry*, **55**, 1283–1286.
20. Simone,R., Balendra,R., Moens,T.G., Preza,E., Wilson,K.M., Heslegrave,A., Woodling,N.S., Niccoli,T., Gilbert-Jaramillo,J., Abdelkarim,S., *et al.* (2018) G-quadruplex-binding small molecules ameliorate C9orf72 FTD/ALS pathology in vitro and in vivo. *EMBO Mol. Med.*, **10**, 22–31.
21. Ursu,A., Baisden,J.T., Bush,J.A., Taghavi,A., Choudhary,S., Zhang,Y.J., Gendron,T.F., Petrucelli,L., Yildirim,I. and Disney,M.D. (2021) A small molecule exploits hidden structural features within the RNA repeat expansion that causes c9ALS/FTD and rescues pathological hallmarks. *ACS Chem. Neurosci.*, **12**, 4076–4089.
22. Ursu,A., Wang,K.W., Bush,J.A., Choudhary,S., Chen,J.L., Baisden,J.T., Zhang,Y.J., Gendron,T.F., Petrucelli,L., Yildirim,I., *et al.* (2020) Structural features of small molecules targeting the RNA repeat expansion that causes genetically defined ALS/FTD. *ACS Chem. Biol.*, **15**, 3112–3123.
23. Wang,Z.F., Ursu,A., Childs-Disney,J.L., Guertler,R., Yang,W.Y., Bernat,V., Rzuczek,S.G., Fuerst,R., Zhang,Y.J., Gendron,T.F., *et al.* (2019) The hairpin form of r(G(4)C(2))(exp) in c9ALS/FTD is repeat-associated and targets for bioactive small molecules. *Cell Chem. Biol.*, **26**, 179–190.
24. Childs-Disney,J.L., Yang,X., Gibaut,Q.M.R., Tong,Y., Batey,R.T. and Disney,M.D. (2022) Targeting RNA structures with small molecules. *Nat. Rev. Drug Discov.*, **21**, 736–762.
25. Warner,K.D., Hajdin,C.E. and Weeks,K.M. (2018) Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.*, **17**, 547–558.
26. Nakatani,K., Natsuhara,N., Mori,Y., Mukherjee,S., Das,B. and Murata,A. (2017) Synthesis of naphthyridine dimers with conformational restriction and binding to DNA and RNA. *Chemistry*, **12**, 3077–3087.
27. Kobori,A., Horie,S., Suda,H., Saito,I. and Nakatani,K. (2004) The SPR sensor detecting cytosine[bond]cytosine mismatches. *J. Am. Chem. Soc.*, **126**, 557–562.
28. Nakatani,K., Sando,S., Kumasawa,H., Kikuchi,J. and Saito,I. (2001) Recognition of guanine-guanine mismatches by the dimeric form of 2-amino-1,8-naphthyridine. *J. Am. Chem. Soc.*, **123**, 12650–12657.
29. Shibata,T., Nagano,K., Ueyama,M., Ninomiya,K., Hirose,T., Nagai,Y., Ishikawa,K., Kawai,G. and Nakatani,K. (2021) Small molecule targeting r(UGGAA)(n) disrupts RNA foci and alleviates disease phenotype in Drosophila model. *Nat. Commun.*, **12**, 236.
30. Nakamori,M., Panigrahi,G.B., Lanni,S., Gall-Duncan,T., Hayakawa,H., Tanaka,H., Luo,J., Otabe,T., Li,J., Sakata,A., *et al.* (2020) A slipped-CAG DNA-binding small molecule induces trinucleotide-repeat contractions in vivo. *Nat. Genet.*, **52**, 146–159.
31. Nakatani,K. (2009) Recognition of mismatched base pairs in DNA. *B Chem Soc Jpn*, **82**, 1055–1069.
32. Nakatani,K., Hagihara,S., Goto,Y., Kobori,A., Hagihara,M., Hayashi,G., Kyo,M., Nomura,M., Mishima,M. and Kojima,C. (2005) Small-molecule ligand induces nucleotide flipping in (CAG)n trinucleotide repeats. *Nat. Chem. Biol.*, **1**, 39–43.
33. Mukherjee,S., Blaszczak,L., Rypniewski,W., Falschlunger,C., Micura,R., Murata,A., Dohno,C., Nakatani,K. and Kiliszek,A. (2019) Structural insights into synthetic ligands targeting A-A pairs in disease-related CAG RNA repeats. *Nucleic Acids Res.*, **47**, 10906–10913.
34. Shibata,T. and Nakatani,K. (2016) Fluorescence probe for detecting CCG trinucleotide repeat DNA expansion and slip-out. *Chembiochem*, **17**, 1685–1688.
35. Zhang,J., Takei,F. and Nakatani,K. (2007) Emission of characteristic fluorescence from the ligand-cytosine complex in U_A/ACU bulged RNA duplex. *Bioorg. Med. Chem.*, **15**, 4813–4817.
