Understanding intercalative modulation of G-rich sequence folding: solution structure of a TINA-conjugated antiparallel DNA triplex

Miguel Garavís [1], Patrick J.B. Edwards[2], Israel Serrano-Chacón [1], Osman Doluca [2], Vyacheslav V. Filichev[2,*] and Carlos González [1,*]

[1]Instituto de Química Física 'Blas Cabrera', (IQF-CSIC), Madrid 28006, Spain
[2]School of Natural Sciences, Massey University, Palmerston North 4412, New Zealand

*To whom correspondence should be addressed. Tel: +34 915619400; Email: cgonzalez@iqfr.csic.es
Correspondence may also be addressed to Vyacheslav V. Filichev. Tel: +64 69517659; Email: v.filichev@massey.ac.nz
Present addresses:
Israel Serrano-Chacón, Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology (BIST), 08028, Barcelona, Spain.
Osman Doluca, Faculty of Engineering, Department of Biochemical Engineering. Izmir University of Economics, Izmir 35330, Turkey.

Abstract

We present here the high-resolution structure of an antiparallel DNA triplex in which a monomer of *para*-twisted intercalating nucleic acid (*para*-TINA: (*R*)-1-*O*-[4-(1-pyrenylethynyl)phenylmethyl]glycerol) is covalently inserted as a bulge in the third strand of the triplex. TINA is a potent modulator of the hybridization properties of DNA sequences with extremely useful properties when conjugated in G-rich oligonucleotides. The insertion of *para*-TINA between two guanines of the triplex imparts a high thermal stabilization ($\Delta T_M = 9°C$) to the structure and enhances the quality of NMR spectra by increasing the chemical shift dispersion of proton signals near the TINA location. The structural determination reveals that TINA intercalates between two consecutive triads, causing only local distortions in the structure. The two aromatic moieties of TINA are nearly coplanar, with the phenyl ring intercalating between the flanking guanine bases in the sequence, and the pyrene moiety situated between the Watson–Crick base pair of the two first strands. The precise position of TINA within the triplex structure reveals key TINA–DNA interactions, which explains the high stabilization observed and will aid in the design of new and more efficient binders to DNA.

Graphical abstract
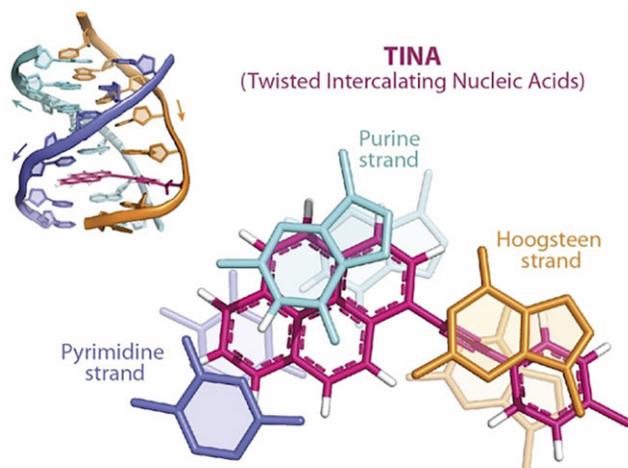
Introduction

Triplex DNA is formed by Hoogsteen base-pairing of a third strand to the purine tract of a DNA duplex (1–4). In the last years, triplex DNA have been detected inside cells through in-cell NMR (5) and using specific triplex antibodies (6–9). This, together with the identification of proteins that bind triplex (10–12), suggests that its formation occurs *in vivo* and has biological significance. The overrepresentation of putative triplex forming sequences in regulatory regions of the genome (13–15) suggests that triplex DNA structures may play roles in gene expression and DNA metabolism. Additionally, triplex DNA formation has been associated with

double strand breaks (DSB) occurring during replication (15,16), which has attracted the interest on these structures as possible tools for novel gene editing technologies such as CRISPR-Cas (17–19).

The formation of a DNA triplex by using exogenous triplex-forming oligonucleotides (TFO) targeting specific regions of a gene has been widely used as a potent strategy to modulate gene expression (4,20–25) and to direct or induce site-specific effects such us damage, mutagenesis or genetic instability on targeted DNA sequences (4,26–28). Interestingly, most of the annotated human genes contain at least one unique poly-purine sequence that could serve as TFO target site (29). On the other hand, triplex formation constitutes the basis of novel nanotechnological devices with multiple applications (30).

Parallel triplex formation by using pyrimidine-rich TFOs can be limited at physiological pH as protonation of cytosines -and thus acidic pH- is required to form Hoogsteen base pairs with the guanines of the target duplex (31) (Figure 1A). On the contrary, guanine-rich TFOs, which form antiparallel triplexes appear as more adequate TFO candidates as they are insensitive to pH (Figure 1A). However, the tendency of these sequences to self-fold into G-quadruplex structures in the presence of monovalent cations (K⁺ and Na⁺) present in the cell context, can completely abolish triplex formation (32,33).

