Nucleic Acids Research, 2025, **53**, gkaf170
https://doi.org/10.1093/nar/gkaf170
Structural Biology

Systematic study of hybrid triplex topology and stability suggests a general triplex-mediated regulatory mechanism

Vito Genna [1,2,†], Guillem Portella [1,3,†], Alba Sala [1,†], Montserrat Terrazas [1,†], Israel Serrano-Chacón [1], Javier González [1], Núria Villegas [1], Lidia Mateo [1], Chiara Castellazzi [1], Mireia Labrador [1], Anna Aviño [4], Adam Hospital [1], Albert Gandioso [1], Patrick Aloy [1], Isabelle Brun-Heath [1], Carlos Gonzalez [5], Ramon Eritja [4], Modesto Orozco [1,6,*]

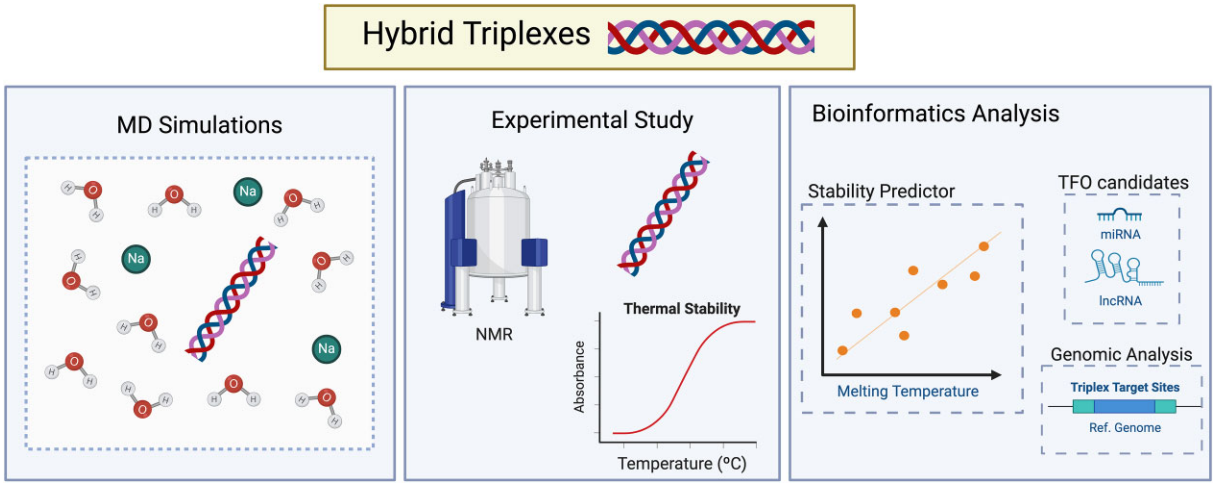
[1]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10-12, E-08028 Barcelona, Spain
[2]Nostrum Biodiscovery, SL. 08029 Barcelona, Spain
[3]Department of Chemistry, University of Cambridge, Cambridge CB2 1TN, UK
[4]Institute for Advanced Chemistry of Catalonia (IQAC), CSIC, Networking Center on Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), E-08034 Barcelona, Spain
[5]Instituto de Química Física Blas Cabrera. CSIC. E-28006. Madrid
[6]Department of Biochemistry and Biomedicine, University of Barcelona, E-08028 Barcelona, Spain

[*]To whom correspondence should be addressed. Email: modesto.orozco@irbbarcelona.org
[†]Equally contributing authors.

Abstract

By combining *in silico*, biophysical, and *in vitro* experiments, we decipher the topology, physical, and potential biological properties of hybrid-parallel nucleic acids triplexes, an elusive structure at the basis of life. We found that hybrid triplex topology follows a stability order: r(Py)-d(Pu)·r(Py) > r(Py)-d(Pu)·d(Py) > d(Py)-d(Pu)·d(Py) > d(Py)-d(Pu)·r(Py). The r(Py)-d(Pu)·d(Py) triplex is expected to be preferred in the cell as it avoids the need to open the duplex reducing the torsional stress required for triplex formation in the r(Py)-d(Pu)·r(Py) topology. Upon a massive collection of melting data, we have created the first predictor for hybrid triplex stability. Leveraging this predictor, we conducted a comprehensive scan to assess the likelihood of the human genome and transcriptome to engage in triplex formation. Our findings unveil a remarkable inclination—of both the human genome and transcriptome—to generate hybrid triplex formation, particularly within untranslated (UTRs) and regulatory regions, thereby corroborating the existence of a triplex-mediated regulatory mechanism. Furthermore, we found a correlation between nucleosome linkers and Triplex-forming sequence (TFS) which agree with a putative role of triplexes in arranging chromatin structure.

Graphical abstract



Received: May 27, 2024. **Revised:** January 18, 2025. **Editorial Decision:** February 12, 2025. **Accepted:** February 24, 2025

Introduction

Triplexes are formed when a polypurine segment of a duplex is recognized by a third oligonucleotide strand (the TFO, triplex forming oligonucleotide) by means of specific hydrogen bond interactions along the major groove (MG) [1–5]. The TFO can be arranged parallel or antiparallel to the purine (Pu) strand. The triads (T-A·T, C^+-G·C and G-G·C) present in parallel triplexes are stabilized by means of Hoogsteen hydrogen bonds, while reverse Hoogsteen hydrogen bond pattern stabilizes triads (A-A·T, G-G·C and T-A·T) in antiparallel triplexes (where "-" refers to Hoogsteen/reverse Hoogsteen and "·" refers to Watson–Crick pairings). Isosteric consideration favors triplexes where the third strand is either homopyrimidine (parallel triplexes; pyrimidine (Py) motif) or homopurine (antiparallel triplexes; purine (Pu) motif) [4–6]. Despite the pH dependence of the C^+-G·C triad, the parallel triplexes are more stable than the antiparallel ones under physiological conditions [6–9].

Early fiber diffraction models suggest an A-type conformation for the DNA triplex [10], but several nuclear magnetic resonance (NMR) experiments and exhaustive molecular dynamics (MD) simulations demonstrated that the DNA triplex shows a "B-like" conformation, with sugars in the South conformation and triads perpendicular to the helix axis [11–18]. The duplex MG is divided by the TFO in two grooves [14, 17, 18]: one very narrow with purine C8 in the bottom (minor Major Groove (mMG) in Fig. 1), and the other very wide (Major Major Groove (MMG) in Fig. 1), covering all the region between the TFO and the third strand of the duplex (Fig. 1). The presence of the TFO blocks the major-groove recognition pattern between transcription factors and the DNA duplex, and while the MMG can be recognized by some proteins [18–20], the general and main effect of triplex formation is the inactivation of DNA transcription, or in some cases the generation of massive mutations in the targeted gene [21–23], the effect being maximized if the triplex is formed in the regulatory regions [4, 21, 24–26]. Very interestingly, promoters in most organisms, including humans, are highly enriched in poly-Pu sequences (triplex target sequences, TTS [27, 28]), suggesting that a large number of genes could be inactivated by triplex formation if a suitable TFO is available [27,29–34]. For example, the long non-coding RNA (lncRNA) HIF1a-AS has been shown to form triplexes with EPHA2 and ADM double stranded DNA (dsDNA) which lead to a decrease of gene expression [35].

The possibility to block DNA expression by adding a DNA-based TFO has been exploited in the so-called antigene therapy [21, 24–26, 36–48]. However, in normal cellular conditions, single-stranded DNA are rare, and putative TFOs are RNA sequences, which act in some cases as regulators of gene expression [31, 49–54]. These experimental findings, combined with bioinformatic analysis [27, 28] suggest the existence of an ancient feedback regulatory mechanism based on the formation of triplexes with TTS in a regulatory gene and TFO defined by the RNA of a regulated gene or an accessory regulatory RNA. Very recently, bioinformatics data have been published showing a significant correlation between human HiC contact maps and regions that can form triplexes with a third strand made of long noncoding RNAs suggesting a correlation between triplex formation and genome structural organization [55, 56].

Understanding the role of hybrid triplexes in gene regulation and chromatin structure first requires a good knowledge of their structural characteristics. Unfortunately, contrary to pure DNA triplexes [1–14, 16–18, 57–59], little and contradictory information exists on the structure of hybrid triplexes [57–60]. We combine here MD simulations and biophysical experiments to explore the stoichiometry, topology, stability, structure, and dynamics of hybrid triplexes. We found a general order of stability: r(Py)-d(Pu)·r(Py) > r(Py)-d(Pu)·d(Py) > d(Py)-d(Pu)·d(Py) > d(Py)-d(Pu)·r(Py), the rest showing little stability, except when the d(Pu) is made of a poly-d(A), in which case the ordering is d(Py)-r(Pu)·r(Py) > d(Py)-d(Pu)·d(Py) > r(Py)-r(Pu)·r(Py) > r(Py)-r(Pu)·d(Py), with little structural differences between the stable hybrid triplexes. We centered our attention in the triplex that is likely to be more prevalent in cellular conditions: r(Py)-d(Pu)·d(Py), developing and validating a stability predictor which allows us to scan for the stability of these triplexes under a range of conditions. Applying this predictor to genomic and transcriptomic data, the likelihood of hybrid triplex formation in human cells is analyzed. A large prevalence of these triplexes is found, being very abundant in regulatory regions (promoters and 5′UTR) and involving mainly miRNAs as TFOs. These findings provide strong support to the hypothesis of an ancient RNA-based triplex-mediated regulatory mechanism. Furthermore, triplexes are located at positions where they can help to fix chromatin structure, both locally and globally.

Materials and methods

Oligonucleotide synthesis and melting experiments

Hairpins **I-XII** were synthesized as previously described [61]. Oligonucleotides **XIII** and **7–9** were synthesized via solid phase synthesis using standard phosphoramidite methods (see Supplementary Methods for details).

Samples containing the required strands were heated to 90°C and slowly cooled down to allow triplex formation in suitable buffers (see Supplementary Methods for details). Melting experiments were performed by heating from low temperature to 100°C at 0.5°C/min, monitoring absorbance at 260 nm every 0.5°C. Experiments were repeated for 5, 8, 12, 18, and 22 μM oligonucleotide concentration to derive melting thermodynamic parameters from Van't Hoff analysis (see Supplementary Methods for details).

NMR spectroscopy

NMR spectra of hairpins were first recorded at a range of temperatures (5–45°C). Later, the TFO was added and the mixture was heated (95°C) and cooled down slowly, collecting spectra in the same temperature range. Spectra were acquired in Bruker spectrometers operating at 600 and 800 Mhz, equipped with cryoprobe and processed with the TOPSPIN software. Water suppression was achieved by including an excitation sculpting module in the pulse sequence [62], see Supplementary Methods for additional details.

