Targeting the ALS/FTD-associated A-DNA kink with anthracene-based metal complex causes DNA backbone straightening and groove contraction

Cyong-Ru Jhan[1,†], Roshan Satange[2,3,†], Shun-Ching Wang[3,†], Jing-Yi Zeng[3], Yih-Chern Horng[4], Peng Jin [5], Stephen Neidle [6] and Ming-Hon Hou [1,2,3,*]

[1]Department of Life Sciences, National Chung-Hsing University, Taichung 402, Taiwan, [2]Ph.D. Program in Medical Biotechnology, National Chung Hsing University, Taichung 402, Taiwan, [3]Institute of Genomics and Bioinformatics; National Chung Hsing University, Taichung 402, Taiwan, [4]Department of Chemistry, National Changhua University of Education, Changhua 50058, Taiwan, [5]Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA and [6]The School of Pharmacy, University College London, London, WC1N 1AX, United Kingdom

ABSTRACT

The use of a small molecule compound to reduce toxic repeat RNA transcripts or their translated aberrant proteins to target repeat-expanded RNA/DNA with a G4C2 motif is a promising strategy to treat C9orf72-linked disorders. In this study, the crystal structures of DNA and RNA–DNA hybrid duplexes with the -GGGCCG- region as a G4C2 repeat motif were solved. Unusual groove widening and sharper bending of the G4C2 DNA duplex A-DNA conformation with B-form characteristics inside was observed. The G4C2 RNA–DNA hybrid duplex adopts a more typical rigid A form structure. Detailed structural analysis revealed that the G4C2 repeat motif of the DNA duplex exhibits a hydration shell and greater flexibility and serves as a 'hot-spot' for binding of the anthracene-based nickel complex, Ni^{II}(Chro)$_2$ (Chro = Chromomycin A3). In addition to the original GGCC recognition site, Ni^{II}(Chro)$_2$ has extended specificity and binds the flanked G:C base pairs of the GGCC core, resulting in minor groove contraction and straightening of the DNA backbone. We have also shown that Chro-metal complexes inhibit neuronal toxicity and suppresses locomotor deficits in a *Drosophila* model of C9orf72-associated ALS. The approach represents a new direction for drug discovery against ALS and FTD diseases by targeting G4C2 repeat motif DNA.

INTRODUCTION

Abnormal expansion of the GGGGCC (G4C2) hexanucleotide DNA repeat within the C9orf72 intron is associated with two neurodegenerative disorders: amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) (1,2). Generally, healthy individuals carry 20–30 hexanucleotide repeat units, while in patients with ALS/FTD, this is expanded to between 800 and >4000 repeats. The proposed pathologic mechanisms of ALS and FTD consist of loss-of-function of the C9orf72 protein and a toxic gain-of-function from C9orf72 repeat RNA or dipeptide repeat proteins (3,4). Many studies have suggested that the expansion of G4C2 hexanucleotide repeats may be attributed to the slippage of DNA strands, forming either quadruplex structures that are stabilized by G-quartets during DNA replication or possibly hairpins stabilized by duplex (5–8). Expansion of the hexanucleotide repeat present in the non-coding region of C9orf72 leads to reduced levels of the C9orf72 protein (9–11). Furthermore, expanded hexanucleotide repeats are transcribed, and the resulting RNA forms nuclear foci that segregate various RNA-binding proteins to cause neurodegeneration (12–14). The final proposed mechanism involves G4C2 repeat-associated non-ATG translation, which may produce toxic dipeptide proteins (15) (Figure 1). Targeting RNA structures of G4C2 hexanucleotide repeats is one of the most widely explored strategies for drug design against ALS and FTD. Several groups have identified small-molecule ligands and antisense oligonucleotides that specifically bind to the expanded G4C2 repeat hairpin RNA to reduce toxic RNA foci and dipeptide proteins to improve ALS/FTD-associated defects (11,16–20). We show here that targeting repeat-specific DNA-containing structures, including DNA duplexes and DNA–RNA hybrid du-
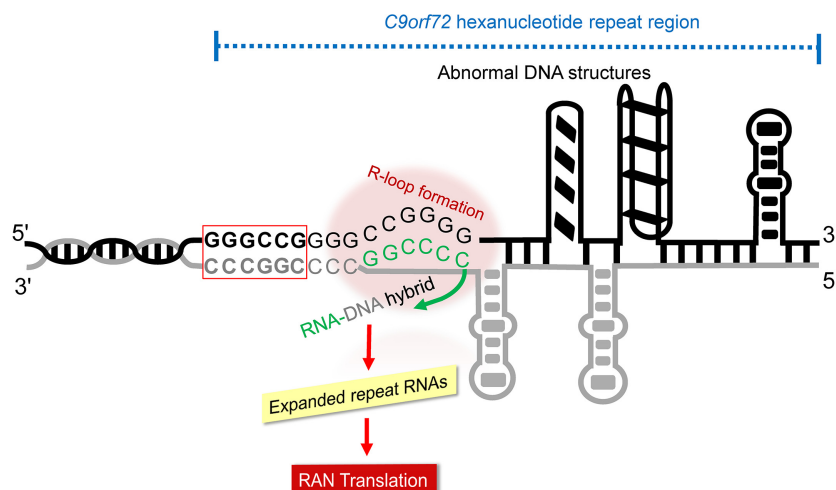
Figure 1. Schematic representation of mechanisms of ALS/FTD pathogenesis induced by abnormal G4C2 repeat expansions on the chromosome 9 open reading frame 72 (*C9orf72*). The hexanucleotide repeat expansion resulted in different types of abnormal secondary structures such as R-loop, hairpins or quadruplexes that interfere with cellular processes such as transcription, replication and DNA repair. The signature -GGGCCG- sequence corresponds to the G4C2 motif is highlighted by a red colour box.

plexes, may serve as an alternative approach to G4C2 hexanucleotide repeat-associated ALS/FTD (21,22).

In this study, the crystal structures of the non-self-complementary DNA duplex, d(GTGGGCCGAC)/ (GTCGGCCCAC) and the analogous RNA–DNA hybrid duplex, r(GUGGGCCGAC)/ d(GTCGGCCCAC) were determined since the central -GGGCCG- region (G4C2 repeat motif) is a core G4C2 repeat sequence. Both duplex structures showed global similarity, including A-DNA-like characteristics, albeit with local structure differences, such as a wider groove and sharper bending angle in the DNA duplex. Moreover, a detailed analysis of crystal packing revealed that the central -GGGCCG- region of the DNA duplex has greater flexibility and a hydration shell that might serve as 'hot-spots' for accommodating small molecule binding, compared to the RNA–DNA hybrid duplex. Several GC-selective DNA-binding agents including anthracene-, anthracycline- and acridine-based compounds, a DNA cross-linker compound and polypeptide antibiotics were screened, and it was found that metal–anthracene complexes, $Ni^{II}(Chro)_2$ and $Co^{II}(Chro)_2$, preferred to bind and stabilize the G4C2 repeat motif compared to other DNA binding ligands. The crystal structure of $Ni^{II}(Chro)_2$-d(TTGGGCCGAA)$_2$ was then solved to investigate the structural basis behind this preference. Importantly, we show that $Ni^{II}(Chro)_2$ and $Co^{II}(Chro)_2$ can suppress G4C2 repeat-induced neuronal toxicity *in vivo* using a *Drosophila* model of G4C2 repeat expansion. These findings highlight the feasibility of utilizing the G4C2 repeat duplex as a target and serve as a basis for the design of novel sequence-specific ligands against ALS/FTD diseases.

MATERIALS AND METHODS

Oligonucleotides and sample preparation

All chemicals used in this study were purchased from Sigma Chemical Co. (St. Louis, MO, USA). DNA oligomers were commercially synthesized and purified by polyacrylamide gel electrophoresis (PAGE) from MDBio, Inc. (New Taipei City, Taiwan), and Genomics (New Taipei City, Taiwan). RNA oligomers were purified using high performance liquid chromatography from MDBio, Inc. Oligonucleotides were prepared in RNase-free water by heating at 95°C for 5 min followed by slow cooling to room temperature (-0.5°C/min) to allow duplex formation. To determine oligonucleotide concentrations, absorbance measurements were performed using a JASCO V-630 spectrophotometer (JASCO International Co. Ltd., Tokyo, Japan) with a quartz cuvette (1 cm path length) set at 260 nm. The concentration of oligonucleotides have been calculated using Beer's law with the extinction coefficients calculated according to tabulated values of monomer and dimer extinction coefficients, with reasonable assumptions (23). Stock solutions of ligands (1 mM) used for melting temperature assays were prepared either in ddH$_2$O or in dimethyl sulfoxide (DMSO, ≥99.9%).

UV spectroscopy

UV melting curves were collected using a JASCO V-630 spectrophotometer. The duplex DNA used for melting temperature analysis was dissolved in buffer containing 50 mM NaCl and 50 mM sodium cacodylate (pH 7.3). A final concentration of 3 μM of the duplex sequences was denatured at 95°C for 5 min and cooled on ice for 30 min for annealing. Ligands were added to oligonucleotide solution at a fixed molar ratio of 1:1 and incubated for 1 h at 4°C to allow complex formation. A 650 μl sample was added to the quartz cuvette with a 1 cm path length and covered with a layer of silicone oil. The sample was then equilibrated at 4°C for 10 min in a spectrophotometer. UV absorbance versus. temperature profiles were acquired by changing the temperature from 4 to 95°C at a rate of 1°C min^{-1} and recording the absorbance at 260 nm every 0.5 min. The T_m (temperature corresponding to half-dissociation of the DNA duplexes) was determined from polynomial fitting of the observed curves

using Varian\Cary WinUV thermal application software v 3.00 (Agilent, Santa Clara, CA, USA).