Several strategies have been devised to overcome these limitation (31,34). One approach involves the use of modulators that alter the hybridisation properties of DNA sequences, such as intercalating nucleic acids (Figure 1B). Within this category, the *para*-twisted intercalating nucleic acid monomer (*para*-TINA: (R)-1-O-[4-(1-pyrenylethynyl)phenylmethyl]glycerol) (Figure 1B) inserted as a bulge into the TFO (Figure 1C), has demonstrated remarkable efficacy in circumventing guanine-mediated self-association of G-rich TFOs, particularly when positioned in the middle of the G-tract (35,36). In addition, for cellular applications it is important that complexes of TINA-containing TFOs with complementary RNAs were significantly destabilised, which makes TINA–TFOs useful molecules targeting genomic DNA (37,38) and not RNA. TINA was designed to provide stabilization to the triplex through stacking interactions with duplex and TFO bases simultaneously. Thus, it was conceived containing two aromatic platforms; a pyrene ring and a phenyl ring aimed to intercalate between two consecutive duplex base pairs and TFO bases, respectively. Additionally, the presence of a linear triple bond connection between both rings provides rotational flexibility allowing the aromatic rings to be in different planes (38). Several studies have shown that TINA increases thermodynamic stability of both parallel (38,39) and antiparallel (35,36) DNA triplexes without compromising sequence-specificity, which agrees with other pyrene-containing DNA probes (40,41). Importantly, TINA-containing TFOs have been proven efficient in regulation of gene expression (36). In contrast, *ortho*-TINA (Figure 1B) in which 1-ethynylpyrene is attached to the *ortho* position of the benzene ring stabilized Watson–Crick-type duplexes and had minimal influence on parallel triplexes (42). The behaviour of *ortho*-TINA was mimicking original intercalating nucleic acids (INA) which stabilized antiparallel duplexes (43) but destabilized triplexes (44). Although some models of *para*-TINA intercalation in duplex and parallel triplex have been published, no high-resolution studies have been reported so far. Here, we use NMR methods to determine the detailed three-dimensional structure of an antiparallel intramolecular triplex with a single insertion of a *para*-

TINA monomer in the Hoogsteen strand (Figure 1D). Since the structural information on triplex is very limited, we focused on one of the few sequences whose structure has been determined (45). To facilitate the assignment of the NMR spectra, the original five-thymine loops were substituted by six ethylene glycol (EG6) linkers (Figure 1D).

Material and methods

Oligonucleotide synthesis and purification

Oligonucleotide synthesis was performed on a Mer-Made-4 automated DNA synthesizer under conditions described previously (35).

CD and UV experiments

Circular dichroism (CD) spectra were recorded on a Jasco J-815 spectropolarimeter. UV spectra were recorded on a Jasco V-730 spectrophotometer. Both instruments are fitted with Peltier thermoelectric cooling modules. Experiments were conducted in a buffer containing 25 mM sodium phosphate and 100 mM NaCl at pH 7. Samples were initially heated at 90°C for 5 min and slowly allowed to cool to room temperature and stored at 4°C until use. Samples were placed in quartz cuvettes with 1 cm pathlength. UV melting curves were recorded at the wavelength of 256 nm, with a heating rate of 0.5°C·min⁻¹. Uncertainties in T_M values are estimated to be ±0.5°C. CD spectra were recorded at various temperatures after 5 min equilibration. Thermodynamic analysis of UV-melting curves was performed according to the published protocol (46) assuming intramolecular equilibrium which is supported by the presence of isosbestic points in CD spectra at various temperatures. Melting point was determined as a temperature in which half of the complex was melted. Complete details regarding the determination of thermodynamic parameters from UV-melting profiles are provided in the Supplementary Data.

NMR experiments

Samples for NMR experiments were suspended in 200 µl of either D₂O or H₂O/D₂O 9:1 in a buffer containing 25 mM sodium phosphate and 100 mM NaCl, pH 7. Final oligonucleotide concentration was around 0.5 mM. NMR spectra were acquired on Bruker Avance spectrometers equipped with cryoprobes and operating at 600 or 800 MHz. NMR data was processed with Bruker Topspin software. TOCSY spectra were recorded with the standard MLEV17 spinlock sequence and with an 80 ms mixing time. NOESY spectra in H₂O were acquired with mixing times ranging from 50 to 250 ms. Water suppression was achieved by including a WATER-GATE (47) module in the pulse sequence prior to acquisition. Two-dimensional experiments were carried out at temperatures ranging from 5 to 25°C. The spectral analysis program NMRFAM-SPARKY (48) was used for peak assignment and semiquantitative NOE analysis.

Experimental NMR constraints

Distance constraints were obtained from an estimation of NOE intensities. In addition to these experimentally derived constraints, hydrogen bond and planarity constraints for the base pairs were used in the initial DYANA (49) calculations. Target values for distances and angles related to hydrogen bonds were set to values obtained from crystallographic data

Figure 1. (**A**) Strand orientation and triads in antiparallel (left) and parallel (right) triplexes: the third strand (yellow) binds to the homopurine sequence (cyan) of a duplex through Hoogsteen or reversed Hoogsteen base pairs (hashed bonds). The two orientations of phosphodiester backbones are indicated by the symbols '⊗' and '⊙'. (**B**) Chemical structure of *para*-TINA, *ortho*-TINA and INA. (**C**) Scheme showing the bulge insertion of TINA into the TFO part of the triplex. (**D**) Scheme of the modified antiparallel triplexes TTa (top) and TTb (bottom) showing three DNA strands linked by two loops of six ethylene glycol (EG) units and the TINA monomer conjugated between the residues G19 and G20 (TTa) and T18 and G19 (TTb).

in related structures (50). Due to the relatively broad line-widths of the sugar proton signals, J-coupling constants were not accurately measured, but only roughly estimated from DQF-COSY cross-peaks. Loose values were set for the sugar dihedral angles ν_0, ν_1, ν_2, ν_3 and ν_4 to constrain the 2′-deoxyribose conformation to the south domain. No backbone angle constraints were employed. Distance constraints with their corresponding error bounds were incorporated into the AMBER potential energy by defining a flat-well potential term.