Electrophoretic mobility shift assays to analyze triplex formation

25 pmoles of DNA oligonucleotide XIII was labeled with ^{32}P ATP using T4 PNK following the manufacturer protocol. The

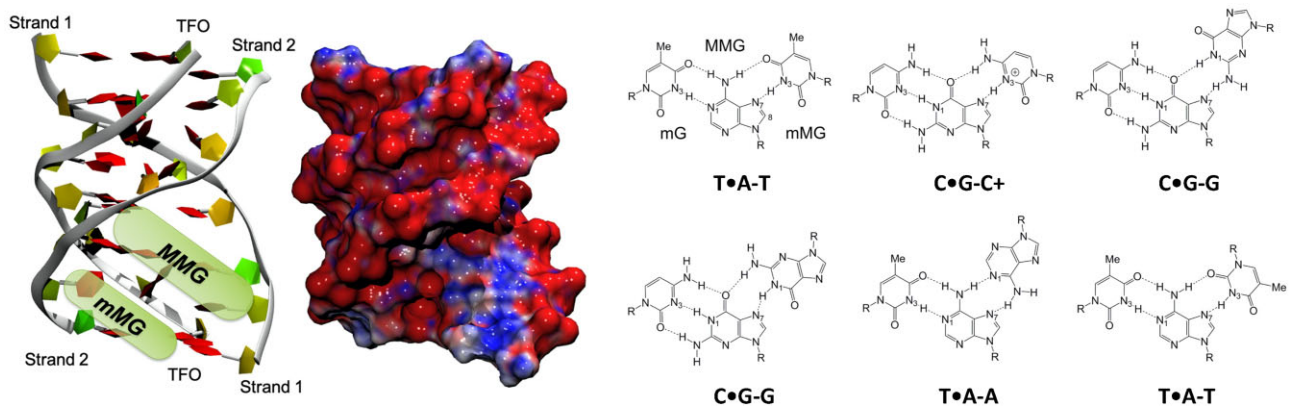


Figure 1. Left panel: Tridimensional representation of a triplex with the two sections of the MG highlighted. Right panel: Schematic representation of T·A-T, C·G-C$^+$, C·G-G, C·G-G, T·A-A, and T·A-T triads in parallel (top row) and antiparallel (bottom row) triplexes.

duplex XIII was formed by heating 32P-labeled DNA XIII at 75°C during 10 min and then quickly cooled on ice. RNA TFO **7** was heated at 75°C for 10 min to prevent self-aggregation and then quickly cooled on ice. Triplex formation was initiated by mixing 2 pmoles of duplex XIII with 2, 20, or 200 pmoles of RNA TFO **7** in RNaseH buffer [20 mM Tris–HCl (pH 7.8), 40 mM KCl, 8 mM $MgCl_2$, and 1 mM DTT) in a final volume of 40 µl. To equilibrate triplex formation, the reaction mixture was incubated at 37°C for 6 h. Then, 0.5 pmoles of triplex was digested with 0.5 units of RNaseH (Thermo Scientific EN0201) or 1 µg of RNaseA (Thermo Scientific EN0531) during 1 h at 37°C. 4 µl of 6× loading buffer (30% glycerol, 0.05% bromophenol, and 0.025% xylene cyanol) was added, and the sample was directly loaded onto a 12% native polyacrylamide gel, prepared in 50 mM Tris-acetate (pH 7.5) and 10 mM $MgCl_2$ buffer. Electrophoresis was performed at 6 V/cm for 14.5 h at 4°C in 50 mM Tris-acetate (pH 7.5) and 10 mM $MgCl_2$ buffer. The gel was dried for 2 h at 64°C and then analyzed by phosphorimaging.

Chromatin preparation and gDNA purification

HeLa cells were grown to 90% confluency in T75 flask with Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. Cells were trypsinized and nuclei were isolated using standard procedures (see Supplementary Methods) and lysed to obtain chromatin, which was then subjected to treatment with Proteinase K and DNAse I to yield fragments with an average size >10 kb (Supplementary Fig. S10A) followed by phenol/chloroform extraction and ethanol precipitation (see Supplementary Methods for details).

In vitro triplex pull-down assay

Purified genomic DNA was sheared into 200–300 bp fragments by sonicating using a Bioruptor Pico (see Supplementary Methods for details). The resulting DNA mixtures were incubated with biotinylated TFO (**8** or **9**) at two different pH values (pH 7.4 and 5.5). TFO-associated DNA was captured by incubation with MyOne Streptavidin DynaBeads followed by treatment with RNase H and elution with suitable buffer (see Supplementary Methods for details).

Parameterization of the nearest neighbor model for RNA–DNA_2 triplex stability

Following Roberts and Crothers [58], we determined the enthalpy of triplex formation by (eq. 1):

$$\Delta H = -\alpha_{cc}(CC) - \alpha_{uc}(UC + CU) - \alpha_{uu}(UU) \quad (1)$$

where (XX) refers to the number or dinucleotide steps of the type XpX in the TFO (CC, UC, CU, or UU), and αs are fitted parameters.

The ΔG is determined as a function of the nucleotide content and the pH (eq. 2):

$$\Delta G(pH) = -\alpha_c(C) - \alpha_u(U) - \alpha'_{cc}(CC) + \delta + (C)(pH - 5.6)\left(\beta - \alpha''_{cc} \cdot (CC)\right) \quad (2)$$

where all symbols in Greek letters are fitted parameters.

From ΔH and ΔG, we can extract the T_m using (eq. 3) [63].

$$Tm = \frac{310 \cdot \Delta H}{\Delta H - \Delta G - 310 \cdot R \cdot ln\left(\frac{4}{C_t}\right)} \quad (3)$$

where R is the ideal gas constant, and C_t is the concentration of the (hairpin) duplex and RNA strands.

The model was parametrized by nonlinear fitting using ΔH and ΔG values obtained from our training set (Supplementary Fig. S8) at different pH and concentrations (see Supplementary Methods for additional details).

Bioinformatics scanning of potential RNA–DNA·DNA formation in humans

We analyzed the triplex potential of annotated lncRNA and microRNA (miRNA) sequences from GENCODE [64] and miRbase [65], respectively. All annotated sequences were scanned with our stability predictor defining potential TFOs with a minimum length of 10 and a maximum of 30 bps. The pH value was set at a default value of 7.0, and the c_{TFO} was set at a value of 12 µM. A T_m of 30°C was set as a threshold to classify stable fragments, while T_m of 45°C was considered to detect highly stable triplexes. In order to detect the formation of potential parallel triplex cores (see Supplementary Methods for additional details), we defined our TTSs as polypurine segments with perfect parallel alignment to the previously found TFOs. The extension of the core was evaluated by the melting predictor (see above) with a penalty equal to 10°C decrease in T_m per mismatch. This is about twice the average decrease

of melting when one nucleotide is eliminated from a triplex based on our melting experiments, which means that we follow a conservative approach to triplex stability, designed to reduce the number of false positives. A more detailed evaluation of mismatches' effects would require a different set-up as that considered in previous studies and a careful calibration of the effect of the mismatch in different triplex environments [66].

The population of potential RNA–DNA·DNA triplexes in the human genome was compared with a random background model. In this model we randomly generated 1 million sequences which followed the base distributions found in the human genome. This allowed us to get a large enough sample for scanning candidate TFOs. We then obtained the target sites from our randomly generated TFOs and used them in the downstream analysis. The potential formation of triplexes in the human genome with our predicted TFOs was validated using two published datasets. First, it was validated with the DNA-associated RNA dataset published by Grummt *et al.* [53] and available in GEO repository under the accession number GSE120849. The data used correspond to the DNA-associated RNA isolated in HeLa cells. The analysis combined the sequences from both the DNA-associated RNA from DNA-IP and SPRI-size selection, which contain 50% and 63% polypyrimidine sequences, respectively. In order to get the percentages of both miRNAs and lncRNAs, we calculated the total number of our candidate TFOs that were found in the isolated RNAs from both samples.

Triplex forming oligonucleotide fragment analysis

Aiming to find the enriched regions where the RNA candidate TFOs would bind, we re-aligned the complementary sequences of our TFOs against the human genome. We used STAR (version 2.5.3a [67]) mapping the candidate TTSs to the hg38 assembly of the human genome. The aligned reads were mapped to the corresponding annotation files and classified accordingly. The annotations for genes, exons, transcripts, and UTR regions were obtained from GENCODE (Release 35 (GRCh38.p13)). Promoters were defined as the regions from the transcription start sites up until 1 kb upstream. Counts for different features at the gene, promoter regions, exon, transcript, and UTR-level resulting from mapping based on the reads were determined separately using the Bioconductor package Rsubread's function featureCounts (version 2.0.1) [68]. To further investigate the role of our TTSs, we mapped the obtained promoter sites to their associated genes and performed a Gene Ontology (GO) Analysis using g:Profiler [69]. A Benjamin–Hochberg False Discovery Rate (FDR) index < 0.05 was set to assess significance corrected from multiple test biases. The GO biological processes of the annotated genes were investigated for terms with size >15 and <2700 in order to avoid mappings to large pathways that are of limited value and increase statistical value when removing small pathways [70].

Triplex formation in the context of chromatin

The nucleosome map in human lymphoblastoid cell lines was obtained analyzing MNase-seq data (accession number: GSE36979) from Gaffney *et al.* [71]. The reads were processed with the nucleR package [72] as follows: mapped fragments were trimmed to 50 bp maintaining the original center and transformed to reads per million. Noise was filtered through Fast Fourier Transform, keeping 2% of the principal components, and peak calling was performed using the following parameters: peak width: 147 bp, peak detection threshold: 35%, maximum overlap of 45 bp, dyad length: 60 bp. The lncRNA and miRNAs expressed in lymphoblastic cells were obtained from RNA-seq and small RNA-seq data (accession numbers: E-MATB-8300 and E-MTAB-8301) [73].

Structural models and molecular dynamics simulations

Starting conformations for the six triplexes considered here: r(Py)-d(Pu)·r(Py), r(Py)-d(Pu)·d(Py), d(Py)-d(Pu)·d(Py), d(Py)-d(Pu)·r(Py), r(Py)-r(Pu)·r(Py), and r(Py)-r(Pu)·d(Py) were built from DNA triplex structures [17, 18, 74, 75]. In order to maximize potential unfolding events, most simulations were done using 8-mer triplexes, but control simulations were performed with 18-mer triplexes. Systems were solvated with waters, neutralized with Na^+ adding 100 mM additional NaCl. The size of the final triclinic box was ∼75 Å × 75 Å × 75 Å for 8-mer and 100 Å × 100 Å × 100 Å for 18-mer, respectively. Simulation systems were optimized and slowly heated and equilibrated for 50 ns prior to production replicas 3 × 500 ns in the isothermal isobaric ensemble (NPT; $T = 35$ °C and $P = 1$ atm). In order to improve sampling temperature, replica exchange simulations (RexMD) were done (8 replicas in the range 30–65°C) with individual trajectories extending for at least 400 ns in the NVT ensemble. For well known technical problems, no Mg^{2+} was introduced in the calculations, which should decrease the stability of the triplexes, favoring again the existence of unfolding events. Finally a rough estimate of the impact of crowding effect of chromatin, one of the triplexes (the 8-mer r(Py)-d(Pu)·d(Py) one) was simulated for 0.5 μs in a box containing c.a. 20 mM of DNA duplex (8 mer) with Na^+ neutralizing the system plus 100 mM NaCl. A control simulation using identical conditions was done using the pure DNA d(Py)-d(Pu)·d(Py) triplex.

Long-range electrostatic interactions were calculated with the particle mesh Ewald method (PME) with a real space cutoff of 12 Å and periodic boundary conditions in the three directions of Cartesian space were used [76]. Constant temperature was imposed using Langevin dynamics [77] with a damping coefficient of 1 ps, while pressure was maintained with Langevin-Piston dynamics [78] with a 200 fs decay period and a 50 fs time constant. LINCS [79] was used to maintain covalent bonds at equilibrium distance, allowing the use of 2 fs integration step. Parmbsc1 was used to describe DNA interactions [80–82], while RNA was described using the chi OL3 force-field [83]. Water molecules were represented by the TIP3P [84] model, while ions were modeled by Dang's parameters [85].

All MD simulations were performed using *AMBER24* code [86]. Coordinates of the systems were collected every 5 ps of the production trajectory. Analyses were carried out using AMBER analysis tools, PyMOL [87], Curves+ [88] and NAFlex [89], and BIGNAsim analysis tools [90, 91]. Trajectories were stored in our BIGNAsim database [90, 91] following FAIR data standards as described elsewhere [91].