36. Das,B., Nagano,K., Kawai,G., Murata,A. and Nakatani,K. (2022) 2-Amino-1,8-naphthyridine dimer (ANP77), a high-affinity binder to the internal loops of C/CC and T/CC sites in double-stranded DNA. *J. Org. Chem.*, **87**, 340–350.
37. Das,B., Murata,A. and Nakatani,K. (2021) A small-molecule fluorescence probe ANP77 for sensing RNA internal loop of C, U and A/CC motifs and their binding molecules. *Nucleic Acids Res.*, **49**, 8462–8470.
38. Cianci,M., Bourenkov,G., Pompidor,G., Karpics,I., Kallio,J., Bento,I., Roessle,M., Cipriani,F., Fiedler,S. and Schneider,T.R. (2017) P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high- and low-energy phasing with variable beam focusing. *J. Synchrotron Radiat.*, **24**, 323–332.
39. Agirre,J., Atanasova,M., Bagdonas,H., Ballard,C.B., Basle,A., Beilsten-Edmands,J., Borges,R.J., Brown,D.G., Burgos-Marmol,J.J., Berrisford,J.M., *et al.* (2023) The CCP4 suite: integrative software for macromolecular crystallography. *Acta Crystallogr. Sect. D, Struct. Biol.*, **79**, 449–461.
40. Evans,P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. Sect. D, Biol. Crystallogr.*, **62**, 72–82.
41. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
42. Afonine,P.V., Grosse-Kunstleve,R.W., Echols,N., Headd,J.J., Moriarty,N.W., Mustyakimov,M., Terwilliger,T.C., Urzhumtsev,A., Zwart,P.H. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D, Biol. Crystallogr.*, **68**, 352–367.
43. Kovalevskiy,O., Nicholls,R.A., Long,F., Carlon,A. and Murshudov,G.N. (2018) Overview of refinement procedures within REFMAC5: utilizing data from different sources. *Acta Crystallogr. Sect. D, Struct. Biol.*, **74**, 215–227.
44. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D, Biol. Crystallogr.*, **66**, 486–501.
45. Brown,J.A. (2020) Unraveling the structure and biological functions of RNA triple helices. *Wiley Interdiscipl. Rev. RNA*, **11**, e1598.
46. Devi,G., Zhou,Y., Zhong,Z., Toh,D.F. and Chen,G. (2015) RNA triplexes: from structural principles to biological and biotech applications. *Wiley Interdiscipl. Rev. RNA*, **6**, 111–128.
47. Müller,P. (2006) In: *Crystal Structure Refinement: A Crystallographers Guide to SHELXL*. Oxford University Press, NY.
48. Kiliszek,A., Kierzek,R., Krzyzosiak,W.J. and Rypniewski,W. (2012) Crystallographic characterization of CCG repeats. *Nucleic Acids Res.*, **40**, 8155–8162.
49. Rypniewski,W., Banaszak,K., Kulinski,T. and Kiliszek,A. (2016) Watson-crick-like pairs in CCUG repeats: evidence for tautomeric shifts or protonation. *RNA*, **22**, 22–31.
50. Costales,M.G., Childs-Disney,J.L., Haniff,H.S. and Disney,M.D. (2020) How we think about targeting RNA with small molecules. *J. Med. Chem.*, **63**, 8880–8900.
51. Tong,Y., Lee,Y., Liu,X., Childs-Disney,J.L., Suresh,B.M., Benhamou,R.I., Yang,C., Li,W., Costales,M.G., Haniff,H.S., *et al.* (2023) Programming inactive RNA-binding small molecules into bioactive degraders. *Nature*, **618**, 169–179.
52. Brown,J.A., Bulkley,D., Wang,J., Valenstein,M.L., Yario,T.A., Steitz,T.A. and Steitz,J.A. (2014) Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat. Struct. Mol. Biol.*, **21**, 633–640.

53. Brown, J.A., Valenstein, M.L., Yario, T.A., Tycowski, K.T. and Steitz, J.A. (2012) Formation of triple-helical structures by the 3'-end sequences of MALAT1 and MENbeta noncoding RNAs. *Proc. Nat. Acad. Sci. U.S.A.*, **109**, 19202–19207.
54. Chen, G., Chang, K.Y., Chou, M.Y., Bustamante, C. and Tinoco, I. Jr (2009) Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of -1 ribosomal frameshifting. *Proc. Nat. Acad. Sci. U.S.A.*, **106**, 12706–12711.
55. Gilbert, S.D., Rambo, R.P., Van Tyne, D. and Batey, R.T. (2008) Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nat. Struct. Mol. Biol.*, **15**, 177–182.
56. Klein, D.J. and Ferre-D'Amare, A.R. (2006) Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science*, **313**, 1752–1756.
57. Nixon, P.L., Rangan, A., Kim, Y.G., Rich, A., Hoffman, D.W., Hennig, M. and Giedroc, D.P. (2002) Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. *J. Mol. Biol.*, **322**, 621–633.
