Nucleic Acids Research, 2024, **52**, 1591–1601
https://doi.org/10.1093/nar/gkae052
Advance access publication date: 31 January 2024
Chemical Biology and Nucleic Acid Chemistry

Dimeric structures of DNA ATTTC repeats promoted by divalent cations

Marko Trajkovski[1,*], Annalisa Pastore[2] and Janez Plavec[1,3,4,*]

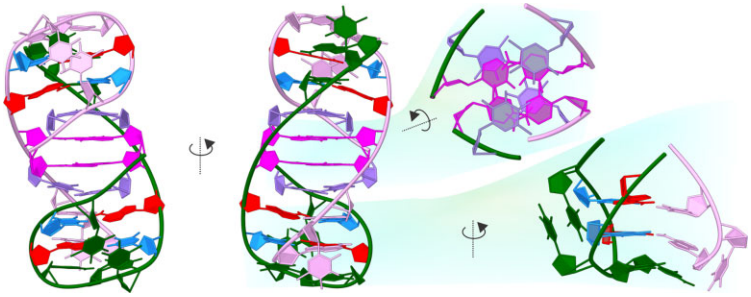
[1]Slovenian NMR Centre, National Institute of Chemistry, 1000 Ljubljana, Slovenia
[2]King's College London, the Maurice Wohl Clinical Neuroscience Institute, London, UK
[3]Faculty of Chemistry and Chemical Technology, University of Ljubljana, 1000 Ljubljana, Slovenia
[4]EN-FIST, Center of Excellence, 1000 Ljubljana, Slovenia

[*]To whom correspondence should be addressed. Tel: +386 1 4760353; Fax: +386 1 4760300; Email: janez.plavec@ki.si
Correspondence may also be addressed to Marko Trajkovski. Email: marko.trajkovski@ki.si

Abstract

Structural studies of repetitive DNA sequences may provide insights why and how certain repeat instabilities in their number and nucleotide sequence are managed or even required for normal cell physiology, while genomic variability associated with repeat expansions may also be disease-causing. The pentanucleotide ATTTC repeats occur in hundreds of genes important for various cellular processes, while their insertion and expansion in noncoding regions are associated with neurodegeneration, particularly with subtypes of spinocerebellar ataxia and familial adult myoclonic epilepsy. We describe a new striking domain-swapped DNA–DNA interaction triggered by the addition of divalent cations, including Mg^{2+} and Ca^{2+}. The results of NMR characterization of $d(ATTTC)_3$ in solution show that the oligonucleotide folds into a novel 3D architecture with two central $C:C^+$ base pairs sandwiched between a couple of T:T base pairs. This structural element, referred to here as the TCCTzip, is characterized by intercalative hydrogen-bonding, while the nucleobase moieties are poorly stacked. The 5′- and 3′-ends of TCCTzip motif are connected by stem-loop segments characterized by A:T base pairs and stacking interactions. Insights embodied in the non-canonical DNA structure are expected to advance our understanding of why only certain pyrimidine-rich DNA repeats appear to be pathogenic, while others can occur in the human genome without any harmful consequences.

Graphical abstract



Introduction

Nucleic acids are characterized by their dynamic and polymorphic nature, which allows them to adopt a variety of 3D structures in addition to the well-known double-stranded helix first described 70 years ago (1). The folding of DNA into the energetically preferred structure is based on the primary sequence with its inherent ability to hybridize through complementary Watson-Crick and other base pairing geometries, as well as various other interactions such as base-base stacking, which may depend on environmental stimuli (2). For instance, in the presence of monovalent cations guanine-rich fragments tend to adopt G-quadruplexes formed by Hoogsteen-type base pairing between four guanines connected into G-quartets. The other well-studied structure adopted by C-rich fragments is the i-motif, which is characterized by a four-stranded intercalated structure stabilized by the formation of hemiprotonated $C:C^+$ base pairs with pronounced pH sensitivity. G-quadruplexes and i-motifs are not only exotic structures formed under special laboratory conditions, but can also form in living cells (3–5). Increasing evidence of their formation *in vivo* and the growing number of nucleotide sequences originating from regulatory genomic regions that can fold into stable, non-canonical structures have led to a reconsideration of fundamental questions, such as how nucleic acids enable, direct, control, and propagate biological events in response to intra- and intercellular conditions.

Findings over the last decades keep underlining large portions of genomic DNA that used to be considered redundant in terms of structure and function (6–8). In fact, the next-generation and long-read sequencing have marked a mile-

Received: October 26, 2023. **Revised:** January 10, 2024. **Editorial Decision:** January 14, 2024. **Accepted:** January 16, 2024

stone in the genome-wide discovery of repetitive sequences far beyond G- and C-rich fragments, culminating in more than a million annotated tandem repeats. While expansions of certain tandem repeats have been associated with a range of abnormalities (9–11), the inherent variability in number of repeats and their sequences is crucial for sustaining normal cell physiological conditions. The origin and cellular roles of repeat motifs appear to be inextricably linked to the formation of non-canonical structures (12), particularly triplexes (13,14), G-quadruplexes (15–18), i-motifs (19), cruciforms and others containing hairpin-like motifs (20–23).

Tandem repeats characterized by up to six nucleotide repeat units have been linked to biologically important regulatory functions (24–26) and their relationships to functions such as DNA replication, repair, and recombination processes are the subject of ongoing studies (6,7,27). Most of the recently discovered pentanucleotide repeat expansions consist of pathogenic motifs that differ from the motifs in the reference and control genomes (28). In particular, insertion and expansion of d(ATTTC)$_n$ in the non-coding region of the DAB1 gene are associated with spinocerebellar ataxia type 37 (SCA37) (29) and several genes related to six subtypes of familial adult myoclonic epilepsies (FAME 1–4, 6–7) (30–33). Interestingly, at least three repeats of d(ATTTC) are found in >100 coding regions of the human genome (34) that play important roles in a number of cellular processes, including transcriptional regulation, intracellular signalling, protein and membrane trafficking, chromatin remodelling, cell adhesion and neuronal migration, all of which go beyond the 'original' role in neurodegenerative diseases. Studying how these potentially pathogenic repeat insertions have arisen and how they are related to discoveries of new repeat expansion disorders have been advanced by structural information on minidumbbells, which were recently resolved at high resolution for d(ATTCT)$_n$ (35–37), d(ATTTT)$_n$ (38) and d(ATTTC)$_n$. The partial topological similarities, like the arrangements of DNA backbone in minidumbbells and in pseudocircular G-hairpins (39,40), which are yet another class of recently discovered non-canonical structures, continues to incite the quest for understanding DNA folding beyond the primary sequence alone. Although the above studies provided structural insights into DNA repeat expansions, they mostly neglected the influence of the environment, which is of great importance and cannot be ignored. These and related achievements were pledged by studying folding of C- and G-rich sequences in response to the changes in solution's pH and concentration of monovalent cations such as K^+ and Na^+. However, the effects of divalent cations such as Mg^{2+} and Ca^{2+}, which may be crucial for DNA structure and function, require further in-depth studies.

At the outset of the current study, we hypothesized that d(ATTTC)$_n$ may adopt non-canonical structures. This was based on our preliminary NMR spectra showing the formation of C:C$^+$ base pairs that could not be predicted based on nucleotide sequence alone (41). Moreover, the same NMR spectra suggested that the structure is based on T:T base pairs. These observations prompted us to investigate the folding of d(ATTTC)$_n$ in more detail, revealing previously unexplored aspects of DNA structural diversity that arise from the presence of divalent cations. The general interest of the studied DNA structural transitions is further extended by the identification of the TCCTzip motif with its T:T and C:C$^+$ base pairs that are being assembled into a unique dimeric structure *via* intriguing loop residues that are themselves capable of forming base pairs and base triads. In particular, the results presented here show that the apparently similar binding of Mg^{2+} and Ca^{2+} cations to DNA induces discrete structural changes, demonstrating overall that homeostasis of divalent cations may be interrelated to cell signalling by previously overlooked feedback loops.

Materials and methods

Sample preparation

DNA oligonucleotides at natural isotope abundance as well as partially (6–8%) residue-specifically ^{13}C- and ^{15}N-isotope labelled ones were synthesized on K&A Laborgeräte GbR DNA/RNA synthesizer H-8 using standard phosphoramidite chemistry in DMT-on mode. After the synthesis the oligonucleotides were deprotected by 30 min incubation at 65°C in AMA (1:1 mixture of aqueous ammonium hydroxide and methylamine). Purification was achieved with the use of Glen-Pak cartridges after which the product was dried on a vacuum centrifuge, dissolved in 200 mM LiCl, desalted with the use of FPLC and lyophilized. Subsequently DNA oligonucleotides were dissolved in 0.5–2 ml Mili-Q water to prepare the stock solutions, which were used for preparing samples for NMR, CD and other measurements. The DNA concentrations were determined by measuring absorbance at 260 nm with the use of Varian CARY-100 UV–VIS instrument.