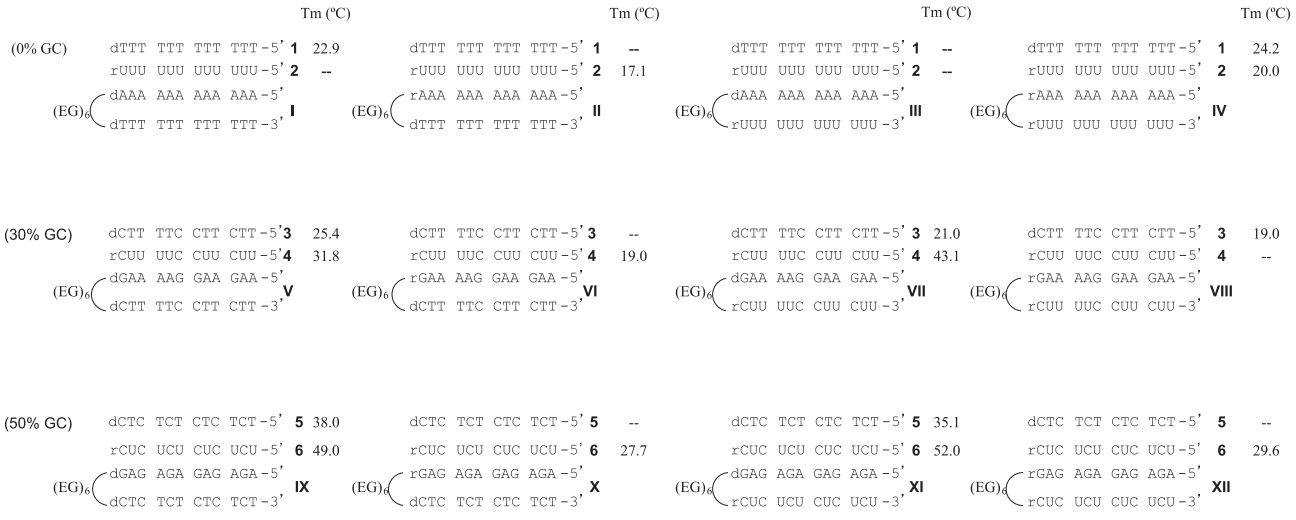


Figure 2. Melting temperatures of d(Py)-d(Pu)·d(Py), r(Py)-d(Pu)·d(Py), d(Py)-r(Pu)·d(Py), r(Py)-r(Pu)·d(Py), d(Py)-d(Pu)·r(Py), r(Py)-d(Pu)·r(Py), d(Py)-r(Pu)·r(Py), and r(Py)-r(Pu)·r(Py) 12mer triplexes of 100%, 70%, and 50% A·T/U content in 10 mM sodium cacodylate buffer (pH 6.0) containing 100 mM NaCl and 10 mM $MgCl_2$ (see "Materials and methods" section for details). Only the T_m of the triplex transition is indicated in all cases. For results at higher pH see below.

Results

Stability of homo-polymers and hybrid triplexes

Melting experiments were first performed using different homopyrimidine triplexes as TFO and homopurine–homopyrimidine hairpins as TTSs. The use of hairpins (polyethylene glycol was used as loop) has the advantage to minimize the formation of other competing structures in the TTS, such as reverse Watson–Crick [92, 93], Hoogsteen duplexes [94–98], quadruplexes, and others [99–101], which will introduce noise in the T_m estimates. We consider three compositions of the hairpin: (100% A·T/U (**I–IV**), 70% A·T/U (**V–VIII**), and 50% A·T/U (**IX–XII**)), we do not consider higher percentages of guanines as this will increase the risk of quadruplex formation. With these compositions, we create all the combinations of DNA and RNA in the hairpin: d(Pu)·d(Py) (**I, V,** and **IX**), r(Pu)·d(Py) (**II, VI**, and **X**), d(Pu)·r(Py) (**III, VII,** and **XI**), and r(Pu)·r(Py) (**IV, VIII,** and **XII**) and incubate them with the corresponding homopyrimidine TFO (DNA with 100%, 70%, or 50% T [1, 3, 5]; RNA with 100%, 70%, or 50% U [2, 4, 6]. Combination of all TFOs with all TTS leads to 24 potential triplexes whose stability was measured by the corresponding melting curves (recorded in all cases at pH 6.0; see "Materials and methods" section). Results (Fig. 2) show melting temperatures in the range $T < 15$°C (not detectable) to 52°C. Triplexes with 100% A·T/U show in general a poor stability with a decreasing order of stability d(Py)-r(Pu)·r(Py)> d(Py)-d(Pu)·d(Py)> r(Py)-r(Pu)·r(Py)> r(Py)-r(Pu)·d(Py), the rest being not detectable (Fig. 2). When the ratio of G·C increases, the triplexes become more stable, showing a quite well-defined order of stability (Fig. 2): r(Py)-d(Pu)·r(Py)> r(Py)-d(Pu)·d(Py)> d(Py)-d(Pu)·d(Py)> d(Py)-d(Pu)·r(Py)> r(Py)-r(Pu)·d(Py) ≈ r(Py)-r(Pu)·r(Py). At the studied pH, the increase in the ratio G·C/A·T implies an increase in the stability of the triplex. For a given G·C/A·T ratio, the two most stable triplexes are those with RNA in the TFO with d(Pu)·r(Py) preferred over d(Pu)·d(Py) in the TTS, the difference being reduced as the percentage of G/C in the TTS increases. Interestingly, the triplex formed by a single RNA recognizing the genomic duplex (r(Py)-d(Pu)·d(Py)) is very stable, even more than the canonical d(Py)-d(Pu)·d(Py) one.

Our melting experiments can be qualitatively compared by stability estimates obtained by Brown's group using gel electrophoresis using U/A rich duplexes as TTS and shorter pyrimidine strands as TFO. Due to the used experimental set-up they were not able to detect all possible RNA/DNA topologies in the triplex [66, 102], but their results agree well with ours: r(Py)-r(Pu)·r(Py) is very unstable, and both r(Py)-d(Pu)·d(Py) and d(Py)-d(Pu)·d(Py) show a similar stability. This is consistent with our results from a triplex with c.a. 9/1 ratio between U and C in the TFO.

In order to confirm that melting experiments were really analyzing triplex→duplex transitions instead of other processes related to strand invasion with R-loop formation, we repeated melting experiments for triplexes d(Py)-d(Pu)·d(Py), r(Py)-d(Pu)·d(Py), d(Py)-d(Pu)·r(Py), and r(Py)-d(Pu)·r(Py) for the case of 70% A·T/U in the TTS at higher pH values (from 6.0 to 8.0; see Fig. 3A–H) finding a reduction of T_m, consistent with triplex formation. The triplex nature of the structures was further confirmed by means of ^{1}H NMR spectroscopy. Thus, ^{1}H-NMR spectra of several hairpins and their equimolar mixes with their corresponding Hoogsteen strands were recorded at different temperatures (see Fig. 3I and Supplementary Figs S1A–4A). In all the hairpin spectra at $T = 5$ °C, the number of imino signals in the 12–13 ppm region and ~14 ppm is consistent with the formation of the expected number of GC and AT/U WC base pairs. As temperature increases, the imino signal intensity decreases. Upon addition of the third strands, the number of exchangeable proton signals increases drastically. At low temperature, NMR signals exhibit a notable broadening, an effect that is more pronounced in triplexes containing more RNA strands. Despite this, the formation of GC Hoogsteen base pairs is demonstrated by the observation of protonated cytosine imino signals ~15 ppm and their corresponding amino signals ~10 ppm. Additional imino signals can be observed

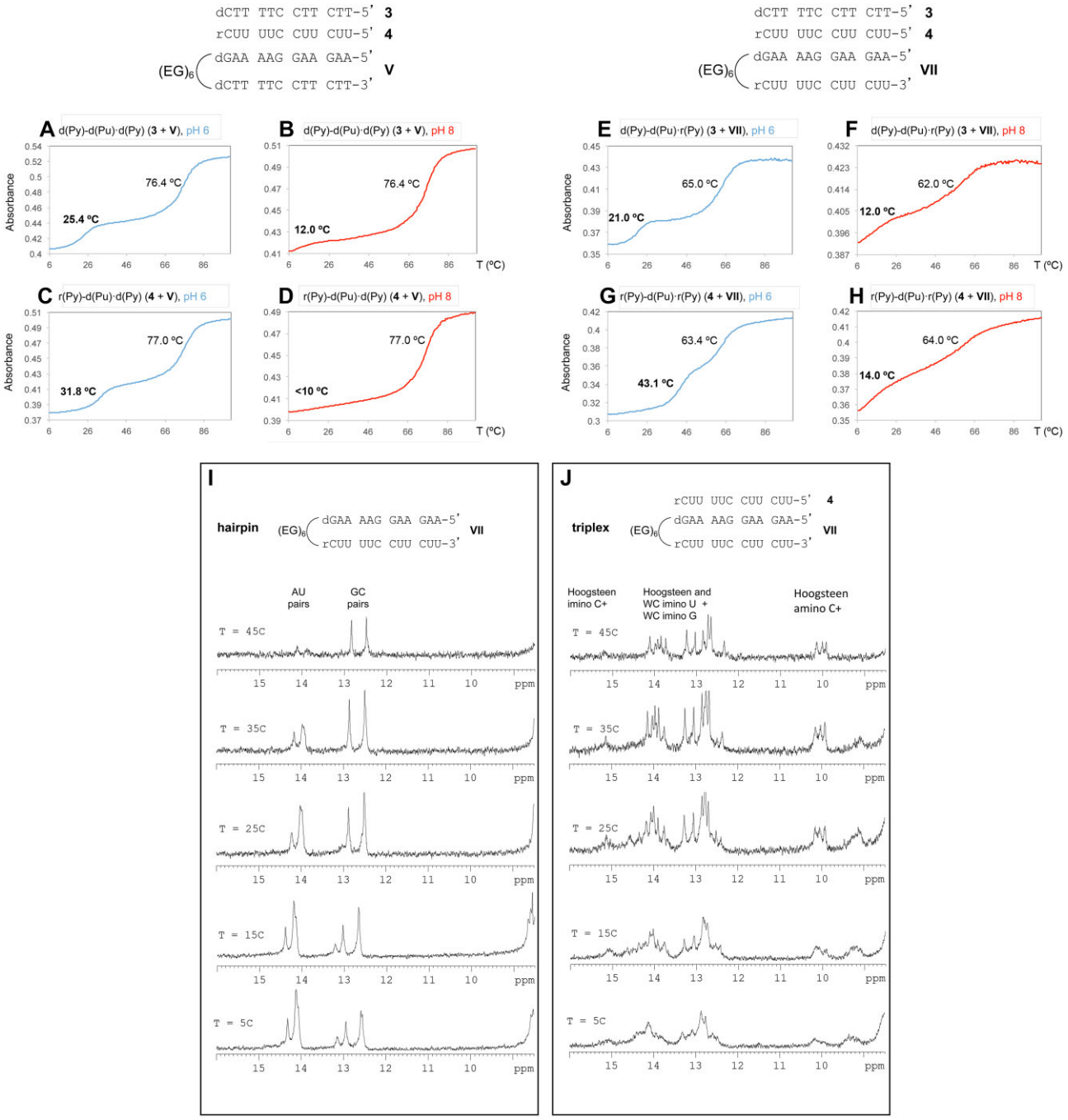


Figure 3. (**A–H**) Thermal stability of d(Py)-d(Pu)·d(Py) (**3** + **V**), r(Py)-d(Pu)·d(Py) (**4** + **V**), d(Py)-d(Pu)·r(Py) (**3** + **VII**), and r(Py)-d(Pu)·r(Py) (**4** + **VII**) 12mer triplexes in 10 mM sodium cacodylate buffer (pH 6.0) containing 100 mM NaCl and 10 mM $MgCl_2$ (A, C, E, and G) and in 10 mM sodium cacodylate buffer (pH 8.0) containing 100 mM NaCl and 10 mM $MgCl_2$ (B, D, F, and H). Melting temperatures (T_m) for the duplex and triplex are indicated in each case (in bold: T_m corresponding to the triplex). (**I** and **J**) ^{1}H NMR spectra of the imino region of D·R hairpin **VII** (**I**) and r(Py)-d(Pu)·r(Py) triplex **4** + **VII** (**J**) acquired at 5, 15, 25, 35, and 45°C in 30 mM phosphate buffer (pH 6.0) containing 100 mM NaCl and 10 mM $MgCl_2$.

in the 13–14 ppm region, supporting the formation of extra AU base pairs. Many of these signals persist at 45°C, indicating the formation of a very stable structure (Fig. 3J and Supplementary Figs S1B–4B). Overall, the NMR spectra of all the mixes are consistent with formation of the expected parallel triplexes depicted in the figures' inserts. Overall, our experiments explain apparently contradictory previous data, agreeing with Crother's estimates [103] obtained for 66% GC triplexes and with Dervan's data [60, 103] collected for 81% AT triplexes.