Structural determination

Structures were calculated with the program DYANA (49) and further refined with the SANDER module of the molecular dynamics package AMBER 18.0 (51). The resulting DYANA structures were used as starting points for the AMBER refinement. The refined structures were first placed in a truncated octahedral box of water molecules with minimal distance between the solute and the box border of 12 Å using the TIP3P model. Eighteen sodium ions were added to neutralize the total charge of the system. The BSC1 force field (52) and suitable parameters for TINA were used to describe the modified oligonucleotide. The specific protocols used for AMBER refinement have been described in detail elsewhere (53). Here, 25 ns restrained molecular dynamics (rMD) were run using experimental NMR constraints involving non-exchangeable protons as well as distance and angle constraints for triads formation (triad constraints). Final structures were obtained extracting ten structures from the rMD trajectories and further relaxation of the structures removing the triad constraints and using the same NMR constraints used during rMD simulations plus NMR constraints involving interchangeable protons. Analysis of the representative structures as well as the MD trajectories was carried out with the programs MOL-MOL (54), X3DNA (55), CURVES (56) and PyMOL (57). Energy decomposition of MD trajectories was performed using the Gromacs-based utility Gromologist (58,59).

Unbiased molecular dynamics calculation without experimental constraints were performed to evaluate the stability and consistency of the NMR structures. These runs consisted of three replicas of a 100 ns simulation starting form a structure derived from NMR data (rMD calculations).

Partial atomic charges for TINA were calculated using the RESP model (60) after geometry optimization of the residue having its 5′ and 3′ oxygens capped with hydrogens. The electrostatic potential energy were carried out at the Hartree–Fock level of theory using the 6–31G(d) basis set for consistency with other atomic charges in the AMBER force field (61). For comparison purposes, the partial atomic charges of TINA were also calculated using the B3LYP/6–31G(d,p) method. Values and differences between charges obtained using HF and B3LYP are shown in Supplementary Table S9. Additional force field parameters needed were obtained from BSC1 (62) or GAFF (63).

Results

TINA insertion increases the thermal stability of the antiparallel triplex

We synthesized the DNA–TINA conjugates TTa and TTb, containing three oligonucleotide tracts of seven residues separated by two EG6 linkers. The oligonucleotide sequences were designed to fold as antiparallel triplexes with the EG6 linkers in the loops (Figure 1D). A TINA monomer was conjugated between the last two guanines (G19 and G20) of TTa and between T18 and G19 of TTb (Figure 1D). According to Doluca *et al.* (35), the affinity constant of TINA-containing oligonucleotides forming antiparallel triplex is highest when TINA is flanked by two purine or two pyrimidine residues. Thus, location of TINA within TTa was conceived to favour a more stable antiparallel triplex compared to TTb.

We used UV melting curves and CD and ^{1}H-NMR spectroscopy to evaluate the thermal stability of these constructs, as well as the unmodified control triplex (UT). UV melting curves of the unmodified and TINA-modified triplexes are shown in Supplementary Figure S1, confirming the strong thermal stabilization induced by TINA with T_M values of 69°C and 78°C for UT and TTa, respectively. The TTb triplex showed several melting transitions indicative of suboptimal position of TINA in the triplex. Presence of isosbestic points in CD spectra of TTa at various temperatures (Supplementary Figure S1) confirms melting of the complex at equilibria. This gives the same T_M value as determined by UV melting, which is expected for melting of unimolecular triplex at different concentrations, i.e. UV melting was performed at 2 μM and CD melting at 20 μM. Considering a unimolecular reversible melting transition, thermodynamic data analysis of the melting curves was conducted, as described in the Supplementary Data. This analysis revealed that triplex stabilisation by TINA ($\Delta\Delta G_{298} = -9.7$ kJ/mol) is attributed to the large enthalpic contribution ($\Delta\Delta H = -41$ kJ/mol) that is partially deprived by the unfavourable entropic component ($\Delta(T\Delta S) = -31.3$ kJ/mol), if comparing data of TTa and UT triplexes.

The 1D NMR spectra at 5°C (Figure 2) show sharp signals in the range of 12–15 ppm, characteristic of imino protons involved in Watson–Crick and Hoogsteen base pairs. Thermal denaturation of UT and TTa was also followed by recording ^{1}H-NMR spectra at different temperatures and observing the decay of imino signal intensities as temperature increases. Whereas imino signals in UT completely disappear at 45°C, in TTa they are still observed at 55°C (Figure 2), indicating that the TTa triplex remains folded at much higher temperature than UT. In order to determine the structural reasons behind this remarkable thermal stabilization induced by TINA, we conducted the structural determination of TTa at atomic resolution.

