Nucleic Acids Research, 2024, **52**, 3390–3405
https://doi.org/10.1093/nar/gkae113
Advance access publication date: 21 February 2024
Structural Biology

Structure of native four-repeat satellite III sequence with non-canonical base interactions

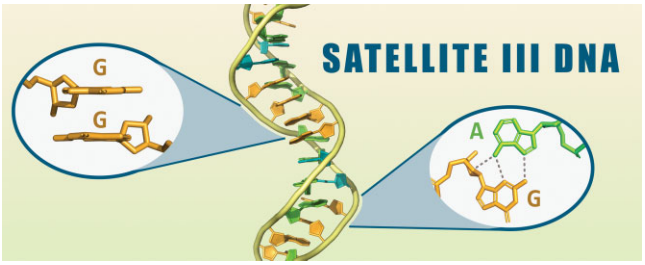
Erin Chen[1], Marko Trajkovski[2], Hyun Kyung Lee[1], Samantha Nyovanie[1], Kailey N. Martin[1], William L. Dean[3], Mamta Tahiliani[4], Janez Plavec[2] and Liliya A. Yatsunyk[1,*]

[1]Department of Chemistry and Biochemistry, Swarthmore College, 500 College Ave, Swarthmore, PA 19081, USA
[2]Slovenian NMR Centre, National Institute of Chemistry, Hajdrihova 19, 1000 Ljubljana, Slovenia
[3]Structural Biology Program JG Brown Cancer Center, University of Louisville, Louisville, KY 40202, USA
[4]Department of Biology, New York University, New York, NY 10003, USA
[*]To whom correspondence should be addressed. Tel: 484 477 8457; Fax: 610 328 8558; Email: lyatsun1@swathmore.edu

Abstract

Tandem-repetitive DNA (where two or more DNA bases are repeated numerous times) can adopt non-canonical secondary structures. Many of these structures are implicated in important biological processes. Human Satellite III (HSat3) is enriched for tandem repeats of the sequence ATGGA and is located in pericentromeric heterochromatin in many human chromosomes. Here, we investigate the secondary structure of the four-repeat HSat3 sequence 5′-ATGGA ATGGA ATGGA ATGGA-3′ using X-ray crystallography, NMR, and biophysical methods. Circular dichroism spectroscopy, thermal stability, native PAGE, and analytical ultracentrifugation indicate that this sequence folds into a monomolecular hairpin with non-canonical base pairing and B-DNA characteristics at concentrations below 0.9 mM. NMR studies at 0.05–0.5 mM indicate that the hairpin is likely folded-over into a compact structure with high dynamics. Crystallographic studies at 2.5 mM reveal an antiparallel self-complementary duplex with the same base pairing as in the hairpin, extended into an infinite polymer. The non-canonical base pairing includes a G–G intercalation sandwiched by sheared A–G base pairs, leading to a cross-strand four guanine stack, so called guanine zipper. The guanine zippers are spaced throughout the structure by A–T/T–A base pairs. Our findings lend further insight into recurring structural motifs associated with the HSat3 and their potential biological functions.

Graphical abstract



Introduction

Satellite DNA are large non-coding multi-megabase sized arrays enriched in tandem repeats found in structural regions of chromosomes such as telomeres, pericentromeres, and centromeres. Repetitive sequences form a significant portion of eukaryotic genomes and range from less than 1% to over 50% of the genome, depending on the species (1). Given their lack of protein coding function, satellite DNA were historically labeled as 'junk DNA,' but are now predicted to play important physiological roles due to their propensity to adopt non-canonical secondary structures. Biological functions of satellite DNA encompass protecting and stabilizing chromatin of telomeres and centromeres (2,3), regulating kinetochore assembly and maintenance (4) and organizing heterochromatin and modulating gene expression (5). Non-canonical DNA structures encompass a highly diverse group of structures, including hairpins, G-quadruplexes, i-motifs and triplex DNA. Non-canonical DNA structures are gaining wide interest due to their biological importance and emerging roles in diverse pathological conditions (6).

Satellite III (HSat3) DNA is the most abundant and highly conserved satellite DNA in humans (7,8). It is found in pericentromeric heterochromatic blocks on numerous chromosomes (9,10). HSat3 is derived from the simple pentameric GAATG repeat and contains tandem repeats of this pentamer $(GAATG)_n$ as well as diverged sequences. Due to repetitive nature of HSat3, it can be also represented by a different register such as ATGGA, GGAAT, etc. The evolutionary conservation of GAATG repeat suggests its important biological role including its link to human diseases. For example, 2.5–3.8 kb

Received: July 5, 2023. **Revised:** January 31, 2024. **Editorial Decision:** February 2, 2024. **Accepted:** February 6, 2024

insertions of the $(GAATG)_n$ motif were discovered in genomes of patients with spinocerebellar ataxia type 31, an adult-onset autosomal-dominant neurodegenerative disorder (11). HSat3 DNA and RNA may also be important for kinetochore assembly playing vital role in mitosis and meiosis (10,12,13). A significant link is reported between the transcription of HSat3 RNA and the heat shock response (14) as well as general stress conditions, aging, and pathology (15). HSat3 attracted our attention because its purine-rich strand alone displays unusually high thermodynamic stability equivalent to that of duplex DNA (8).

Extensive biophysical and structural studies of the purine-rich strand of HSat3 indicate that this DNA can form self-complementary duplexes (16–19) as well as a variety of hairpins (20–22), whose specific conformation depends on DNA concentration, register (frame shift), number of pentameric repeats, and co-solutes. All reported HSat3 structures contain stabilizing G–C base pair(s) installed at termini to prevent fraying of DNA ends (16–20,22). The presence of these G–C base pairs may influence the resulting structure. In additions, all structures contain only one or two pentameric repeats (e.g. $(GAATG)_1$ or $(TGGAA)_2$); there is only one hairpin structure with three pentameric repeats, but this sequence also includes G-to-C mutation in the middle repeat and G–C stabilizing base pair at termini (22). Finally, the structures of HSat3 with four and six pentameric repeats without any mutations have been determined by NMR (21). These structures have been criticized (17), and have not been deposited into any public data bank.

Each published HSat3 structure contains a sheared G–A base pair (17–19,22), one of the most well studied non-canonical structural elements. G–A base pair was observed for the first time via NMR in 1983 and via X-ray crystallography in 1986 (23,24). In 1991, the Wilson group characterized four different G-A base pairs, with the most frequent being the G(*anti*)-A(*anti*) head-to-head conformation (25). The conformation most relevant to our discussion is the sheared G(*anti*)-A(*anti*) pair (Figure 1). The specific conformation depends on the sequence and environmental conditions. The diversity of structural configuration for the G-A base pair is largely a result of the ability of guanine and adenine bases to form hydrogen bonds with the Hoogsteen and Watson–Crick–Franklin faces as well as the sugar edge of the bases. Interestingly, while one G–A base pair destabilizes B-DNA, the presence of two G-A base pairs causes cross-strand purine-purine base stacking that does not perturb B-DNA structure or stability (25). HSat3 structures with multiple pentameric repeats ($n \geq 2$) contain another important structural element, a guanine zipper – a pair of stacked guanines sandwiched by sheared G-A base pairs leading to a cross-strand four guanine stack (17,18,20,22). One face of each guanine is free and exposed to solvent. Thus, a stack of four guanines can provide a 'sticky patch' for interactions with ligand, proteins or complementary DNA (forming a DNA triplex).

Due to lack of biologically relevant mutation free HSat3 structures, we wanted to determine the structure of four perfect pentameric repeats of the 5′-ATGGA ATGGA ATGGA ATGGA-3′ sequence, thereafter referred to as 'S3' or $d(ATGGA)_4$. We employed X-ray crystallography, NMR, and biophysical studies including circular dichroism spectroscopy, thermal difference spectra, thermal stability, analytical ultracentrifugation, and gel electrophoresis. We succeeded in obtaining a 1.57 Å resolution crystal structure of S3 which shows an infinite polymer of a self-complementary duplex. At the same time, our NMR studies at 0.05–0.5 mM of S3 indicate interconversion between two monomolecular compact hairpin structures. We confirmed the monomolecular nature of S3 using biophysical studies at concentration of < 0.9 mM. X-ray and NMR structures are maintained by extensive non-canonical base pairing where each pentameric repeat contains one guanine zipper separated by two canonical T-A/A-T base pairs.

Hairpin and duplex secondary structures formed by the $(GAATG)_n$ repeat of HSat3 and the dynamic interconversion between them have been suggested to cause or contribute to disease-associated repeat expansion (20). Thus, understanding of the structure of the purine-rich strand of HSat3 facilitated by our work will bring us closer to elucidating the biological functions of these sequences in the human centromere and will also provide insight which could lead to the treatment of associated diseases.

Materials and methods

Lyophilized S3 oligonucleotide with the sequence 5′-ATGGA ATGGA ATGGA ATGGA-3′ was purchased from Integrated DNA Technologies (IDT; Coralville, IA) with standard desalting purification. DNA was hydrated in double-distilled water to 1–3 mM and stored at -80°C. This DNA was used for biophysical experiments and crystallization. For NMR experiments, S3 was synthesized on H-8 synthesizer (K&A LaborGeräte) using standard phosphoramidite chemistry. Extinction coefficient of 217.4 mM^{-1} cm^{-1} was obtained using IDT's OligoAnalyzer 3.1. DNA concentration was determined from UV-vis spectra collected at 95°C. The buffer for all studies, but NMR, consisted of 10 mM Tris-acetate pH 8.3, 50 mM potassium acetate, and 5 mM magnesium acetate (TSB buffer). DNA was prepared in TSB buffer at the desired concentration and annealed by heating at 95°C for 5 min followed by slow cooling to room temperature over a period of 4–5 h. These samples were stored overnight at 4°C before further use. Proper folding was verified using circular dichroism (CD) spectroscopy for all samples. Biophysical data were processed using Origin 9.1.

Circular dichroism (CD) spectroscopy

CD spectra were collected on samples ∼4 μM in 1-cm quartz cuvettes at 20°C on an AVIV-435 instrument with 2 nm bandwidth, 1 s averaging time, 1 nm step size over a 220–330 nm window. Five scans were recorded and averaged. CD concentration studies were conducted in the concentration range of 3–850 μM DNA in 10, 2, 1 and 0.11 mm pathlength cuvettes. Note, in our concentration studies we were limited by 850 μM (using 0.11 mm pathlength) as samples at higher concentration absorb too much signal to obtain any interpretable CD data. We chose the appropriate cuvette such that the UV-vis absorbance of each sample was, if possible, close to 0.8. Data were processed as previously described (26).

Thermal stability monitored via CD

The stability of DNA was determined in CD melting experiments, monitoring 267 nm, the wavelength of maximum CD signal while varying temperature from 20 to 95°C. The temperature was changed in a 1°C step with a 1°C/min temperature rate, 15 s averaging time, and 5 s equilibration time.

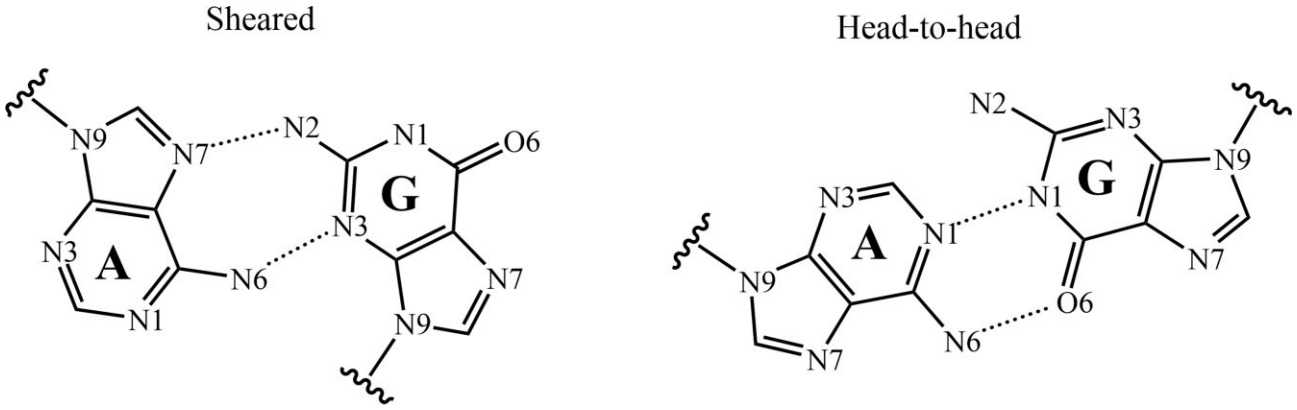


Figure 1. Sheared and head-to-head G-A base pairs with nitrogen atoms in adenine and guanine numbered.