Very interestingly, hybrid triplexes d(Py)-d(Pu)·r(Py) are quite unstable compared to the r(Py)-d(Pu)·d(Py) ones, which suggests that triplexes with a 2:1 (DNA:RNA) stoichiometry show a topology r(Py)-d(Pu)·d(Py). Furthermore, triplexes r(Py)-d(Pu)·r(Py), which (to our knowledge) were never explored before are very stable. However, as its formation requires strand invasion, untwisting of the DNA, and the generation of an unpaired d(Py) strand, we expect them to be unlikely in the cell, except in the context of very active transcription, where the formation of R-loop can facilitate them.

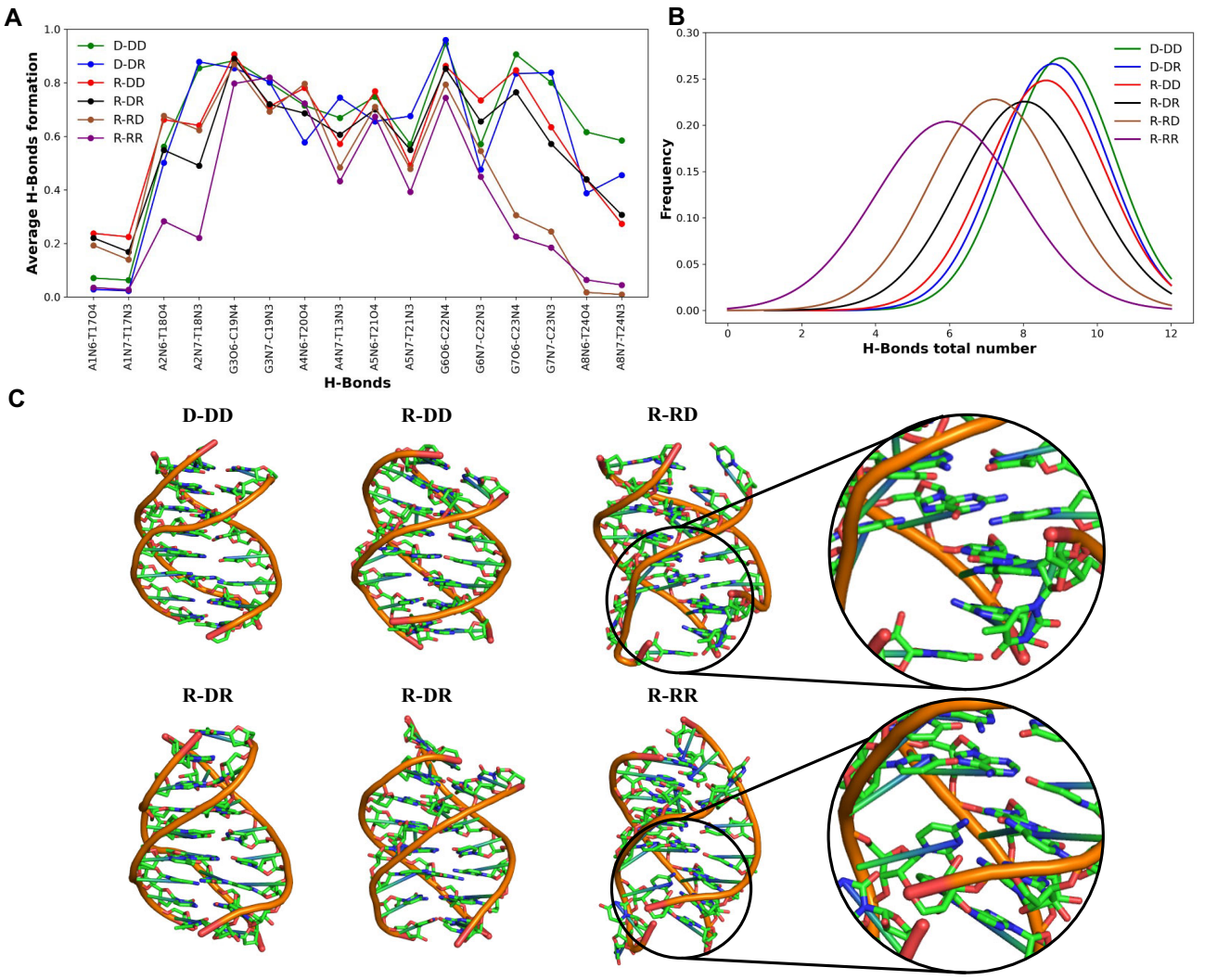


Figure 4. Selected results of the MD simulation of different hybrid triplexes. Values are obtained at 35°C and are the average of three independent replicas (total simulation time 1.5 μs) (**A**). Average Hoogsteen H-bonds along the sequence of the different 8-mer triplexes, TFO sequence: 5′-TTCTTCCT-3′; (**B**). Histograms of the Hoogsteen H-bonds on the central 6-mer of the triplexes. (**C**) Examples of structures sampled at the end of the simulation for the different triplexes are shown in panel.

Further work is required to verify this exciting possibility. At this time, our results suggest that r(Py)-d(Pu)·d(Py), with TTS being the genomic DNA and the TFO being expressed RNAs will be the most stable triplex topology in the cell.

The structure and dynamics of the hybrid triplexes

To gain structural and mechanistic insights on the structure and stability of hybrid triplexes, we performed a set of extensive MD simulations (see "Materials and methods" section) of six triplexes in aqueous solution: r(Py)-d(Pu)·r(Py), r(Py)-d(Pu)·d(Py), d(Py)-d(Pu)·d(Py), d(Py)-d(Pu)·r(Py), r(Py)-r(Pu)·r(Py), and r(Py)-r(Pu)·d(Py). In the stable portion of the trajectories, all hybrid triplexes show similar structures, close to the canonical DNA triplex [10–17, 61], with the DNA and RNA strands showing South and North puckerings respectively. As expected (see "Materials and methods" section) by the length of the triplexes, the temperature, and the absence of Mg^{2+}, significant unfolding is detected, which becomes very evident in the disruption of hydrogen bonds, especially the Hoogsteen ones (see Fig. 4). Better conservation of H-bonds is found for the Watson–Crick pairings (see Supplementary Fig. S5), in good agreement with the expected three state unfolding mechanism of triplexes (see Fig. 3).

Very interestingly, the loss of H-bond is not the same for all the triplexes, which help us to obtain qualitative information about their relative stability. Thus, the r(Py)-r(Pu)·d(Py), and specially the r(Py)-r(Pu)·r(Py) triplex shows a massive distortion of the geometry of the Hoogsteen strand, which in fact, remains attached to the duplex just by a few G–C$^+$ interactions (see Fig. 4). This agrees with the very poor stability found for these triplexes in our experimental measurements (Fig. 2), as well as with previous experimental data from Brown's group [102], supporting the reliability of our simulations [61, 80–82]. The remaining triplexes show a similar stability with the triplex with potential impact in gene regulation r(Py)-d(Pu)·d(Py) appearing quite stable (Fig. 4), again in good agreement with our experimental measurements and with previous data by Brown's group [102].

The use of short triplexes offered us the possibility to detect significant distortions in short simulations but raises con-

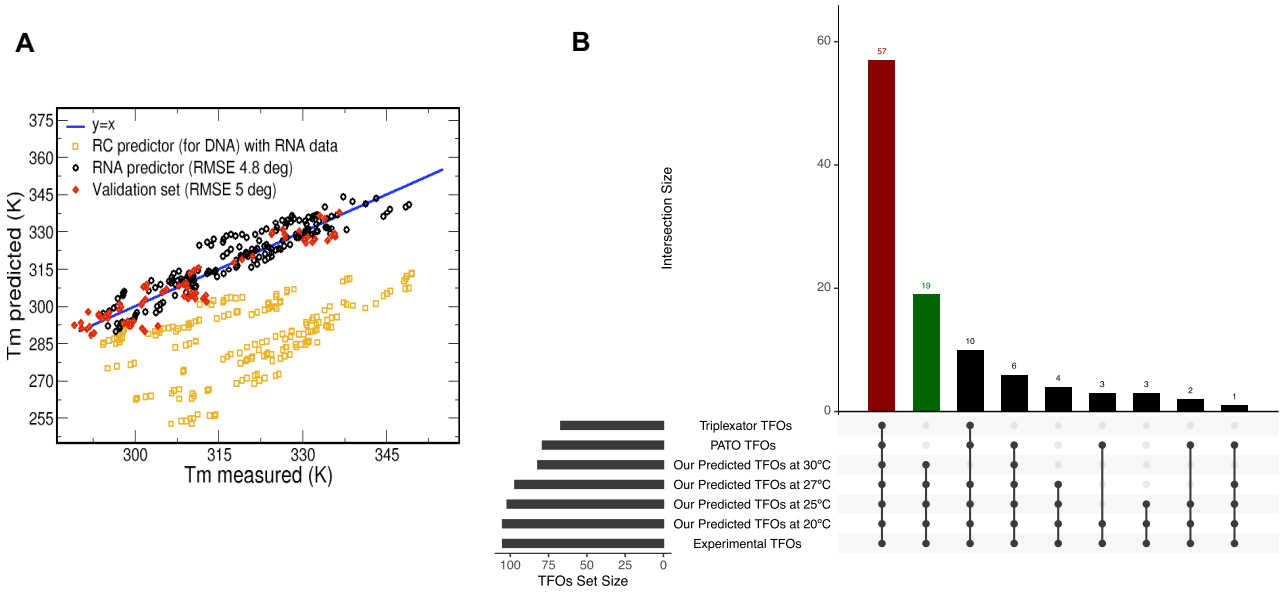


Figure 5. (**A**) Predicted versus measured melting temperatures of triplexes. Values that correspond to estimates obtained using Roberts–Crothers method for triplexes (in yellow). Values of the data for the training set and data for a completely independent validation set (in black and red respectively) (see "Materials and methods" section for details). (**B**) Predicted TFOs from our model in comparison to Triplexator and PATO in the evaluation of stable triplexes at various temperatures. Results are shown as an intersecting upset plot, where the red bar denotes the candidate sequences from each method and the green bar denotes the sequences only recapitulated by our predictor.

cerns about the value of the results for longer oligos. Thus, we decided to explore 18-mer triplexes, using RexMD to increase sampling, and analyzing data at one of the highest temperatures (65°C) explored, which allowed us to enrich the simulation of partially unfolding events. Results in Supplementary Fig. S6 confirms those obtained for shorter oligos, with r(Py)-r(Pu)·d(Py) and r(Py)-r(Pu)·r(Py) triplexes showing poor stability, as expected the differences are narrower with respect to those found for shorter oligos. Again, the triplex of interest r(Py)-d(Pu)·d(Py) shows a stability similar to that of the reference d(Py)-d(Pu)·d(Py) one and much higher than that of the r(Py)-r(Pu)·r(Py) one.