Assignment of NMR spectra

Assignment of non-exchangeable proton signals of TTa was performed using NOESY and TOCSY spectra acquired in D_2O. In general, due to the presence of long poly-purine tracks, NMR spectra of antiparallel triplexes are poorly dispersed and difficult to assign. However, the presence of TINA provokes a significant dispersion of the NMR signals, facilitating their assignments. This signal dispersion effect has also been observed in G-quadruplex structures having substitutions of the pyrene-containing nucleobase U^{Py} (5-(pyren-1-yl-ethynyl)-dUMP) (64). Connectivity between consecutive residues along the first strand can be followed from T1 to C5 through H2′/H2″(i) – H6 (i + 1) and H1′(i) – H6(i + 1) cross-peaks (Figure 3A). The lack of sequential cross-peaks between C5 and C6 is consistent with the intercalation of TINA. Connectivity is again recovered between C6 and terminal T7 (Figure 3A). Further sequential cross-peaks of medium inten-
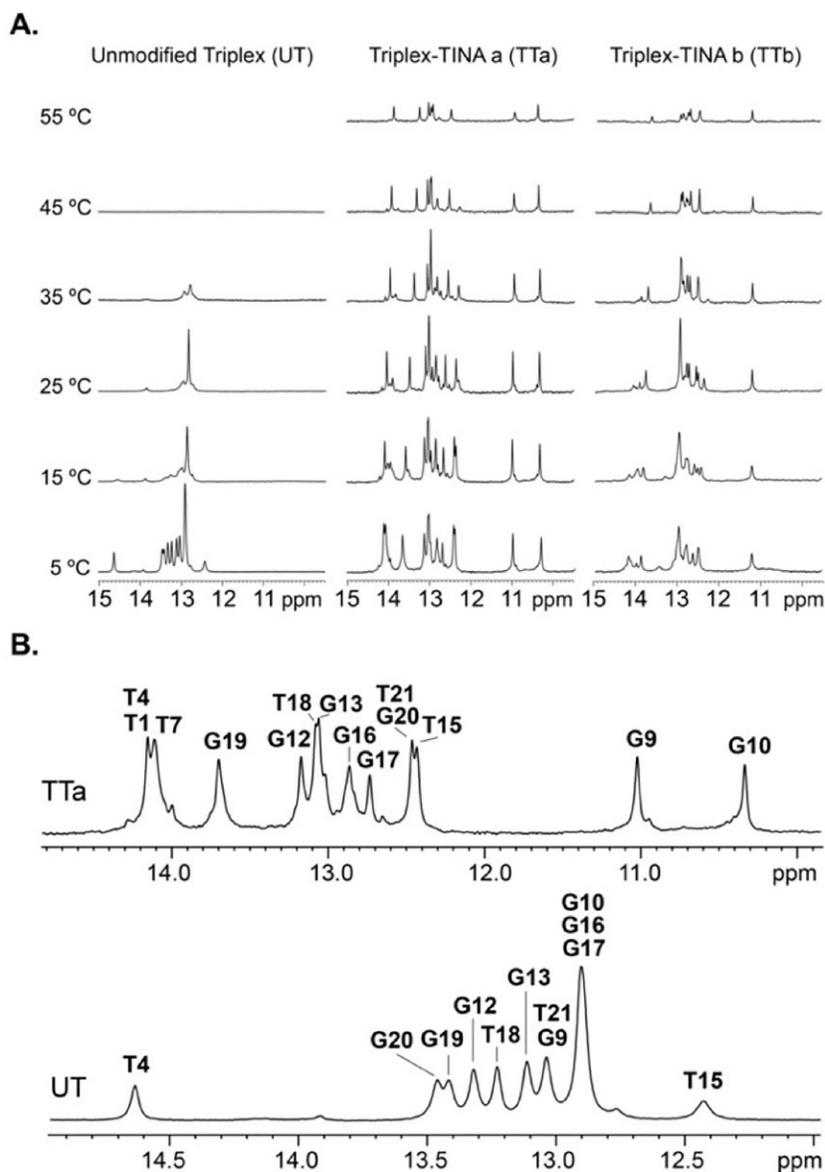
Figure 2. (**A**) Exchangeable proton regions of the 1D ^{1}H-NMR spectra of UT (left), TTa (middle) and TTb (right) at 400 μM acquired at different temperatures. (**B**) Expanded view of the imino region of UT and TTa at 5°C showing the assignment of the imino protons. Buffer conditions: 25 mM sodium phosphate, 100 mM NaCl, pH 7, H_2O/D_2O 90:10.

sity along the first strand such as H6(i)–CH$_3$(i + 1) in steps C–T (Supplementary Figure S2) and H6(i)–H5(i + 1) in steps C2–C3 and T4–C5 are observed and indicative of a right-handed helix formation.