Preparation of crystals

For crystallization, PAGE-purified oligonucleotides were annealed by heating at 95°C for 5 min and then cooled slowly to room temperature (-0.5°C/min). Crystallization trials were performed using the sitting-drop vapour-diffusion method, and crystals were obtained at 4°C after 2–7 days. To grow G4C2 DNA crystals, 1 mM DNA duplex solution was mixed with buffer containing 50 mM 2-(N-morpholino)ethanesulfonic acid (pH 6.5), 100 mM lithium chloride, 10 mM manganese (II) chloride tetrahydrate and 2.6 M sodium malonate dibasic monohydrate at a ratio of 1:3, and equilibrated against 200 µl solution. Small, colourless, rhombic crystals appeared after 3 days of crystallization and grew to a maximum size after 1–2 weeks. To obtain G4C2 RNA–DNA hybrid crystals, equimolar concentrations (0.25 mM) of DNA and RNA oligonucleotides were mixed in a solution containing 5 mM magnesium sulphate hydrate, 50 mM Tris hydrochloride (pH 8.5) and 2.9 M 1,6-hexanediol. Square-shaped, colourless crystals appeared after 1 week of crystallization. The $Ni^{II}(Chro)_2$–DNA complex crystals were obtained from a solution of 1 mM DNA duplex, 3 mM Chro, 6 mM $NiSO_4$, 40 mM sodium cacodylate (pH 7), 2.5 mM spermine and 6% polyethylene glycol (PEG) 400, and equilibrated against 500 µl of 6% PEG 400. Yellow-coloured, needle-shaped crystals were grown after 4–7 days. Each condition produced several crystals, from which a single crystal was used to collect the diffraction data.

X-ray data collection, phasing and structure refinement

X-ray diffraction data from a single crystal of each of G4C2 DNA, G4C2 RNA–DNA hybrid and the $Ni^{II}(Chro)_2$-DNA complex were collected at the synchrotron radiation facility of the National Synchrotron Radiation Research Center, Taiwan. Diffraction data integration and reduction were performed using the HKL-2000 package (24). Phases for the G4C2 DNA structure were determined in the $P3_121$ space group by molecular replacement using Phaser MR in a python-based hierarchical environment for integrated xtallography (*PHENIX)* (version 1.10.1) and using a modelled A-DNA structure with WinCoot (version 0.8.9.1) as a template (25). The G4C2 DNA structure was then used to determine phases for the G4C2 RNA–DNA hybrid in the $P2_12_12_1$ space group, using a molecular replacement protocol. The phases for the $Ni^{II}(Chro)_2$–DNA complex structure were determined by molecular replacement in the $P6_522$ space group with Phaser MR in the *PHENIX* suite using the partial structure of the $[Mg^{2+}-(Chro)_2]$–d(TTGGCCAA)_2 complex (PDB ID: 1VAQ) as a starting template. The resulting well-defined electron density maps were used to build initial models using the molecular graphics programs MIFit and crystallographic object-oriented toolkit (COOT). Initial models of the G4C2 DNA, G4C2 RNA–DNA hybrid duplexes and the $Ni^{II}(Chro)_2$–DNA complex were rebuilt and refined with COOT and *PHENIX*. Structure refinements were performed using phenix.refine

in *PHENIX* (version 1.10.1) and Refmac5 in CCP4i (Oxon, UK); the crystallographic and refinement statistics of the structures are listed in Supplementary Table S1 (25,26). The final $2F_o – F_c$ electron density maps were created using the fast Fourier transform in CCP4i and PyMOL (version 4.5), which was used to draw graphical representations of the refined structures.

DNA parameter analysis

A standard A-DNA structure with the same sequence as G4C2 DNA and G4C2 RNA–DNA hybrid DNA duplexes was generated using Discovery Studio Client software (version-19.1.0.18287) (Dassault Systems Biovia Corp, Vélizy-Villacoublay, France) for comparison. DNA structural parameters, including helix, base and base-pair parameters were analysed with the Web 3DNA 2.0 server and the CURVES+ program (27,28). RMSDs were calculated in PyMOL and van der Waals interactions were monitored using LigPlot+ (29). DNA parameter values for typical A- and B-DNA were used for comparison (30).

MD simulation protocol

Structures of metal$^{II}(Chro)_2$–DNA complexes obtained from the crystal structures were used for MD simulations. These were performed using the Discovery Studio software v19.1.0.18287 and the CHARMm force field (31). A 1000 step smart-minimizer algorithm was used, and the models were immersed in an orthorhombic water box built based on a modified TIP3P system (32). The dimensions of the water box were selected such that the distances of non-hydrogen atoms in a nucleic acid duplex were at least 7 Å from the edge of the box. The total charge on each system under study was neutralized by randomly adding sodium ions. The system was then subjected to 5000-step Steepest Descent (SD) and 10 000-step Conjugate Gradient (CG) minimizations followed by 20 ps heating from 50 to 300 K with position constraints imposed on the nucleic acid duplexes. The following 20 ps equilibration run was performed with position constraints on nucleic acid duplexes. The production run was carried out with an NPT ensemble with periodic boundary conditions. During the entire simulation, all covalent bonds involving hydrogen were constrained using the SHAKE algorithm, where the time step was set to 2 fs. The particle mesh Ewald method was used for treating long-range electrostatic interactions (33). The production simulations was extended up to 500 ps for the $Ni^{II}(Chro)_2$–DNA and $Mg^{II}(Chro)_2$–DNA complex. RMSF values were analysed using the 'Analyse Trajectory' function in Discovery Studio software v19.1.0.18287. Final graphs were plotted using Origin Pro8 v8.0721 (OriginLab Corp., Northampton, MA, USA) (34).

Drosophila experiments

The following fly lines (gmr-GAL4, UAS-GGGGCC_30-EGFP), (elav-GAL4, UAS-GGGGCC_3-EGFP), (elav-GAL4, UAS-GGGGCC_30-EGFP), (ok371-GAL4, UAS-GGGGCC_3-EGFP), (ok371-GAL4, UAS-GGGGCC_3-EGFP) and (gmr-GAL4, UAS-CGG_90-EGFP) have been

generated in previously published work ([14]). All *Drosophila* lines were maintained, and crosses were performed in standard medium at 25°C ([35]). $Ni^{II}(Chro)_2$ and $Co^{II}(Chro)_2$ were dissolved in DMSO (Sigma), and then added to green coloured food (Kroger), an indicator for better mixture with fly food, which was administered to the fly food. The same amount of DMSO was added to the fly food as an internal control. The progeny were collected and aged to 7 days. The screen was performed by scoring the eye phenotype, which was visualized using light microscopy. For the *Drosophila* activity assay, equal amounts of males and females were individually placed in *Drosophila* Activity Monitoring System (TriKinetics Inc, Waltham, MA, USA) testing chambers that had been capped with regular food at one end. The flies were grown on a 12-h:12-h light: dark cycle at 25°C for 4 weeks. Locomotor data were collected in the light cycle at 25°C. Locomotor activity averages of each day during the fourth week were calculated. Data were assayed in eight flies from each group at a time.

RESULTS

The crystal structures of the ALS/FTD-associated G4C2 repeat DNA motif and RNA–DNA hybrid duplexes reveal that these duplexes adopt A-DNA-like conformations

To provide insight into higher-order structure formation of the G4C2 motif, the crystal structures of d(GTGGGCCGAC)/d(GTCGGCCCAC) (G4C2 DNA) and r(GUGGGCCGAC)/d(GTCGGCCCAC) (G4C2 RNA-DNA hybrid) were solved at resolutions of 1.58 Å and 1.78 Å, respectively. The signature $(GGGCCG)_n$ sequence corresponds to the G4C2 motif, which comprises a central d(GGCC) motif flanked by two G:C base pairs embedded in the DNA. The structures include all atoms in the final refinement with no obvious indication of disorder, as shown by the clear electron density map (Supplementary Figure S1). The oligonucleotides self-assemble into a single anti-parallel duplex for the G4C2 DNA and G4C2 RNA–DNA hybrid duplex structures in each asymmetric unit. Duplexes were numbered from G1 to C10 in one strand and G11 to C20 in the complementary strand (Figure 2A,B). The crystal structures of both duplexes exhibit A-type-like DNA duplex structures with similar widths at the major and minor grooves, which are also revealed in backbone conformational angles and helical parameters ([36]). For example, the average glycosyl torsion angles (χ) of the DNA and RNA–DNA hybrid duplexes have values of -161° and -152° respectively, with the sugar pucker for all bases being well preserved in the C3′-*endo* conformation (Supplementary Table S2). The roll angles in the G4C2 motif pairs of the G4C2 DNA and G4C2 RNA–DNA hybrid duplexes were approximately 4–8°, which resulted in a clear kink of the DNA structures toward the major groove. The average helical twist values of the G4C2 DNA and G4C2 RNA–DNA hybrid duplexes were calculated to be 33° and 31°, respectively. This average twist value corresponds to unwinding angles from B-DNA of 3° and 5°, respectively ([36]). To compare the crystal structures, a uniform model of the ideal A-form was built using the Discovery studio client package v19.1 with the

same sequence. The all-atom root-mean-square deviation (RMSD) of ideal A-form DNA to the G4C2 DNA and G4C2 RNA-DNA hybrid duplexes was 1.3 and 1.6 Å, respectively. Superimposing the G4C2 DNA duplex with that of the hybrid duplex resulted in an all-atom RMSD of 1.4 Å. These results suggested that the three structures were globally identical, albeit with significant local differences (Figure 2C). The crystal structure of the G4C2 DNA duplex is remarkable in that the structure exists as a hybrid A-form with some B-form character, having average twist angles and rise steps of 35° and 3.3 Å, respectively, in the central d(GGCC) core (Figure 2D), which is characteristic of a B-DNA structure and looking down the centre of the helix, it has a B-like crowded appearance. Interestingly, the GpG and CpG side steps of the d(GGGCCG) core in the G4C2 DNA duplex have helical twists of 25° and 26° respectively, an unwinding of ~10° from B-DNA. Compared to the G4C2 DNA duplex, significant unwinding of base pairs is also observed in the GpC and CpC steps of the central d(GGCC) core for the G4C2 RNA–DNA hybrid duplex, with lower twist angles of 32° and 29°, respectively. Local differences in translational parameters were also observed in the central core of both duplexes. Values of the stagger parameter for the DNA duplex are close to those for B-DNA (Supplementary Figure S2). In contrast, the stagger values of the G4C2 RNA–DNA hybrid were negative and are closer to those for A-DNA. In order to better understand the effect of the central GGGCCG core on major and minor groove geometry, average widths of the major and minor grooves were calculated as interstrand *P–P* distances using Web 3DNA 2.0 analysis (Figure 2E) ([27]). Both structures have the minor groove wider than the major groove in the central core, which is closer to typical A-DNA. It is notable that the G4C2 DNA duplex is bent to a greater extent than that of the G4C2 RNA–DNA hybrid, exposing more minor groove surface area. We suggest that this feature could act as a hot spot and may be important for DNA recognition by various types of ligands.