NMR experiments

NMR spectra were acquired on Bruker AVANCE NEO 600 and 800 MHz spectrometers equipped with QCI and TCI cryogenic probes at 5°C, if not stated differently. NMR samples were prepared from DNA stock solutions at final 0.0125–0.8 mM DNA oligonucleotide concentrations in 90%/10% $H_2O/^2H_2O$ or 100% ^2H_2O, 20 mM NaPi buffer (pH 6.0, 6.5 or 7.2). The volume of the 20 mM aqueous solutions of $MgCl_2$ and $CaCl_2$ added to the NMR samples did not exceed 2% of the sample total volume. Water suppression in ^{1}H and 2D ^{1}H–^{1}H NOESY ($\tau_m = 100$ ms) experiments was achieved by using excitation sculpting method. 2D ^{1}H–^{1}H TOCSY ($\tau_m = 50$ ms) and DQF COSY spectra were acquired in 100% ^2H_2O. ^{15}N- and ^{13}C-edited HSQC experiments were recorded on partially (6–8%) residue-specifically ^{15}N- and ^{13}C-isotopically labelled DNA oligonucleotides. Sixteen different gradient strengths (1−51 G × cm^{-1}) were used in DOSY experiments. ^{1}H NMR chemical shift were referenced with respect to the signal at δ 0.0 ppm for external standard DSS. NMR spectra were processed and analysed by using TopSpin (Bruker) and Sparky (UCSF) software.

CD experiments

CD experiments were carried out on an Applied Photophysics Chirascan CD spectrometer at temperatures between 4 and 90°C, over the 200−320 nm wavelength range by using 0.1 cm pathlength quartz cells. Each of the CD spectra corresponds to the average of three repeated measurements performed on samples with a DNA concentration of 10 μm, 20 mM $MgCl_2$ and 20 mM NaPi buffer, pH 6.0.

Structural restraints and calculations

Starting structure was generated with the use of leap module of AMBER 20 (42) and 3DNA (43) software packages. The distance restraints were derived from the integral values of cross-peaks in 2D ^{1}H–^{1}H NOESY (τ_m = 100 ms) by classifying them as strong (1.8−3.6 Å), medium (2.6−5.0 Å) and weak (3.5−6.5 Å), whereby relying on the reference distance of 2.5 Å corresponding to the average value obtained for cytosines' H5–H6 correlations. The glycosidic bond torsion angle (χ) was restrained to *anti* conformation (240 ± 70°) for all residues according to the analysis of NOE interactions between intra-residual H6/H8 and H1' protons. Hydrogen-bond distance restraints were used corresponding to N1–H3 and H61–O4 in each of the $A6_{a/b}$:$T12_{a/b}$ and $A11_{a/b}$:$T7_{a/b}$ base pairs, H41–O2, O2–H41 and H3–N3 in each of $C5_{a/b}$:$C15_{b/a}$ base pairs, as well as to H3–O4 and O4–H3 in each of $T4_{a/b}$:$T14_{b/a}$ base pairs. Accounting for the Watson–Crick A:T as well as the non-canonical C:C^{+} and T:T base pairing, 16 planarity restraints were used, i.e. two per base pair. The dimeric $d(ATTTC)_3$ high-resolution structure was calculated with the use of Amber 20 software by relying on parmbsc1 force field, Born implicit solvent model and random starting velocities. The partial charges for protonated cytosine residues were derived with the use of RESP ESP charge Derive (R.E.D.) server. The calculations included two rounds of restrained 1000 ps long simulated annealing (SA) protocol. In the first SA round the force constants were 1 kcal mol^{-1} Å^{-2} for NOE-based distance, glycosidic torsion angles (χ), hydrogen-bond and planarity of base pair restraints, while they were 100 kcal mol^{-1} Å^{-2} for chirality restraints. Restraints corresponding to the intermolecular interactions were not used in the first SA round. In the second and final SA round, the force constants were 20 kcal mol^{-1} Å^{-2} for NOE-based distances and glycosidic torsion angles (χ) restraints, 2 kcal mol^{-1} Å^{-2} for hydrogen-bond and planarity of base pair restraints, while they were 100 kcal mol^{-1} Å^{-2} for chirality restraints. In both rounds of calculations the force constants for restraints were scaled from the initial value 0.1 to the final value 1.0 in the first 50 ps and held constant until the end of the calculation. The protocols included heating over the first 50 ps from 0 to 1000 K, followed by 150 ps equilibration at 1000 K, cooling over 700 ps from 1000 to 0 K and additional 100 ps at 0 K. 100 structures were calculated of which 10 structures with the lowest energy were subjected to energy minimization with a maximum of 10 000 steps to yield representative ensemble.

Results and discussion

We first explored the effect of different d(ATTTC) repeat lengths on folding into a defined structure. ^{1}H NMR spectra of oligonucleotides consisting of two to nine d(ATTTC) repeats show a broad imino signal at δ ∼11.0 ppm and broad and overlapping signals dispersed over a similar chemical shift range in aromatic, sugar and methyl regions (Supplementary Figures S1A–C). The observed spectral features are consistent with an equilibrium of different conformations peculiar to the repeating pentanucleotide unit. Attempts to examine the effects of bulk solution pH (Supplementary Figures S2 and S3), ionic strength (Supplementary Figure S4) and phosphate buffer counterions, i.e. sodium and potassium (Supplementary Figure S5) did not result in reduction in the

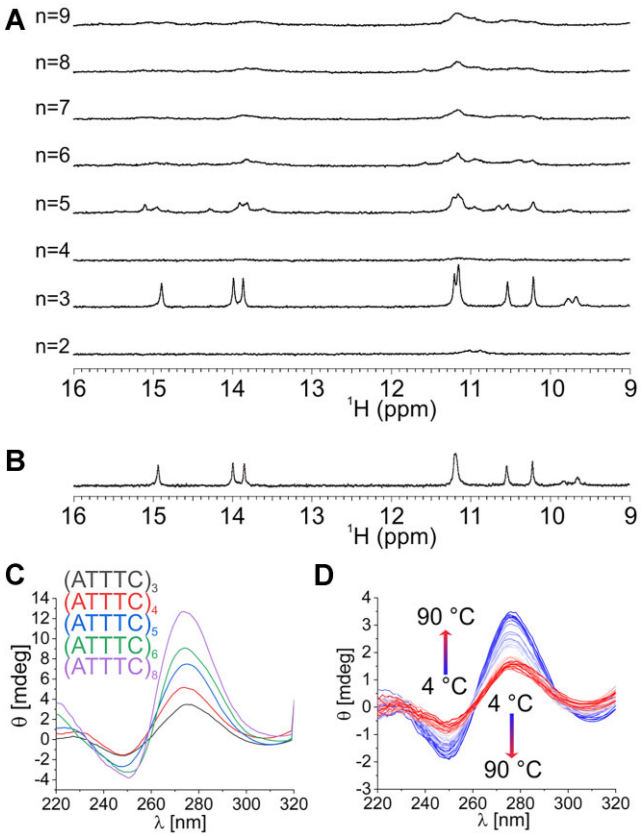


Figure 1. ^{1}H NMR and CD spectral analysis of DNA oligonucleotides with different number of d(ATTTC) repeats. Imino ^{1}H NMR spectral region of (**A**) $d(ATTTC)_{2-9}$ in the presence of 20 mM $MgCl_2$ and (**B**) $d(ATTTC)_3$ in the presence of 20 mM $CaCl_2$. (**C**) CD spectra of $d(ATTTC)_{3-8}$ and (**D**) temperature-dependent CD spectra of $d(ATTTC)_3$. ^{1}H NMR spectra were recorded in 90%/10% $H_2O/^{2}H_2O$ at 5°C, 0.3 mM DNA concentration, 20 mM NaPi buffer pH 6.0. CD spectra were recorded at 10 μm DNA concentration, 20 mM $MgCl_2$, 20 mM NaPi buffer, pH 6.0 and (C) 5°C or (D) between 4 and 90°C.

structural polymorphism and/or a better resolution of the ^{1}H NMR spectra. On the other hand, addition of $MgCl_2$ to each of $d(ATTTC)_{2-9}$ resulted in remarkable ^{1}H NMR spectral changes, with the appearance of signals in the imino region, consistent with the formation of secondary structures *via* hydrogen-bonding in Watson–Crick (δ ∼13.9 and 14.0 ppm) and non-canonical base pair (δ from 9.8 to 11.2 and ∼14.9 ppm) geometries (Figure 1A and Supplementary Figure S1). Assessment of the folding of $d(ATTTC)_3$ as a function of increasing the Mg^{2+} ion concentration revealed the gradual appearance of a specific set of ^{1}H NMR signals at 5°C. The plateau of the folded structure was reached at $MgCl_2$ concentrations between 20 and 30 mM (Supplementary Figure S6).