Reverse scans were also collected to determine the hysteresis, (i.e. difference in melting temperatures extracted from melting and cooling curves). The data were fit using a two-state model which assumes that enthalpy of unfolding, ΔH, is temperature-independent. This fitting yielded melting temperature, T_m, and ΔH values and associated fitting errors (27). Reported results represent the average of 2–3 trials.

Thermal stability monitored via UV–vis

The stability of DNA was also determined using UV–vis melting experiments, monitoring absorbance at 260 and 335 nm using 1 cm pathlength quartz cuvettes. Absorbance at 260 nm reflected the DNA folding state and absorbance at 335 nm served as the baseline for the instrument's performance. Temperature was increased from 4 to 95°C. Reverse scans were collected to determine hysteresis. The temperature was changed in a 1°C step with a 0.5°C/min temperature rate, 2 s averaging time and 2 nm spectral bandwidth. The data was baseline corrected for absorbance at 335 nm and processed as described above. Reported results represent the average of 2–3 trials.

Thermal difference spectra (TDS)

UV–vis spectra were collected at 4 and 90°C after 10 min of equilibration. In principle, the low and high temperature limits are defined by the temperatures at which the DNA is (mostly) folded and unfolded, respectively. TDS signatures were obtained by subtracting high temperature spectrum from the low temperature spectrum.

Analytical ultracentrifugation (AUC)

AUC was carried out to determine the molecularity and purity of S3 samples. The samples were prepared at 5, 50, 100, 200 and 720 μM in TSB buffer from three different stocks. Sedimentation velocity measurements were carried out in a Beckman Coulter ProteomeLab XL-A analytical ultracentrifuge (Beckman Coulter Inc., Brea, CA) at 20.0°C and 40 000 rpm in standard 2 sector cells. Sample volumes were 0.45 ml. Two hundred scans were collected over a 10-h centrifugation period at either 260, 298 or 305 nm. An off maximum wavelength was used due to high concentration of the samples. Data were analyzed using Sedfit in the continuous *c*(*s*) mode (28). The sedimentation coefficient is denoted as *s*. Buffer density was determined on a Mettler/Paar Calculating Density Meter DMA 55A at 20.0°C and buffer viscosity was measured on an Anton Paar Automated Microviscometer AMVn. For the calculation of molecular weight, 0.55 ml/g was used for the partial specific volume.

Native polyacrylamide gel electrophoresis (PAGE)

Native polyacrylamide gels were prepared at 15% in Tris–borate–EDTA (TBE) buffer supplemented with 10 mM KCl. TBE with 10 mM KCl was used as a running buffer. The gel was pre-migrated for 30 min at 150 V at room temperature. Annealed DNA samples in TSB buffer at 0.3 μg/μl were weighted with 9% sucrose, loaded onto the gel, and run for 2.8 h at 150 V. Oligothymidylate markers 5′ dTn (where $n = 15$, 24, 30 and 57) were used as migration standards. DNA bands were visualized using StainsAll and the gel was photographed with a smartphone camera.

X-ray crystallography

S3 samples were annealed and equilibrated in TSB buffer at 2.5 mM. Crystallization was achieved at room temperature using the hanging-drop vapor diffusion method. Trays were set using the Mosquito Liquid Handling Robot system (TTP Labtech) at 0.1 μl sample mixed with 0.1 μl crystallization condition. The latter consisted of 0.1 M magnesium acetate tetrahydrate, 0.05 M MES monohydrate pH 5.6, and 20% v/v 2-methyl-2,4-pentanediol (MPD). Crystallization condition was optimized using pH range 5.6–6.5 and MPD range 20–35%. Crystals were most consistently produced in the condition that contained 30% MPD and pH of 5.6. Large cubic crystals grew within 1–2 weeks and were cryoprotected either with 15% glycerol or with 12.5% glycerol/12.5% ethylene glycol added to base conditions and flash frozen in liquid nitrogen.

Data sets were collected at the Advanced Photon Source 24 ID-C synchrotron facility to a maximum resolution of 1.57 Å. Raw diffraction data were processed using XDS (29), Pointless (30), and Aimless (31). The structure was solved via molecular replacement (MR) using Phenix.Phaser (version 1.13_2998) (32). The model for MR was based on an antiparallel self-complementary duplex 5′-GCGAATGAGC-3′ (PDB ID: 175D) which was trimmed to contain only the AATG core (underligned). An initial MR solution contained only four DNA chains (labeled A-D); two more DNA chains, E and F, were added during model building although their

Table 1. Data collection and refinement statistics

Resolution range	32.54–1.57
(Highest resolution shell)	1.63–1.57
Space group	$P2_12_12_1$
Unit cell dimensions	
a, *b*, *c* (Å)	39.52, 40.22, 55.38
α, β, γ (°)	90, 90, 90
Unique reflections	12 855 (643)
Redundancy	5.9 (5.1)
Completeness (%)	99.7 (99.0)
I/sigma	15.1 (2.5)
R-merge (%)	0.042 (0.398)
R_{work}/R_{free} (%)	0.2206/0.2551
Number of atoms	715
DNA (no hydrogens)	636
Ligand	4
Solvent	74
Magnesium	1
Copies in asymmetric unit	3*
Overall *B*-factor (Å^2)	48.43
RMS deviations	
Bond length (Å)	0.017
Bond angles (°)	1.668
PDB ID	7JLH

* ASU contains three antiparallel duplexes that are only five nucleotides long, for six total DNA chains.

density remained weak. The final model was produced via manual model building cycles in Coot 0.8.8 (33) followed by Phenix.Refine (34). The resultant asymmetric unit (ASU) contains six DNA chains (A-F), one acetate molecule, one hexaaquamagnesium ion, $[Mg(H_2O)_6]^{2+}$ and 74 waters. The structure was deposited into the PDB database and assigned PDB ID: 7JLH. Data collection and refinement statistics are presented in Table 1.

Determination and refinement of ASU content

Analysis of the unit cell and sequence length yielded Matthew's coefficient of ∼3.5 for one copy of S3 (20 nt) and ∼1.6 for two copies of S3 (40 nt) in the ASU. Neither value agrees with the most commonly observed Matthew's coefficients of ∼2–2.5 (35). This discrepancy was ultimately explained by the presence of 30 nt in the ASU in the form of three 5-nt duplex fragments, which is equivalent to 1.5 copies of S3 in the ASU.

The repetitive nature of DNA led to the possibility of building these five-nt fragments in any phase-shift (ATGGA, AATGG, GAATG, GGAAT, TGGAA). We began structure building using ATGGA repeat in accordance with the oligonucleotide sequences used to grow crystals. In this case, the two A–T base pairs were constructed such that one base was built and another was symmetry-generated. After a round of refinement, the nucleotides in the ASU that were too close to the symmetry-generated nucleotides were pushed away from each other. This rearrangement did not fit well into the electron density leading to increased R values. Therefore, we switched to the AATGG repeat. In this case, only one (intercalated) guanine (here G5) interacted with the symmetry-generated guanine (G5', where 'indicates a symmetry generated molecule). While the repulsion between ASU and symmetry-generated nucleotides still existed (and led to a shortening C3'–O3' bond for G5) it is still within acceptable limits.

Analysis of crystallographic data

DNA backbone torsion angles, local base-pair and local base-pair step parameters were calculated using 3DNA (36). Groove widths were calculated manually via an average of C3'–C3' sugar atom distances. RMSD was calculated in PyMol. B-factors were calculated both manually from the PDB and with Baverage in CCP4i. Metal geometry was checked using Metal Binding Site Validation Server (https://cmm.minorlab.org/) (37). G5–G5' base stacking distances were measured with an in-lab MATLAB script that calculated the distance between centroids of guanine bases (using 11 atoms of the guanine base).

NMR spectroscopy

NMR experiments were performed on S3 samples prepared in 90% H_2O/10% ^2H_2O (v/v) or 100% ^2H_2O, in 25 mM sodium-phosphate buffer at pH 6.5 and S3 concentration in the range 0.05–0.5 mM per strand. The experiments were conducted on Bruker AVANCE NEO 600 MHz spectrometers equipped with TCI cryoprobe. NMR spectra were processed and analyzed with Topspin 4.1.3 (Bruker, Germany) and NMRFAM-SPARKY (38) software. Assignment of the ^{1}H NMR resonances relied on ^{15}N- and ^{13}C-edited HSQC experiments acquired on partially (5%) site-specifically ^{15}N- and ^{13}C-isotope labelled oligonucleotides. 2D NOESY spectra were acquired at mixing times (τm) between 80 and 300 ms. Excitation sculpting water suppression method was used in 1D and 2D NMR experiments.

Restrained simulated annealing calculations

The calculations of the high-resolution structural models were based on restrained 1000 ps long simulated annealing (SA) protocol, using Amber 20 software, parmbsc1 force field, Born implicit solvent model, and random starting velocities. Details regarding the restraints used in the calculations are given in Supplementary Table S1. The force constants in SA calculations were 10 kcal mol^{-1} Å^{-2} for distance and hydrogen-bonds restraints, 20 kcal mol^{-1} Å^{-2} for chirality restraints and planarity restraints of A–T and G–A base pairs, while 100 kcal mol^{-1} Å^{-2} for glycosidic torsion angles (χ) restraints. The force constants for restraints were scaled from the initial value of 0.1 to the final value of 1.0 in the first 50 ps and held constant until the end of the calculation. SA protocol included heating from 0 to 1000 K over the first 50 ps, followed by 150 ps equilibration at 1000 K, cooling over 700 ps from 1000 to 0 K and equilibration for 100 ps at 0 K. The starting structure was generated with the use of 3DNA software and leap model of AMBER 20. Twenty structures of each, the major and minor forms, were calculated. Two representative structures were selected for each, the major and minor form, that had the lowest energy and the least violations with respect to the structural restraints used in the calculations.