Crystal packing, a hydration network, and the flexibility of the G4C2 DNA duplex minor groove have created hot spots for ligand binding

The 'hot-spot' feature for ligand binding has been further confirmed in the crystal packing of the G4C2 DNA duplex, which exhibits a familiar 'base-pair into minor groove' type packing that has been previously reported for DNA duplexes (Supplementary Figure S3) ([37]). In this crystal packing environment, one end of an asymmetric unit duplex interacts with the bases of other duplexes into the minor groove. Interestingly, the asymmetric unit duplex in G4C2 DNA showed direct interactions with the central G4-C17, G5-C16, C6-G15 and C7-G14 base-pair core region through hydrogen bonding, water-mediated H-bonds and van der Waals interactions (Figure 3A). Terminal G1* and C10* bases form H-bonding and extensive water-mediated interactions with the G5-C16 and C17 bases. Additionally, residues G11* and C20* from another asymmetric unit formed direct hydrogen bonds and water-mediated contacts with the C6-G15 and C7-G14 base pairs. In con-
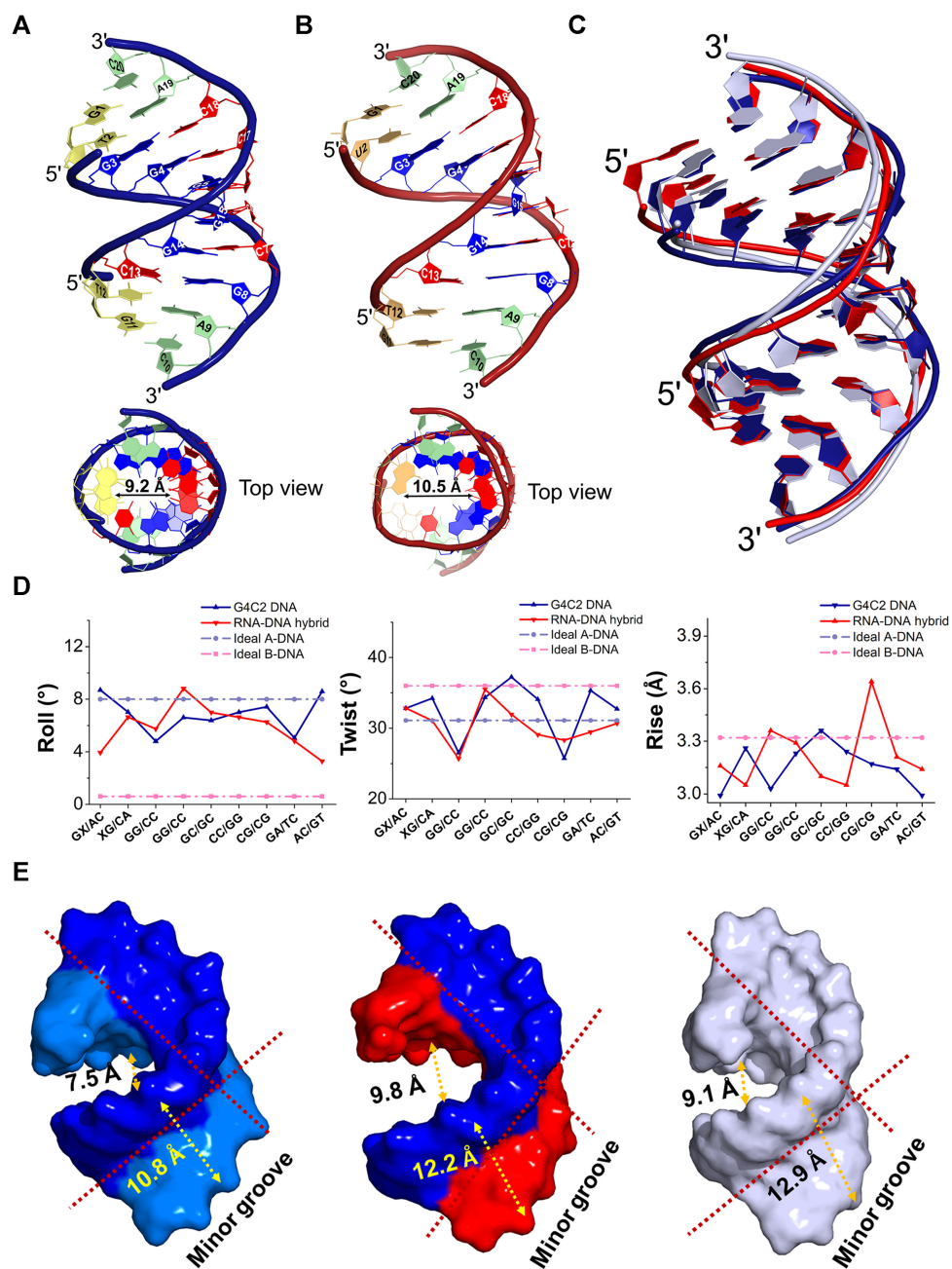
Figure 2. Overall structure features of the G4C2 motif-containing DNA and RNA-DNA hybrid duplexes in comparison with typical A-form duplex. A crystal structure overview of (**A**) d(GTGGGCCGAC)/ d(GTCGGCCCAC) DNA duplex (G4C2 DNA) represented in density blue cartoon and (**B**) r(GUGGGCCGAC)/d(GTCGGCCCAC) hybrid duplex (G4C2 RNA–DNA hybrid) represented in red cartoon as viewed from the side. The two duplexes are numbered from G1 to C10 in one strand and G11 to C20 in the complementary strand. Central -GGGCCG- core residues are shown in blue and red stick representation while the terminal residues are coloured in green and yellow sticks. The top view of each duplex is shown below with the inner diameter of central hole. (**C**) Superimposition of crystal structures of G4C2 DNA duplex (density blue), G4C2 RNA–DNA hybrid (red) and a typical A-form DNA duplex (light-blue) showing the overlap between these structures. (**D**) Prominent variations in the roll (°), twist (°) and rise (Å) DNA parameters are observed in the crystal structure of G4C2 DNA (blue) and G4C2 RNA-DNA hybrid (red) duplexes compared to typical A- and B-form DNA duplexes. (**E**) Surface representation of minor and major grooves of G4C2 DNA duplex, G4C2 RNA–DNA hybrid and a typical A-DNA duplex showing bending of backbone. The DNA alone model is represented by a blue surface, RNA–DNA hybrid as red (RNA strand) and blue (DNA strand) surface and a typical A-form DNA in light-blue surface representation.
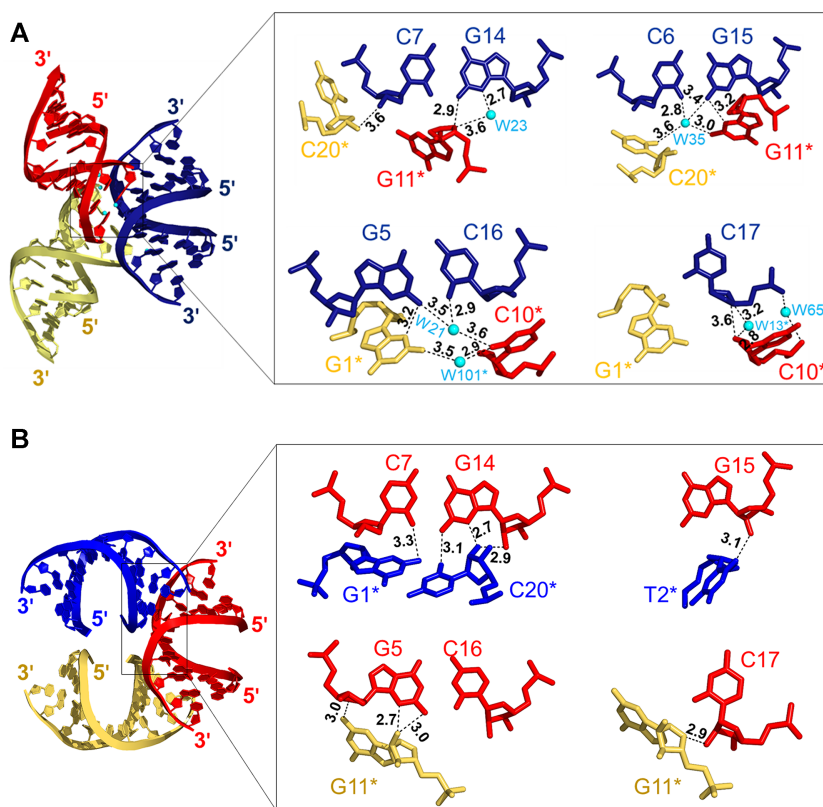
Figure 3. The detailed crystal packing shows a familiar 'base-pair into minor groove' type of packing in both G4C2 DNA and RNA–DNA hybrid duplexes. (**A**) The two-fold symmetry of G4C2 DNA shows strong crystal contacts with the central G4-C17, G5-C16, C6-G15 and C7-G14 base-pairs which resulted into the 'end-to-central' type of packing environment. The DNA duplex is shown in density blue cartoon representation while the asymmetric unit duplexes are shown by red and yellow cartoons. (**B**) The two-fold symmetry of the G4C2 RNA–DNA hybrid duplex shows a distinct crystal packing environment where the duplexes in an asymmetric unit are oriented away from the central core. The RNA–DNA hybrid duplex is shown in red colour while the asymmetric unit duplexes are shown in blue and yellow colours. The residues in an asymmetric unit are highlighted with an asterisk (*) sign. The hydrogen bonds and water-mediated contacts are shown in black dashed lines and the water molecules are shown as cyan spheres. Note that some symmetry related duplexes are not shown for clarity.