58. Qiao, F. and Cech, T.R. (2008) Triple-helix structure in telomerase RNA contributes to catalysis. *Nat. Struct. Mol. Biol.*, **15**, 634–640.
59. Shefer, K., Brown, Y., Gorkovoy, V., Nussbaum, T., Ulyanov, N.B. and Tzfati, Y. (2007) A triple helix within a pseudoknot is a conserved and essential element of telomerase RNA. *Mol. Cell. Biol.*, **27**, 2130–2143.
60. Theimer, C.A., Blois, C.A. and Feigon, J. (2005) Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol. Cell*, **17**, 671–682.
61. Ulyanov, N.B., Shefer, K., James, T.L. and Tzfati, Y. (2007) Pseudoknot structures with conserved base triples in telomerase RNAs of ciliates. *Nucleic Acids Res.*, **35**, 6150–6160.
62. Kiliszek, A., Pluta, M., Bejger, M. and Rypniewski, W. (2022) Structure and thermodynamics of a UGG motif interacting with Ba²⁺ and other metal ions: accommodating changes in the RNA structure and the presence of a G(syn)-G(syn) pair. *RNA*, **29**, 44–54.
63. Bowman, J.C., Lenz, T.K., Hud, N.V. and Williams, L.D. (2012) Cations in charge: magnesium ions in RNA folding and catalysis. *Curr. Opin. Struct. Biol.*, **22**, 262–272.
64. Gray, D.M., Hung, S.H. and Johnson, K.H. (1995) Absorption and circular dichroism spectroscopy of nucleic acid duplexes and triplexes. *Methods Enzymol.*, **246**, 19–34.
65. Sugimoto, N., Wu, P., Hara, H. and Kawamoto, Y. (2001) pH and cation effects on the properties of parallel pyrimidine motif DNA triplexes. *Biochemistry*, **40**, 9396–9405.
66. Szabat, M., Kierzek, E. and Kierzek, R. (2018) Modified RNA triplexes: thermodynamics, structure and biological potential. *Sci. Rep.*, **8**, 13023.
67. Benabou, S., Avino, A., Eritja, R., Gonzalez, C. and Gargallo, R. (2014) Fundamental aspects of the nucleic acid i-motif structures. *RSC Adv.*, **4**, 26956–26980.
68. Snoussi, K., Nonin-Lecomte, S. and Leroy, J.L. (2001) The RNA i-motif. *J. Mol. Biol.*, **309**, 139–153.
69. Mooren, M.M., Pulleyblank, D.E., Wijmenga, S.S., Blommers, M.J. and Hilbers, C.W. (1990) Polypurine/polypyrimidine hairpins form a triple helix structure at low pH. *Nucleic Acids Res.*, **18**, 6523–6529.
70. Holland, J.A. and Hoffman, D.W. (1996) Structural features and stability of an RNA triple helix in solution. *Nucleic Acids Res.*, **24**, 2841–2848.
71. Moody, E.M., Lecomte, J.T. and Bevilacqua, P.C. (2005) Linkage between proton binding and folding in RNA: a thermodynamic framework and its experimental application for investigating pKa shifting. *RNA*, **11**, 157–172.
72. Jones, E.L., Mlotkowski, A.J., Hebert, S.P., Schlegel, H.B. and Chow, C.S. (2022) Calculations of pK(a) values for a series of naturally occurring modified nucleobases. *J. Phys. Chem. A*, **126**, 1518–1529.
73. Tang, C.L., Alexov, E., Pyle, A.M. and Honig, B. (2007) Calculation of pKas in RNA: on the structural origins and functional roles of protonated nucleotides. *J. Mol. Biol.*, **366**, 1475–1496.
74. Ferre-D'Amare, A.R. and Doudna, J.A. (2000) Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J. Mol. Biol.*, **295**, 541–556.
75. Ferre-D'Amare, A.R., Zhou, K. and Doudna, J.A. (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature*, **395**, 567–574.
76. Nixon, P.L., Cornish, P.V., Suram, S.V. and Giedroc, D.P. (2002) Thermodynamic analysis of conserved loop-stem interactions in P1-P2 frameshifting RNA pseudoknots from plant Luteoviridae. *Biochemistry*, **41**, 10665–10674.
77. Nixon, P.L. and Giedroc, D.P. (2000) Energetics of a strongly pH dependent RNA tertiary structure in a frameshifting pseudoknot. *J. Mol. Biol.*, **296**, 659–671.
78. Su, L., Chen, L., Egli, M., Berger, J.M. and Rich, A. (1999) Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nat. Struct. Biol.*, **6**, 285–292.
79. Geng, Y., Liu, C., Cai, Q., Luo, Z., Miao, H., Shi, X., Xu, N., Fung, C.P., Choy, T.T., Yan, B., et al. (2021) Crystal structure of parallel G-quadruplex formed by the two-repeat ALS- and FTD-related GGGGCC sequence. *Nucleic Acids Res.*, **49**, 5881–5890.