Sequential assignment along the second strand starts by identifying the H8 aromatic proton of A8, which shows cross-peaks between 3.0 and 3.6 ppm with methylene protons of the first EG6 loop. Weak inter-residual cross-peaks lead to the identification of G9H8. The connectivity is lost between G9 and G10, supporting also the intercalation of TINA between these residues of the second strand. G10 is identified by the characteristic NOE between its aromatic proton and the methyl protons of the thymine in the next triad (T18) (Supplementary Figure S2). The analogous NOE between G13H8 and T15CH$_3$ is observed for the other G-A step in the second strand (Supplementary Figure S2). From G10, the

connectivity is maintained along the rest of the residues in the strand. As observed for T7H6 and A8H8, A14H8 also shows cross-peaks with protons of EG units linking the second and the third strand. The H2 protons of adenines were identified by their cross-peaks with imino protons of the first strand (Supplementary Figure S3) and with EG6 protons of the first and second EG6 linkers in the case of the terminal adenines A8 and A14, respectively.

The aromatic proton of T15 shows cross-peaks with protons of the second EG6 loop and a particularly intense intra-residual cross-peak with H2′ which appears rather shifted to high-field (Supplementary Figure S2). The connectivity between T15 and G16 is detected through a weak sugar-base T15H2′–G16H8 cross-peak (Supplementary Figure S2) and aromatic-aromatic NOE T15H6-G16H8 (Supplementary Figure S3). Sugar-base connectivity is lost between G16
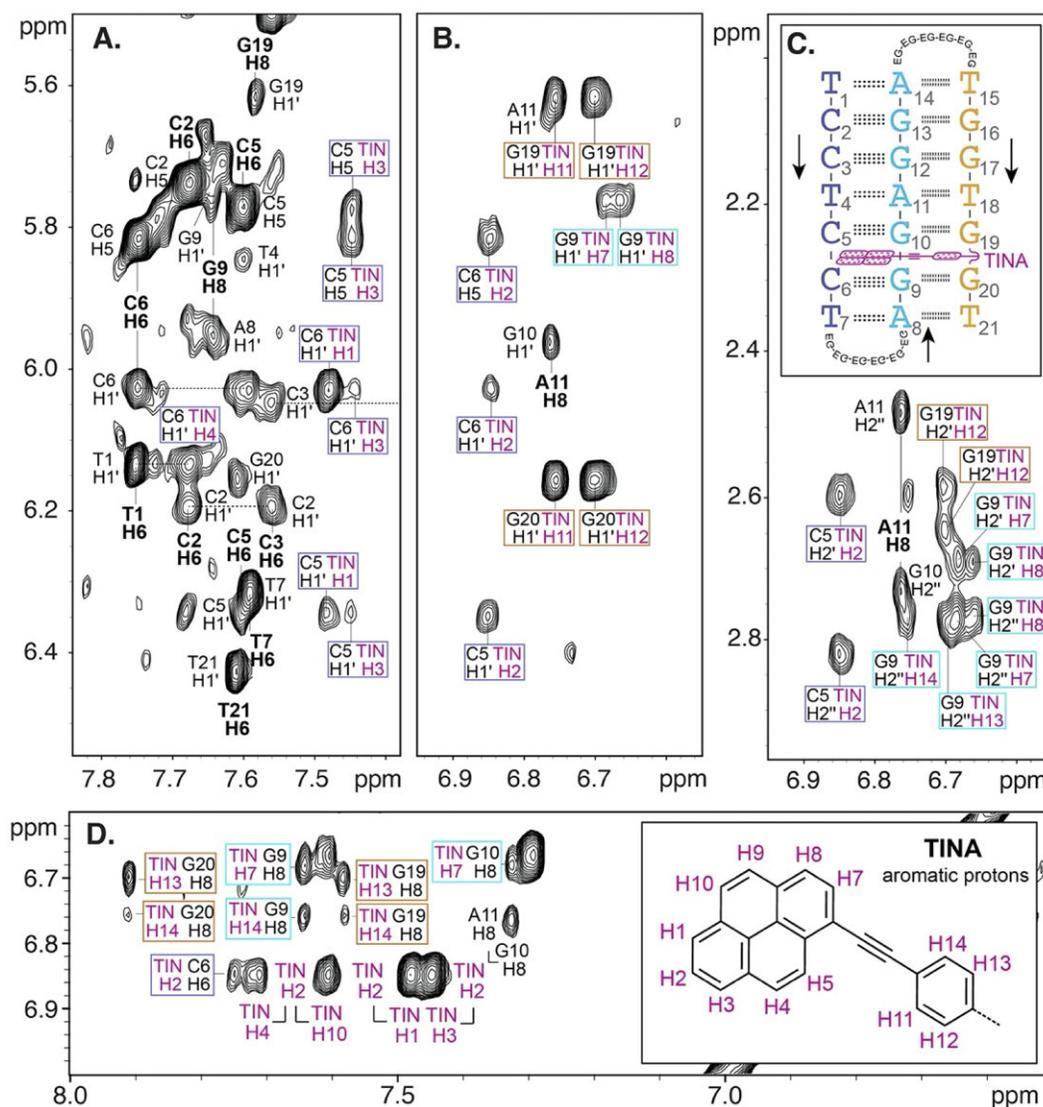
Figure 3. Different regions of the NOESY spectra of TTa in D_2O at 5 °C showing sequential assignment of first DNA strand of TTa (panel A, dashed lines) and cross-peaks between TINA protons and: (**A** and **B**) H5 of C5 and C6 and H1′ protons of C5, C6, G9, G19 and G20; (**C**) H2′ and H2″ of C5, G9 and G19; and (**D**) aromatic protons of C6, G9, G10, G19 and G20. Label frames are coloured in blue slate (C5 and C6), cyan (G9 and G10) and orange (G19 and G20). Aromatic protons of the nucleotides are labelled in bold. Some aromatic-aromatic cross-peaks of neighbour aromatic TINA protons are also shown (magenta interconnected labels in panel D). A scheme of the antiparallel triplex and the chemical structure of TINA with labelled aromatic hydrogens are shown in the inset of panels C and D, respectively.