trast, the crystal packing interactions in the G4C2 RNA–DNA hybrid were oriented away from the central core (Figure 3B). Although the residues in the G4C2 RNA–DNA hybrid showed hydrogen-bonding interactions, there were no water-mediated interactions. The total number of crystal packing interactions in the G4C2 DNA duplex, RNA–DNA hybrid and other typical A-DNA-like structures is given in Supplementary Table S3. Intriguingly, among these structures, only the G4C2 DNA duplex showed a peculiar 'end-to-central' type packing environment. This type of crystal packing has arisen in the G4C2 DNA duplex because the DNA is more bent at the central core (the sum of roll angles is, ∼32.2°), which exposes the residues in the central core to the minor groove for terminal base pair binding. In addition to crystal packing interactions, the minor groove hydration shell includes five distinct water molecules that stabilize the base pairs in the duplex and mediated interstrand hydrogen bonding, which was not observed in the G4C2 RNA–DNA hybrid (Supplementary Figure S4). These results also suggest that the minor groove of the G4C2 DNA duplex is a hydrophilic environment, which has been widely exploited for ligand binding (38,39).

Metal[II](Chro)$_2$ complexes specifically stabilized the G4C2 DNA duplex and not the RNA–DNA hybrid

To determine the stabilization effects and selectivity towards the G4C2 repeats, several well-known DNA binding compounds with G4C2 repeat-containing DNA duplex selectivity were screened at a fixed compound:DNA stoichiometry of 1:1 (Figure 4A–D). By measuring the difference in melting temperature (ΔT_m) of the G4C2 repeat sequence in the presence of these compounds, it was found that the anthracene-based metal complexes, Ni[II](Chro)$_2$ and Co[II](Chro)$_2$ best stabilized the G4C2 repeat DNA motif duplex with ΔT_m value of 8.0°C, while the other compounds had a ΔT_m of <4°C under similar conditions (Figure 4E). Chro binds to divalent metal ions (such as Ni[II] or Co[II]) to form a dimer that shows specificity to the 5′-GGCC- binding core (40). Therefore, the hexanucleotide repeat sequence provided an excellent binding motif for metal[II](Chro)$_2$ complexes, causing strong stabilization effects. Furthermore, Ni[II](Chro)$_2$ did not show any melting temperature change when titrated with the same G4C2 RNA–DNA hybrid sequence, suggesting that Ni[II](Chro)$_2$ was unable to stabilize the hybrid duplex (Supplementary Figure S5). These results suggest that Ni[II](Chro)$_2$ and
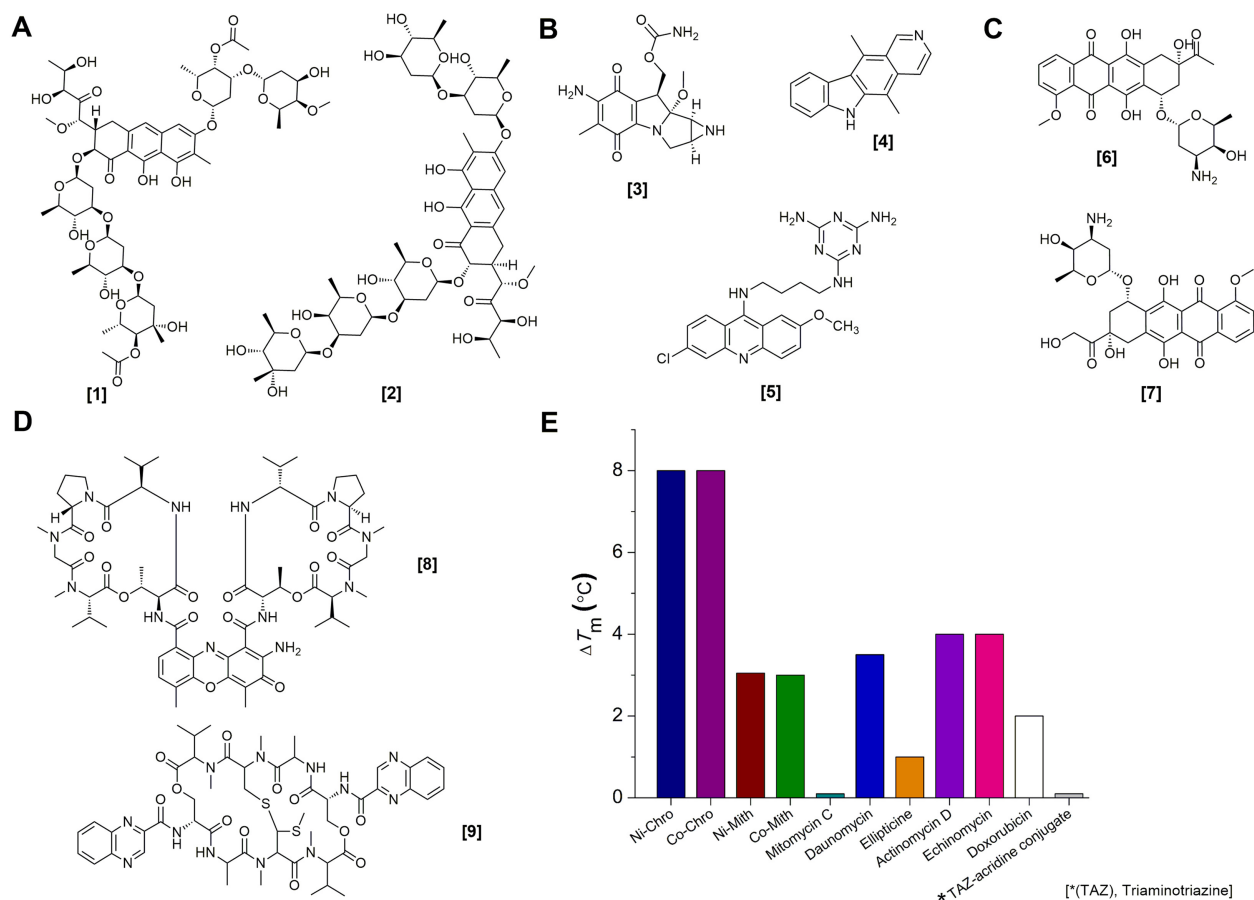
Figure 4. Chemical structures of the DNA binding ligands use for melting temperature assay with G4C2 DNA duplex. Equimolar concentrations (1:1 ratio of DNA:ligand) of DNA binding compounds including (**A**) anthracene-based chromomycin A3 [1], mithramycin [2]; (**B**) aziridine-based mitomycin C [3], heterotetracyclic ellipticine [4] and acridine-based triaminotriazine-acridine conjugate [5]; (**C**) anthracycline-based daunorubicin [6] and doxorubicin [7]; (**D**) cyclic peptide antibiotics actinomycin D [8] and echinomycin [9] are used for melting temperature assay. (**E**) Change in DNA melting temperatures (ΔT_m) for G4C2 DNA duplex in the presence of Ni^{II}(chromomycin A3)$_2$ (Ni-Chro), Co^{II}(chromomycin A3)$_2$ (Co-Chro) complex, Ni^{II}(mithramycin)$_2$ (Ni-Mith), Co^{II}(mithramycin)$_2$ (Co-Mith) complexes, mitomycin C, daunomycin, ellipticine, actinomycin D, echinomycin, doxorubicin and a triaminotriazine-acridine conjugate.

Co^{II}(Chro)$_2$ are able to distinguish between the specific structural differences of G4C2 DNA and hybrid duplexes with the same sequence.