Final concerns arise on the applicability of our dilute simulations in the context of chromatin, where the triplex should face a very crowded environment, which is experimentally suggested to stabilize it [104]. It is very difficult to define a realistic model of the nuclei environment, as inert crowders might be a bad mimic of cellular environments [105], so we decide to create a dense nucleic acid environment (compatible with that existing in human nuclei) by surrounding the triplex by 8-mer duplexes with their associated cationic environment (see "Materials and methods" section). Results suggest that the r(Py)-d(Pu)·d(Py) triplex remains stable in crowded environments, with a maintenance of the structure that is even larger than that found in dilute conditions (see Supplementary Fig. S7).

Development and validation of a predictor of the stability of r(Py)-d(Pu)·d(Py) triplexes

As discussed above, the most stable triplex (r(Py)-d(Pu)·r(Py)) is not expected to have a large prevalence in the cell out of R-loop constructs. However, the second most stable triplex: r(Py)-d(Pu)·d(Py) can be easily formed by pairing an RNA segment with genomic DNA. Following Robert and Crothers approach [58], we trained a simple nearest neighbor model for r(Py)-d(Pu)·d(Py) to reproduce experimental data in a variety of triplexes (see "Materials and methods" section and Supplementary Figs S8 and S9) in several conditions. The refined method predicts experimental melting observables with root mean square errors around: 4.8 degrees (T_m), and 0.7 kcal/mol (melting free energy), improving dramatically the accuracy obtained by transferring Roberts–Crothers DNA triplex method (see Fig. 5A). Our predictions also outperform the widely used PATO and Triplexator softwares [106, 107] which are unable to detect all stable triplexes at a given temperature; see green bar at Fig. 5B.

To further validate the predictive power of our model, we designed a 50 nt polypyrimidine TFO (TFO **7**; Fig. 6A) which, according to our method, should form stable triplexes (T_m values = 57°C at pH 6.5 and 48°C at pH 7.0) in the promoter region of the BRD7 gene. As shown in Fig. 6A, synthetic TFO **7** interacts with a radiolabeled synthetic double-stranded DNA hairpin (**XIII**) comprising the target BRD7 polypurine sequence, forming a low-mobility complex resistant to RNaseH but sensitive to RNaseA. To validate the triplex nature of this complex, chromatin extracted from HeLa cells (see "Materials and methods" section for details) was treated with DNase I and proteinase K and sonicated to yield fragments of 200–300 nt (see Supplementary Fig. S10). Two aliquots of this purified DNA were incubated with TFO **8**, a biotinylated version of TFO **7**, at pH 5.5 and 7.0 in the presence of RNase H to digest putative R-loops, or r(Py)-d(Pu)·r(Py) triplexes. The streptavidin-retained DNA was eluted and identified by qPCR amplification, finding significant DNA recovery when using BRD7 promoter-specific primers amplifying a 92 nt region just around the target triplex region (92Nt primers; Fig. 6B, red panel; Supplementary Fig. S11 and Supplementary Table S1). Interestingly, a decrease in the pH led to an increase in DNA recovery, which is consistent with pH-dependent stability in C-C·G triplex formation as captured by our predictor.

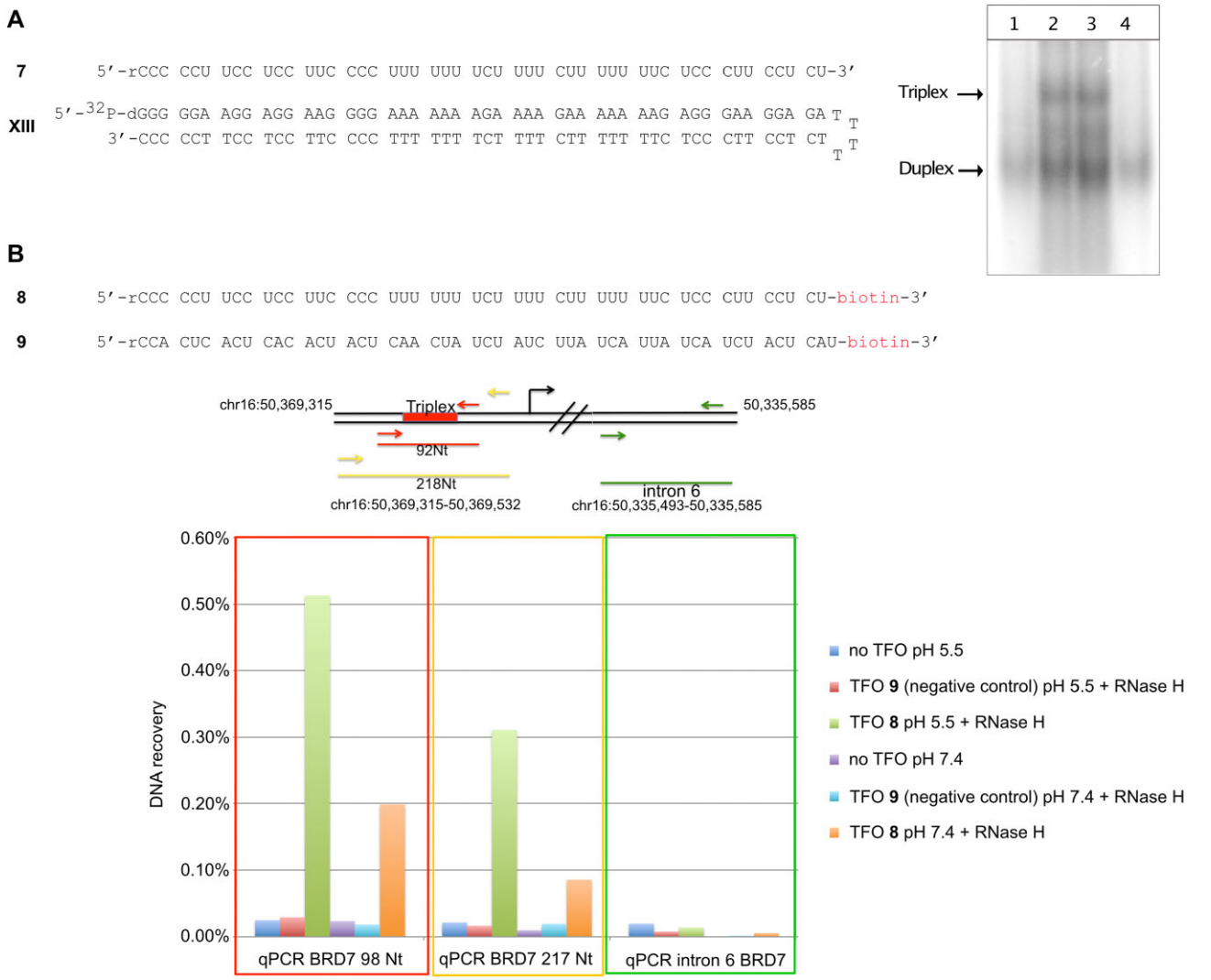


Figure 6. (**A**) Electrophoretic mobility shift assays to analyze triplex formation between TFO **7** and hairpin **XIII** before (lane 2) and after digestion with RNaseH (lane 3) or RNaseA (lane 4). Lane 1 represents the folded hairpin **XIII** without TFO **7**. (**B**) Biotinylated TFOs used in this study (**8**: TFO targeting the polypurine A/G site of BRD7 promoter region (chr16:50 369 387–50 369 436); **9**: negative control). Schematic of human chr16 from nucleotide 50 369 315 to 50 335 585 depicting the position of the target triplex forming region and the primers used for DNA amplification. Upon binding to streptavidin beads, associated DNA was analyzed by qPCR using promoter-specific [92 NT (in red) or 217 NT (in yellow)] or intronic-specific (in green) primers.

Similar results were obtained when using promoter-specific primers amplifying a larger region (217 nt) around the target triplex region (217Nt primers; Fig. 6B, yellow panel). On the contrary, no recovery was observed when using intronic-specific primers (Fig. 6B, green panel), or when genomic DNA fragments were incubated with a TFO lacking the sequence matching the target region (TFO **9**), confirming the specificity of the above-described results and the triplex nature of the complex predicted by our model.

Potential formation of r(Py)-d(Pu)·d(Py) triplexes in human cells

We used our predictor to screen for potential TFOs amongst annotated human lncRNAs and miRNAs from the gencode and miRbase [64, 65] databases respectively (see "Materials and methods" section). We found a strong enrichment of TFO candidates (triplex $T_m > 30°$) in both dataset in comparison to the population of expected TTSs from a random distribution (see random model in "Materials and methods" section) (Fig. 7, and Supplementary Tables S2 and S3), suggesting potential r(Py)-d(Pu)·d(Py) triplex formation *in vivo*. In order to further validate our results, we decided to compare our predicted candidates with two published studies [53, 108]. When compared with the DNA-associated RNA isolated by Grummt *et al.* [53], we found that 44% or our predicted TFOs from lncRNAs and 51% from miRNAs were indeed found associated with DNA in a triplex structure. In addition, when we intersected our predicted TFOs with those found by Maldonado *et al.* [108], 35% of our predicted miRNA candidates and 31% of the lncRNA candidates were found in this published dataset. The overall distribution of distances to an *in vivo* triplex from this study is clustered at 0bp with distances mostly <100 bps (see Supplementary Fig. S12).

Both sets of TTS (from miRNA and lncRNA TFOs) were mapped to the human genome, and we observed an over-representation in promoters when analyzing stable triplexes ($T_m > 30°C$), and in 5′UTR (in the case of miRNA's

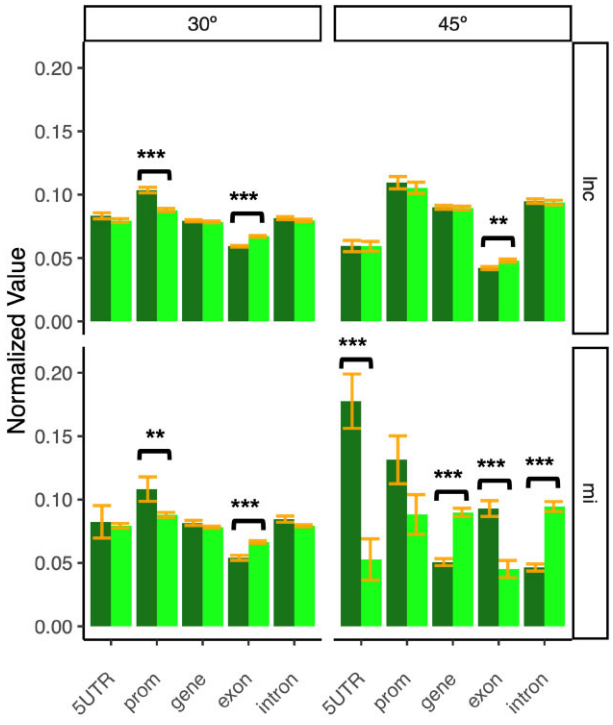


Figure 7. Overall representation across different genomic classes in lncRNAs (upper plots: dark green) and miRNAs (lower plots: dark green) against a random background model (light green), with the corresponding standard error bars (orange). RNA–DNA·DNA triplexes are shown as stable at 30°C (left-most plots) or 45°C (right-most plots) in the different regions. When a *P*-value is <0.05, it is flagged with one star (*). If a *P*-value is ≤0.01, it is flagged with two stars (**). If a *P*-value is ≤0.001 it is flagged with three stars (***).