Result

Characterization of S3 via X-ray crystallography

We successfully obtained diffraction-quality crystals (Figures 2A and S1) and solved the structure of S3 in space group $P2_12_12_1$ to a resolution of 1.57 Å. We were only able to grow crystals from 2.5 mM solution of S3. Concentration of S3 below 1.0 mM did not produce any crystals, indicating that high DNA concentration is critical for the struc-

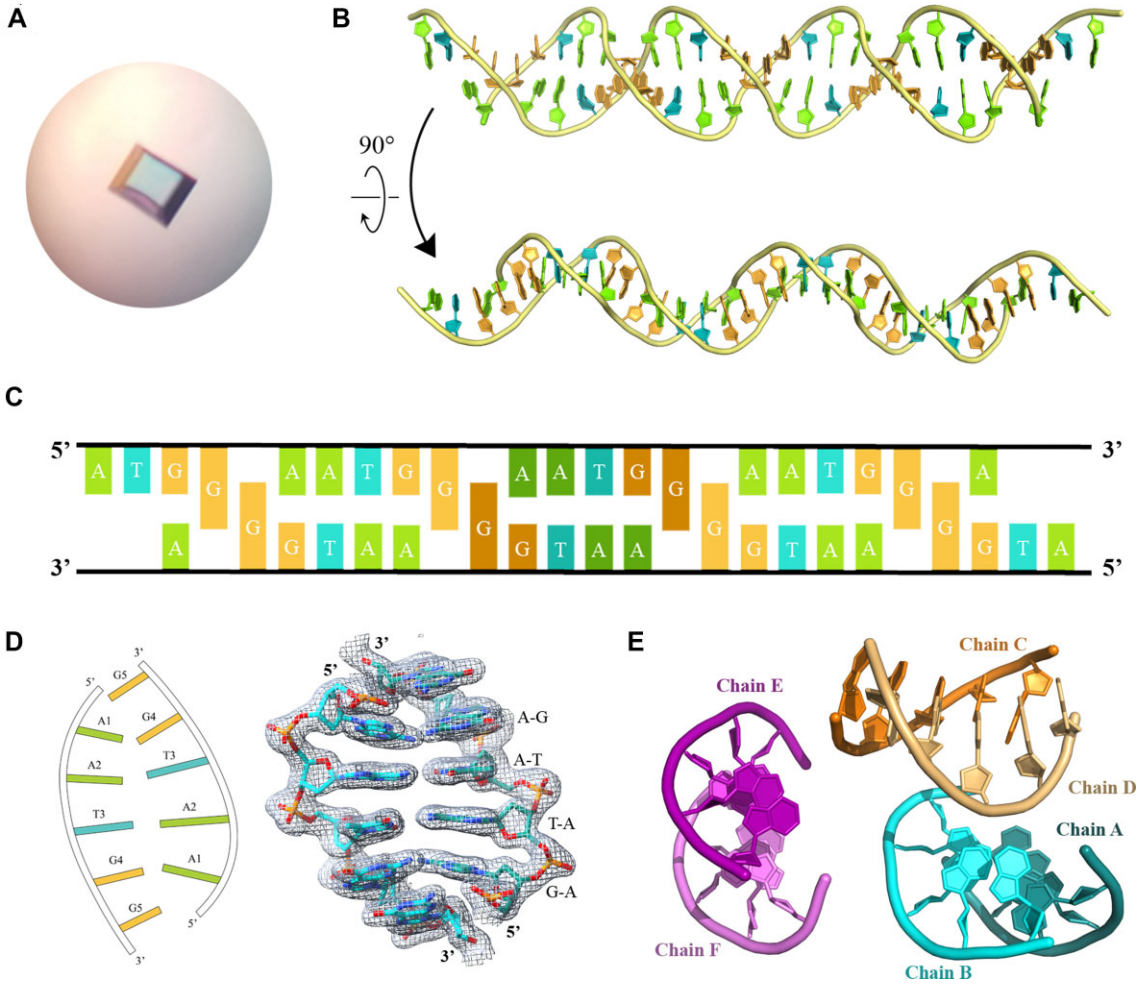


Figure 2. The structure of S3. (**A**) A representative S3 crystal. (**B**) Cartoon representation of four merged AATGG repeats, with major (top) and minor (bottom) grooves facing the reader. Four guanine stacks in orange are clearly visible in the bottom image. (**C**) Schematics of 5′-ATGGA ATGGA ATGGA ATGGA-3′ DNA duplex. Adenines are green, thymines are blue, and guanines are orange. Darker nucleotides represent a segment from the asymmetric unit. (**D**) Schematics (colored by nt) and sticks model (colored by atom) of a 5-nt double stranded-fragment, 5′-AATGG-3′. Electron density is at $I/\sigma = 1.0$. (**E**) ASU content composed of three double-stranded 5′-AATGG-3′ segments, labeled A–B, C–D and E–F.

ture. It rapidly became clear based on the unit cell dimensions and Matthew's coefficient that S3 forms a bimolecular antiparallel self-complementary duplex in the crystalline state (Figure 2B, C).

Overall architecture, base pairing and stacking patterns in S3 structure

Due to the repetitive nature of S3 and the high symmetry of the structure, the 20 nt sequence 5′-ATGGA ATGGA ATGGA ATGGA-3′ is best represented in the asymmetric unit (ASU) by a pentameric double stranded-fragment, 5′-AATGG-3′, Figure 2D. This fragment is phase-shifted as compared to the DNA sequence used in the crystallization studies. For clarity, we will number nucleotides according to the ASU model, A1–A2–T3–G4–G5, and not the actual sequence. The ASU contains three such double-stranded fragments formed by six strands in total, labeled A–B, C–D and E–F, Figure 2E. There are only minor differences between these fragments with RMSD of 0.618 Å for A–B versus C–D; 0.775 Å for A–B versus E–F; and 0.887 Å for C–D versus E–F. Due to great similarity of the double-stranded fragments, we will mostly discuss the A–B duplex as a representative high-quality example.

The structure of S3 is maintained by a classical Watson–Crick–Franklin A–T base pairing, non-canonical sheared G–A base pairing (Figure 2D), and a G–G intercalation. The order of pairing from the 5′ end is: A1–G4, A2–T3, T3–A2 and G4–A1. Every T in the structure is involved in a Watson–Crick–Franklin base pair with A. The last guanine (G5) is involved in stacking interactions with G5' (' indicates a symmetry generated nucleotide). The G5–G5' stack is intercalated between two G–A base pairs, an arrangement known as a guanine zipper. Moving along one chain, there is a significant base overlap for A1–A2, A2–T3 and T3–G4, poor base overlap for G4–G5, and a nearly perfect base overlap for G5–G5', Figure 3.

Sheared G4–A1 base pairs are stabilized by hydrogen bonds between N6 and N7 of adenine and N3 and N2 of guanine, respectively (Figure 1). The two bases are highly non-planar with the largest contribution coming from a buckle deformation of 32 ± 2°. Both bases participate in several other important stabilizing interactions depicted in detail in Supplementary Figure S2: (i) N6 of A1 hydrogen bonds with O2 of T3 and O4' of the G4 sugar; (ii) O4' of the G5 sugar interacts with six-membered ring of A1 via a cation–π interaction and (iii) the phosphate of A1 hydrogen bonds with N2 of G5 (distances are shown in Supplementary Table S2). The

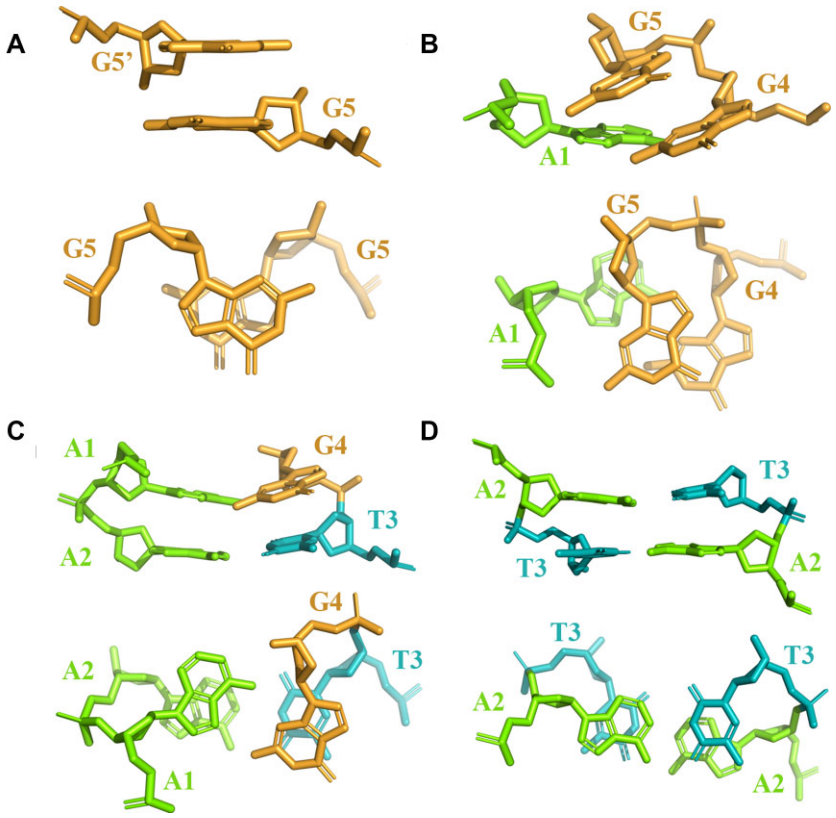


Figure 3. Stacking and base pairing patterns in the A–B duplex shown as a side and top view. (**A**) G5-G5′ stacking. (**B**) G5 stacking with a G4–A1 base pair. (**C**) G4–A1 base pair stacking with a T3–A2 base pair. (**D**) T3–A2 base pair stacking with an A2–T3 base pair. The color scheme is the same as in Figure 2.

adenine rich sequence, 5′-GCGAAAAGCG-3′, which has a central A–A–A–A intercalation in place of the G–G intercalation, has been shown to form a similar intercalated structure (39). Due to the lack of N2 in adenine, the (iii) hydrogen bond cannot form, but this does not prevent formation of the duplex. Thus, we speculate that the hydrogen bond between N2 of G5 and the phosphate of A1 contributes to duplex stability, but is not required for its formation. In the cation-π interaction (the (ii) hydrogen bond), the distance from the G5 sugar to the six-member ring of A1 is 3.0–3.10 Å, consistent with 3.02 ± 0.02 Å obtained from the analysis of ten Z-DNA dinucleotides, which display similar bonding and are deposited into the Rutgers Nucleic Acid Database (40). Interestingly, the packing of the O4' sugar against a nucleobase is a major stabilizing force in the CpG step of Z-DNA that also displays very little base stacking (41).

All nucleotides adopt *anti* glycosidic conformations. Torsion angle analysis of S3 backbone indicates five outliers (Supplementary figure S3). All but one of these outliers are found in duplex E–F. It is likely that the E–F outliers result from the weaker electron density in this region (see Materials and methods). Three outliers (including the one not from duplex E–F) are associated with the A2 nucleotide, likely because A2 is a terminal residue 25% of the time and a linking residue 75% of the time (more on this below). All sugars except for those in G5 adopt typical C2'-endo conformation. Sugars for G5 from Chains A-C and E–F adopt the less common C3'-endo conformation described in other reported structures of the human centromeric repeat (17,18,20,22). The G5 sugar in Chain D adopts C2'-endo conformation. It was suggested that C3'-endo conformation is required to accommodate stretching of the DNA backbone due to G-G intercalation (17,18). In an ongoing study from our lab, we crystallized a G-quadruplex DNA in complex with an intercalated ligand (data not published). Due to this intercalation, the G-quadruplex displays a stretched backbone with the distance between the bases of two consecutive nucleotides of 6.7 Å. However, all eight guanines π-stacked with the ligand display a C2'-endo conformation suggesting that perhaps the C3'-endo conformation is not required to accommodate backbone stretching.

G–G intercalation and a four guanine stack

Initially, we expected to detect the presence of a non-canonical G–G base pair as reported in the NMR structure of $(AATGG)_4$ (21). This repeat is nearly identical to the DNA sequence used in our study, differing only in a frameshift of one nucleotide. However, instead we observed a G–G intercalation seen in several two and three repeat structures of HSat3 derived oligonucleotides (17,18,20,22). The G5 and G5' bases π–π stack such that they are nearly parallel, with an average distance of 3.5 ± 0.2 Å, which is consistent with known distances for efficient π–π stacking (42,43). The aromatic rings of G5 and G5' overlap nearly perfectly such that the six-membered rings overlap with five-membered rings, Figure 3A. The G–G intercalation is sandwiched between two G-A base pairs leading to a continuous cross-strand four guanine stack with near-parallel orientation of all four guanine bases as is clearly seen in Figure 2B (bottom). Four guanine stacks are separated along the infinite duplex by canonical A–T/T–A base pairs and together, they strongly stabilize the whole assembly. The T–A

pairs are nearly planar with only 10 ± 3° buckle and 9 ± 1° propeller deformations (Supplementary Table S3).

Groove width and helical twist in S3

Just like canonical duplexes, the S3 structure contains a major and minor groove (Figure 2B). The A–T base pairs form the floor of the major groove with an average width of 14.8 ± 0.4 Å. The groove narrows at the G–A base pair to 11.5 ± 0.4 Å, and then narrows further to 7.0 ± 0.7 Å at the intercalated guanines. The presence of the G–A pair brings the backbone closer together allowing for G–G intercalation. The values of groove width for each duplex are presented in Supplementary Table S4. Due to the non-planarity of the G–A base pairs and the offset of the sugars in the intercalated stack, the values of groove width likely overestimate the space available for binding (e.g. with small molecules or proteins) by at least 0.5 Å.