The Ni^{II}(Chro)$_2$ complex induces conformational changes in the G4C2 DNA duplex

To understand the structural basis for the preference of Ni^{II}(Chro)$_2$ (Figure 5A) for G4C2 DNA, the crystal structure of the Ni^{II}(Chro)$_2$-d(TTGGGCCGAA/TTCGGCCCAA) complex was solved at 2.95 Å resolution (Supplementary Figure S6). One asymmetric unit contains two independent Chro-DNA complexes that are packed through van der Waals contacts between two Chro moieties, as shown in Figure 5B; these complexes are labelled CPX1 and CPX2, respectively. There are significant differences between the conformations of CPX1 and CPX2, as indicated by the 3.2 Å RMSD between them. CPX2 has two terminal adenine base pairs (A9 and A10) extruding out of the duplex structure that are further stabilized by intermolecular triplet hydrogen bonding between (A:T):T bases from the two independent strands,

as observed in the crystal packing. Although the two complexes in this structure are significantly different, the central GGGCCG tetranucleotide core, upon Ni^{II}(Chro)$_2$ binding, adopts A-DNA-like characteristics, as reflected by the glycosyl torsion angle values that are in the range -156° to -179° with C3'-*endo* sugar pucker conformations for most of the residues and show a high degree of similarity, with an RMSD of 0.5 Å (Supplementary Figure S7, Supplementary Table S4). This is very distinct from the previously-reported Mg^{II}(Chro)$_2$-d(TTGGCCAA)$_2$ complex structure [Protein Data Bank (PDB): 1VAQ], which adopts a B-like character with a pronounced kink at the central 4-bp segment (41). The overall DNA structure of the complexes preserves right-handed helical features with significant unwinding of base pairs at the GpG and CpC steps of the GGCC central core. This is in contrast to the structure of G4C2 DNA, which shows unwinding at the side steps of the GGGCCG duplex core. Upon ligand binding to the central core of the DNA duplex, the roll angles among the GGGCCG core steps are lower than those of the G4C2 DNA alone structure, which causes straightening of the DNA duplex towards the minor groove and a consequential contraction
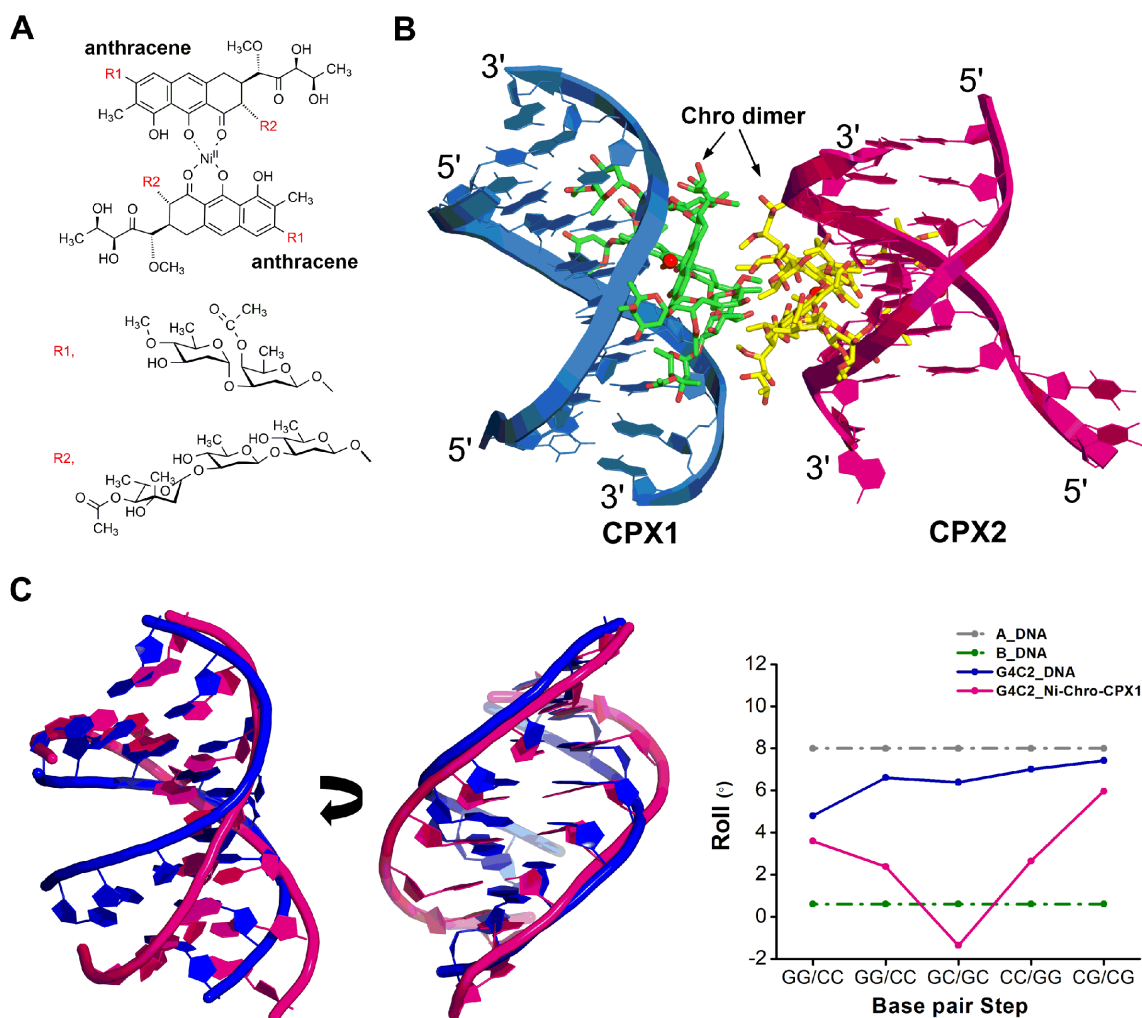
Figure 5. Structural overview of the Ni[II](Chro)$_2$-d(TTGGGCCGAA)$_2$ complex. (**A**) Chemical structure of the anthracene-based chromomycin A3 (Chro) dimer formation in the presence of Nickel(II) ions (Ni[II](Chro)$_2$). (**B**) Each asymmetric unit contains two independent complexes, CPX1 and CPX2 as shown in light teal and pink cartoons, respectively. Two Chro dimers are represented by green and yellow sticks while Ni[II] ions are shown as red spheres. (**C**) Superimposition of the overall structures of G4C2 DNA duplex (blue) and CPX1 of Ni[II](Chro)$_2$-DNA (pink) shows the straightening of a DNA backbone and narrowing of minor groove due to Ni[II](Chro)$_2$ binding. The two structures thus exhibit significant variations in the roll (°) DNA parameter, a comparison shown by blue (G4C2 DNA) and pink lines (Ni[II](Chro)$_2$-d(TTGGGCCGAA)$_2$, CPX1), respectively.

of the minor groove (Figure 5C). The average minor-groove width for each complex is 9.3 Å at the base pairs of the GGGCCG duplex core, whereas the major groove widens up to 17.2 Å in each complex, which differs substantially from the minor and major groove widths of the G4C2 DNA alone structure, at 10.8 and 7.5 Å, respectively.

The structure of Ni[II](Chro)$_2$ shows that the Ni[II] metal ion is tetrahedrally coordinated to the O1 and O9 atoms of the two anthracenes and the A-B disaccharide and C-D-E trisaccharide moieties attached to the anthracenes extended in opposite directions in the dimer conformation (Supplementary Figure S8). The anthracene moieties of the Chro dimer is sandwiched and mutually stacked between the C-D glycosidic linkage of the other Chro monomer and the ribose of C6, bringing the backbone of the DNA duplex towards the ligand direction in order to stabilize the ligand–DNA complex (Supplementary Figure S9). This is similar to the packing interaction resulting from the native G4C2

DNA molecules abutting their terminal base pairs onto the minor groove of other DNA duplexes in the native G4C2 DNA structure. The binding of Ni[II](Chro)$_2$ to the central GGCC sequence shows specific intermolecular hydrogen-bonding interactions between O8 of the anthracene and the N2 atom of the G5 (G15) base in CPX1 and CPX2 (Supplementary Figure S10A, B). Likewise, the two EO1 oxygen atoms of the trisaccharide E ring are hydrogen bonded with the N2 atoms of the G4 and G14 bases. Unlike previous observations, a dihedral angle of 71° between two anthracene monomers positions the O3 of the B ring in an orientation that enables it to form hydrogen bonds with OP1 of base G15 in CPX1 and 2 (Supplementary Table S5). Similarly, O3 of the B ring also forms a hydrogen bond with OP1 of G5 in CPX1, while the distance between O3 and OP1 of G5 in CPX2 is slightly greater than the cut-off value assigned to the hydrogen bonds (Supplementary Figure S10A, B). In addition, the terminal intermolecular (A:T):T triplet

base pair adds overall stabilization of the complex structure (Supplementary Figure S10C). The coordination of N7 of guanines G5 and G15 with Ni^{II} ions in both complexes further stabilizes the duplexes (Supplementary Figure S11).

In addition to these interactions in the central GGCC core, Chro also shows extended specificity for the flanking G:C base pairs of the central GGCC region in the G4C2 motif, as suggested by MD simulations of 500 ps for both complexes. The dynamic behaviour of these individual base pairs in both complexes was determined in terms of RMSF values. The RMSF plot indicates a similar residue fluctuation profile for the central GGGCCG motif with an average RMSF of 1.1 Å, indicating higher stability due to (Chro)$_2$ binding (Supplementary Figure S12A). The terminal residues T1:A20 and T2:A19 show extreme fluctuations, suggesting that Chro binding did not affect the terminal residues. Interestingly, this specificity was not observed in the previously reported structure of Mg^{II}(Chro)$_2$-d(TTGGCCAA) (PDB ID- 1VAQ), where the flanking base pair is A:T (Supplementary Figure S12B). This suggests that the Ni^{II}(Chro)$_2$ complex has extended specificity due to the indirect binding effects of the flanking G3:C18 and G8:C13 base pairs in the G4C2 sequence. Larger stagger and propeller twist values might cause these base pairs to interact with Ni^{II}(Chro)$_2$ (Supplementary Figure S13). The base stacking interactions and steric clashes between Chro sugars might also allow Chro to stabilize the overall structure. In addition, LigPlot+ analysis shows that the E ring of the trisaccharide moiety forms a close van der Waals contact with C1 of the deoxyribose sugar of C18 and G18 from CPX1 and CPX2, respectively (Supplementary Figure S14). The unusually low value of the base-pair rise results in increased stacking of the flanked G3:C18 and G8:C13 pairs in the G4C2 complexes, as indicated by the hydrophobic interactions. Overall, these interactions strengthen the specific recognition of the GGGCCG repeat motif by the Ni^{II}(Chro)$_2$ complex.