Role of Mg^{2+} and Ca^{2+} ions

$d(ATTTC)_3$ exhibits a very similar imino region of ^{1}H NMR spectrum in the presence of $MgCl_2$ and $CaCl_2$ (Figure 1B). These results show that even though Mg^{2+} and Ca^{2+} ions exhibit distinctly different ionic radii and other properties, their binding to DNA is crucial for folding and likely involves analogous interactions. Hence, they may act *via* modulation of interactions between different parts of the $d(ATTTC)_3$ structure, *e.g.* by neutralizing the electrostatic repulsion between negatively charged sugar-phosphate backbone(s) in a multi-

stranded structure. When the DNA concentration is lowered while the $MgCl_2$ concentration is maintained at 20 mM, the gradual unfolding of an intermolecular structure is evident from the broadening of the corresponding ^{1}H NMR signals (Supplementary Figure S7). In parallel, the relative intensity of the second set of ^{1}H NMR signals, which matches the one observed for d(ATTTC)$_3$ in the absence of the divalent cations, increases (Supplementary Figure S8). These results show that the folding of d(ATTTC)$_3$ depends on both the concentrations of Mg^{2+} ions and DNA. Analysis of the relative intensities of the ^{1}H NMR signals corresponding to the two species at 20 mM $MgCl_2$ suggests that the dimeric structure predominates above 0.3 mM DNA concentration.

Temperature-dependent changes in the ^{1}H NMR spectra of d(ATTTC)$_3$ in the presence of Mg^{2+} ions (Supplementary Figure S9) revealed a gradual broadening of the imino, aromatic and methyl signals of the dimer until their complete disappearance above 20°C. In parallel, another set of aromatic and methyl ^{1}H NMR signals corresponding to the monomeric species gradually intensifies as the temperature is increased from 5 to 20°C. These signals initially appear broad, consistent with a slow exchange of partially folded DNA, while complete unfolding at 55°C is indicated by the observed single set of narrow signals.

Comparative CD spectral analysis in the presence of $MgCl_2$ revealed similar profiles for d(ATTTC)$_{3–8}$, consistent with their common structural features (Figure 1C). The spectral maximum and minimum near 275 and 250 nm, respectively, are consistent with hairpin secondary structures (44–46). The temperature-dependent CD spectra of d(ATTTC)$_3$ showed two isosbestic points at 260 and 295 nm (Figure 1D), suggesting temperature induced denaturation of two intermediate structures. Analysis of temperature induced unfolding of d(ATTTC)$_3$ monitored at the maximum and minimum of the CD spectrum affords the apparent T_m of ca. 28°C (Supplementary Figure S10). However, the melting profile at 250 nm indicates that unfolding may involve multiple structural transitions and reaches a plateau at temperatures above 50°C. Since the spectral melting profiles correspond to the unfolding of monomeric and dimeric structures, and considering that lower DNA concentrations were used in the CD experiments compared to the NMR experiments, the apparently higher T_m values observed in the CD analysis suggest that the monomeric structure may have higher thermodynamic stability compared to the dimeric structure.

The amount of dimeric d(ATTTC)$_3$ structure at equilibrium decreases when the pH of the solution is changed from 6.0 to 6.5, and decreases to a minimum when the pH is further increased to pH 7.0, favoring the formation of a monomeric structure (Supplementary Figure S11).

NMR assignment

Deciphering the structures adopted by d(ATTTC)$_3$ required unambiguous assignment of the NMR resonances, which relied on the use of partial residue-specific ^{13}C- and ^{15}N-isotopically labelled oligonucleotides combined with ^{13}C- and ^{15}N-edited HSQC experiments. Indeed, the increased spectral resolution of the heteronuclear NMR experiments enabled assignment of the ^{1}H NMR signals corresponding to the monomeric structure (Supplementary Figures S12 and S13) and the dimeric structure formed in the presence of Mg^{2+} ions (Supplementary Table S1 and Supplementary Figures S14–S18). Of particular note, only fifteen sets of NMR signals are observed for the dimeric d(ATTTC)$_3$ structure, consistent with a symmetrical arrangement of the two DNA strands in which equivalent positions have identical chemical shielding environments.

Assessment of hydrodynamic properties by diffusion ordered spectroscopy (DOSY) analysis afforded the translation diffusion coefficients (D_t) of 0.8 (± 0.1) × 10^{-10} m^2 s^{-1} for both, the monomeric and dimeric structures adopted by d(ATTTC)$_3$ in the absence and in the presence of Mg^{2+} ions, respectively (Supplementary Figure S19). The similar D_t values suggest that the hairpin-like d(ATTTC)$_3$ monomers upon addition of Mg^{2+} ions self-associate and form a compact dimeric structure. Noteworthy, addition of the complementary d(GAAAT)$_3$ to d(ATTTC)$_3$ leads to the formation of a duplex, which prevails over the non-canonical structures regardless on the presence of $MgCl_2$ in the sample. At 20 mM $MgCl_2$, the d(ATTTC)$_3$/d(GAAAT)$_3$ duplex exhibits a D_t of 0.6 (± 0.1) × 10^{-10} m^2 s^{-1}, consistent with a rod-like topology, which is more elongated in comparison to the monomeric and dimeric structures adopted by d(ATTTC)$_3$ itself in the absence and presence of $MgCl_2$, respectively (Supplementary Figure S20).

Remarkably, the signals observed in the ^{15}N HSQC spectra of d(ATTTC)$_3$ in the presence of Mg^{2+} ions at 5°C indicate that the imino protons of all thymine residues are protected from exchange with the solvent (Figure 2A and Supplementary Figure S16). Specifically, the ^{15}N HSQC spectra of d(ATTTC)$_3$ with a single partially ^{15}N-isotopically labelled thymine at positions 2, 8 or 9 exhibit a signal with a very similar ^{1}H and ^{15}N chemical shift, i.e. at δ 11.17 and 156 ppm, respectively. These results suggest that the imino protons of T2, T8 and T9 may not be hydrogen-bonded but are in slow exchange with the bulk solvent due to the experimental conditions of relatively low temperature and pH (6.0). The imino protons of T7 and T12 resonate in a range of chemical shift characteristic of Watson–Crick A:T base pairs, *i.e.* at ^{1}H δ 13.87 and 13.99 ppm, respectively. On the other hand, T3, T4, T13 and T14 exhibit ^{1}H NMR imino signals at δ 9.80, 10.22, 10.55 and 11.22 ppm, respectively, consistent with their hydrogen-bonding in a (de)shielded environment due to pronounced ring current effects. Noteworthy, the ^{1}H NMR signal observed at δ 14.90 ppm corresponds to C5 and C15, suggesting that they form a hemi-protonated C:C^+ base pair.

Analysis of the NOESY spectra of d(ATTTC)$_3$ in the presence of Mg^{2+} ions suggested that the dimeric structure includes a stem-loop segment in which the Watson-Crick base paired A6:T12 and T7:A11 are bridged by T8–T9–C10 loop (Supplementary Figure S21). The NOESY cross-peaks between imino and methyl protons of T4 and T14 (Figure 2 and Supplementary Figure S22), as well as between amino protons of C5 and C15 suggested formation of non-canonical T4:T14 and C5:$C15^+$ base pairs, respectively. Interestingly, the observed imino-imino and imino-aromatic proton NOE correlations show that the consecutive nature of the T7:A11 and A6:T12 base pairs extends across the non-canonical T4:T14 and C5:$C15^+$ base pairs (Figures 2A–C). Furthermore, the mutual orientations of T4 and C5 and of T14 and C15 give rise to 9 and 6 inter-residual proton-proton NOE interactions, respectively (Supplementary Table S2), that can be interpreted consistently only in the context of a peculiar four-stranded interface constituting the centre of a dimeric struc-