TFO) when analyzing very stable triplexes (T_m > 45°C), suggesting that parallel r(Py)-d(Pu)·d(Py) triplexes form preferentially in regions important for the control of gene expression.

GO analysis of the genes potentially controlled by RNA–DNA·DNA triplex formation with miRNAs and lncRNAs showed that these genes are frequently related to complex processes, such as development (see Supplementary Fig. S13), with very significant hits in the development of the nervous system. It is tempting to speculate that stable triplexes generated by the binding of transcribed RNAs with genomic DNA can be involved in a fine-tuning regulatory mechanism, which was inherited from an ancient triplex-mediated DNA–RNA regulatory network. Note that this finding agrees well with the work from Pasquier *et al.* in Drosophila that showed that the genes targeted by TFOs were involved in development and morphogenesis [109].

Final comment on role in chromatin structure

To investigate triplex formation in the context of chromatin, we predicted the putative TFOs from lncRNA and miRNAs expressed in lymphoblastic cells [73] and compared the location of their target sites along the human genome with a genome-wide map of nucleosome occupancy in human lymphoblastoid cell line [71]. We observed a local minimum which coincides with the nucleosome dyads (Fig. 8), suggesting a correlation between chromatin accessibility and triplex formation. These results which agree with previous findings by Maldonado *et al.* [108] showed that triplexes could form away from the dyad, at the entry–exit site of the nucleosome, helping to fix the nucleosome array and in the case of very long lncRNA helping to approach in the space distant regions as suggested by Marti *et al.* [55].

Discussion

A variety of parallel triplexes can be formed mixing complementary DNA and RNA strands, and a significant number of them can be stable under physiological conditions as predicted by state-of-the-art atomistic MD simulations and confirmed by melting and NMR experiments. We have studied here the preferred parallel triplexes, where the third strand is polypyrimidine and runs parallel to the DNA polypurine strand of the DNA duplex. We consider all possible combinations of DNA and RNA strands leading to six different triplexes whose relative stability was explored by melting experiments. In general, for a 50% A/G duplex composition, the r(Py)-d(Pu)·r(Py) (i.e. a polypyrimidine RNA as TFO and a hybrid DNA(Pur)·RNA(Pyr) as TTS) leads to the most stable structures, followed very closely by the r(Py)-d(Pu)·d(Py) triplex and the canonical DNA homotriplex (d(Py)-d(Pu)·d(Py)). Other triplexes, such as the RNA homotriplex (r(Py)-d(Pu)·d(Py)) appears unstable in agreement with previous studies by Brown's group [66]. In cellular conditions, r(Py)-d(Pu)·r(Py) is expected to be disfavored, except perhaps in situations of high transcriptional stress, due to the need to disrupt the DNA duplex and the intrinsic instability of an unpaired pyrimidine DNA strand. On the contrary, the r(Py)-d(Pu)·d(Py) triplex can be easily formed without the need for disruption of the DNA duplex, taking as TFO an expressed RNA sequence complementary with the purine strand of the duplex and shows stability in dilute and (at least *in silico*) crowded conditions similar to those at the nuclei.

A massive experimental effort allowed us to develop the first predictor for hybrid parallel triplexes. The method could be improved by incorporating mismatches (now very penalized), longer triplexes and different conditions closer to those at nuclei, but even with its current simplicity it shows quantitative accuracy and a speed that allows genome-scale analysis. The predictor was used to determine all the potential parallel triplexes in human lnc and miRNAs (TFO: expressed RNAs and TTS: genomic DNA). Despite the conservative threshold used and the penalty assumed for mismatches, calculations show a very large number of possible stable triplexes, much more than those predicted by random models. Potential triplexes are concentrated in regulatory regions and UTRs, quite interestingly in genes that are related to the development, morphology and functioning of central nervous system, suggesting a potential role of triplexes in a RNA←→DNA mediated regulatory network. This work also suggests that, despite the fact that miRNAs are commonly known as post transcriptional regulators, their nuclear function as transcription regulators via triplex formation is more widespread than at first thought [110]. Furthermore, mapping potential triplex formation with chromatin structure, we found evidence suggesting a role of triplex formation in fixing nucleosome arrays probably protecting nucleosome from eviction and in the case of lncRNA helping, as suggested by others, to compact chromatin [55].

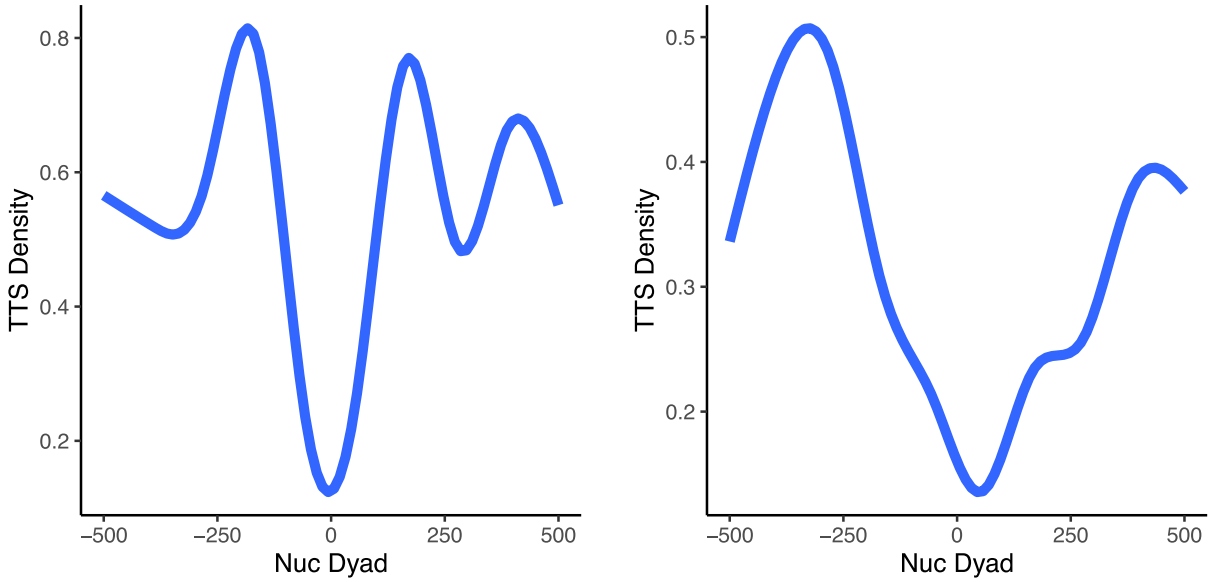


Figure 8. TTS densities from candidate TFOs centered at nucleosome dyads for lncRNAs (left) and miRNAs (right). The nucleosome maps are obtained from lymphoblastoid cells [71] and TFOs originate from lncRNAs and miRNAs expressed in lymphoblastic cells.

Acknowledgements

VG thanks the European Molecular Biology Organization (EMBO) for financial support (ALTF 103-2018) and "Juan De La Cierva Fellowship" (IJC2019-040468-I/25A04100). MO thanks Spanish Ministry of Science [PID2021-122478NB-I00]; BioExcel-3: Centre of Excellence for Computational Biomolecular Research [European Union: 101093290; Ministerio de Ciencia e Innovación: PCI2022-134976-2]; Instituto de Salud Carlos III – Instituto Nacional de Bioinformatica, Fondo Europeo de Desarrollo Regional [ISCIII PT 17/0009/0007]; European Regional Development Fund, ERFD Operative Programme for Catalunya, the Catalan Government AGAUR [SGR2021 00863]. The IRB Barcelona is the recipient of a Severo Ochoa Award of Excellence from the MINECO.The authors thankfully acknowledge the NMR resources and the technical support provided by the LRB and LMR of the Spanish ICTS Red de Laboratorios de RMN de Biomoléculas (R-LRB).

Author contributions: Conceptualization: MT, IBH, MO; Data curation: AS, GP; Formal Analysis: VG, GP, AS, ISC, PA, IBH; Funding acquisition: MO; Investigation: VG, GP, AS, MT, ISC, JG, NV, LM, CC, ML, AA, AH, AG, PA, IBH, CG, RE; Methodology: GP, MT; Project administration: IBH, MO; Resources: MO; Software: GP, AS, AH; Supervision: IBH, CG, RE, MO; Visualization: GP, MT, ISC, IBH; Writing – original draft: VG, MT, MO; Writing – review & editing: AS, ISC, IBH, MO.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

Funding to pay the Open Access publication charges for this article was provided by the European Molecular Biology Organization (EMBO) (ALTF 103-2018) and "Juan De La Cierva Fellowship" (IJC2019-040468-I/25A04100) to V.G. M.O. thanks Spanish Ministry of Science [RTI2018-096704-B-100]; European Research Council (ERC SimDNA), MINECO Severo Ochoa Award of Excellence (Government of Spain) (awarded to IRB Barcelona); the Biomolecular and Bioinformatics Resources Platform (ISCIII PT 13/000/0030 co-funded by the Fondo Europeo de Desarrollo Regional [FEDER]) and the H2020 BioExcel Center of Excellence.

Data availability

All relevant data supporting the key findings of this study are available within the article and the Supplementary Information. The MD simulations generated and analyzed during the current study are available in the open access MDDB repository (https://irb.mddbr.eu/) under the following accession "Hybrid triplex 2025." The code for the bioinformatics stability analyses are available in Zenodo under DOI:10.5281/zenodo.14882595.

References

1. Felsenfeld G, Rich A. Studies on the formation of two- and three-stranded polyribonucleotides. *Biochim Biophys Acta* 1957;**26**:457–68. https://doi.org/10.1016/0006-3002(57)90091-4
2. Frank-Kamenetskii MD, Mirkin SM. Triplex DNA structures *Annu Rev Biochem* 1995;**64**:65–95. https://doi.org/10.1146/annurev.bi.64.070195.000433
3. Pauling L, Corey RB. A proposed structure for the nucleic acids. *Proc Natl Acad Sci USA* 1953;**39**:84–97. https://doi.org/10.1073/pnas.39.2.84