Metal^{II}(Chro)$_2$ suppressed G4C2 repeat toxicity in a *Drosophila* model of G4C2 repeat expansion

Given the specific recognition of G4C2 repeats by Ni^{II}(Chro)$_2$ and Co^{II}(Chro)$_2$ G4C2 repeat toxicity was tested using a previously established *Drosophila* model of G4C2 repeats (14). In this model, 30 G4C2 repeats are expressed using a tissue- or cell type-specific GAL4 driver and their expression leads to neuronal cell death and locomotor deficits. Ni^{II}(Chro)$_2$ and Co^{II}(Chro)$_2$ were mixed with fly food at different concentrations and tested in gmr-GAL4 and UAS-GGGGCC$_{30}$-EGFP strains. As described previously (14), the neuronal toxicity caused by 30 G4C2 repeats is classified into four categories (I = normal; II = slight, < 5%; III = moderate, 5%–30%; IV = severe > 30%) (Figure 6A). At 500 nM, both Ni^{II}(Chro)$_2$ and Co^{II}(Chro)$_2$ inhibited the neuronal toxicity caused by the expanded G4C2 repeats in the fly eye (Figure 6B), with the Co^{II}(Chro)$_2$ complex almost completely suppressing the most severe toxicity. To further determine the impact of Ni^{II}(Chro)$_2$ and Co^{II}(Chro)$_2$ on motor neurons expressing the G4C2 repeat that were most affected in patients, both normal and expanded G4C2 repeats were
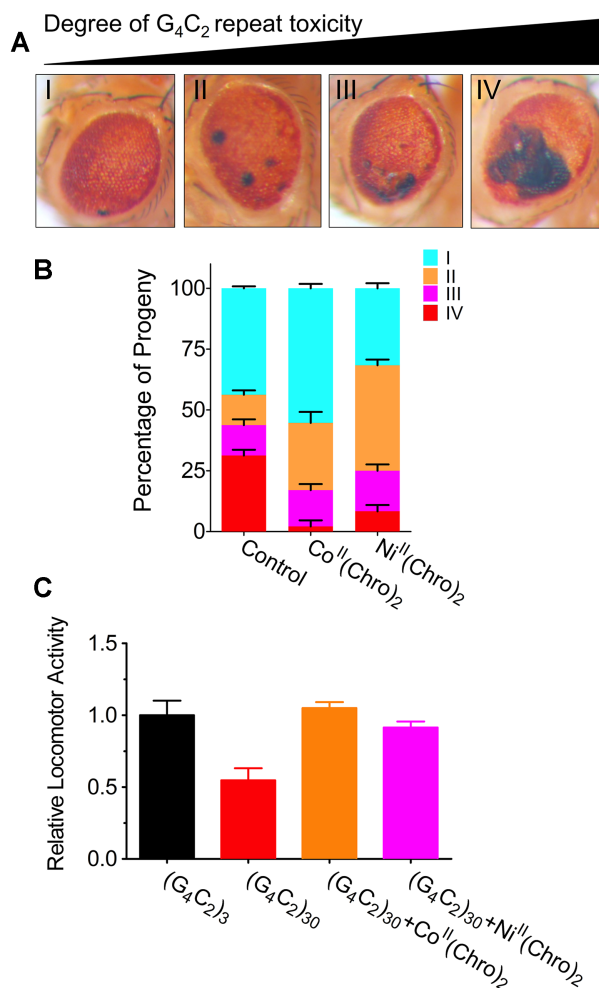

Figure 6. Effects of metal^{II}(Chro)$_2$ on G4C2 repeat toxicity in the *Drosophila* model. (**A**) The expression of (G$_4$C$_2$)$_{30}$ driven by the Gal4 driver causes neuronal toxicity in the fly eye that is classified into four categories (I = normal; II = slight, <5%; III = moderate, 5%-30%; IV = severe >30%) as shown under the light microscope. (**B**) Ni^{II}(Chro)$_2$ and Co^{II}(Chro)$_2$ inhibits the neuronal toxicity caused by expanded G4C2 repeats in the fly eye. (**C**) Ni^{II}(Chro)$_2$ and Co^{II}(Chro)$_2$ suppresses the locomotor deficits caused by the expression G4C2 repeats in ok371-GAL4, UAS-GGGGCC$_{30}$-EGFP and ok371-GAL4, UAS-GGGGCC$_3$-EGFP fly models. Data were assayed in eight flies from each group at a time and the graphs were plotted with GraphPad Prism software (version 9.0.0).

expressed in motor neurons using ok371-GAL4, and the locomotor activity of the following flies, ok371-GAL4, UAS-GGGGCC$_{30}$-EGFP and ok371-GAL4, and UAS-GGGGCC$_3$-EGFP was determined. Both Ni^{II}(Chro)$_2$ and Co^{II}(Chro)$_2$ suppressed the locomotor deficits caused by the expression of G4C2 repeats (Figure 6C).

DISCUSSION

Many studies support a gain of toxicity in which the C9orf72 repeat RNAs play a crucial role in ALS/FTD pathogenesis (42–44). Consequently, reducing toxic repeat RNA transcripts or their translated aberrant proteins by targeting C9orf72 repeat RNAs with small molecules is a promising strategy, and the most advanced in terms of eventual clinical development (20,45,46). However, target-

ing RNA molecules may possess several challenges; most of the target RNAs are expressed at a low level with a short half-life and RNA conformational dynamics in cells, making it difficult for small molecule ligands to preferentially recognize their target. However, the development of small molecules to modulate repeat expansions and related instabilities or inhibit the production of toxic transcripts and proteins by targeting specific repeat DNA-containing duplexes has been shown as another potential therapeutic approach against neurological diseases (47–49). For example, Nguyen *et al*. have reported several triaminotriazine-based small molecule ligands that inhibit abnormal expansion of CTG repeats associated with myotonic dystrophy type 1 pathobiology by inhibiting transcription through DNA binding, as well as by binding to the toxic RNA to inhibit aberrant protein binding (50). Siboni *et al*. have shown that actinomycin D is able to reduce CUG repeat-associated toxic RNA in *in vitro* and *in vivo* models by inhibiting the transcription of CTG repeat expansion (51). Therefore, understanding the structural details of sequence-specific binding modes in repeat-associated DNA including the G4C2 DNA duplex and G4C2 RNA–DNA hybrid might provide useful leads for designing small molecules targeting these special sequences. In the current study, two relevant crystal structures of the G4C2 repeat motif containing DNA and RNA–DNA hybrid duplexes were solved. These show that the G4C2 DNA duplex is more flexible and exhibits an A-type-like DNA duplex, albeit with a number of base morphology parameter characteristics typical of B-DNA, such as twist and base-pair rise steps.

This structure resembles an intermediate in the A-B transition with a prominent kink of over 32°, a water shell in the minor groove, indicating a hot-spot for toxic RNA transcription (52). Previous studies have shown assemblies of structured water molecules with cyclic and linear water arrangements that are observed in the major groove and minor groove of B-DNA, respectively (39,53). However, an A-type double helix usually forms when DNA is dehydrated, but is also found in DNA/RNA hybrids (54). Compared to most studied mixed DNA/RNA hybrids that are found in the A form, the G4C2 RNA–DNA hybrid duplex is more rigid, although it is also in a typical A form; the DNA strand adopts a conformation close to that of duplex RNA, with *C3′-endo* sugar puckers and a typical A-type helical backbone conformation (55–57). The crystal packing interactions for the current G4C2 DNA and G4C2 RNA–DNA hybrid were analysed and compared to other A-form DNA duplexes. The crystal packing interactions of G4C2 DNA and G4C2 RNA-DNA hybrid duplexes shows a general 'base-pair into the minor groove' type packing for the A-form duplexes. This crystal packing environment may be further categorised into two types for DNA and RNA–DNA hybrid duplexes. In the first type, the G4C2 DNA duplex shows 'end-to-centre' type packing interactions, where one end of a symmetry-related duplex formed strong intermolecular interactions with the central G5 and C6 bases of the DNA (Figure 7A). Second, the G4C2 RNA–DNA hybrid duplex showed 'end-to-side' type packing, in which symmetry-related duplexes interacted with G4 and C7 bases that are one step away from the central G5pC6 step bases (Figure 7B). This type of packing has also been found in

other A-form DNA duplexes containing the central GGCC motif (Figure 7C and D) (58). In one of the A-DNA duplex structures, the end-to-side type packing environment interacts two base pairs away from the central GpC step (Figure 7E) (59). The crystal packing interactions in G4C2 DNA are mainly stabilised through direct hydrogen bonding, water-mediated interactions, and van der Waals contacts. Interestingly, the G4C2 RNA–DNA hybrid possesses an extra hydroxyl moiety on the RNA strand, which results in a greater number of hydrogen-bonded interactions at the crystal packing interface. Other A-DNA duplexes showed less hydrogen bonding and van der Waals interactions at the crystal packing interface (Supplementary Table S3). The G4C2 DNA duplex is more bendable and flexible than that of the G4C2 RNA–DNA hybrid, which has more surface area of the minor groove exposed to accommodate cogent ligands, such as proteins or small molecules, to interact with the minor groove and make it an ideal target for appropriately-shaped small molecule binding.