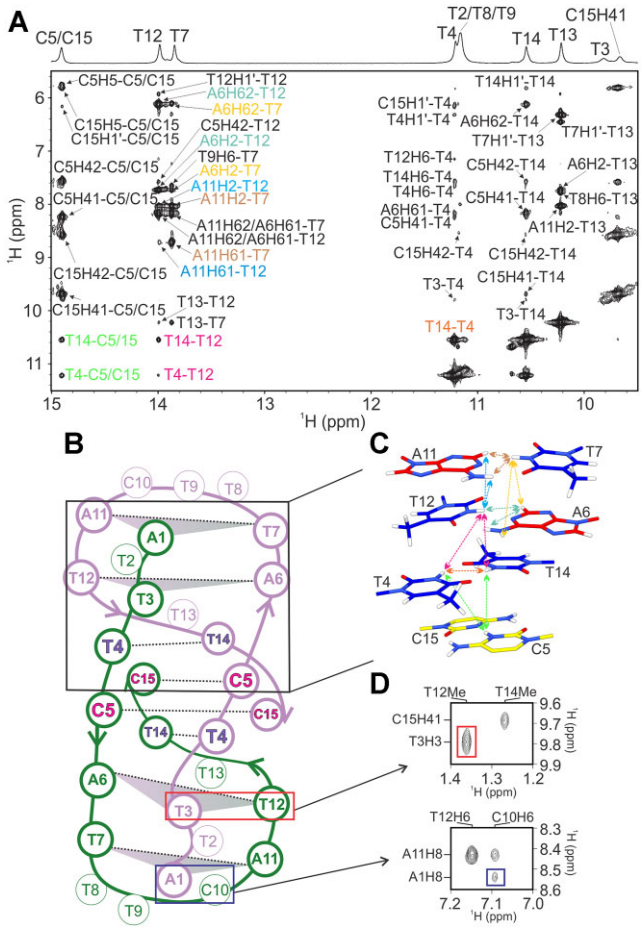


Figure 2. Region from the 2D NOESY spectrum and schematic representation of the dimeric structure of d(ATTTC)$_3$ in the presence of $MgCl_2$. (**A**) NOESY spectrum of d(ATTTC)$_3$ (τ_m = 100 ms) and the corresponding region of the 1D ^{1}H NMR spectrum (on top) with the respective assignments. The key NOESY cross-peaks reflecting the sequential arrangement of T7:A11, A6:T12, T4:T14 and C5:C15$^+$ base pairs are color-coded to match the corresponding representation in panel C. (**B**) Folding topology with the two strands of the dimer distinguished by green and pink colours. The circles depict residues, with the thicker outlines used for those in base pairs and base triads, while thinner lines mark residues that are not base paired. Hydrogen-bonds within the A6:T12 and T7:A11 Watson-Crick base pairs and the C5:C15$^+$ and T4:T14 non-canonical base pairs are depicted by dotted black lines. The shaded triangles highlight T7:A11:A1 and A6:T12:T3 base triads. Red and blue squares highlight the proximity of T3 and T12, and of A1 and C10, respectively. (**C**) Spatial orientation of the four base pairs with inter-residue distances corresponding to the coloured cross-peaks (*cf.* panel A) in the NOESY spectrum. (**D**) Sections of the NOESY spectrum showing the proximity of T3 and T12 and of A1 and C10. The spectra were recorded at 5°C in 90%/10% $H_2O/^2H_2O$, 20 mM NaPi, pH 6.0, 0.8 mM DNA concentration and 20 mM $MgCl_2$.

ture (Figure 2). NOE interactions corresponding to the intercalation of two neighbouring C5:C15$^+$ base pairs sandwiched between a couple of T4:T14 base pairs include T4$_a$Me-C5$_b$H41, T4$_a$Me-T14$_b$H3, T14$_a$Me-C15$_b$H41 and T14$_a$Me-T4$_b$H3 (Supplementary Table S2 and Supplementary Figure S22), where '$_a$' and '$_b$' can be interchanged because of the symmetrical arrangement of the two DNA strands. Such intertwinement of two DNA strands *via* two central C5:C15$^+$ base pairs positioned between T4:T14 base pairs on each side and further extending to stem–loop segments is corroborated by NOESY cross-peaks between imino protons of inter- and intra-stranded T4:T14 and A6:T12 base pairs, respectively. The intermolecular NOE interactions of A1 with T7, T9, C10 and A11, as well as of T3 with A6, A11 and T12 (Figure 2D and Supplementary Figures S21-S22 and Supplementary Table S2) demonstrate intermolecular interactions between the 5′-end residues of one of the strands and residues in the stem-loop segment of the other.

3D Structure insights

Structure calculations based on simulated annealing and experimentally derived restraints provided high-resolution insights into the unique dimer formed by d(ATTTC)$_3$ (Figure 3A and Supplementary Figure S23). Inspection of the details of the stem-loop segment and the 5′-end shows that the orientation of their corresponding residues is defined by electrostatic and stacking interactions. In detail, the sequentially related A6 and T7 as well as A11 and T12 are stacked tightly. The T7:A11 and A6:T12 base pairs are further stabilized by stacking of T9 on T7 (Figure 3B). The calculated structure reveals stacking between the pyrimidine moieties of T8 and T13 that together cap the groove formed by T7:A11 and A6:T12 base pairs, whereby T13 is close to the sugar edge of the A:T base pairs (Figure 3B). This disposition is related to the interactions between T7 O2 and T13 H3. Electrostatic interactions are indicated by the proximities of amino protons of C10$_b$ and T2$_a$ O4, and to a lesser extent of C10$_b$ N3 and T2$_a$ H3, where a classical hydrogen bond is refuted by the corresponding nucleobase planes inclined at nearly 90°. The interactions between T2 and C10 are confirmed by the observation of the destabilization of the dimeric structure when thymine is replaced by an abasic residue at position 2 (Supplementary Figure 24). Orientation of C10$_b$ is additionally stabilized by stacking to the pyrimidine moiety of A1$_a$. The stacked C10$_b$, A1$_a$ and T3$_a$ nucleobases appear crucial for inter-strand interactions between the stem-loop and the 5′-end segments. This arrangement is coupled with a near coplanar alignment of the T3$_a$ nucleobase with respect to the A6$_b$:T12$_b$ base pair, suggesting the formation of a hydrogen-bonded base triad based on the juxtaposition of T3$_a$ H3 and T12$_b$ O4 as well as of T3$_a$ O4 and A6$_b$ H62. An additional base triad is suggested by nearly coplanar arrangement of A1$_a$ nucleobase with respect to the T7$_b$:A11$_b$ base pair, with hydrogen-bonding between A1$_a$ and A11$_b$ indicated by the proximity of their H62 and N7 atoms, respectively. The formation of T7$_b$:A11$_b$:A1$_a$ base triad is corroborated by the observed downfield ^{1}H NMR chemical shift of A11 amino (H61 and H62) protons (Supplementary Figure S18). Consideration of the structural details indicates that the interactions between the residues at the 5′-end and the stem-loop segment are crucial for stability (Supplementary Figure S25), which is corroborated also by the observation of lower ratios of dimeric structure for the oligonucleotides with truncated 5′-end with respect to d(ATTTC)$_3$ (Supplementary Figure S26).

In the dimeric d(ATTTC)$_3$ structure, T9 H4' and C10 H5'/H5'' are very close to each other and interestingly show remarkably upfield-shifted ^{1}H NMR signals (Supplementary Figure S27). These observations suggest a peculiar chemical shielding environment reflecting interactions between DNA and divalent cations or alternatively specific arrangement of the T8–T9–C10 loop and nearby T13, which is stacked on T8. To address this question, we examined the structural fea-

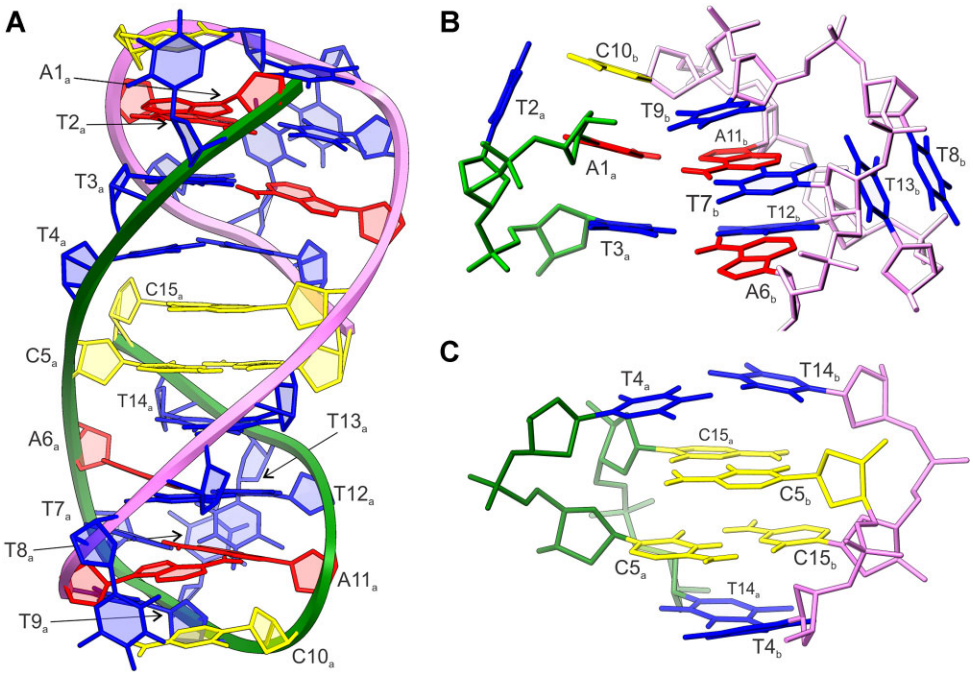


Figure 3. Solution-state dimeric d(ATTTC)$_3$ structure in the presence of $MgCl_2$ (PDB ID: 8Q5Q). (**A**) Side view on the lowest energy structure. (**B**) Details of the stem-loop segment and its interactions with the proximal 5′-end residues showing the following characteristic elements: (i) continuous stack of $C10_b$, $A1_a$ and $T3_a$, (ii) intermolecular interactions in $T7_b$:$A11_b$:$A1_a$ and $A6_b$:$T12_b$:$T3_a$ base triads, and (iii) intramolecular stacking of $T8_b$ and $T13_b$, as well as of $T9_b$ and $T7_b$. (**C**) Details of the core of the structure corresponding to block of four intercalated non-canonical base pairs, referred to here as TCCTzip, that includes symmetrically arranged C5:C15$^+$ base pairs sandwiched between T4:T14 base pairs. Adenine, cytosine and thymine residues are shown in red, yellow and blue, respectively, with the sugar-phosphate backbones of the two strands delineated by green and pink colours. NMR restrains and structural statistics are presented in Supplementary Table S3.