and G17 and appears again between G17 and T18. The notably intense G17H8-T18CH$_3$ (Supplementary Figure S2) is evidential of the right-handed helical nature of the third strand. The analogous NOE in the other G–T step, G20H8–T21CH$_3$, is also detected (Supplementary Figure S2). Likewise, moderately intense cross-strand NOEs G10H8–T18CH$_3$ and G13H8–T15CH$_3$ (Supplementary Figure S2) confirm the reverse-Hoogsteen base-pairing between adenines and thymines in the second and third strand, respectively (Figure 1). As previously observed for the steps C5–C6 and G9–G10, the connectivity is also lost between G19 and G20 which is consistent with the intercalation of TINA between the triads C5:G10*G19 and C6:G9*G20. Finally, G20 also shows sugar-base connections with the last residue of the triplex, T21.

The exchangeable signals observed between 12 and 15 ppm in the NOESY spectra in H$_2$O were assigned to imino

protons involved in the formation of all the seven triads of the triplex. Imino protons of thymines of the first strand were identified by their intense cross-peaks with the H2 of their Watson–Crick-paired adenines (Supplementary Figure S4). In the case of the internal A11:T4 base pair, the cross peaks between T4H3 and A11 amino protons are also detected (Supplementary Figure S4). Imino protons of the second strand (G9H1, G10H1, G12H1 and G13H1) were assigned attending to their strong NOEs with amino protons of their base-paired cytosines of the first strand (C6, C5, C3 and C2, respectively) (Supplementary Figures S4 and S5). Interestingly, the imino protons of G9 and G10 resonate and are very shifted to high field with respect to the rest of imino protons (Supplementary Figure S5). Finally, the imino protons of thymines and guanines of the third strand show the characteristic strong NOEs with the H8 aromatic protons of their Hoogsteen base-paired purines of the second strand

(Supplementary Figure S4). Moreover, imino protons of these residues exhibit cross-peaks of weak or medium intensity with H2′, H2″ and H8 protons of the nucleotide preceding their base-paired partner in the second strand (Supplementary Figures S4 and S6). The chemical shifts of all the DNA assigned protons are shown in Supplementary Table S1.

NMR signals of TINA aromatic protons (TIN H) were assigned by analysing TOCSY and NOESY spectra in D$_2$O (Supplementary Figure S3, Supplementary Table S2). Identification of TIN H2 proton was unambiguous due to its intense cross-peaks with TIN H1 and TIN H3 observed in both NOESY and TOCSY spectra. Consequently, TIN H3 shows two cross-peaks in the NOESY that were assigned according to their intensity as TIN H4 (stronger) and TIN H5 (weaker) which show, in turn, intense TOCSY and NOESY cross-peaks with each other. TIN H5 is further correlated in the NOESY to the proton of the phenyl ring TIN H11. On the other hand, TIN H1 shows two NOEs, the stronger of them with TIN H10 and the weaker with TIN H9. The two later protons are corelated in the NOESY and TOCSY spectra. TIN H9 shows an intense cross-peak with TIN H8, which in turn resonates very close to its neighbour TIN H7, so the NOE and TOCSY cross-peaks correlating them lay very close to the diagonal of the spectra. The protons of the phenyl ring TIN H11 and TIN H12 resonate at the same chemical shift and are correlated in the NOESY and TOCSY (peaks close to the diagonal) with TIN H13 and TIN H14 which are also degenerated.

NOESY spectra exhibit numerous short distances between TINA and DNA protons

A total of 45 cross-peaks between DNA and TINA protons were identified (Supplementary Table S3). All the residues of the triads flanking the intercalated TINA (C5:G10*G19 and C6:G9*G20) show contacts with TINA protons. The precise position of TINA between the triads is defined by the presence and intensity of NOEs between TINA protons and DNA protons of the six surrounding nucleotides. On the one hand, several NOEs correlating sugar protons of C5 and C6 with the TINA protons H1, H2, H3, H4 and H10 (Figure 3A–C) strongly constraint the position of the TINA pyrene ring. Additionally, the H5 of C5 and C6 shows moderately intense cross-peaks with H2 and H3 of TINA (Figure 3A and B) which further inform about the position of that side of the pyrene ring. Consistently, sugar protons of G9 show NOEs with the pyrene ring TINA protons H7 and H8 (Figure 3B and C). The aromatic protons of G9 and G10 also show relatively intense cross-peaks with TINA H8 (Figure 3D). The protons of the TINA phenyl ring correlate with protons of the flanking residues in the sequence G19 and G20. The particularly intense NOEs between H1′ of the two guanines and TINA H11 and H12 (Figure 3B) are indicative of the penetration degree of TINA into the triplex. The position of the phenyl ring is further determined by NOEs correlating the aromatic protons of G19 and G20 with the other two protons of the ring H13 and H14 (Figure 3D).