The extensive molecular dynamics and simulations studies by Sagui *et al*. reported that the G4C2 related sequences from different reading frames formed various secondary structures containing G:G mismatches in the sense and C:C mismatches in the antisense strand (7,8). We report here that the anthracene-based metal complexes ($Ni^{II}(Chro)_2$ and $Co^{II}(Chro)_2$) selectively target the G4C2 repeat DNA motif duplexes, which provides the GGCC and CCGG tracts and preferred recognition site for these compounds (60). Therefore, we analysed the stabilizing effects of Ni^{II}-$(Chro)_2$ on the G4C2 repeat related motifs while hairpin structure formation that contains the central GGCC tract flanked by two G:C base-pairs, two G:G mismatch pairs, one G:C base-pair and one G:G mismatch pair for the sense strand, and the central CCGG tract flanked by two C:C mismatch pairs for the antisense strand (Supplementary Figure S15A). The melting temperature analysis suggested that the Ni^{II}-$(Chro)_2$ complex can firmly stabilize all the sense G4C2 repeat related motifs with the GGCC tract while the Chro complex had minor stabilizing effects on the antisense G4C2 repeat motif with the CCGG tract (Supplementary Figure S15B). However, in contrast to the same ligand-free DNA structure, the binding of $Ni^{II}(Chro)_2$ to the GGGCCG sequence showed increased additional specificity through hydrophobic interactions between $Ni^{II}(Chro)_2$ and the flanked G:C base pairs of the central GGCC tract. It was proposed that the binding of $Ni^{II}(Chro)_2$ to the GGGCCG sequence induced contraction of the minor groove and straightening of the DNA backbone by the induced fit paradigm. Previously, the recognition of $Ni^{II}(Chro)_2$ of CCG trinucleotide repeats had been identified as following the classic induced-fit mechanism by triggering large-scale DNA deformations, with the cytosine extruded out from the helix (61). The binding of the $Ni^{II}(Chro)_2$ complex to the nascent G4C2 expanded strand might therefore inhibit toxic transcript formation in the resulting abnormal G4C2 expansion sequence. *Drosophila* is a widely used organism for the modelling of various neurological diseases, such as Parkinson's and Huntington's diseases and ALS/FTD (62,63). Therefore, a *Drosophila* model was used that expressed up to 30 G4C2 repeats, in order to evaluate the possi-
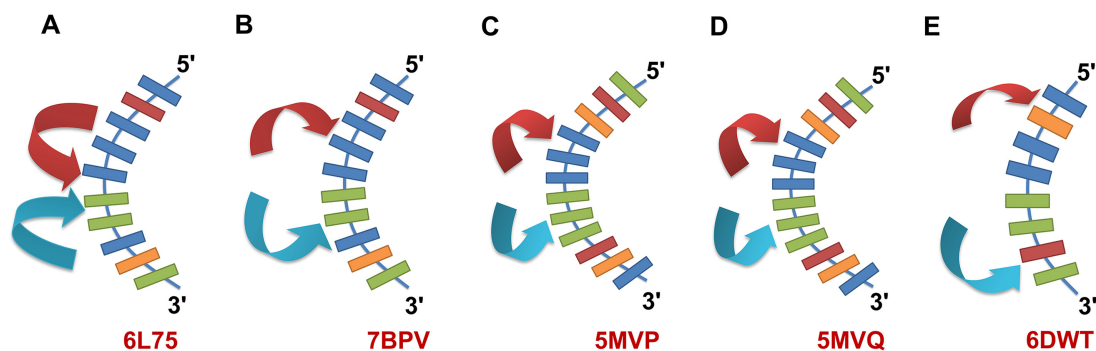
Figure 7. Schematic representation of 'base-pair into minor groove' type crystal packing interactions in different A-form DNA duplex structures. (**A**) The G4C2 DNA duplex showed 'end-to-center' type of packing interactions where the ends of DNA duplexes from different symmetry unit interact with the central GGCC residue core in the DNA. (**B**) G4C2 RNA-DNA hybrid showed 'end-to-side' type of packing environment where the interactions of different duplexes from crystal symmetry are away from the central GGCC core. (**C–E**) Other typical A-DNA duplexes showed 'end-to-side' type of crystal packing interactions. Arrows indicate symmetry related DNA duplexes. The DNA bases are colored as adenine-orange, thymine-brown, cytosine-green and guanine-blue. The respective PDB IDs for each structure are given in bold, red letters.

bility of anthracene-metal complexes targeting ALS/FTD repeat-related sequences. These results demonstrate that $Ni^{II}(Chro)_2$ and $Co^{II}(Chro)_2$ inhibit neuronal toxicity and suppress locomotor deficits caused by the expression of G4C2 repeats in the fly model and raises the possibility of using a DNA-binding compound to treat G4C2-related disease pathogenesis.

In summary, it has been demonstrated that the G4C2 motifs of C9orf72 repeat-expanded DNA provide an appropriate structural environment with considerable flexibility, serving as hot spots for selective drug binding. It is notable that of the various structurally diverse DNA-binding compounds evaluated for their ability to stabilize the repeat motif, only the two anthracene-based metal complexes showed substantial effects. At this point even though it is not yet possible to definitively state whether the biological responses observed here are due to the metal complexes interacting with hairpin, duplex or quadruplex forms of the G4C2 repeat expansion, we suggest that this has revealed an alternative drug discovery model, directly targeting G4C2 repeat DNAs, that could increase the repertoire of lead compounds for developing new therapeutic regimes against ALS and FTD diseases.

DATA AVAILABILITY

The atomic coordinates and structure factors for the reported crystal structures have been deposited in the PDB under the accession numbers 6L75 for d(GTGGGCCGAC/GTCGGCCCAC) (G4C2 DNA duplex), 7BPV for r(GUGGGCCGAC)/d(GTCGGCCCAC) (G4C2 RNA-DNA hybrid duplex) and 6L76 for the $Ni^{II}(Chro)_2$-d(TTGGGCCGAA/TTCGGCCCAA) complex.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We sincerely thank the National Synchrotron Radiation Research Center (Taiwan) staff for X-ray data collection. We

REFERENCES

1. DeJesus-Hernandez,M., Mackenzie,I.R., Boeve,BF., Boxer,A.L., Baker,M., Rutherford,N.J., Nicholson,A.M., Finch,NiCA., 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.
2. Renton,AE., Majounie,E., Waite,A., Simón-Sánchez,J., Rollinson,S., Gibbs,J.R., Schymick,JC., Laaksovirta,H., van Swieten,J.C., Myllykangas,L. *et al.* (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*, **72**, 257–268.
3. Babić,L., Župunski,M.V., Kirincich,J., Smilović,D., Hortobágyi,T., Hof,P.R. and Šimić,G. (2019) Molecular mechanisms of neurodegeneration related to C9orf72 hexanucleotide repeat expansion. *Behav. Neurol.*, **2019**, 2909168.
4. Jiang,J. and Ravits,J. (2019) Pathogenic mechanisms and therapy development for C9orf72 amyotrophic lateral sclerosis/frontotemporal dementia. *Neurotherapeutics*, **16**, 1115–1132.
5. Šket,P., Pohleven,J., Kovanda,A., Štalekar,M., Župunski,V., Zalar,M., Plavec,J. and Rogelj,B. (2015) Characterization of DNA G-quadruplex species forming from C9ORF72 G4C2-expanded repeats associated with amyotrophic lateral sclerosis and frontotemporal lobar degeneration. *Neurobiol. Aging*, **36**, 1091–1096.
6. Brčić,J. and Plavec,J. (2018) NMR structure of a G-quadruplex formed by four d(G4C2) repeats: insights into structural polymorphism. *Nucleic Acids Res.*, **46**, 11605–11617.
7. Zhang,Y., Roland,C. and Sagui,C. (2017) Structure and dynamics of DNA and RNA double helices obtained from the GGGGCC and CCCCGG hexanucleotide repeats that are the hallmark of C9FTD/ALS diseases. *ACS Chem. Neurosci.*, **8**, 578–591.
8. Zhang,Y., Roland,C. and Sagui,C. (2018) Structural and dynamical characterization of DNA and RNA quadruplexes obtained from the GGGGCC and GGGCCT hexanucleotide repeats associated with

C9FTD/ALS and SCA36 diseases. *ACS Chem. Neurosci.*, **9**, 1104–1117.

9. Waite,A.J., Bäumer,D., East,S., Neal,J., Morris,H.R., Ansorge,O. and Blake,D.J. (2014) Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion. *Neurobiol. Aging*, **35**, 1779.

10. Frick,P., Sellier,C., Mackenzie,I.R.A., Cheng,C.-Y., Tahraoui-Bories,J., Martinat,C., Pasterkamp,R.J., Prudlo,J., Edbauer,D., Oulad-Abdelghani,M. *et al.* (2018) Novel antibodies reveal presynaptic localization of C9orf72 protein and reduced protein levels in C9orf72 mutation carriers. *Acta Neuropathol. Commun.*, **6**, 72.

11. Su,Z., Zhang,Y., Gendron,TF., Bauer,PO., Chew,J., Yang,W.-Y., Fostvedt,E., Jansen-West,K., Belzil,VV., 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.

12. Barker,H.V., Niblock,M., Lee,Y.-B., Shaw,C.E. and Gallo,J.-M. (2017) RNA Misprocessing in C9orf72-Linked Neurodegeneration. *Front. Cell Neurosci.*, **11**, 195.

13. Lee,Y.-B., Chen,H.-J., Peres,JN., Gomez-Deza,J., Attig,J., Štalekar,M., Troakes,C., Nishimura,A.L., Scotter,E.L., Vance,C. *et al.* (2013) Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep.*, **5**, 1178–1186.

14. Xu,Z., Poidevin,M., Li,X., Li,Y., Shu,L., Nelson,D.L., Li,H., Hales,C.M., Gearing,M., Wingo,T.S. *et al.* (2013) Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proc. Natl. Acad. Sci.*, **110**, 7778–7783.

15. Nguyen,L., Cleary,J.D. and Ranum,L.P.W. (2019) Repeat-associated non-ATG translation: molecular mechanisms and contribution to neurological disease. *Annu. Rev. Neurosci.*, **42**, 227–247.

16. Zamiri,B., Reddy,K., Macgregor,R.B. Jr and Pearson,C.E. (2014) TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)n repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. *J. Biol. Chem.*, **289**, 4653–4659.

17. 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(G4C2)exp in c9ALS/FTD is repeat-associated non-ATG translated and a target for bioactive small molecules. *Cell Chem. Biol.*, **26**, 179–190.

18. Donnelly,CJ., Zhang,P.-W., Pham,JT., Haeusler,AR., Mistry,NA., Vidensky,S., Daley,E.L., Poth,EM., Hoover,B., Fines,D.M. *et al.* (2013) RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron*, **80**, 415–428.