4. Potaman VN, Sinden RR. Stabilization of triple-helical nucleic acids by basic oligopeptides. *Biochemistry* 1995;**34**:14885–92. https://doi.org/10.1021/bi00045a033
5. Robles J, Grandas A, Pedroso E *et al.* Nucleic acid triple helices: stability effects of nucleobase modifications. *COC* 2002;**6**:1333–68. https://doi.org/10.2174/1385272023373482
6. Waring MJ. *DNA-targeting Molecules as Therapeutic Agents*. The Royal Society of Chemistry, London, 2018.
7. Chandler SP, Fox KR. Specificity of antiparallel DNA triple helix formation. *Biochemistry* 1996;**35**:15038–48. https://doi.org/10.1021/bi9609679
8. Jaumot J, Avina A, Eritja R *et al.* Resolution of parallel and antiparallel oligonucleotide triple helices formation and melting processes by multivariate curve resolution. *J Biomol Struct Dyn* 2003;**21**:267–78. https://doi.org/10.1080/07391102.2003.10506922
9. Scaria PV, Shafer RH. Calorimetric analysis of triple helices targeted to the d(G3A4G3).d(C3T4C3) duplex. *Biochemistry* 1996;**35**:10985–94. https://doi.org/10.1021/bi960966g
10. Arnott S, Bond PJ, Selsing E *et al.* Models of triple-stranded polynucleotides with optimised stereochemistry. *Nucleic Acids Res* 1976;**3**:2459–70. https://doi.org/10.1093/nar/3.10.2459
11. Bornet O, Lancelot G. Solution structure of a selectively 13C-labeled intramolecular DNA triplex. *J Biomol Struct Dyn* 1995;**12**:803–14. https://doi.org/10.1080/07391102.1995.10508777
12. Howard FB, Miles HT, Liu K *et al.* Structure of d(T)n.d(A)n.d(T)n: the DNA triple helix has B-form geometry with C2'-endo sugar pucker. *Biochemistry* 1992;**31**:10671–7. https://doi.org/10.1021/bi00159a005
13. Macaya RF, Schultze P, Feigon J. Sugar conformations in intramolecular DNA triplexes determined by couping constants obtained by automated simulation of P.COSY cross peaks. *J Am Chem Soc* 1992;**114**:781–3. https://doi.org/10.1021/ja00028a067
14. Radhakrishnan I, Patel DJ. Solution structure and hydration patterns of a pyrimidine.purine.pyrimidine DNA triplex containing a novel T.CG base-triple. *J Mol Biol* 1994;**241**:600–19. https://doi.org/10.1006/jmbi.1994.1534
15. Radhakrishnan I, Patel DJ. Hydration sites in purine.purine.pyrimidine and pyrimidine.purine.pyrimidine DNA triplexes in aqueous solution. *Structure* 1994;**2**:395–405. https://doi.org/10.1016/S0969-2126(00)00041-1
16. Raghunathan G, Miles HT, Sasisekharan V. Symmetry and molecular structure of a DNA triple helix: d(T)n.d(A)n.d(T)n. *Biochemistry* 1993;**32**:455–62. https://doi.org/10.1021/bi00053a009
17. Shields GC, Laughton CA, Orozco M. Molecular dynamics simulations of the d(T·A·T) triple helix. *J Am Chem Soc* 1997;**119**:7463–9. https://doi.org/10.1021/ja970601z
18. Soliva R, Laughton CA, Luque FJ *et al.* Molecular dynamics simulations in aqueous solution of triple helices containing d(G·C·C) trios. *J Am Chem Soc* 1998;**120**:11226–33. https://doi.org/10.1021/ja981121q
19. Guieysse AL, Praseuth D, Helene C. Identification of a triplex DNA-binding protein from human cells. *J Mol Biol* 1997;**267**:289–98. https://doi.org/10.1006/jmbi.1997.0884
20. Jimenez-Garcia E, Vaquero A, Espinas ML *et al.* The GAGA factor of Drosophila binds triple-stranded DNA. *J Biol Chem* 1998;**273**:24640–8. https://doi.org/10.1074/jbc.273.38.24640
21. Giovannangeli C, Hélène C. Triplex technology takes off *Nat Biotechnol* 2000;**18**:1245–1246. https://doi.org/10.1038/82348
22. Giovannangeli C, Helene C. Triplex-forming molecules for modulation of DNA information processing. *Curr Opin Mol Ther* 2000;**2**:1245–6.
23. Vasquez KM, Narayanan L, Glazer PM. Specific mutations induced by triplex-forming oligonucleotides in mice. *Science* 2000;**290**:530–3. https://doi.org/10.1126/science.290.5491.530
24. Bacolla A, Wang G, Vasquez KM. New Perspectives on DNA and RNA Triplexes as effectors of biological activity *PLoS Genet* 2015;**11**:1–12. https://doi.org/10.1371/journal.pgen.1005696
25. Buske FA, Mattick JS, Bailey TL. Potential in vivo roles of nucleic acid triple-helices. *RNA Biol* 2011;**8**:427–439. https://doi.org/10.4161/rna.8.3.14999
26. Knauert MP, Glazer PM. Triplex forming oligonucleotides: sequence-specific tools for gene targeting. *Hum Mol Genet* 2001;**10**:2243–2251. https://doi.org/10.1093/hmg/10.20.2243
27. Goni JR, de la Cruz X, Orozco M. Triplex-forming oligonucleotide target sequences in the human genome. *Nucleic Acids Res* 2004;**32**:354–60. https://doi.org/10.1093/nar/gkh188
28. Goni JR, Vaquerizas JM, Dopazo J *et al.* Exploring the reasons for the large density of triplex-forming oligonucleotide target sequences in the human regulatory regions. *BMC Genomics* 2006;**7**:63. https://doi.org/10.1186/1471-2164-7-63
29. Ali T, Rogala S, Krause NM *et al.* Fendrr synergizes with wnt signalling to regulate fibrosis related genes during lung development via its RNA:dsDNA triplex element. *Nucleic Acids Res* 2023;**51**:6227–37. https://doi.org/10.1093/nar/gkad395
30. Mondal T, Subhash S, Vaid R *et al.* MEG3 long noncoding RNA regulates the TGF-beta pathway genes through formation of RNA-DNA triplex structures. *Nat Commun* 2015;**6**:7743. https://doi.org/10.1038/ncomms8743
31. Postepska-Igielska A, Giwojna A, Gasri-Plotnitsky L *et al.* LncRNA Khps1 regulates expression of the proto-oncogene SPHK1 via triplex-mediated changes in chromatin structure. *Mol Cell* 2015;**60**:626–36. https://doi.org/10.1016/j.molcel.2015.10.001
32. Trembinski DJ, Bink DI, Theodorou K *et al.* Aging-regulated anti-apoptotic long non-coding RNA Sarrah augments recovery from acute myocardial infarction. *Nat Commun* 2020;**11**:2039. https://doi.org/10.1038/s41467-020-15995-2
33. Yari H, Jin L, Teng L *et al.* LncRNA REG1CP promotes tumorigenesis through an enhancer complex to recruit FANCJ helicase for REG3A transcription. *Nat Commun* 2019;**10**:5334. https://doi.org/10.1038/s41467-019-13313-z
34. Leisegang MS, Warwick T, Stotzel J *et al.* RNA-DNA triplexes: molecular mechanisms and functional relevance. *Trends Biochem Sci* 2024;**49**:532–44. https://doi.org/10.1016/j.tibs.2024.03.009
35. Leisegang MS, Bains JK, Seredinski S *et al.* HIF1alpha-AS1 is a DNA:DNA:RNA triplex-forming lncRNA interacting with the HUSH complex. *Nat Commun* 2022;**13**:6563. https://doi.org/10.1038/s41467-022-34252-2
36. Alvarez-Salas LM. Nucleic acids as therapeutic agents. *CTMC* 2008;**8**:1379–404. https://doi.org/10.2174/156802608786141133
37. Besch R, Giovannangeli C, Schuh T *et al.* Characterization and quantification of triple helix formation in chromosomal DNA. *J Mol Biol* 2004;**341**:979–89. https://doi.org/10.1016/j.jmb.2004.05.079
38. Conde J, Oliva N, Atilano M *et al.* Self-assembled RNA-triple-helix hydrogel scaffold for microRNA modulation in the tumour microenvironment. *Nature Mater* 2016;**15**:353–63. https://doi.org/10.1038/nmat4497
39. Cooney M, Czernuszewicz G, Postel EH *et al.* Site-specific oligonucleotide binding represses transcription of the human c-myc gene in vitro. *Science* 1988;**241**:456–9. https://doi.org/10.1126/science.3293213
40. Devi G, Zhou Y, Zhong Z *et al.* RHA triplexes: from structural principles to biological and biotech applications. *Wiley Interdisciplinary Reviews: RNA* 2015;**6**:111–128.
41. Duval-Valentin G, Thuong NT, Helene C. Specific inhibition of transcription by triple helix-forming oligonucleotides. *Proc Natl Acad Sci USA* 1992;**89**:504–8. https://doi.org/10.1073/pnas.89.2.504
42. Fox KR In: Sugimoto N (ed.), *Handbook of Chemical Biology of Nucleic Acids*. Springer Nature Singapore, Singapore, 2023, pp. 2613–43.