tures of oligonucleotides analogous to d(ATTTC)$_3$, in which thymine at position 13 is replaced by cytosine, or thymine at position 8 is replaced by either cytosine or abasic residue corresponding to a sugar-phosphate linker without nucleobase. ^{1}H NMR analysis showed that the dimeric structure is formed when T8 is substituted with C8, whereas it does not form when T8 is substituted with an abasic residue or T13 is modified to C13 (Supplementary Figure S27). These results suggest that π-π stacking interactions between the pyrimidine nucleobases at positions 8 and 13 contribute but are not crucial for the formation of the dimeric structure, which on the contrary may rely on the specific arrangement of T13 O4 and T8 (or C8) O2. The fact that in the dimeric d(ATTTC)$_3$ structure T13 O4 and T8 O2 are spatially close to each other and furthermore appear in the vicinity of T9 H4' and C10 H5'/H5' observed at notably low ^{1}H NMR chemical shift suggests that Mg^{2+} ions potentially bind in the vicinity of T9 and C10 sugar-phosphate backbone. Yet, the unambiguous identification of potential Mg^{2+} ion binding sites in the dimeric d(ATTTC)$_3$ structure requires further in-depth studies, which could indeed be worthwhile and decipher new DNA-cation interfaces.

TCCTzip motif with intercalated T:T and C:C+ base pairs

In the solution-state structure of d(ATTTC)$_3$, the block of four consecutive intercalated base pairs comprising two central C5:C15$^+$ base pairs sandwiched between the non-canonical T4:T14 base pairs (Figure 3C) represents a peculiar structural element that has not been reported previously. It constitutes a distinctive interface as the DNA strands interact *via* hydrogen-bonding, while the nucleobase moieties are poorly stacked. Hereafter, we have named it TCCTzip to facilitate comparison of its details with respect to previously reported structures in which C:C$^+$ base pairs are capped by G:T:G:T quartet (47,48), G:C:G:T quartet (49) and G:G base pairs (50). A stretch of intercalated C:C$^+$ base pairs flanked on one side by T:T base pair and by a G:T:G:T quartet on the other was recently reported by the Gonzalez group(51). Notably, it has been well established that a T:T base pair can increase the stability of i-motif not only as a capping element(41,52–54), but also when intercalated in the core of the structure between two C:C$^+$ base pairs(55,56).

With the exception of C5, the sugar puckering of all residues is well converged and consistent with the predominance of South-type conformation for A1-T2, T4 and A6-T14, and a North-type conformation for T3 and C15. In the ensemble of the final structures, C5 exhibits sugar puckering in the range between C1'-exo and the more unusual O4'-endo conformations. This flexibility continues along the backbone between C5 and A6, resulting in notable variations in the corresponding epsilon ($C5_{C4'\text{-}C3'\text{-}O3'}$-$A6_P$) and zeta ($C5_{C3'\text{-}O3'}$-$A6_{P\text{-}O5'}$) torsion angles, ranging from 190° to 278° and from 186° to 260°, respectively. Despite the noted plasticity of the backbone at the C5-A6 step, two clearly different groove geometries are defined at the inter- and intra-strand segments. The phosphate-phosphate distances of ca. 15 Å for $T4_a$-$A6_b$ and ca. 6.5 Å for $A6_b$-$C15_b$ indicate that the dimensions of the wide and the narrow grooves at this four-stranded part of the dimeric d(ATTTC)$_3$ structure are even more extreme than in the previously reported i-motifs (57). The local over-winding at the central part of the dimeric d(ATTTC)$_3$ structure is also reflected in the geometrical parameters, in particular rise of ca.

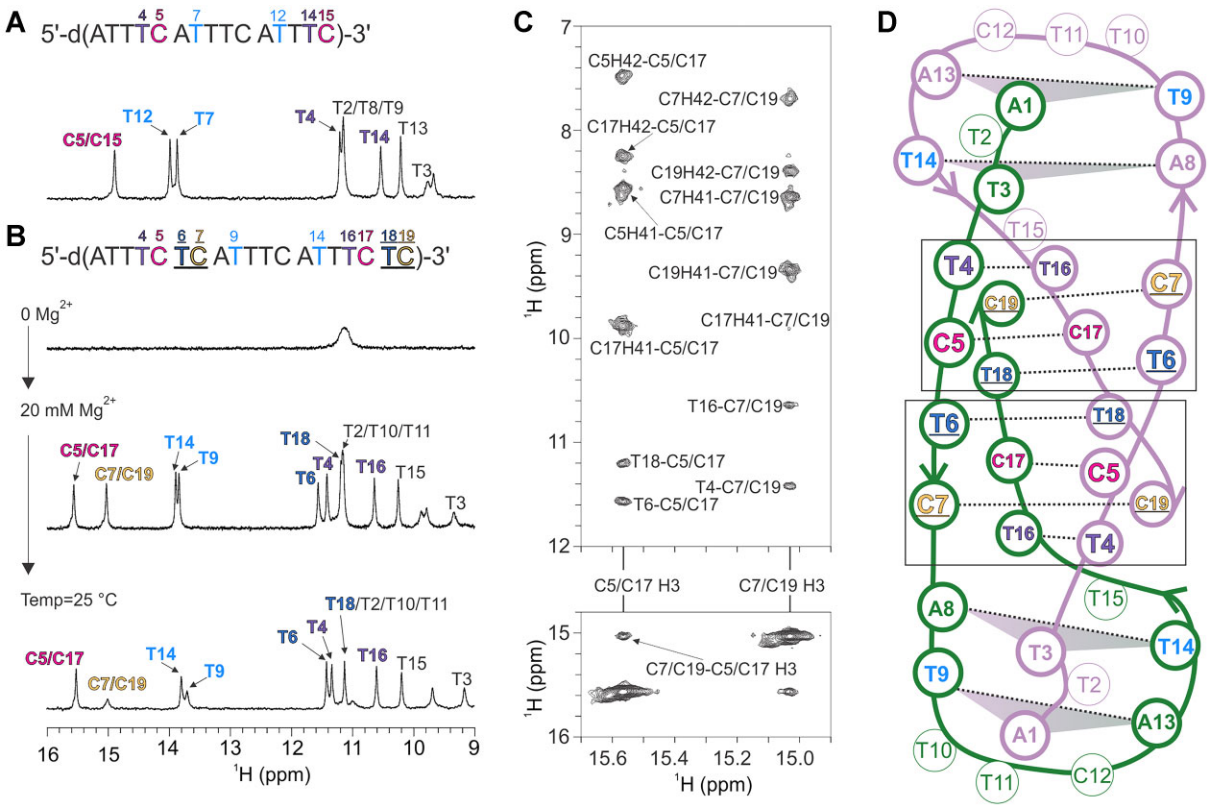


Figure 4. ^{1}H NMR spectra of (**A**) d(ATTTC)$_3$ and (**B**) of d(ATTTCTCATTTCATTTCTC), 2PyZip. The additional 'TC' dinucleotide segments in 2PyZip with respect to d(ATTTC)$_3$ are underlined in the primary sequence shown above the corresponding ^{1}H NMR spectra. (**C**) Regions from the 2D NOESY spectrum (τ_m = 100 ms) of 2PyZp showing the key dipole-dipole interactions consistent with eight intercalated base pairs: T4:T16, C7:C19$^+$, C5:C17$^+$, T6:T18, T6:T18, C5:C17$^+$, C7:C19$^+$ and T4:T16. (**D**) Folding topology of 2PyZip with the two DNA strands distinguished by green and pink colours. The circles depict residues, with the thicker outlines used for those in base pairs and base triads, while thinner lines mark residues that are not base paired. Hydrogen-bonding within the A:T Watson–Crick and the non-canonical C:C$^+$ and T:T base pairs are depicted with dotted black lines. The shaded triangles indicate T9:A13:A1 and A8:T14:T3 base triads. Black rectangles highlight the two blocks of TCCTzip structural motif. The NMR spectra were recorded in 90%/10% $H_2O/^2H_2O$, at 0.2 mM DNA, 20 mM NaPi, pH 6.0 and at 5°C and 20 mM $MgCl_2$, if not indicated differently.

6 Å and twist between 41° and 44° (Supplementary Table S4). This peculiar local arrangement of DNA strands is strongly coupled to the relative orientation of C5:C15$^+$ and T4:T14 base pairs in the TCCTzip structural motif, while it may be, at least in part, promoted by arrangement of residues in other parts of the structure. Indeed, it has been previously noted that the stacking intervals between intercalated base pairs in a core of a DNA structure depend on the conformation of residues in the neighbouring parts (50,58).