The values of base step parameters are presented in Supplementary Table S3. The average helical twist between G–A and T–A base pairs is 60 ± 2°, much higher than the 50.5° twist observed in the two repeat HSat3 5′-G(TGGAA)$_2$C-3′ structure (PDB ID 5GUN) (20). Conversely, the average helical twist between T–A and A–T base pairs is only 24 ± 3° lower than ∼34.5°, the typical value in B-DNA. Thus, the S3 duplex is overwound at the G-G intercalation and underwound at the A-T base pair. Over multiple units, such winding leads to an overall straight flat duplex that twists every five nucleotides at the G–G intercalation (Figure 2B). Overall, helical twist is the only prominent base pair step character, with other characters (shift, slide, rise, tilt and roll) displaying low values.

Formation of an infinite DNA duplex

The ASU is best represented by the self-complementary pentameric fragment 5′-AATGG-3′. Base pairing of the two fragments and application of the symmetry elements creates an infinite duplex. How this infinite duplex was created by the 5′-ATGGA ATGGA ATGGA ATGGA-3′ sequence used to grow crystals is ambiguous. Four arrangements are possible starting with the duplex that contains a maximum number of base pairs and 2 nucleotide sticky ends at 5′ and 3′ termini (Figure 2C). When one strand is moved over by one, two or three pentameric registers, new duplexes will have 7, 8 or 3 nucleotide sticky ends, respectively (Supplementary Figure S4). The presence of sticky ends should allow these duplexes to hybridize into an infinite polymer. In any of these arrangements, any two adenines are connected by phosphates in three out of four cases (i.e. 75% of the time); in the fourth case these adenines are terminal as reflected in the DNA sequence used to grow crystals. We indeed observed a weaker electron density between adjacent adenines and modeled one phosphorous and two bridging oxygens at 0.75 occupancy to represent the physical reality.

3D arrangement of the infinite DNA duplexes and intermolecular interactions

S3 crystals are formed by two alternating planes of DNA duplexes. The first plane contains alternating A–B and E–F duplexes and the second plane contains spaced out C–D duplexes, oriented perpendicular to the A–B/E–F duplexes (Figure 4A). Duplex A–B displays many close contacts with duplex C–D: they interact directly via four hydrogen bonds between the G4 phosphate in chain D, and N1 of G4 and N2 of G5 in chain B (Figure 4B). In addition, they share a well-defined and expansive water network (Supplementary Figure S5A). Interestingly, many of the water molecules appeared immediately after molecular replacement. Contrary, only a small number of waters are found between duplexes C–D and E'–F' (note, C–D duplex interacts with E'–F' duplex from the neighboring ASU and not from the same ASU). In most cases, water molecules bridge duplexes indirectly through interactions with other water molecules (Supplementary Figure S5B).

The ASU also contains one hexaaquamagnesium ion and one acetate ion. Both ions originate from the crystallization condition, which contained 0.1 M magnesium acetate tetrahydrate. Mg^{2+} ion is in perfect octahedral environment of six waters which interact with two phosphate, one acetate and three nitrogen atoms of DNA bases in chains B–D (Supplementary Figure S5A). The acetate also forms hydrogen bonds with nitrogen atoms of G4 in chain A. Thus Mg^{2+} and acetate bring together chains A–D and are responsible for stabilizing the S3 duplex. In the published structure of human centromeric repeat, Mg^{2+} is typically fully or partially hydrated in a near perfect octahedral geometry, as is observed here, and interacts with phosphates in the backbone either directly or through its primary water coordination sphere (16). The structure of 5′-G(TGGAA)$_2$C-3′ HSat3 sequence (PDB ID 5GUN) contains Co^{2+} ion which interacts primarily with the nitrogen atoms in DNA bases (20).

The extensive network of stabilizing interactions between chains A–D coupled with well-defined water networks lead to low average *B*-factors of 38 ± 2 Å^2 (Supplementary Figure S6). Chains E and F display a significantly higher average *B*-factor of 67 and 72 Å^2, respectively. In agreement with these values, Molecular Replacement procedure placed only A–B and C–D duplexes into the original electron density. The density of the E–F duplex became apparent only after a few cycles of refinement.

NMR-based structural characterization of S3 in aqueous solution

Next, we explored the structure of S3 using NMR to assure that crystallization conditions (30% MPD and 0.1 M magnesium acetate) did not affect S3 folding. NMR studies have the advantage of accessing the structure of S3 at a lower concentration range (0.05–0.5 mM per strand) compared with the crystallization studies (done at 2.5 mM).

The imino region of the 1D ^{1}H NMR spectrum at 25°C clearly reveals three groups of resolved signals corresponding to (i) A-T base pairs at δ 13–14 ppm, (ii) G–G intercalation at δ 9.7 ppm, and (iii) G–A base pairs at δ 10.5 ppm (Figure 5A). Their chemical shifts are defined by the ring current effects related to the stacking of nucleobases and hydrogen bonding. The spectral resolution for G–G and G–A regions is low, suggesting the similarity of their respective local shielding environments. We assigned ^{1}H NMR chemical shifts using X-edited HSQC spectra acquired at 25 and 0°C on partial residue-specifically ^{13}C- and ^{15}N-isotope labelled oligonucleotides (Supplementary Figures S7–S10, Supplementary Table S5). Based on the assignment, the G–A spectral region encompasses G3, G8, G13 and G18, while the G–G spectral region includes G4, G9, G14 and G19. The match between our ^{1}H NMR fingerprint region of S3 and previous data for the two repeat sequence 5′-TGGAA TGGAA-3′ (17) indicates that G4–G14 and G9–G19 form intercalated pairs, while G3,

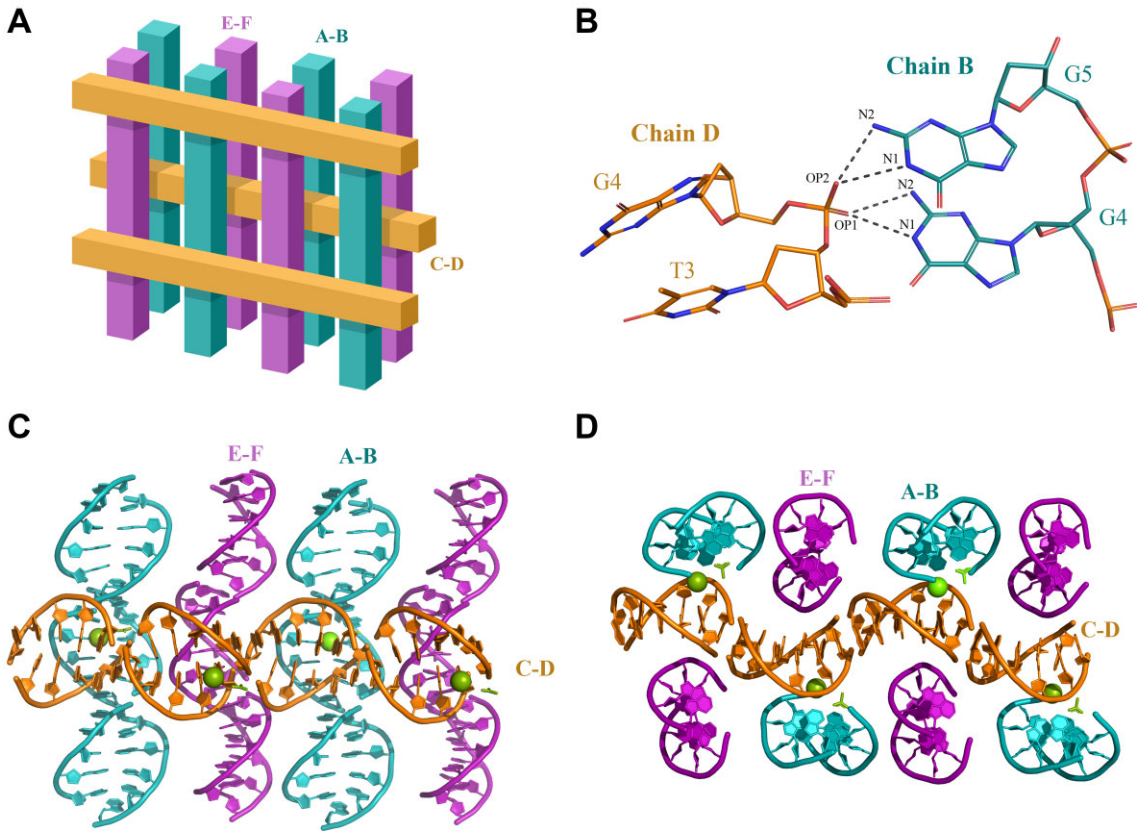


Figure 4. 3D arrangement of S3 DNA. (**A**) Schematic representation of the crystal lattice. (**B**) Direct interaction between A–B and C–D duplexes via backbone of chain D and guanine bases of chain B. Nucleotides are colored by atom: oxygen is red, nitrogen is blue, phosphorus is orange and carbon is teal in Chain B and orange in Chain D. (**C**) Front view of cartoon representation of two back rows in (A). (**D**) Bird's eye view of cartoon representation of three back rows in (A). The magnesium ions are shown as green spheres and acetate is colored light green. ASU contains unmerged five-nucleotides segments of DNA, leading to missing links between the segments.

G8, G13 and G18 form non-canonical sheared G-A base pairs. The formation of sheared and not head-to-head G-A base pairs (Figure 1) is substantiated by the lack of NOE interactions between guanine imino protons and adenine aromatic protons as well as the observed weak-to-medium intraresidual H1'-H6/H8 NOE correlations, consistent with *anti* glycosidic bond for all residues (also observed in the crystal structure of S3). Heating of 0.05–0.2 mM S3 samples leads to gradual unfolding of the secondary structure that is independent of the concentration consistent with a monomeric fold. The NMR data at high concentration do not display any evidence of dimer formation (Supplementary Figure S11). The monomeric nature of S3 is also observed in our biophysical studies (see below) but contradicts our X-ray observation of self-complementary duplex.

The NMR spectral analysis reveals the presence of a major and a minor form of S3 (peaks labeled with * on Figure 5B, C) that are found in dynamic equilibrium, with approximate ratio of 70:30. This ratio remains constant with increasing temperature (Supplementary Figures S11). NOESY spectra at 25°C display cross-peak T7H3-T12H3 corresponding to the major form and cross-peak T7H3*–T2H3* corresponding to the minor form (Figure 5B). Therefore, both forms exhibit A–T base pairs but engage A and T from different pentameric segments (registers). Specifically, the major form contains A1–T17, A6–T12, A11–T7 and A16–T2 base pairs, while the minor form contains A1–T7, A6–T2, A11–T17 and A16–T12 base pairs. In all cases, each A–T base pair is sandwiched by a T–A base pair on one side and a A–G base pair on the other side leading to the extremely narrow range of H2 signals for A1, A6, A11 and A16, δ 7.70–7.78 ppm (Figures 5C and S8). Interestingly, the A16H8 signal appears at different positions in the major and minor forms but the H8 signals for A1, A6 and A11 are found at the same positions in the major and minor forms (Supplementary Figure S8). This observation suggests that the orientation and/or dynamics of A16–T17–G18–G19–A20 segment in the major and minor forms may be different. At the same time, the H8 signal for A10, A15 and A20 engaged in G–A base pairing differs significantly for the major and the minor form, consistent with structural variations at the G–A base pairs and intercalated G–Gs. A schematic model of the major and minor forms consistent with the above observations is shown in Figure 5D and can be best described as a folded-over hairpin.