43. Grigoriev M, Praseuth D, Robin P *et al.* A triple helix-forming oligonucleotide-intercalator conjugate acts as a transcriptional repressor via inhibition of NF kappa B binding to interleukin-2 receptor alpha-regulatory sequence. *J Biol Chem* 1992;**267**:3389–95. https://doi.org/10.1016/S0021-9258(19)50743-8
44. Guntaka RV, Varma BR, Weber KT. Triplex-forming oligonucleotides as modulators of gene expression *Int J Biochem Cell Biol* 2003;**35**:22–31. https://doi.org/10.1016/S1357-2725(02)00165-6
45. Joseph J, Kandala JC, Veerapanane D *et al.* Antiparallel polypurine phosphorothioate oligonucleotides form stable triplexes with the rat alpha1(I) collagen gene promoter and inhibit transcription in cultured rat fibroblasts. *Nucleic Acids Res* 1997;**25**:2182–8. https://doi.org/10.1093/nar/25.11.2182
46. Postel EH, Flint SJ, Kessler DJ *et al.* Evidence that a triplex-forming oligodeoxyribonucleotide binds to the c-myc promoter in HeLa cells, thereby reducing c-myc mRNA levels. *Proc Natl Acad Sci USA* 1991;**88**:8227–31. https://doi.org/10.1073/pnas.88.18.8227
47. Rogers FA, Lloyd JA, Glazer PM. Triplex-forming oligonucleotides as potential tools for modulation of gene expression. *CMCACA* 2005;**5**:319–326. https://doi.org/10.2174/1568011054222300
48. van Dongen MJ, Doreleijers JF, van der Marel GA *et al.* Structure and mechanism of formation of the H-y5 isomer of an intramolecular DNA triple helix. *Nat Struct Biol* 1999;**6**:319–26.
49. Grote P, Wittler L, Hendrix D *et al.* The tissue-specific lncRNA fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell* 2013;**24**:206–14. https://doi.org/10.1016/j.devcel.2012.12.012
50. Li T, Mo X, Fu L *et al.* Molecular mechanisms of long noncoding RNAs on gastric cancer. *Oncotarget* 2016;**7**:8601–12. https://doi.org/10.18632/oncotarget.6926
51. Li Y, Syed J, Sugiyama H. RNA-DNA Triplex formation by Long Noncoding RNAs. *Cell Chem Biol* 2016;**23**:1325–1333.
52. Martianov I, Ramadass A, Serra Barros A *et al.* Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* 2007;**445**:666–70. https://doi.org/10.1038/nature05519
53. Senturk Cetin N, Kuo CC, Ribarska T *et al.* Isolation and genome-wide characterization of cellular DNA:RNA triplex structures. *Nucleic Acids Res* 2019;**47**:2306–21. https://doi.org/10.1093/nar/gky1305
54. Sridhar B, Rivas-Astroza M, Nguyen TC *et al.* Systematic mapping of RNA–chromatin interactions *in vivo*. *Curr Biol* 2017;**27**:602–9. https://doi.org/10.1016/j.cub.2017.01.011
55. Farabella I, Di Stefano M, Soler-Vila P *et al.* Three-dimensional genome organization via triplex-forming RNAs. *Nat Struct Mol Biol* 2021;**28**:945–54. https://doi.org/10.1038/s41594-021-00678-3
56. Soibam B, Zhamangaraeva A. LncRNA:DNA triplex-forming sites are positioned at specific areas of genome organization and are predictors for topologically associated domains. *BMC Genomics* 2021;**22**:397. https://doi.org/10.1186/s12864-021-07727-7
57. Mergny JL, Sun JS, Rougee M *et al.* Sequence specificity in triple-helix formation: experimental and theoretical studies of the effect of mismatches on triplex stability. *Biochemistry* 1991;**30**:9791–8. https://doi.org/10.1021/bi00104a031
58. Roberts RW, Crothers DM. Prediction of the stability of DNA triplexes. *Proc Natl Acad Sci USA* 1996;**93**:4320–5. https://doi.org/10.1073/pnas.93.9.4320
59. Singleton SF, Dervan PB. Equilibrium association constants for oligonucleotide-directed triple helix formation at single DNA sites: linkage to cation valence and concentration. *Biochemistry* 1993;**32**:13171–9. https://doi.org/10.1021/bi00211a028
60. Han H, Dervan PB. Sequence-specific recognition of double helical RNA and RNA.DNA by triple helix formation. *Proc Natl Acad Sci USA* 1993;**90**:3806–10. https://doi.org/10.1073/pnas.90.9.3806
61. Terrazas M, Genna V, Portella G *et al.* The origins and the biological consequences of the pur/pyr DNA·RNA asymmetry. *Chem* 2019;**5**:1619–31. https://doi.org/10.1016/j.chempr.2019.04.002
62. Stott K, Keeler J, Hwang TL *et al.* Excitation Sculpting in High-Resolution Nuclear Magnetic Resonance Spectroscopy: Application to Selective NOE Experiments. *J Am Chem Soc* 1995;**117**:4199–4200 . https://doi.org/10.1021/ja00119a048
63. Marky LA, Breslauer KJ. Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* 1987;**26**:1601–20. https://doi.org/10.1002/bip.360260911
64. Frankish A, Diekhans M, Ferreira AM *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* 2019;**47**:D766–73. https://doi.org/10.1093/nar/gky955
65. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res* 2019;**47**:D155–62. https://doi.org/10.1093/nar/gky1141
66. Kunkler CN, Schiefelbein GE, O'Leary NJ *et al.* A single natural RNA modification can destabilize a U*A-T-rich RNA*DNA-DNA triple helix. *RNA* 2022;**28**:1172–84. https://doi.org/10.1261/rna.079244.122
67. Dobin A, Davis CA, Schlesinger F *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**:15–21. https://doi.org/10.1093/bioinformatics/bts635
68. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* 2019;**47**:e47–. https://doi.org/10.1093/nar/gkz114
69. Raudvere U, Kolberg L, Kuzmin I *et al.* g:profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* 2019;**47**:W191–8. https://doi.org/10.1093/nar/gkz369
70. Reimand J, Isserlin R, Voisin V *et al.* Pathway enrichment analysis and visualization of omics data using g:profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc* 2019;**14**:482–517. https://doi.org/10.1038/s41596-018-0103-9
71. Gaffney DJ, McVicker G, Pai AA *et al.* Controls of nucleosome positioning in the human genome. *PLoS Genet* 2012;**8**:e1003036. https://doi.org/10.1371/journal.pgen.1003036
72. Flores O, Orozco M. nucleR: a package for non-parametric nucleosome positioning. *Bioinformatics* 2011;**27**:2149–50. https://doi.org/10.1093/bioinformatics/btr345
73. Florian RT, Kraft F, Leitao E *et al.* Unstable TTTTA/TTTCA expansions in MARCH6 are associated with familial Adult myoclonic Epilepsy type 3. *Nat Commun* 2019;**10**:4919. https://doi.org/10.1038/s41467-019-12763-9
74. Gotfredsen CH, Schultze P, Feigon J. Solution structure of an intramolecular pyrimidine−purine−pyrimidine triplex containing an RNA third strand. *J Am Chem Soc* 1998;**120**:4281–9. https://doi.org/10.1021/ja973221m
75. Ruszkowska A, Ruszkowski M, Hulewicz JP *et al.* Molecular structure of a U*A-U-rich RNA triple helix with 11 consecutive base triples. *Nucleic Acids Res* 2020;**48**:3304–14. https://doi.org/10.1093/nar/gkz1222
76. Darden T, York D, Pedersen L. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J Chem Phys* 1993;**98**:10089–92. https://doi.org/10.1063/1.464397
77. Grest GS, Kremer K. Molecular dynamics simulation for polymers in the presence of a heat bath. *Phys Rev A* 1986;**33**:3628–31. https://doi.org/10.1103/PhysRevA.33.3628
78. Feller SE, Zhang Y, Pastor RW *et al.* Constant pressure molecular dynamics simulation: the Langevin piston method. *J Chem Phys* 1995;**103**:4613–21. https://doi.org/10.1063/1.470648
79. Hess B, Bekker H, Berendsen HJC *et al.* LINCS: a linear constraint solver for molecular simulations. *J Comput Chem*

1997;**18**:1463–72. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12%3c1463::AID-JCC4%3e3.0.CO;2-H
80. Dans PD, Ivani I, Hospital A *et al.* How accurate are accurate force-fields for B-DNA?*Nucleic Acids Res* 2017;**45**:4217–30.
81. Dans PD, Walther J, Gómez H *et al.* Multiscale simulation of DNA *Curr Opin Struct Biol* 2016;**37**:29–45. https://doi.org/10.1016/j.sbi.2015.11.011
82. Ivani I, Dans PD, Noy A *et al.* Parmbsc1: a refined force field for DNA simulations. *Nat Methods* 2016;**13**:55–8. https://doi.org/10.1038/nmeth.3658
83. Zgarbova M, Otyepka M, Sponer J *et al.* Refinement of the Cornell *et al*. Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J Chem Theory Comput* 2011;**7**:2886–902. https://doi.org/10.1021/ct200162x
84. Jorgensen WL, Chandrasekhar J, Madura JD *et al.* Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;**79**:926–35. https://doi.org/10.1063/1.445869
85. Dang LX. Mechanism and thermodynamics of ion selectivity in aqueous solutions of 18-crown-6 ether: a molecular dynamics study. *J Am Chem Soc* 1995;**117**:6954–60. https://doi.org/10.1021/ja00131a018
86. Páll S, Abraham MJ, Kutzner C *et al.* *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, Vol. **8759**.Springer International Publishing. Cham.
87. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph* 1996;**14**, 33-38:27–38.
88. Blanchet C, Pasi M, Zakrzewska K *et al.* CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res* 2011;**39**:W68–73. https://doi.org/10.1093/nar/gkr316
89. Hospital A, Faustino I, Collepardo-Guevara R *et al.* NAFlex: a web server for the study of nucleic acid flexibility. *Nucleic Acids Res* 2013;**41**:W47–55. https://doi.org/10.1093/nar/gkt378
90. Hospital A, Andrio P, Cugnasco C *et al.* BIGNASim: a NoSQL database structure and analysis portal for nucleic acids simulation data. *Nucleic Acids Res* 2016;**44**:D272–8. https://doi.org/10.1093/nar/gkv1301
91. Hospital A, Battistini F, Soliva R *et al.* Surviving the deluge of biosimulation data. *Wiley Interdiscipl Rev* 2020;**10**:e1449
92. Cubero E, Luque FJ, Orozco M. Theoretical studies of d(A:t)-based parallel-stranded DNA duplexes. *J Am Chem Soc* 2001;**123**:12018–25. https://doi.org/10.1021/ja011200t
93. Otto C, Thomas GA, Rippe K *et al.* The hydrogen-bonding structure in parallel-stranded duplex DNA is reverse Watson–Crick. *Biochemistry* 1991;**30**:3062–9. https://doi.org/10.1021/bi00226a012
94. Abrescia NG, Thompson A, Huynh-Dinh T *et al.* Crystal structure of an antiparallel DNA fragment with Hoogsteen base pairing. *Proc Natl Acad Sci USA* 2002;**99**:2806–11. https://doi.org/10.1073/pnas.052675499
95. Aishima J, Gitti RK, Noah JE *et al.* A Hoogsteen base pair embedded in undistorted B-DNA. *Nucleic Acids Res* 2002;**30**:5244–5252. https://doi.org/10.1093/nar/gkf661
96. Cubero E, Abrescia NG, Subirana JA *et al.* Theoretical study of a new DNA structure: the antiparallel Hoogsteen duplex. *J Am Chem Soc* 2003;**125**:14603–12. https://doi.org/10.1021/ja035918f
97. Cubero E, Avino A, de la Torre BG *et al.* Hoogsteen-based parallel-stranded duplexes of DNA. Effect of 8-amino-purine derivatives. *J Am Chem Soc* 2002;**124**:3133–42. https://doi.org/10.1021/ja011928+
98. Radwan MM, Wilson HR. Fibre and molecular structure of thymidylyl-3′,5′-deoxyadenosine. *Int J Biol Macromol* 1982;**4**:145–9. https://doi.org/10.1016/0141-8130(82)90043-5
99. Brahms S, Brahms J, Van Holde KE. Nature of conformational changes in poly[d(A-T)-d(A-T)] in the premelting region. *Proc Natl Acad Sci USA* 1976;**73**:3453–7. https://doi.org/10.1073/pnas.73.10.3453
100. Cubero E, Luque FJ, Orozco M. Theoretical study of the Hoogsteen–Watson–Crick junctions in DNA. *Biophys J* 2006;**90**:1000–8. https://doi.org/10.1529/biophysj.105.059535
101. Spiegel K, Rothlisberger U, Carloni P. Duocarmycins binding to DNA investigated by molecular simulation. *J Phys Chem B* 2006;**110**:3647–60. https://doi.org/10.1021/jp0548265
102. Kunkler CN, Hulewicz JP, Hickman SC *et al.* Stability of an RNA*DNA-DNA triple helix depends on base triplet composition and length of the RNA third strand. *Nucleic Acids Res* 2019;**47**:7213–22. https://doi.org/10.1093/nar/gkz573
103. Roberts RW, Crothers DM. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science* 1992;**258**:1463–6. https://doi.org/10.1126/science.1279808
104. Endoh T, Tateishi-Karimata H, Sugimoto N. In: Sugimoto N (ed.), *Handbook of Chemical Biology of Nucleic Acids*. Singapore: Springer Nature Singapore, 2023, 1127–71.
105. Candotti M, Orozco M. The differential response of proteins to macromolecular crowding. *PLoS Comput Biol* 2016;**12**:e1005040. https://doi.org/10.1371/journal.pcbi.1005040
106. Amatria-Barral I, Gonzalez-Dominguez J, Tourino J. PATO: genome-wide prediction of lncRNA–DNA triple helices. *Bioinformatics* 2023;**39**:btad134. https://doi.org/10.1093/bioinformatics/btad134
107. Buske FA, Bauer DC, Mattick JS *et al.* Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res* 2012;**22**:1372–81. https://doi.org/10.1101/gr.130237.111
108. Maldonado R, Schwartz U, Silberhorn E *et al.* Nucleosomes stabilize ssRNA–dsDNA triple helices in Human cells. *Mol Cell* 2019;**73**:1243–1254.e6. https://doi.org/10.1016/j.molcel.2019.01.007
109. Pasquier C, Agnel S, Robichon A. The mapping of predicted triplex DNA:RNA in the drosophila genome reveals a prominent location in development- and morphogenesis-related genes. *G3* 2017;**7**:2295–304. https://doi.org/10.1534/g3.117.042911
110. Paugh SW, Coss DR, Bao J *et al.* MicroRNAs form triplexes with double stranded DNA at sequence-specific binding sites; a eukaryotic mechanism via which microRNAs could directly alter gene expression. *PLoS Comput Biol* 2016;**12**:e1004744. https://doi.org/10.1371/journal.pcbi.1004744

Received: May 27, 2024. **Revised:** January 18, 2025. **Editorial Decision:** February 12, 2025. **Accepted:** February 24, 2025