Robustness of TCCTzip motif

The mutual orientations of T4:T14 and A6:T12 base pairs that constitute the interface between the central part, i.e. TCCTzip and the stem-loop segments of the dimeric d(ATTTC)$_3$ structure indicate that nucleobase moieties and the two motifs are poorly stacked. We then reasoned that, if the TCCTzip is an isolated module embedded in the core of a dimeric structure, it should be possible to design an oligonucleotide that can form a dimeric structure with two central TCCTzip motifs. A DNA oligonucleotide with the corresponding primary sequence 5′-d(ATTTC TC ATTTC ATTTC TC)-3′, designated '2PyZip', contains two additional 'TC' dinucleotide segments with respect to d(ATTTC)$_3$ (Figure 4). The ^{1}H NMR spectrum of 2PyZip in the absence of divalent cations exhibits broad signals, suggesting an equilibrium of secondary structures in which the imino protons are in fast exchange with the bulk solvent (Figure 4B). More importantly, the spectra of 2PyZip change dramatically upon addition of $MgCl_2$, with the striking appearance of resolved imino ^{1}H NMR signals corresponding to A:T, C:C$^+$ and T:T base pairs (Figure 4B). The assignment of the imino, aromatic, sugar and methyl ^{1}H NMR resonances of 2PyZip in the presence of Mg^{2+} ions was based on NOESY spectra (Figure 4C). Analysis revealed that 2PyZip adopts a dimeric structure with a core containing two tandem TCCTzip motifs, each comprising a block of intercalated T4:T16, C7:C19$^+$, C5:C17$^+$ and T6:T18 intermolecular base pairs. The inter-residual NOESY cross-peaks between sugar protons, including H1'-H1' correlations for T4-C19, C5-C19, C5-T18, T6-T18, T6-C17, C7-C17 and C7-T16 pairs demonstrated that the central part of the structure exhibits narrow grooves between the segments T4-C5-T6-C7 and T16-C17-T18-C19 of the same strands. The analysis is furthermore consistent with the formation of a stem-loop segment comprising A8:T14 and T9:A13 base pairs that are bridged by T10-T11-C12 loop. The NOE cross-peaks corresponding to the interactions of A1H8 with C12H6 and T3H3 with T14Me are consistent with intermolecular interactions, with the 5′-end of one of the two strand proximal to the stem-loop segment of the other. Altogether, the NMR results demonstrate that 2PyZip forms a well-defined dimeric structure, that differs from the analogue dimeric d(ATTTC)$_3$ structure by its core, consist-

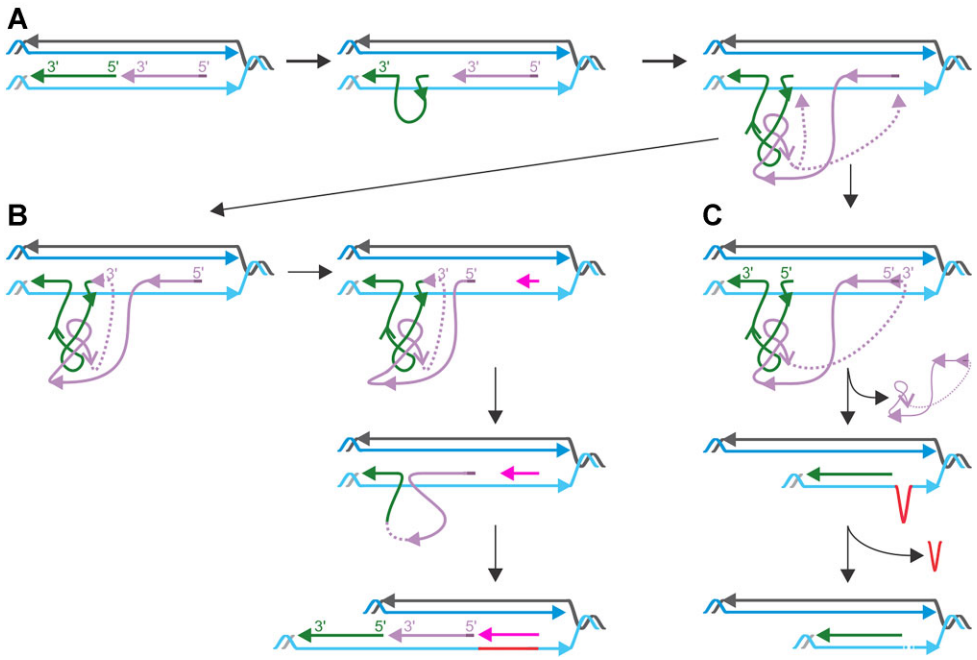


Figure 5. Proposed mechanisms of DNA replication events that include (**A**) formation of dimeric d(ATTTC)$_3$ structure that leads to (**B**) expansion or (**C**) contraction of d(ATTTC) repeats. The parental d(ATTTC)$_n$ and d(GAAAT)$_n$ strands are depicted in dark grey and light blue, respectively. The nascent leading strand is depicted in dark blue. The short nascent strands (Okazaki fragments) involved in formation of the dimeric d(ATTTC)$_3$ structure are shown in green and pink. The segments of the parental DNA strand depicted in red correspond to the parts that are synthesized or alternatively removed during expansion or contraction. The expanded repeats in the short nascent strand are depicted in magenta.

ing of two tandem TCCTzip motifs (Figure 4D). The demonstrated robustness of the TCCTzip structural motif incited the hypothesis of whether it can also form intra-molecularly and whether such a non-canonical structure can be embedded in the DNA duplex (Supplementary Figure S28). In this regard, it may be crucial to uncover yet unexplored aspects of susceptibility of the d(ATTTC)$_n$ folding to the nature and concentration of divalent cations in solution.

Functional implications in repeat instability

It is noteworthy that the two stem–loop segments that are part of the dimeric d(ATTTC)$_3$ structure resemble the minidumbbells characterized earlier, such as that of d(ATTCT)$_2$ (37) (Supplementary Figure S29). Li and Guo *et al.* recently reported the detailed structural characterization of DNA oligonucleotides with 3, 4 or 5 repeats of d(ATTTC) and found that they all adopt intramolecular minidumbbell (38). However, our results show that addition of Mg^{2+} ions in up to physiologically relevant concentrations completely changes folding of d(ATTTC)$_3$ and leads to the formation of an unprecedented dimeric structure.

The susceptibility of the folding of d(ATTTC)$_n$ oligonucleotides to the presence of divalent cations shown here and in particular the high-resolution details of the dimeric interface, are a good complement to biological studies on the variability of certain tandem repeats, including aspects of neurological disorders in SCA37 and FAMEs. The molecular model for the formation of repeat expansions is based on the slippage of the DNA strand during replication.(9,12,59–61) The combination of the strand-slippage model with the formation of an intramolecular analogue of the dimeric d(ATTTC)$_3$ structure may help to interpret repeat expansion or contraction (Supplementary Figure S28). The putative structural models could also potentially help to explain genomic variations that include fork reversal and template switching.(62,63) If the template strand or the elongated strand took this structure during DNA replication, the length of the repeat sequence would change. Additionally, the hypothetical mechanisms for the expansion and contraction are grounded in the slippage of one strand and the dissociation of the 3′-end segment of the other strand with simultaneous formation of the dimeric d(ATTTC)$_3$ structure (Figure 5A and Supplementary Figure S30A). The distinctive events arise upon different joining of the Okazaki fragments. According to the first scenario, a repeat expansion occurs when the Okazaki fragments that form the dimeric structure join at the 5′-end of the slipped strand and at the 3′-end of the strand that has partially detached from the parental strand (Figure 5B). This interaction drives the slippage of the second strand and leads to the synthesis of a new short nascent strand, resulting in an expansion of repeats. The repeat contraction occurs when the strand whose 3′ end has separated from the parental strand is joined at its 5′ and 3′ ends. This creates circular DNA that separates from the parental strand, making it appear too long, and DNA repair mechanisms remove the redundant segment (Figure 5C). The key intermediates of the proposed expansion and contraction mechanisms are shown in Supplementary Figure S30. Admittedly, additional studies are warranted to elucidate if structures analogous to the ones examined in this study may contribute to variability in the number of repeats.

As a general conceptual advance and broad appeal, our results show that either Mg^{2+} or Ca^{2+} ions promote discrete changes in DNA folding of d(ATTTC)$_n$, leading to the formation of previously unknown yet robust TCCTzip structural motif. By scrutinizing the discrete non-canonical DNA struc-

ture promoted by seemingly common DNA-cation interactions we might have explored merely an example of previously overlooked roles of divalent cations and biological events important for cellular homeostasis. Considering cation-induced structural switches, we note that C-rich DNA oligonucleotides originating from human telomeric repeat can adopt i-motif and hairpin structures in the presence of Mg^{2+} (64) or Cu^{2+} (65) cations. Interestingly, a variety of physiological functions in humans and other eukaryotes are carried out by extrachromosomal circular DNA (66) that arises from excision of chromosomal sequences in a process that requires residual amounts of Mg^{2+} ions, while ATP and sequence-specific enzymes are not needed (67). In addition, $d(ATTTC)_n$ and similar AT-rich repeats are considered as DNA unwinding elements (DUE) (35,36,68) at replication origins, wherein they promote separation of the strands, i.e. unwinding of double-stranded DNA helix. More particularly, the insights into the 3D structures of $d(ATTTC)_n$ repeats, although not representing the sequence of pathogenic repeats alone, may enable molecular diagnosis of specific types of SCA and FAMEs and development of new therapeutic strategies. Besides, d(ATTTC) repeats are enriched in a range of different genes, including MED12L that encodes a subunit of kinase module of mediator complex, vital for transcriptional coactivation of numerous RNA polymerase II-dependent genes. Clearly, further studies are needed to evaluate the potential biological roles of the herein identified TCCTzip and other details of the divalent-promoted formation of dimeric $d(ATTTC)_3$ structure. The insights may potentially be applicable also in bionanotechnological contexts, e.g. as novel tools to control DNA scaffolding.