We gained a clearer understanding of the differences between the major and minor forms and their dynamic exchange by comparing the ^{1}H NMR signals for (i) methyl groups in A–T base pairs; (ii) H2 in G–A base pairs and (iii) H8 in the G–G intercalation. These ^{1}H signals are typically overlooked as they originate from protons that are exposed to grooves rather than engage in base-base interactions. The NOESY cross-peaks between the signals corresponding to methyl groups in A–T base pairs of the major and minor forms are consistent with exchange between the two species

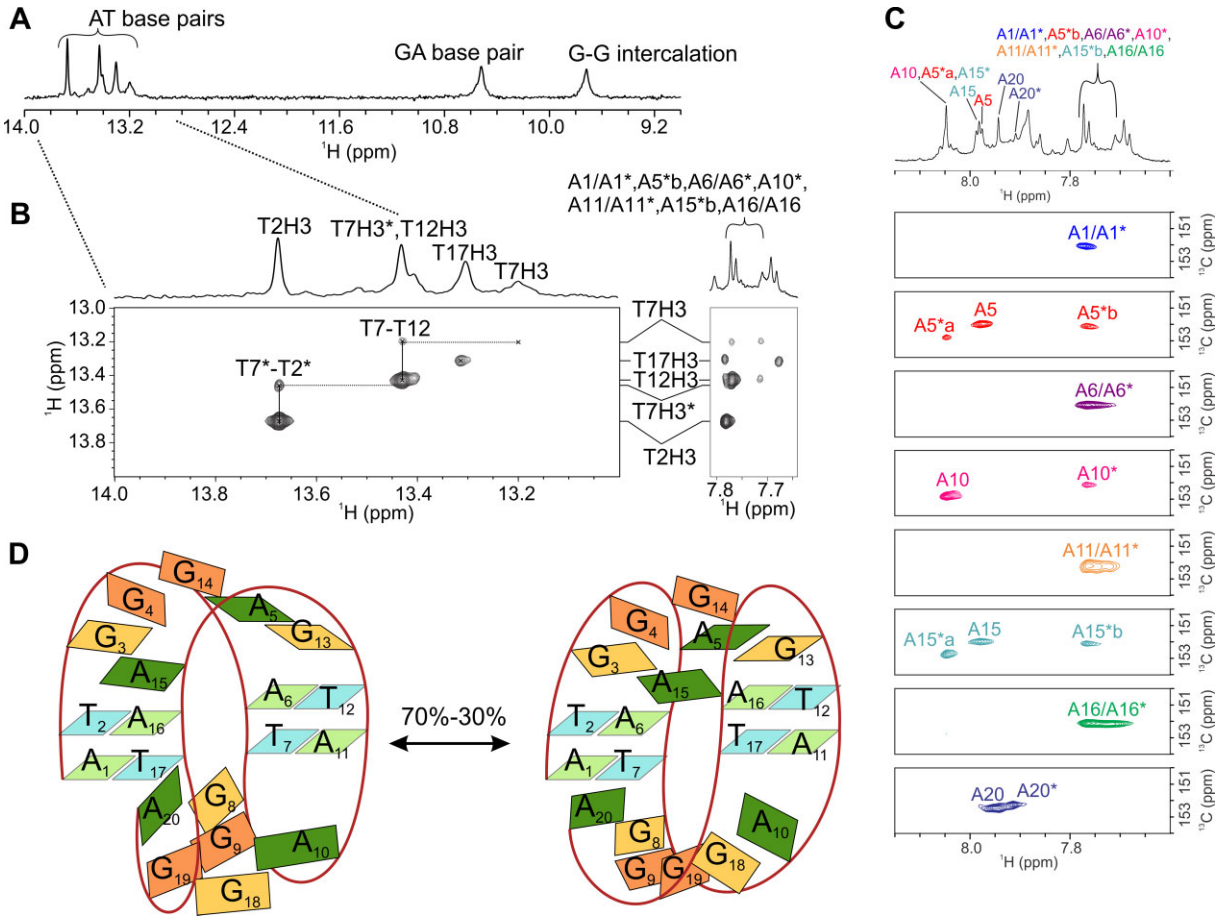


Figure 5. Solution-state NMR analysis of A–T base pairing and proposed structure of S3. (**A**) Imino-imino region of ^{1}H NMR spectrum. (**B**) Imino-imino (left) and imino-aromatic (right) regions of ^{1}H–^{1}H NOESY NMR spectrum (τm = 200 ms) with the corresponding ^{1}H NMR spectrum shown at the top. (**C**) ^{13}C-edited HSQC NMR spectra of S3 showing C2–H2 cross-peaks acquired on oligonucleotides with partial site-specifically ^{13}C-isotope labelled adenine residues. The spectra were recorded in 90%/10% $H_2O/^2H_2O$ at 25°C, 25 mM sodium-phosphate buffer pH 6.5 and 0.2–0.5 mM S3 per strand. Resolved signals for the minor form are labelled with *, while **a* and **b* denote ambiguous assignments that might relate to minor form signals or the background signals at natural ^{13}C-isotope abundance or signals corresponding to unfolded species. (**D**) Schematic depiction of equilibrium between the major and minor form of S3. The major form is displayed on the left. A and T nucleotides are in light green and cyan, respectively. A and G nucleotides in G–A base pairs are in dark green and yellow, respectively. Intercalated Gs are in orange. Numbering corresponds to the full length S3 sequences and does not correspond to the numbering of the X-ray structure.

(Supplementary Figure S12). Even more interesting, the ^{1}H NMR signals corresponding to H8 of the G–G intercalation in the major and minor form resonate at notably different positions. At 25°C, the signals for the major form G4 and G14 are found at δ 7.34 and 7.36 ppm, respectively, while signals for G9 and G19 are found at δ 7.69 and 7.70 ppm, respectively. These positions are nearly reversed for the minor form: δ 7.71 and 7.70 ppm for G4* and G14* H8, respectively, and at δ 7.33 and 7.31 ppm for G9* and G19*, respectively (Supplementary Figure S13). This data is consistent with G-G intercalation in the major and minor form exposed to alternating local environments. On the other hand, only one set of ^{1}H NMR signals is observed for H8 of guanines that participate in G-A base pairing (G3, G8, G13 and G18) implying that these residues exhibit more similar conformations in both the major and minor forms.

The spectral analysis of the equilibrium between the major and minor forms is convoluted by distinct local flexibilities. Such flexibilities lead to different number of ^{1}H NMR signals observed for aromatic protons of adenine residues in G–A base pairs. In particular, the numbers of H2/H8 signals at 0°C are 1/3, 2/2, 2/3 and 3/2 for A5, A10, A15 and A20, respectively. Hence, the degrees of freedom for an individual residue or segment are coupled with the overall features of a particular form. It is also interesting to compare widths of the thymine imino ^{1}H NMR signals upon temperature decrease from 25 to 0°C. One can observe severe broadening of T7H3 and T12H3, while T2H3 and T17H3 remain well resolved. Therefore, the only imino-imino NOE cross-peak observed at 0°C corresponds to interaction of T2H3 and T17H3 (Supplementary Figure S14).

In order to provide more structural insights into the major and minor forms, we performed restrained simulated annealing calculations using restraints presented in Supplementary Table S1. The resulting structural models of the major and minor forms (Supplementary Figures S15 and S16) corroborate the above noted features and reveal addition details related to the flexibility of certain segments, especially G18–G19–A20. In particular, we compared two models of the minor form, one that has the lowest energy and another that violates the least restraints. This comparison reveals conformational variations of intercalated G9 and G19 with their imino protons

either far from any potential hydrogen bond acceptor (>5.2 Å) or in the proximity of the phosphate groups of A20 and A10, respectively (<2.0 Å, Supplementary Figure S17). This short distance suggests existence of hydrogen bonds. The different dispositions of G9–A20 base pair, intercalated G9–G19 and G18–A10 base pair in the models of the minor form is consistent with two sets (higher and lower intensity) of NMR signals observed at 0°C for imino protons of G9, G18 and G19.

In sum, our NMR data suggest that S3 in solution exists in a form of two compact (i.e. folded-over) hairpins with identical type of base pairing and stacking but engaging nucleotides from different registers. These hairpins are in dynamic equilibrium with each other and likely with basic hairpins (Figure 8A) and an unfolded structure. The canonical and non-canonical bonding observed in NMR is fully consistent with the bonding observed in the crystal structure.

Biophysical characterization of S3

We examined S3 in TSB buffer under dilute conditions from 4 to 850 μM using circular dichroism (CD), thermal stability (Figure 6 and Supplementary Tables S6), native PAGE (Figure 7A), thermal difference spectra, TDS (Supplementary Figure S18), and analytical ultracentrifugation, AUC (Supplementary Table S7). The TDS signature consists of a single peak at 257 nm, which suggests G–A pairing (44). The CD signature contains a major peak at 270 nm, a trough at 244 nm and a shoulder at ∼280 nm indicating the clear presence of a secondary structure. The CD signature is concentration independent (in the 4–850 μM range) suggesting that the structure formed is likely monomolecular. This observation is corroborated by thermal stability experiments, which demonstrate that melting transitions are concentration independent and reversible (Figure 6C, D, Supplementary Table S6). In TSB buffer S3 melts at 62.0 ± 0.7°C with the enthalpy of unfolding of 222 ± 9 kJ/mol and low hysteresis (<3°C). Using these values, we can calculate the entropy and Gibbs free energy (at 25°C) for the folding process: –660 ± 30 J/(mol × K) and –25 ± 1 kJ/mol, respectively. The observed values indicate that folding of S3 in solution is characterized by favorable Gibbs free energy that results from favorable enthalpic and unfavorable entropic contributions. Favorable enthalpy results from base pairing and base stacking interaction during folding, while unfavorable entropy is due to ordering of the unfolded DNA strands as well as likely organization of ions and solvent around DNA backbone.

PAGE demonstrates that S3 runs as a clean fast-moving band at both low and high concentrations (Figure 7A) suggesting that the secondary structure is homogeneous, monomolecular and compact in agreement with the folded-over hairpin model from NMR. AUC data indicates that in the 5–720 μM range S3 is mostly monomolecular (>80%). AUC samples also contain a small amount of dimeric species (e.g. duplex) at higher concentrations. It is likely that the major and minor conformations observed in NMR behave rather similar in the biophysical experiments and, thus are indistinguishable.

Effect of buffer and its components on S3 folding and stability

S3 samples for X-ray crystallographic studies were prepared in TSB buffer (10 mM Tris-acetate pH 8.3, 50 mM potassium acetate, and 5 mM magnesium acetate), while the NMR samples were prepared in 25 mM NaPi pH 6.5 buffer. Therefore, we directly compared the folding and stability of S3 in both buffers as well as tested the effect of ionic strength and Mg^{2+} using CD scans and UV–vis thermal stability studies. We find that the identity of the buffer (and its pH in the range 6.5–8.3) does not affect the fold of S3 (Figure 6B). We altered the ionic strength by adding increasing amounts of potassium acetate (0–150 mM) to 10 mM Tris-acetate, pH 8.3 buffer or by increasing amount of NaPi pH 6.5 (25–150 mM). We observed that the CD signature of S3 is independent of the ionic strength (Supplementary Figure S19A, B), while the stability increases with ionic strength (Supplementary Figure S19D). Interestingly, the stability of S3 was nearly identical in Tris-acetate pH 8.3 and in NaPi pH 6.5 buffers of a similar ionic strength (at 150 mM $T_m \sim 63$°C Supplementary Figure S19E). Finally, the presence of Mg^{2+} did not affect the CD signature (Supplementary Figure S19C) and only slightly improved the stability of S3 from 59.1 (0 mM Mg^{2+}) to 61.2°C (10 mM Mg^{2+}). In summary, these results suggest that differences in structure and molecularity observed in NMR and X-ray crystallographic studies are not related to buffer composition or pH. It is likely, that S3 concentration may be responsible for the observed differences.