The unmodified version of the same sequence (UT) was also studied by two-dimensional NMR. Most of the chemical shifts of UT protons (Supplementary Table S4) are similar to the ones assigned for TTa (Supplementary Table S1). However, the presence of TINA within the sequence affects remarkably the chemical shift of some protons. Consistently with the position of TINA, the highest chemical shift differ-

ences between UT and TTa protons corresponds to those of C5:G10*G19 and C6:G9*G20 triads (Supplementary Table S5 and Supplementary Figure S7) some of which differ in more than 1 (C5H41, C6H41 and G20H1) and 2 ppm (G9H1 and G10H1).

Structural features of TTa

On the basis of the experimental constraints derived from NMR spectra (Supplementary Table S6), we carried out the structural determination of TTa by restrained molecular dynamics methods (rMD). Two different views of the structure (PDB ID: 8PWR) are shown in Figure 4A. The structures were calculated as a trimolecular complex since the EG6 connectors were not assigned in the NMR spectra and, therefore, were not considered in the calculation. All final structures exhibit reasonably low energy values and no average distance violation larger than 0.25 Å (Supplementary Table S6). All the residues are very well defined with overall RMSD values around 0.6 Å (Supplementary Table S6). The terminal residues are equally well-defined as the rest of the structure (Supplementary Figure S8). The consistency of the resulting structures was further verified by running longer unbiased molecular dynamics trajectories. The analysis of RMSD variation along three 100 ns trajectories exhibit good convergence for both the overall structures and the TINA interaction site, with RMSD values sustained around 2 Å along the trajectory (Supplementary Figure S9).

The structure of TTa (Figure 4A) is a right-handed triple helix with the geometrical parameters shown in Supplementary Tables S7 and S8. Sugar puckers of most residues are in the general south domain, except T18, C5, T15 and G20 that are in the east domain, and G19 that adopts north or south puckering in different structures (Supplementary Table S7). Glycosidic angles of most nucleotides are *anti* with χ values in the range of −110° to −150° except for G9, G10, G16, T18 and G19 that adopt a high *anti* glycosidic conformation (χ values from −50° to −103°) (Supplementary Table S8). General helical parameters (see Supplementary Figure S10) are consistent with expected values for an antiparallel triplex, except in the TINA insertion region, where notably high rise and reduced twist values are observed.

The structure is stabilized by three T:A*T and four C:G*G (Figure 1A) triads and a TINA monomer intercalated between the triads C5:G10*G19 and C6:G9*G20. The phenyl and the pyrene rings of TINA exhibit a slight tilt with respect to each other while remaining equidistant to the two triads (Figure 4A). The pyrene ring lays between the Watson–Crick base-pairs C5:G10 and C6:G9 (Figure 4B) while the phenyl ring intercalates between the G19 and G20 bases, exhibiting a larger stacking with the former (Figure 4B).