19. 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.

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. Lu,Y., Dohno,C. and Nakatani,K. (2020) Recognition of expanded GGGGCC hexanucleotide repeat by synthetic ligand through interhelical binding. *Biochem. Biophys. Res. Commun.*, **531**, 56–61.

22. Shibata,T., Murakami,E. and Nakatani,K. (2018) 1,3-Di(quinolin-2-yl)guanidine binds to GGCCCC hexanucleotide repeat DNA in C9ORF72. *Bioorg. Med. Chem. Lett.*, **28**, 2364–2368.

23. Murphy,J.H. and Trapane,T.L. (1996) Concentration and extinction coefficient determination for oligonucleotides and analogs using a general phosphate analysis. *Anal. Biochem.*, **240**, 273–282.

24. Otwinowski,Z. and Minor,W. (1997) In: *Methods in Enzymology*. Academic Press, Vol. 276, pp. 307–326.

25. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallograph. Section D, Biol. Crystallograph.*, **66**, 213–221.

26. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallograph. Section D, Biol. Crystallograph.*, **53**, 240–255.

27. Li,S., Olson,W.K. and Lu,X.-J. (2019) Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.*, **47**, W26–W34.

28. Blanchet,C., Pasi,M., Zakrzewska,K. and Lavery,R. (2011) CURVES+ web server for analyzing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res.*, **39**, W68–W73.

29. Laskowski,R.A. and Swindells,M.B. (2011) LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.*, **51**, 2778–2786.

30. Olson,W.K., Bansal,M., Burley,S.K., Dickerson,R.E., Gerstein,M., Harvey,S.C., Heinemann,U., Lu,X.-J., Neidle,S., Shakked,Z. *et al.* (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.*, **313**, 229–237.

31. Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, **4**, 187–217.

32. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.

33. Darden,T., York,D. and Pedersen,L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.

34. May,R.A. and Stevenson,K.J. (2009) Software Review of Origin 8. *J. Am. Chem. Soc.*, **131**, 872–872.

35. Jin,P., Zarnescu,D.C., Zhang,F., Pearson,C.E., Lucchesi,J.C., Moses,K. and Warren,S.T. (2003) RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila. *Neuron*, **39**, 739–747.

36. Neidle,S. (2008) In: *Principles of Nucleic Acid Structure*. Academic Press, NY, pp. 38–80.

37. Wang,A.H.-J. and Teng,M.K. (1988) Crystallization and crystal packing analysis of DNA oligonucleotides. *J. Cryst. Growth*, **90**, 295–310.

38. Liu,Y., Kumar,A., Depauw,S., Nhili,R., David-Cordonnier,M.-H., Lee,M.P., Ismail,M.A., Farahat,A.A., Say,M., Chackal-Catoen,S. *et al.* (2011) Water-mediated binding of agents that target the DNA minor groove. *J. Am. Chem. Soc.*, **133**, 10171–10183.

39. Wei,D., Wilson,W.D. and Neidle,S. (2013) Small-molecule binding to the DNA minor groove is mediated by a conserved water cluster. *J. Am. Chem. Soc.*, **135**, 1369–1377.

40. Hsu,C.-W., Chuang,S.-M., Wu,W.-L. and Hou,M.-H. (2012) The crucial role of divalent metal ions in the DNA-acting efficacy and inhibition of the transcription of dimeric chromomycin A3. *PLoS One*, **7**, e43792.

41. Hou,M.H., Robinson,H., Gao,Y.G. and Wang,A.H.J. (2004) Crystal structure of the [Mg2+-(chromomycin A3)2]–d(TTGGCCAA)2 complex reveals GGCC binding specificity of the drug dimer chelated by a metal ion. *Nucleic Acids Res.*, **32**, 2214–2222.

42. Kumar,V., Hasan,G.M. and Hassan,M.I. (2017) Unraveling the role of RNA mediated toxicity of C9orf72 repeats in C9-FTD/ALS. *Front. Neurosci.*, **11**, 711–711.

43. Reddy,K., Zamiri,B., Stanley,S.Y., Macgregor,R.B. Jr and Pearson,C.E. (2013) The disease-associated r(GGGGCC)n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. *J. Biol. Chem.*, **288**, 9860–9866.

44. Fratta,P., Mizielinska,S., Nicoll,A.J., Zloh,M., Fisher,E.M.C., Parkinson,G. and Isaacs,A.M. (2012) C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci. Rep.*, **2**, 1016.

45. Jiang,J., Zhu,Q., Gendron,TF., Saberi,S., McAlonis-Downes,M., Seelman,A., Stauffer,JE., Jafar-nejad,P., Drenner,K., Schulte,D. *et al.* (2016) Gain of toxicity from ALS/FTD-linked repeat expansions in C9ORF72 is alleviated by antisense oligonucleotides targeting GGGGCC-containing RNAs. *Neuron*, **90**, 535–550.

46. Meyer,S.M., Williams,C.C., Akahori,Y., Tanaka,T., Aikawa,H., Tong,Y., Childs-Disney,J.L. and Disney,M.D. (2020) Small molecule recognition of disease-relevant RNA structures. *Chem. Soc. Rev.*, **49**, 7167–7199.

47. Satange,R., Chang,C.-k. and Hou,M.-H. (2018) A survey of recent unusual high-resolution DNA structures provoked by mismatches, repeats and ligand binding. *Nucleic Acids Res.*, **46**, 6416–6434.

48. Hou,M.-H., Satange,R. and Chang,C.-K. (2018) In: Waring,M.J. (ed). *DNA-targeting Molecules as Therapeutic Agents*. The Royal Society of Chemistry, pp. 144–174.

49. Chien,C.-M., Wu,P.-C., Satange,R., Chang,C.-C., Lai,Z.-L., Hagler,L.D., Zimmerman,S.C. and Hou,M.-H. (2020) Structural basis for targeting T:T mismatch with triaminotriazine-acridine conjugate induces a U-Shaped head-to-head four-way junction in CTG repeat DNA. *J. Am. Chem. Soc.*, **142**, 11165–11172.

50. Nguyen,L., Luu,L.M., Peng,S., Serrano,J.F., Chan,H.Y.E. and Zimmerman,S.C. (2015) Rationally designed small molecules that target both the DNA and RNA causing myotonic dystrophy type 1. *J. Am. Chem. Soc.*, **137**, 14180–14189.

51. Siboni,R.B., Nakamori,M., Wagner,S.D., Struck,A.J., Coonrod,L.A., Harriott,SA., Cass,D.M., Tanner,M.K. and Berglund,J.A. (2015) Actinomycin D specifically reduces expanded CUG repeat RNA in myotonic dystrophy models. *Cell Rep.*, **13**, 2386–2394.

52. Vargason,J.M., Henderson,K. and Ho,P.S. (2001) A crystallographic map of the transition from B-DNA to A-DNA. *Proc. Natl. Acad. Sci.*, **98**, 7265–7270.

53. Minasov,G., Tereshko,V. and Egli,M. (1999) Atomic-resolution crystal structures of B-DNA reveal specific influences of divalent metal ions on conformation and packing11Edited by T. Richmond. *J. Mol. Biol.*, **291**, 83–99.

54. Ivanov,V.I. and Krylov,D.Y. (1992) In: *Methods in Enzymology*. Academic Press, Vol. **211**, pp. 111–127.

55. Conn,G.L., Brown,T. and Leonard,G.A. (1999) The crystal structure of the RNA/DNA hybrid r(GAAGAGAAGC)·d(GCTTCTCTTC) shows significant differences to that found in solution. *Nucleic Acids Res.*, **27**, 555–561.

56. Han,G.W., Kopka,M.L., Langs,D., Sawaya,M.R. and Dickerson,R.E. (2003) Crystal structure of an RNA·DNA hybrid reveals intermolecular intercalation: dimer formation by base-pair swapping. *Proc. Natl. Acad. Sci.*, **100**, 9214–9219.

57. Wahl,M.C. and Sundaralingam,M. (2000) B-form to A-form conversion by a 3′-terminal ribose: crystal structure of the chimera d(CCACTAGTG)r(G). *Nucleic Acids Res.*, **28**, 4356–4363.

58. Hardwick,J.S., Ptchelkine,D., El-Sagheer,A.H., Tear,I., Singleton,D., Phillips,S.E.V., Lane,A.N. and Brown,T. (2017) 5-Formylcytosine does not change the global structure of DNA. *Nat. Struct. Mol. Biol.*, **24**, 544–552.

59. Hou,C. and Tsodikov,O.V. (2019) Utilizing guanine-coordinated Zn(2+) ions to determine DNA crystal structures by single-wavelength anomalous diffraction. *Acta Crystallograph. Sect. D, Struct. Biol.*, **75**, 32–40.

60. Hou,M.-H., Lu,W.-J., Lin,H.-Y. and Yuann,J.-M.P. (2008) Studies of sequence-specific DNA binding, DNA cleavage, and topoisomerase I inhibition by the dimeric chromomycin A3 complexed with feII. *Biochemistry*, **47**, 5493–5502.

61. Tseng,W.-H., Chang,C.-k., Wu,P.-C., Hu,N.-J., Lee,G.-H., Tzeng,C.-C., Neidle,S. and Hou,M.-H. (2017) Induced-fit recognition of CCG trinucleotide repeats by a nickel–chromomycin complex resulting in large-scale DNA deformation. *Angew. Chem. Int. Ed.*, **56**, 8761–8765.

62. Hirth,F. (2010) Drosophila melanogaster in the study of human neurodegeneration. *CNS Neurol. Disord. Drug Targets*, **9**, 504–523.

63. Yuva-Aydemir,Y., Almeida,S. and Gao,F.-B. (2018) Insights into C9ORF72-related ALS/FTD from Drosophila and iPSC models. *Trends Neurosci.*, **41**, 457–469.