Molecularity of S3

We further addressed the question of molecularity and the effect of S3 concentration on its fold using native PAGE, Figure 7A. We observed that 50 μM and 2.5 mM S3 as well as S3 crystals dissolved in TSB buffer ran similarly on native gels, suggesting either that the concentration does not affect the molecularity or that self-complementary duplexes observed in the crystalline state dissociate during electrophoresis. We designed two complementary pyrimidine stands, one that should produce blunt ends with S3 (S3D, $d(TCCAT)_4$) and another that should lead to 5′- and 3′-overhangs of two nucleotides (S3O, $d(ATTCC)_4$). If our design works, we expect duplex formation between S3 and S3D and formation of an infinite duplex between S3 and S3O relying on adhesion of the overhangs, similar to the situation in our crystal structure. On native gels S3D and S3O ran at a significantly retarded rate as compared to S3 (Figure 7A) confirming that S3 folds into a compact secondary structure, while S3O and S3D remained unfolded. In addition, we observed formation of Watson–Crick–Franklin S3:S3D and S3:S3O duplexes which migrated more slowly than unfolded strands due to their approximately doubled molecular weight. Importantly, S3:S3D and S3:S3O duplexes migrated at a similar position on the native gels indicating that S3:S3O is not an infinite duplex, suggesting either that base pairing in overhangs does not occur under these conditions or that individual duplexes in the infinite polymer are separated by electrophoretic forces. Our PAGE results are fully consistent with those from Bradbury laboratory, which demonstrated increased electrophoretic mobility for two different phase-shifted variants of purine strand of HSat3, $d(AATGG)_4$ and $d(GGAAT)_4$, compared with the complementary pyrimidine strand, $d(CCATT)_4$, and even slower mobility for a Watson–Crick–Franklin duplex, $d(AATGG)_4$:$d(CCATT)_4$ (21).

We complemented PAGE study with CD scans and thermal stability studies. S3D and S3O displayed a single broad CD peak at 273 (Supplementary Figure S20) and poorly defined melting transitions (Figure 7B). S3:S3D and S3:S3O

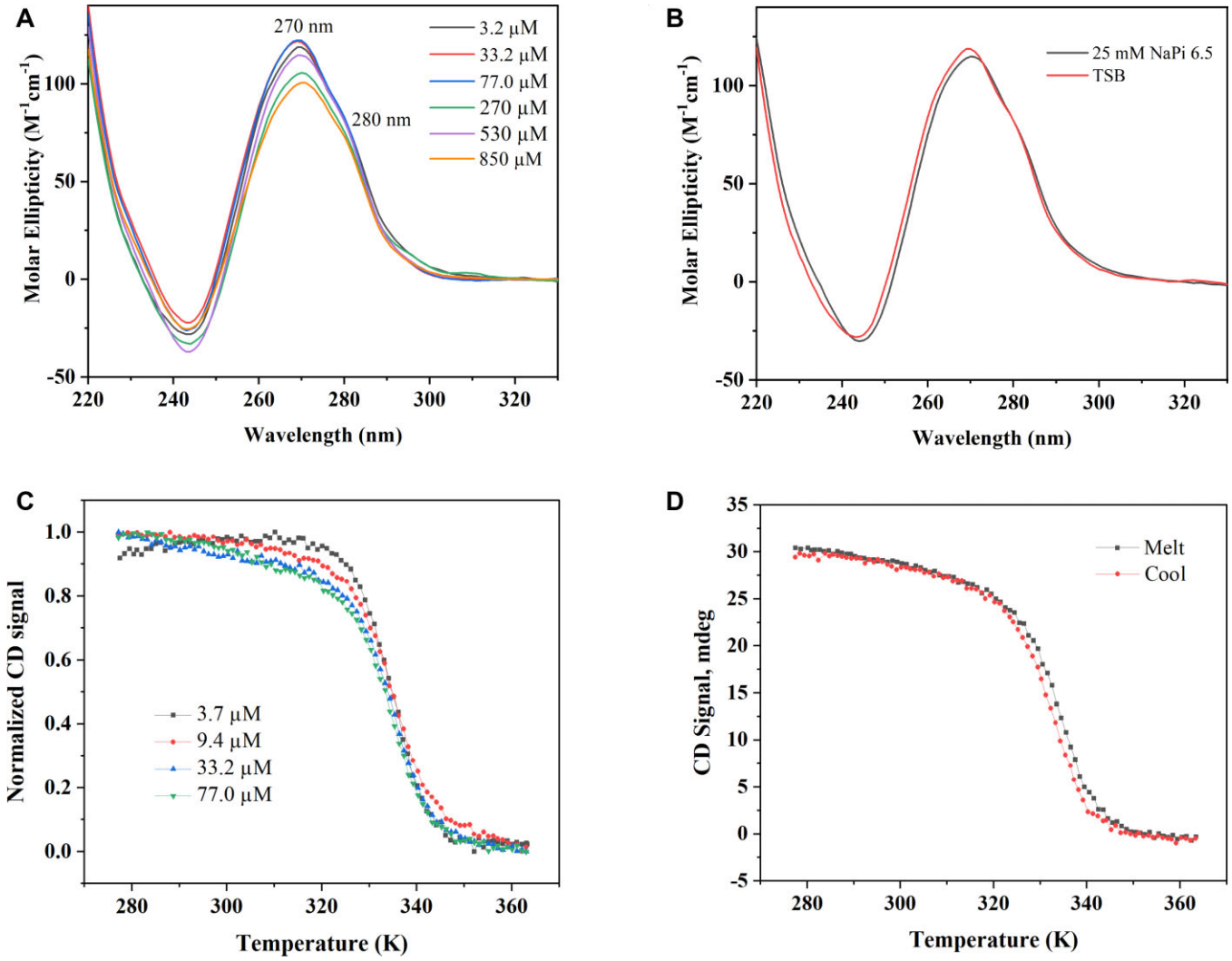


Figure 6. Biophysical characterization of S3. (**A**) CD spectra of S3 samples at 4–850 μM in TSB buffer. (**B**) Comparison of CD signature of S3 in two buffers used for X-ray and NMR structural studies. (**C**) Concentration dependence of the melting transition. Melting curves overlay well, suggesting that the structure is monomolecular and the melting transition is a two-state process. (**D**) Overlay of melting and cooling curves for S3 at 77.0 μM. The hysteresis for this data set is 1.8°C.

Watson–Crick–Franklin duplexes, on the other hand, displayed well-defined melting curves with T_m of 64.6 ± 0.3, and 63.8 ± 0.3°C, respectively (Figures 7B and Supplementary Table S8) due to the extensive base pairing between the two oligonucleotides. Significantly, the melting curve and melting temperature (T_m = 62.0 ± 0.7°C) for purine-rich S3 strand, but not pyrimidine-rich S3O and S3D strands, are almost identical to that of the duplexes providing further evidence for extensive intramolecular base pairing within S3 (Figures 7B and Supplementary Table S8). Interestingly, the enthalpy values differ greatly between the S3 hairpin (with ΔH of 222 ± 9 kJ/mol) and S3:S3D duplex (with ΔH of 352 ± 5 kJ/mol) and S3:S3O duplex (with ΔH of 328 ± 1 kJ/mol). These higher enthalpy values indicate a higher number or higher efficiency of base pairing interactions in the duplexes compared with the S3 hairpin. Our data are consistent with similar studies on the six-repeat HSat3 sequence $d(GGAAT)_6$, where the purine-rich strand and Watson–Crick–Franklin duplex displayed a similar thermal stability of ~65°C, while the pyrimidine strand did not display a clear melting transition (8). Taken together, our biophysical data demonstrates that S3 folds into a monomolecular structure at concentrations below 0.85 mM in TSB buffer, providing additional support for our NMR data.

Importance of the guanine zipper element tested via point mutations

Guanine zipper is composed of the intercalated G–G pair bracketed on each side by the G–A non-canonical base pairs. Its most prominent feature is the four guanines stack best seen in Figure 2B. To test the importance of this structural element for S3 folding and stability, we systematically replaced G9 (using numbering shown in Figure 8A) from the intercalated G–G pair with either A, T or C. The list of mutants and their thermodynamic parameters are listed in Supplementary Table S8. No major perturbation of the S3 structure was observed according to PAGE (Figure 8B), TDS, CD scans or thermal stability studies (Supplementary Figure S21). This finding indicates that either G9 does not intercalate and instead forms a single nucleotide loop as suggested by the earlier NMR study (21) or that A, G, C and T are all able to intercalate.

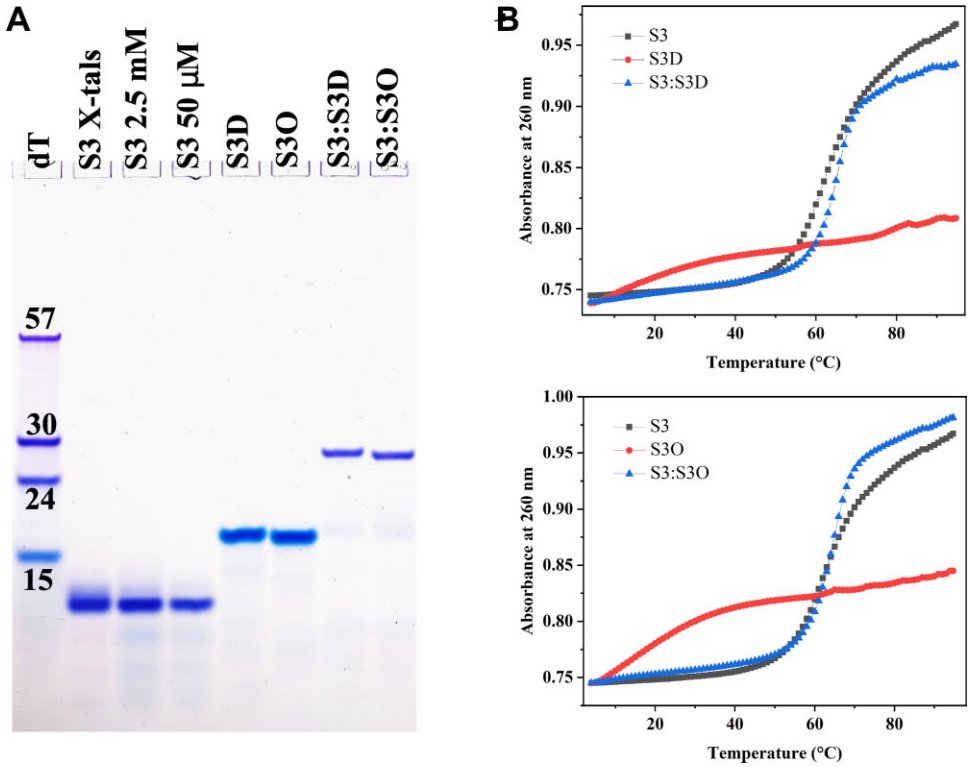


Figure 7. Homogeneity, molecularity, and conformation of S3 with the sequence d(ATGGA)$_4$, and its two different complementary strands, S3D with the sequence d(TCCAT)$_4$ and S3O with the sequence d(ATTCC)$_4$. (**A**) Fifteen percent native PAGE gel prepared in 1 × TBE supplemented with 10 mM KCl. All DNA samples were prepared at 50 μM except for S3 X-tals (crystals washed in crystallization buffer and dissolved in TSB buffer), S3 2.5 mM (the high concentration sample of S3; this sample was diluted to 50 μM right before loading) and S3:S3D and S3:S3O that were prepared at 25 μM of each strand. (**B**) UV–vis melting curves for S3, S3D and S3O strands and S3:S3D and S3:S3O duplexes. All experiments were performed in TSB buffer.

In the next set of mutants, we probed the role of the sheared G–A base pair in the guanine zipper. We hypothesized that replacing the G–A base pair with Watson–Crick–Franklin pairs should lead to greater stability of the resulting structures. We replaced the two sheared G–A/A–G base pairs with either (i) T–A and A–T (S3-AT1 mutant); (ii) A–T and T–A (S3_AT2 mutant); (iii) G–C and C–G (S3_GC1 mutant) or (iv) C–G and G–C (S3_GC2 mutant). The biophysical data (PAGE in Figure 8C and CD, TDS and melt in Supplementary Figure S22) indicate that for each mutant either PAGE mobility (S3_AT1 and S3_AT2), CD signature (S3_AT2 and S3_GC1), TDS (S3_AT1 and S3_GC1) or thermal stability (S3_AT2 and S3_GC2) changed dramatically as compared to S3 indicating that such mutations are not tolerated.

Discussion

Stabilization of S3 structure by non-canonical stacking and base pairing

The key structural element in S3 is a guanine zipper, which is formed by G–A base pairs bracketing each side of intercalated guanines leading to a cross-strand four guanine stack. Such a structural element is also observed in other structures of HSat3 DNA containing at least two pentanucleotide repeats, Supplementary Table S9 (17,18,20,22). Each guanine in the four guanine stack is oriented such that the face of the base containing N2–N1–O6 is exposed resulting in a four nucleotide long 'sticky patch' available for interactions with centromere proteins, complementary DNA strands, or small molecule ligands. Similar elements featuring four intercalated adenines sandwiched by sheared G–A base pairs are known as adenine zippers (39). Guanine zippers in S3 are spaced by canonical T–A/A–T base pairs.