In summary, we highlight here a striking domain-swapped like interaction triggered by the presence of divalent cations and elucidate the details of the dimeric structure that $d(ATTTC)_3$ adopts in the presence of Mg^{2+} ions. Particularly intriguing is the unique TCCTzip motif in its core, in which two $C:C^+$ base pairs are flanked on each side by T:T base pairs, whereby the base pair twists between the sequential pairs of intercalated T:T and $C:C^+$ base pairs are close to 90°. The peculiar details together with the investigated divalent cations-promoted structural transitions reveal hitherto unexplored DNA folding.

Data availability

Data has been deposited in the Protein Data Bank (https://www.rcsb.org/) and Biological Magnetic Resonance Data Bank (https://bmrb.io/) under the accession numbers PDB 8Q5Q and 34845, respectively.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

The authors acknowledge the contributions of David Bezlaj and Žan Fortuna to the preliminary experimental results, and would like to thank CERIC-ERIC consortium for the access to experimental facilities and financial support.

Funding

Slovenian Research and Innovation Agency [P1-0242, J1-1704]. Funding for open access charge: Slovenian Research and Innovation Agency [P1-0242].

Conflict of interest statement

None declared.

References

1. Watson,J.D. and Crick,F.H.C. (1953) Molecular structure of nucleic acids. *Nature*, **171**, 737–740.
2. Tateishi-Karimata,H. and Sugimoto,N. (2012) A-T base pairs are more stable than G-C base pairs in a hydrated ionic liquid. *Angew. Chem. Int. Ed.*, **51**, 1416–1419.
3. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.
4. Zeraati,M., Langley,D.B., Schofield,P., Moye,A.L., Rouet,R., Hughes,W.E., Bryan,T.M., Dinger,M.E. and Christ,D. (2018) I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.*, **10**, 631–637.
5. Summers,P.A., Thomas,A.P., Kench,T., Vannier,J.B., Kuimova,M.K. and Vilar,R. (2021) Cationic helicenes as selective G4 DNA binders and optical probes for cellular imaging. *Chem. Sci.*, **12**, 14624–14634.
6. Iglesias,A.R., Kindlund,E., Tammi,M. and Wadelius,C. (2004) Some microsatellites may act as novel polymorphic cis-regulatory elements through transcription factor binding. *Gene*, **341**, 149–165.
7. Li,Y.C., Korol,A.B., Fahima,T. and Nevo,E. (2004) Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.*, **21**, 991–1007.
8. Horton,C.A., Alexandari,A.M., Hayes,M.G.B., Marklund,E., Schaepe,J.M., Aditham,A.K., Shah,N., Suzuki,P.H., Shrikumar,A., Afek,A., *et al.* (2023) Short tandem repeats bind transcription factors to tune eukaryotic gene expression. *Science*, **381**, eadd1250.
9. Mirkin,S.M. (2007) Expandable DNA repeats and human disease. *Nature*, **447**, 932–940.
10. Nazaripanah,N., Adelirad,F., Delbari,A., Sahaf,R., Abbasi-Asl,T. and Ohadi,M. (2018) Genome-scale portrait and evolutionary significance of human-specific core promoter tri- and tetranucleotide short tandem repeats. *Hum. Genomics*, **12**, 17.
11. Malik,I., Kelley,C.P., Wang,E.T. and Todd,P.K. (2021) Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat. Rev. Mol. Cell Biol.*, **22**, 589–607.
12. Wang,G. and Vasquez,K.M. (2023) Dynamic alternative DNA structures in biology and disease. *Nat. Rev. Genet.*, **24**, 211–234.
13. Campuzano,V., Montermini,L., Molto,M.D., Pianese,L., Cossee,M., Cavalcanti,F., Monros,E., Rodius,F., Duclos,F., Monticelli,A., *et al.* (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science*, **271**, 1423–1427.
14. Mariappan,S.V., Catasti,P., Silks,L.A. 3rd, Bradbury,E.M. and Gupta,G. (1999) The high-resolution structure of the triplex formed by the GAA/TTC triplet repeat associated with Friedreich's ataxia. *J. Mol. Biol.*, **285**, 2035–2052.
15. Brcic,J. and Plavec,J. (2015) Solution structure of a DNA quadruplex containing ALS and FTD related GGGGCC repeat stabilized by 8-bromodeoxyguanosine substitution. *Nucleic Acids Res.*, **43**, 8590–8600.
16. Brcic,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.