The necessity of G–A sheared base pairs to interface between G–G intercalation and T–A base pairs arrives from poor compatibility of their groove widths. Specifically, the G–G intercalation has a backbone-to-backbone groove width of only 7.0 ± 0.7 Å, while the T–A base pairs have groove width of 14.8 ± 0.4 Å (Supplementary Table S4). Significant changes in groove width likely add strain to the structure, especially when accounting for the stretch of the backbone introduced by the G–G intercalation. Incorporation of a sheared G–A base pair, whose backbone phosphates are spaced by 11.5 ± 0.4 Å, provides a smooth transition. A head-to-head G–A base pair would not be able to accommodate such transition as the two backbones are positioned farther apart than even in a Watson–Crick–Franklin base pair (16.9 Å) (45). The importance of the sheared G–A base pair is emphasized further through our mutational studies. Replacement of the G–A base pairs with canonical A–T or C–G base pairs results in a disrupted fold and stability despite the expected stabilization due to introduction of highly efficient Watson–Crick–Franklin interactions.

The G–G intercalation is also accompanied by the non-planarity of the G–A pairs (a buckle deformation of 32°, Supplementary Table S3). The local twisting of the duplex happens in such a way that the infinitely long polymer adopts a geometry of a straight flat ribbon that is underwound at A–T/T–A location (helical twist of 24°) and highly overwound at G–G intercalation sites (helical twist of 60°; for comparison, the helical twist in B-DNA is ∼34.5°) (Figure 2B).

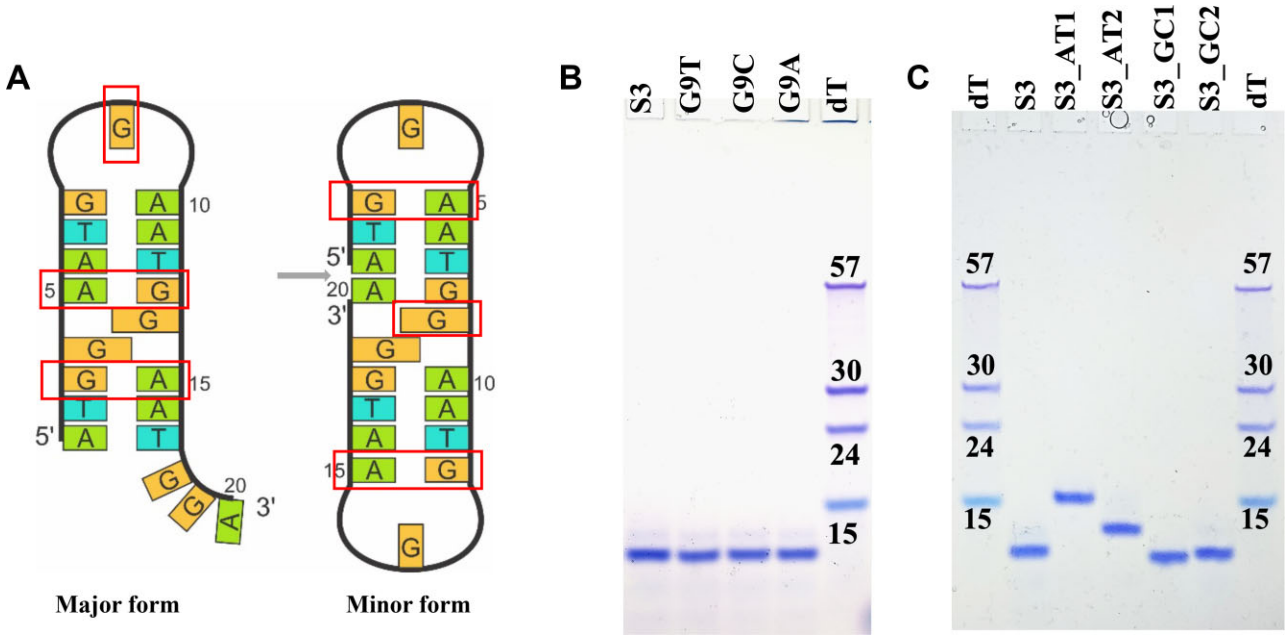


Figure 8. Effect of mutations in the guanine zipper element on S3 structure in solution. (**A**) Schematics of the basic S3 hairpin structures in dilute conditions. These structures were obtained by 'unfolding' of the folded-over hairpin structures obtained via NMR and presented in Figure 5D. The regions probed by mutations are highlighted. (**B**) Fifteen percent native PAGE gel showing migration of oligonucleotides bearing G9 mutations. (**C**) Fifteen percent native PAGE for G–A mutants. Size markers correspond to dT_n sequences. All DNA samples were prepared at ~50 μM in TSB buffer. Gel was prepared in 1 × TBE buffer supplemented with 10 mM KCl.

Comparison of S3 to other HSat3 structures

Careful examination of all HSat3 sequences studied via NMR and X-ray crystallography indicates that only one unique single-repeat sequence (16,19), one unique two-repeat sequence (17,18,20), and one unique three-repeat sequence with a G-to-C mutation in the central repeat (22) have been structurally characterized (Supplementary Table S9). In addition, there are two studies of 3–6, 8 and 10 repeat sequences with different registers, $(AATGG)_n$ using NMR (21) and $(TGGAA)_n$ via single molecule FRET (20). Structural work was pioneered by the Reid laboratory who used NMR to demonstrate that DNA sequences with one GAATG repeat form self-complementary duplexes with tandem G–A base pairs; while the two- and three-repeat sequences of a phase-shifted variant, TGGAA, also form self-complementary duplexes with a guanine zipper instead (17,18). The absence of the guanine zipper in the single repeat DNA was likely a result of the register choice. Both one- and two-repeat structures were later confirmed in X-ray structural studies (16,20). In all cases, the sequences under consideration contained one or two non-native G–C terminal base pairs installed to prevent fraying of DNA ends. There is also one structure of the three-repeat 5′-TGGAA TGCAA TGGAA-3′ sequence with a single G-to-C mutation in the middle repeat (underlined). This sequence forms a hairpin with a one nucleotide C-loop stacked on a sheared G–A base pair (22). More generally the Reid laboratory demonstrated that sequences with GNA triplets form (i) predominantly a self-complementary duplex when N is G (which is the case in this work); (ii) a mixture of duplex/hairpin with hairpin being a more prominent structure when N is either A or T and (iii) a hairpin when N is C (46). Cytosine is particularly good at making turns in hairpins, while guanines are the least capable of such turns. Our structure is the first deposited into a public data bank of a human

Table 2. RMSD between S3 and previously published HSat3 structures. S3 is compared to either AATG or, where available, to AATGG fragment. Results are averaged for all three copies of S3

PDB ID	RMSD vs. AATG sequence, Å	RMSD vs. AATGG sequence, Å
1XUE	0.91 ± 0.08	-
5GUN	0.72 ± 0.10	0.78 ± 0.12
1D8X	1.1 ± 0.1	-
1D9R	0.9 ± 0.1	-
1DCR	0.95 ± 0.07	-
103D	0.6 ± 0.2	0.7 ± 0.2
175D	1.1 ± 0.1	-

HSat3-derived four-repeat sequence without any mutations. It is highly similar to other published HSat3 structures and has a RMSD value under 1.1 Å (Tables 2 and S10). Such low values of RMSD suggests that the presence of G–C stabilizing base pairs in all previously published structures does not affect the HSat3 fold.

We are not the first to study HSat3-derived repeats with $n \geq 4$ containing only native nucleotides. The Bradbury laboratory investigated $(AATGG)_n$, where $n = 2$, 3, 4 and 6 via NMR (21). Their results suggest that depending on n, DNA and salt concentration, the sequences can adopt a monomeric hairpin or end-stacked dimeric hairpin. The monomeric hairpin typically dominates at low DNA concentrations. At higher DNA concentrations and for $n = 3$, 4 and 6, the dimeric hairpin was observed with a base paired stem and two loops (21). The loops are composed of GGA nucleotides in kinetic exchange with a one-nucleotide G loop stacked on a sheared G–A base pair that extends the stem. Note, careful examination of our NMR data does not suggest any S3 dimer formation at the concentrations studied (up to 0.5 mM). Brad-

bury's NMR data were interpreted to suggest a G–G base pair in place of G–G intercalation observed in our and Reid's structures, although this and some other structural elements have been criticized by the Reid laboratory (17). Unfortunately, the Bradbury structures were not deposited into any public databank.

More recently, $(TGGAA)_n$ (where $n = 3–6$, 8 and 10) was investigated using single molecule FRET (20). The data indicate that DNA adopts a stable monomeric hairpin with either blunt ends for an odd n or with a 3′ overhang equal to one pentameric register for an even n. It is important to note, that such conformations are valid only for the chosen register, otherwise, the length of the stem and overhangs will vary. Interestingly, the presence of the long overhang was suggested to lead to strand slippage and may potentially explain the origin of repeat expansion diseases involving HSat3 sequences, such as spinocerebellar ataxia type 31 (20).

Concentration dependence of HSat3 structures

In our hands, S3 with the sequence $d(ATGGA)_4$ adopts either a self-complementary duplex structure (Figure 2, X-ray) or compact, folded-over hairpin structure (Figure 5D, NMR). Duplex structure is promoted at concentrations of 2.5 mM under crystallization conditions, while folded-over hairpin structure is formed at concentrations below 0.5 mM in 25 mM NaPi buffer pH 6.5. The folded-over hairpin is highly dynamic and is in equilibrium with another folded-over hairpin that uses nucleotides from a different pentameric register as well as, likely, the basic hairpin (Figure 8A). Both arrangements—duplex and hairpins—contain guanine zippers spaced by T–A/A–T base pairs and the principal difference is the presence of a G (or also possible GGA) loop in the hairpins, disrupting the extended structural pattern.

Our work is in full agreement with previous structural and biophysical studies. All NMR studies in Reid's laboratory were performed at high DNA concentration, ≥5 mM, which can explain why they observed self-complementary duplex DNA (17–19), save the three-repeat hairpin structure with the G-to-C mutation in the central repeat (22). Similarly, two other crystallographic studies which started with 1.5 mM DNA also observed self-complementary duplexes (16,20). In contrast, the single molecule FRET study used very low DNA concentrations (DNA was annealed at 5 μM and diluted to 10 pM before use) (20). This study observed only hairpins. Similarly, d(GCGAAAGCT) DNA forms a self-complementary duplex with an adenine zipper in a crystalline state (39), but a hairpin under more dilute conditions (47). In addition to DNA concentration, duplex-hairpin equilibrium is also influenced by salt identity and concentration (21), mutations (46), buffer, and temperature.

Biological implications

Our data and that of others indicate that the purine-rich strand of HSat3 under biologically relevant conditions, and when dissociated from its complementary strand, can adopt a hairpin structure. The hairpin can involve as few as 2–3 pentameric repeats or can be much longer, displaying properties of a self-complementary duplex (such as those observed in our X-ray structural studies). HSat3 duplexes and the stem of a hairpin are anchored by the same set of canonical and non-canonical interactions: A–T base pairs, sheared G–A base pairs, and a G–G intercalation. Previous studies suggest the HSat3 hairpin contains a GGA loop that likely forms a sheared G–A base pair capped by a stacked guanine (21). It is also possible that all three nucleotides in the loop are unstructured. Single molecule FRET data suggests that longer sequences may tolerate longer loops, e.g. of eight nucleotides (GGA repeat plus one pentameric register) (20). If the repeat is imperfect, G-to-C is the most common mutation in HSat3 with cytosine having high propensity for the tight turn (46) thus facilitating hairpin formation (22). The robustness of the HSat3 fold derives from the simplicity of its structural elements and the fact that the HSat3 hairpin can be formed by any number of pentameric repeats (for $n > 2$), while displaying similar structural and thermodynamic characteristics. This robustness is demonstrated by the fact that the introduction of stabilizing terminal G–C mutations into a number HSat3 variants did not disrupt this fold (16–22). As shown previously, repeat numbers greater than six can lead to hairpin slippage which can fuel the onset and progression of repeat expansion diseases (20). Intriguingly for the biology of HSat3 sequences in normal cells, one face of the four guanine stack in HSat3 hairpins has a number of exposed hydrogen bond donors and acceptors that can interact with (i) proteins or other biological molecules as part of kinetochore complex; (ii) with small molecules as potential therapy against repeat expansion diseases or (iii) with complementary DNA strands resulting in the formation of triplex DNA.

Conclusion

The tandem repeat sequence of HSat3 is highly conserved in humans and is typically known as $(GAATG)_n$, although it could be equally represented by different registers. In this work, we crystallized a four-repeat sequence 5'-ATGGA ATGGA ATGGA ATGGA-3'. Despite its 20 nucleotide length, the ASU contains only a pentameric double stranded fragment with four base pairs (two A–T and two G–A) and a G overhang at the 5′ and 3′ ends which leads to G–G intercalation. The pentameric fragments are arranged into highly symmetric infinite polymers. The high-resolution (1.57 Å) crystal structure described in this work is the first structure of HSat3-derived four-repeat sequence without any mutations, deposited into public data bank. Our NMR structural studies suggest that the same sequence forms at least two monomolecular folded-over hairpins in dynamic equilibrium with each other. The hairpins display base interactions similar to those found in the self-complementary duplex. Our biophysical and structural data shed light on the possible non-canonical biologically relevant arrangement of HSat3 purine-rich strand and, thus, serve as a basis for elucidation of HSat3 biological functions and involvement in repeat expansion diseases.

Data availability

S3 structure was deposited into PDB and assigned PDB ID: 7JLH.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We would like to thank Bryce Stanhope and Sawyer McCarthy (Swarthmore) for conducting preliminary biophysical experiments and PAGE. We would also like to thank Dr Kay Perry, a Staff Scientist at NE-CAT beamline at APS for her help and advice with data collection and processing. The authors acknowledge the CERIC-ERIC consortium for the access to experimental facilities and financial support.

Funding

L.Y. acknowledges National Institutes of Health [1R15CA253134]; Swarthmore startup funds award; Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health [P30 GM124165]; the Eiger 16M detector on 24-ID-E is funded by a NIH-ORIP HEI grant [S10OD021527]; Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory [DE-AC02-06CH11357]; M.T. acknowledges funding from the Packard Fellowship for Science and Engineering; M.T. and J.P. acknowledge the Slovenian Research Agency [P1-0242 and J1-1704]. Funding for open access charge: NIH [1R15CA253134].

Conflict of interest statement

None declared.

References

1. Garrido-Ramos,M. (2017) Satellite DNA: an evolving topic. *Genes*, **8**, 230.
2. Shatskikh,A.S., Kotov,A.A., Adashev,V.E., Bazylev,S.S. and Olenina,L.V. (2020) Functional significance of satellite DNAs: insights from Drosophila. *Front. Cell Dev. Biol.*, **8**, 312.
3. Plohl,M., Meštrović,N. and Mravinac,B. (2014) Centromere identity from the DNA point of view. *Chromosoma*, **123**, 313–325.
4. Holmquist,G.P. and Dancis,B. (1979) Telomere replication, kinetochore organizers, and satellite DNA evolution. *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 4566–4570.
5. Pezer,Ž., Brajkovic,J., Feliciello,I. and Ugarkovc,Đ. (2012) Satellite DNA-mediated effects on genome regulation. In: Garrido-Ramos, M.A. (ed.) *Genome Dynamics*. S. KARGER AG, Basel, Vol. **7**, pp. 153–169.
6. Bansal,A., Kaushik,S. and Kukreti,S. (2022) Non-canonical DNA structures: diversity and disease association. *Front. Genet.*, **13**, 959258.
7. Maio,J.J., Brown,F.L. and Musich,P.R. (1981) Toward a molecular paleontology of primate genomes. *Chromosoma*, **83**, 103–125.
8. Grady,D.L., Ratliff,R.L., Robinson,D.L., McCanlies,E.C., Meyne,J. and Moyzis,R.K. (1992) Highly conserved repetitive DNA sequences are present at human centromeres. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 1695–1699.
9. Warburton,P.E., Hasson,D., Guillem,F., Lescale,C., Jin,X. and Abrusan,G. (2008) Analysis of the largest tandemly repeated DNA families in the human genome. *Bmc Genomics [Electronic Resource]*, **9**, 533.
10. Altemose,N., Logsdon,G.A., Bzikadze,A.V., Sidhwani,P., Langley,S.A., Caldas,G.V., Hoyt,S.J., Uralsky,L., Ryabov,F.D., Shew,C.J., *et al.* (2022) Complete genomic and epigenetic maps of human centromeres. *Science*, **376**, eabl4178.
11. Sato,N., Amino,T., Kobayashi,K., Asakawa,S., Ishiguro,T., Tsunemi,T., Takahashi,M., Matsuura,T., Flanigan,K.M., Iwasaki,S., *et al.* (2009) Spinocerebellar ataxia type 31 is associated with "inserted" penta-nucleotide repeats containing (TGGAA)n. *Am. J. Hum. Genet.*, **85**, 544–557.
12. Miga,K.H. (2017) The promises and challenges of genomic studies of Human centromeres. *Prog. Mol. Subcell. Biol.*, **56**, 285–304.
13. Miga,K.H. and Sullivan,B.A. (2021) Expanding studies of chromosome structure and function in the era of T2T genomics. *Hum. Mol. Genet.*, **30**, R198–R205.
14. Goenka,A., Sengupta,S., Pandey,R., Parihar,R., Mohanta,G.C., Mukerji,M. and Ganesh,S. (2016) Human satellite-III non-coding RNAs modulate heat shock-induced transcriptional repression. *J. Cell Sci.*, **129**, 3541–3552.
15. Porokhovnik,L.N., Veiko,N.N., Ershova,E.S. and Kostyuk,S.V. (2021) The role of Human satellite III (1q12) copy number variation in the adaptive response during aging, stress, and pathology: a pendulum model. *Genes*, **12**, 1524.
16. Gao,Y.-G., Robinson,H., Sanishvili,R., Joachimiak,A. and Wang,A.H.-J. (1999) Structure and recognition of sheared tandem G·A base pairs associated with Human centromere DNA sequence at atomic resolution. *Biochemistry*, **38**, 16452–16460.
17. Chou,S.-H., Zhu,L. and Reid,B.R. (1994) The unusual structure of the Human centromere (GGA)2 motif: unpaired guanosine residues stacked between sheared G·A pairs. *J. Mol. Biol.*, **244**, 259–268.
18. Zhu,L., Chou,S.-H. and Reid,B.R. (1995) The structure of a novel DNA duplex formed by Human centromere d(TGGAA) repeats with possible implications for chromosome attachment during mitosis. *J. Mol. Biol.*, **254**, 623–637.
19. Chou,S.H., Cheng,J.-W. and Reid,B.R. (1994) DNA sequence GCGAATGAGC containing the human centromere core sequence GAAT forms a self-complementary duplex with sheared G.A pairs in solution. *J. Mol. Biol.*, **241**, 467–479.
20. Huang,T.-Y., Chang,C., 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.
21. Catasti,P., Gupta,G., Garcia,A.E., Ratliff,R., Hong,L., Yau,P., Moyzis,R.K. and Bradbury,E.M. (1994) Unusual structures of the tandem repetitive DNA sequences located at Human centromeres. *Biochemistry*, **33**, 3819–3830.
22. Zhu,L., Chou,S.H. and Reid,B.R. (1996) A single G-to-C change causes human centromere TGGAA repeats to fold back into hairpins. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 12159–12164.
23. Brown,T., Hunter,W.N., Kneale,G. and Kennard,O. (1986) Molecular structure of the G.A base pair in DNA and its implications for the mechanism of transversion mutations. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 2402–2406.
24. Kan,L.S., Chandrasegaran,S., Pulford,S.M. and Miller,P.S. (1983) Detection of a guanine-adenine base pair in a decadeoxyribonucleotide by proton magnetic resonance spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 4263–4265.
25. Li,Y., Zon,G. and Wilson,W.D. (1991) Thermodynamics of DNA duplexes with adjacent G–A mismatches. *Biochemistry*, **30**, 7566–7572.
26. Nicoludis,J.M., Barrett,S.P., Mergny,J.-L. and Yatsunyk,L.A. (2012) Interaction of human telomeric DNA with N- methyl mesoporphyrin IX. *Nucleic Acids Res.*, **40**, 5432–5447.
27. Ramsay,G.D. and Eftink,M.R. (1994) Analysis of multidimensional spectroscopic data to monitor unfolding of proteins. *Methods Enzym*, **240**, 615–645.
28. Schuck,P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.*, **78**, 1606–1619.
29. Kabsch,W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.*, **26**, 795–800.

30. Evans,P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 72–82.
31. Evans,P.R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 282–292.
32. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) *Phaser* crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
33. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of *Coot*. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
34. Afonine,P.V., Grosse-Kunstleve,R.W., Echols,N., Headd,J.J., Moriarty,N.W., Mustyakimov,M., Terwilliger,T.C., Urzhumtsev,A., Zwart,P.H. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with *phenix.Refine*. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 352–367.
35. Kantardjieff,K.A. and Rupp,B. (2003) Matthews coefficient probabilities: improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci.*, **12**, 1865–1871.
36. Lu,X.-J. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
37. Zheng,H., Cooper,D.R., Porebski,P.J., Shabalin,I.G., Handing,K.B. and Minor,W. (2017) *CheckMyMetal*: a macromolecular metal-binding validation tool. *Acta Crystallogr. Sect. Struct. Biol.*, **73**, 223–233.
38. Lee,W., Tonelli,M. and Markley,J.L. (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, **31**, 1325–1327.
39. Shepard,W., Cruse,W.B., Fourme,R., de la Fortelle,E. and Prangé,T. (1998) A zipper-like duplex in DNA: the crystal structure of d(GCGAAAGCT) at 2.1 å resolution. *Structure*, **6**, 849–861.
40. Sponer,J., Gabb,H.A., Leszczynski,J. and Hobza,P. (1997) Base-base and deoxyribose-base stacking interactions in B-DNA and Z-DNA: a quantum-chemical study. *Biophys. J.*, **73**, 76–87.
41. Wang,A.H.-J., Quigley,G.J., Kolpak,F.J., van der Marel,G., van Boom,J.H. and Rich,A. (1981) Left-handed double helical DNA: variations in the backbone conformation. *Science*, **211**, 171–176.
42. Hobza,P. and Šponer,J. (1999) Structure, energetics, and dynamics of the nucleic acid base pairs: nonempirical *abinitio* calculations. *Chem. Rev.*, **99**, 3247–3276.
43. Neidle,S. and Balasubramanian,S. (eds). (2007) Fundamentals of quadruplex structures. In: *RSC Biomolecular Sciences*. Royal Society of Chemistry, Cambridge, pp. 1–30.
44. Mergny,J.-L. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.*, **33**, e138.
45. Privé,G.G., Heinemann,U., Chandrasegaran,S., Kan,L.S., Kopka,M.L. and Dickerson,R.E. (1987) Helix geometry, hydration, and G.A mismatch in a B-DNA decamer. *Science*, **238**, 498–504.
46. Chou,S.-H., Zhu,L. and Reid,B.R. (1996) On the relative ability of centromeric GNA triplets to form HairpinsversusSelf-paired duplexes. *J. Mol. Biol.*, **259**, 445–457.
47. Hirao,I., Nishimura',Y., Naraoka,T., Watanabe,K. and Aratal,Y. (1989) Extraordinary stable structure of short single-stranded DNA fragments containing a specific base sequence: d(GCGAAAGC). *Nucleic Acids Res.*, **17**, 2223–2231.

Received: July 5, 2023. **Revised:** January 31, 2024. **Editorial Decision:** February 2, 2024. **Accepted:** February 6, 2024