17. Sissi,C., Gatto,B. and Palumbo,M. (2011) The evolving world of protein-G-quadruplex recognition: a medicinal chemist's perspective. *Biochimie*, **93**, 1219–1230.
18. Jana,J., Mohr,S., Vianney,Y.M. and Weisz,K. (2021) Structural motifs and intramolecular interactions in non-canonical G-quadruplexes. *RSC Chem. Biol.*, **2**, 338–353.
19. 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.
20. Depienne,C. and Mandel,J.L. (2021) 30 years of repeat expansion disorders: what have we learned and what are the remaining challenges?*Am. J. Hum. Genet.*, **108**, 764–785.
21. Sakamoto,N., Chastain,P.D., Parniewski,P., Ohshima,K., Pandolfo,M., Griffith,J.D. and Wells,R.D. (1999) Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia. *Mol. Cell*, **3**, 465–475.
22. Huang,T.Y., Chang,C.K., Kao,Y.F., Chin,C.H., Ni,C.W., Hsu,H.Y., Hu,N.J., Hsieh,L.C., Chou,S.H., Lee,I.R., *et al.* (2017) Parity-dependent hairpin configurations of repetitive DNA sequence promote slippage associated with DNA expansion. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 9535–9540.
23. Tateishi-Karimata,H. and Sugimoto,N. (2021) Roles of non-canonical structures of nucleic acids in cancer and neurodegenerative diseases. *Nucleic Acids Res.*, **49**, 7839–7855.
24. Subramanian,S., Mishra,R.K. and Singh,L. (2003) Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biol.*, **4**, R13.
25. Balzano,E., Pelliccia,F. and Giunta,S. (2021) Genome (in)stability at tandem repeats. *Semin. Cell Dev. Biol.*, **113**, 97–112.
26. Gharesouran,J., Hosseinzadeh,H., Ghafouri-Fard,S., Taheri,M. and Rezazadeh,M. (2021) STRs: ancient architectures of the genome beyond the sequence. *J. Mol. Neurosci.*, **71**, 2441–2455.
27. Fondon,J.W., Hammock,E.A.D., Hannan,A.J. and King,D.G. (2008) Simple sequence repeats: genetic modulators of brain function and behavior. *Trends Neurosci.*, **31**, 328–334.
28. Loureiro,J.R., Castro,A.F., Figueiredo,A.S. and Silveira,I. (2022) Molecular mechanisms in pentanucleotide repeat diseases. *Cells*, **11**, 205.
29. Seixas,A.I., Loureiro,J.R., Costa,C., Ordonez-Ugalde,A., Marcelino,H., Oliveira,C.L., Loureiro,J.L., Dhingra,A., Brandao,E., Cruz,V.T., *et al.* (2017) A pentanucleotide ATTTC repeat insertion in the non-coding region of DAB1, mapping to SCA37, causes spinocerebellar ataxia. *Am. J. Hum. Genet.*, **101**, 87–103.
30. Ishiura,H., Doi,K., Mitsui,J., Yoshimura,J., Matsukawa,M.K., Fujiyama,A., Toyoshima,Y., Kakita,A., Takahashi,H., Suzuki,Y., *et al.* (2018) Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. *Nat. Genet.*, **50**, 581–590.
31. Corbett,M.A., Kroes,T., Veneziano,L., Bennett,M.F., Florian,R., Schneider,A.L., Coppola,A., Licchetta,L., Franceschetti,S., Suppa,A., *et al.* (2019) Intronic ATTTC repeat expansions in STARD7 in familial adult myoclonic epilepsy linked to chromosome 2. *Nat. Commun.*, **10**, 4920.
32. Yeetong,P., Pongpanich,M., Srichomthong,C., Assawapitaksakul,A., Shotelersuk,V., Tantirukdham,N., Chunharas,C., Suphapeetiporn,K. and Shotelersuk,V. (2019) TTTCA repeat insertions in an intron of YEATS2 in benign adult familial myoclonic epilepsy type 4. *Brain*, **142**, 3360–3366.
33. Florian,R.T., Kraft,F., Leitao,E., Kaya,S., Klebe,S., Magnin,E., van Rootselaar,A.F., Buratti,J., Kuhnel,T., Schroder,C., *et al.* (2019) Unstable TTTTA/TTTCA expansions in MARCH6 are associated with familial Adult myoclonic Epilepsy type 3. *Nat. Commun.*, **10**, 4919.
34. O'Leary,N.A., Wright,M.W., Brister,J.R., Ciufo,S., McVeigh,D.H.R., Rajput,B., Robbertse,B., Smith-White,B., Ako-Adjei,D., Astashyn,A., *et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, **44**, D733–D745.
35. Potaman,V.N., Bissler,J.J., Hashem,V.I., Oussatcheva,E.A., Lu,L., Shlyakhtenko,L.S., Lyubchenko,Y.L., Matsuura,T., Ashizawa,T., Leffak,M., *et al.* (2003) Unpaired structures in SCA10 (ATTCT)n.(AGAAT)n repeats. *J. Mol. Biol.*, **326**, 1095–1111.
36. Cherng,N., Shishkin,A.A., Schlager,L.I., Tuck,R.H., Sloan,L., Matera,R., Sarkar,P.S., Ashizawa,T., Freudenreich,C.H. and Mirkin,S.M. (2011) Expansions, contractions, and fragility of the spinocerebellar ataxia type 10 pentanucleotide repeat in yeast. *Proc. Natl. Acad. Sci*, **108**, 2843–2848.
37. Guo,P. and Lam,S.L. (2020) Minidumbbell structures formed by ATTCT pentanucleotide repeats in spinocerebellar ataxia type 10. *Nucleic Acids Res.*, **48**, 7557–7568.
38. Li,J., Wan,L., Wang,Y., Chen,Y., Lee,H.K., Lam,S.L. and Guo,P. (2023) Solution nuclear magnetic resonance structures of ATTTT and ATTTC pentanucleotide repeats associated with SCA37 and FAMEs. *ACS Chem. Neurosci.*, **14**, 289–299.
39. Gajarsky,M., Zivkovic,M.L., Stadlbauer,P., Pagano,B., Fiala,R., Amato,J., Tomaska,L., Sponer,J., Plavec,J. and Trantirek,L. (2017) Structure of a stable G-hairpin. *J. Am. Chem. Soc.*, **139**, 3591–3594.
40. Zivkovic,M.L., Gajarsky,M., Bekova,K., Stadlbauer,P., Vicherek,L., Petrova,M., Fiala,R., Rosenberg,I., Sponer,J., Plavec,J., *et al.* (2021) Insight into formation propensity of pseudocircular DNA G-hairpins. *Nucleic Acids Res.*, **49**, 2317–2332.
41. Ghezzo,M., Trajkovski,M., Plavec,J. and Sissi,C. (2023) A screening protocol for exploring loop length requirements for the formation of a three cytosine-cytosine(+) base-paired i-motif. *Angew. Chem. Int. Ed.*, **62**, e202309327.
42. Case,D.A., Belfon,K., Ben-Shalom,I.Y., Brozell,S.R., Cerruti,D.S., Cheatham,T.E. 3rd, Cruzeiro,V.W.D., Darden,T.A., Duke,R.E., Giambasu,G., *et al.* (2020) *AMBER 2020*. University of California, San Francisco.
43. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
44. Kypr,J., Kejnovská,I., Renciuk,D. and Vorlícková,M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.*, **37**, 1713–1725.
45. Vorlickova,M., Kejnovska,I., Bednarova,K., Renciuk,D. and Kypr,J. (2012) Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes. *Chirality*, **24**, 691–698.
46. Iaccarino,N., Cheng,M.P., Qiu,D.H., Pagano,B., Amato,J., Di Porzio,A., Zhou,J., Randazzo,A. and Mergny,J.L. (2021) Effects of sequence and base composition on the CD and TDS profiles of i-DNA. *Angew. Chem. Int. Ed.*, **60**, 10295–10303.
47. Gallego,J., Chou,S.H. and Reid,B.R. (1997) Centromeric pyrimidine strands fold into an intercalated motif by forming a double hairpin with a novel T:G:G:T tetrad: solution structure of the d(TCCCGTTTCCA) dimer. *J. Mol. Biol.*, **273**, 840–856.
48. Escaja,N., Viladoms,J., Garavis,M., Villasante,A., Pedroso,E. and Gonzalez,C. (2012) A minimal i-motif stabilized by minor groove G:T:G:T tetrads. *Nucleic Acids Res.*, **40**, 11737–11747.
49. Mir,B., Serrano,I., Buitrago,D., Orozco,M., Escaja,N. and Gonzalez,C. (2017) Prevalent sequences in the Human genome can form mini i-motif structures at physiological pH. *J. Am. Chem. Soc.*, **139**, 13985–13988.
50. Chen,Y.W., Jhan,C.R., Neidle,S. and Hou,M.H. (2014) Structural basis for the identification of an i-motif tetraplex core with a parallel-duplex junction as a Structural motif in CCG triplet repeats. *Angew. Chem. Int. Ed.*, **53**, 10682–10686.
51. Serrano-Chacon,I., Mir,B., Escaja,N. and Gonzalez,C. (2021) Structure of i-motif/duplex junctions at neutral pH. *J. Am. Chem. Soc.*, **143**, 12919–12923.
52. Lieblein,A.L., Furtig,B. and Schwalbe,H. (2013) Optimizing the kinetics and thermodynamics of DNA i-motif folding. *ChemBioChem*, **14**, 1226–1230.
53. Garavis,M., Escaja,N., Gabelica,V., Villasante,A. and Gonzalez,C. (2015) Centromeric alpha-satellite DNA adopts dimeric i-motif

structures capped by AT Hoogsteen base pairs. *Chem. Eur. J.*, **21**, 9816–9824.
54. El-Khoury,R., Macaluso,V., Hennecker,C., Mittermaier,A.K., Orozco,M., Gonzalez,C., Garavis,M. and Damha,M.J. (2023) i-motif folding intermediates with zero-nucleotide loops are trapped by 2'-fluoroarabinocytidine via F…H and O…H hydrogen bonds. *Commun. Chem.*, **6**, 31.
55. Canalia,M. and Leroy,J.L. (2005) Structure, internal motions and association-dissociation kinetics of the i-motif dimer of d(5mCCTCACTCC). *Nucleic Acids Res.*, **33**, 5471–5481.
56. Canalia,M. and Leroy,J.L. (2009) [5mCCTCTCTCC]4: an i-motif tetramer with intercalated T*T pairs. *J. Am. Chem. Soc.*, **131**, 12870–12871.
57. Leroy,J.L. and Gueron,M. (1995) Solution structures of the i-motif tetramers of d(TCC), d(5methylCCT) and d(T5methylCC): novel NOE connections between amino protons and sugar protons. *Structure*, **3**, 101–120.
58. Abou Assi,H., Garavis,M., Gonzalez,C. and Damha,M.J. (2018) i-motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.*, **46**, 8038–8056.
59. Streisinger,G., Okada,Y., Emrich,J., Newton,J., Tsugita,A., Terzaghi,E. and Inouye,M. (1966) Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. *Cold Spring Harb. Symp. Quant. Biol.*, **31**, 77–84.
60. Kunkel,T.A. (1993) Nucleotide repeats. Slippery DNA and diseases. *Nature*, **365**, 207–208.
61. Ellegren,H. (2004) Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.*, **5**, 435–445.
62. Branzei,D. and Foiani,M. (2010) Maintaining genome stability at the replication fork. *Nat. Rev. Mol. Cell Biol.*, **11**, 208–219.
63. Neelsen,K.J. and Lopes,M. (2015) Replication fork reversal in eukaryotes: from dead end to dynamic response. *Nat. Rev. Mol. Cell Biol.*, **16**, 207–220.
64. Saxena,S., Joshi,S., Shankaraswamy,J., Tyagi,S. and Kukreti,S. (2017) Magnesium and molecular crowding of the cosolutes stabilize the i-motif structure at physiological pH. *Biopolymers*, **107**, e23018.
65. Day,H.A., Wright,E.P., MacDonald,C.J., Gates,A.J. and Waller,Z.A. (2015) Reversible DNA i-motif to hairpin switching induced by copper(ii) cations. *Chem. Commun.*, **51**, 14099–14102.
66. Zuo,S., Yi,Y., Wang,C., Li,X., Zhou,M., Peng,Q., Zhou,J., Yang,Y. and He,Q. (2021) Extrachromosomal circular DNA (eccDNA): from chaos to function. *Front. Cell Dev. Biol.*, **9**, 792555.
67. Cohen,Z. and Lavi,S. (2009) Replication independent formation of extrachromosomal circular DNA in mammalian cell-free system. *PLoS One*, **4**, e6126.
68. Khristich,A.N. and Mirkin,S.M. (2020) On the wrong DNA track: molecular mechanisms of repeat-mediated genome instability. *J. Biol. Chem.*, **295**, 4134–4170.

Received: October 26, 2023. **Revised:** January 10, 2024. **Editorial Decision:** January 14, 2024. **Accepted:** January 16, 2024