A detailed view of the structure confirms all the DNA-TINA contacts observed in the NOESY spectra. Thus, TINA H1, H2, H3 and H10 are close to the residues of the first strand (C5 and C6) (Figure 4C); TINA protons H7 and H8 are near bases of the second strand (G9 and G10); and protons H11 and H12 of the phenyl ring are at close distance to residues of the third strand (G19 and G20) (Figure 4D). This TINA location explains the changes in chemical shifts with respect to the unmodified triplex (Supplementary Figure S7). Especially evident are the highly shifted imino signals of G9 and G10, observed between 10 and 11 ppm in TTa (Figure 2 and Supplementary Figure S5) and at 13.05 ppm and 12.92 ppm
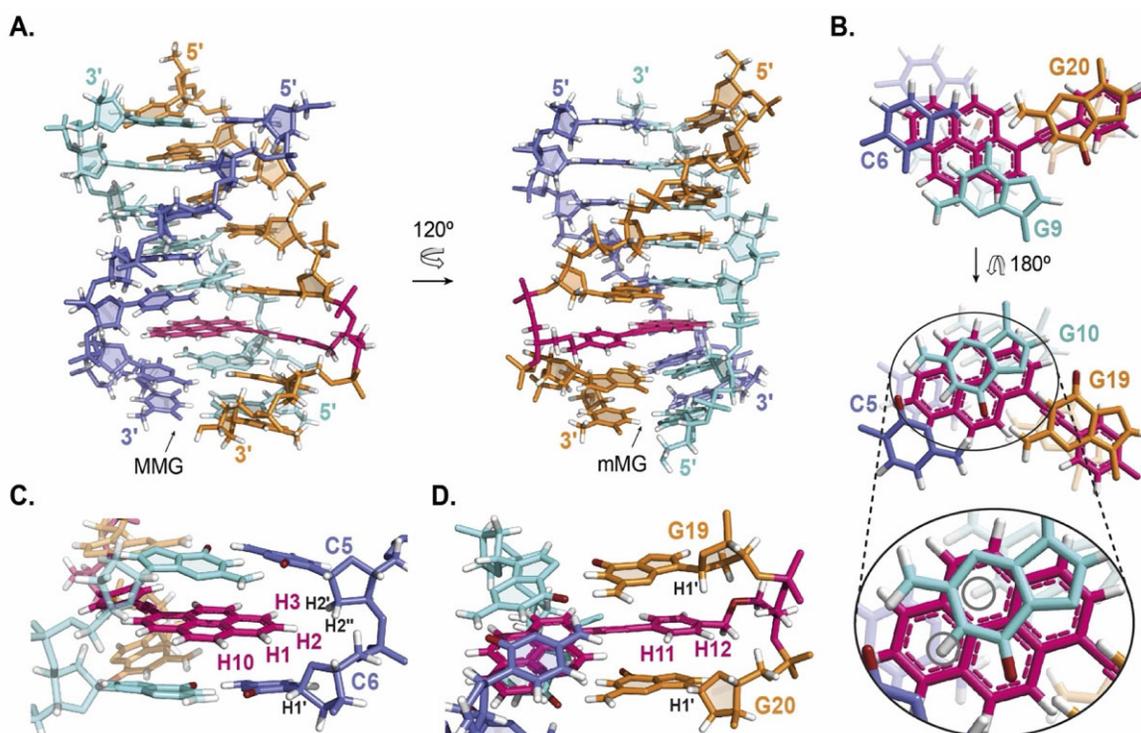
Figure 4. (**A**) Structure of TTa viewed from the major-major groove (MMG), and minor-major groove (mMG) perspectives. (**B**) Detailed view of the triads C6:G9*G20 (top) C5:G10*G19 (middle) showing the stacking of TINA aromatic rings on the flanking triads. Expanded view of G10 laying above the pyrene ring of TINA (B bottom) showing the position of G9 and G10 imino protons (inside grey circles) under the influence of the pyrene ring of TINA. (**C**) Detailed lateral view from the minor groove (mG) illustrating the close position of TINA H1, H2, H3 and H10 protons to sugar protons of C5 and C6. (**D**) Detailed lateral view from the MMG groove showing the proximity of TINA H11 and H12 to the H1′ of their flanking guanines G19 and G20.

in UT. These protons are notably influenced by ring-current effects, owing to their close proximity to the TINA pyrene ring, as shown in Figure 4B. The same effect is observed in the NMR spectra of the sequence TTb, where TINA is conjugated between T18 and G19 instead of G19 and G20. In this case TINA intercalates between the T4:A11*T18 and C5:G10*G19 triads so the pyrene ring only stands above one imino proton of the second strand (G10H1). Indeed, only one signal is observed between 10 and 12 ppm which presumably corresponds to G10H1 (Figure 2). This characteristic chemical shift of the guanine imino protons flanking TINA in the purine strand could be used as a sign to confirm TINA–TFOs binding by NMR.

Discussion

Triplex formation is a promising tool for selectively influencing gene expression (2,65). Antiparallel triplexes offer several advantages over parallel ones because of their stability under physiological conditions. Despite this, there is a striking scarcity of structural information regarding this specific triplex conformation. To date, only one high-resolution structure of an antiparallel DNA triplex is available in the PDB, and it was reported 30 years ago (45). In the present work we demonstrate that a single insertion of TINA is enough to significantly increase the thermal stability of an antiparallel triplex in comparison to its unmodified analogue. We determine the structure of the modified triplex in solution, which represents the first high-resolution structure of a TINA-containing DNA molecule and unveils the binding mode of

this modification and the structural reasons behind its stabilization effect.

Our structural determination shows that the position of TINA between triads mostly reproduces the one anticipated by its rational design, with the pyrene ring stacking on the base pairs of the duplex and the phenyl ring sandwiched by the flanking bases attached to it. As was predicted by molecular modelling and fluorescence experiments on a similar antiparallel triplex (36), the location of the pyrene ring is closer to the duplex moiety. The structure shows that, specifically, the pyrene ring stacking occurs on the bases of the purine strand G9 and mostly G10 (Figure 4B). Given so, the use of polyaromatic platforms larger than pyrene may also fit well between triads and might further enhance the stability of the triplex through stacking interactions covering more surface of the duplex base pairs. Likewise, some analogous substituents of the phenyl ring would enlarge the stacking on the contiguous bases of the third strand. In this regard, Bomholt et al. (66) reported that the T_M of a parallel triplex conjugated with a TINA having a naphthalene ring instead of a phenyl ring was 2°C higher than the one of the same triplex modified with a TINA having a phenyl ring. By the virtue of the chemical design, *ortho*-TINA and particularly INA (Figure 1B) cannot position pyrene optimally for intercalation between the duplex part of the triplex as *para*-TINA does, thus explaining contrasting properties of these molecules.

The triple bond of TINA was conceived to allow the fitting of the two aromatic platforms on bases having different tilt. Indeed, the bases of the Hoogsteen strand run coplanar along the strand but in a different plane with respect to the base pairs of the duplex moiety. The structure shows that, both
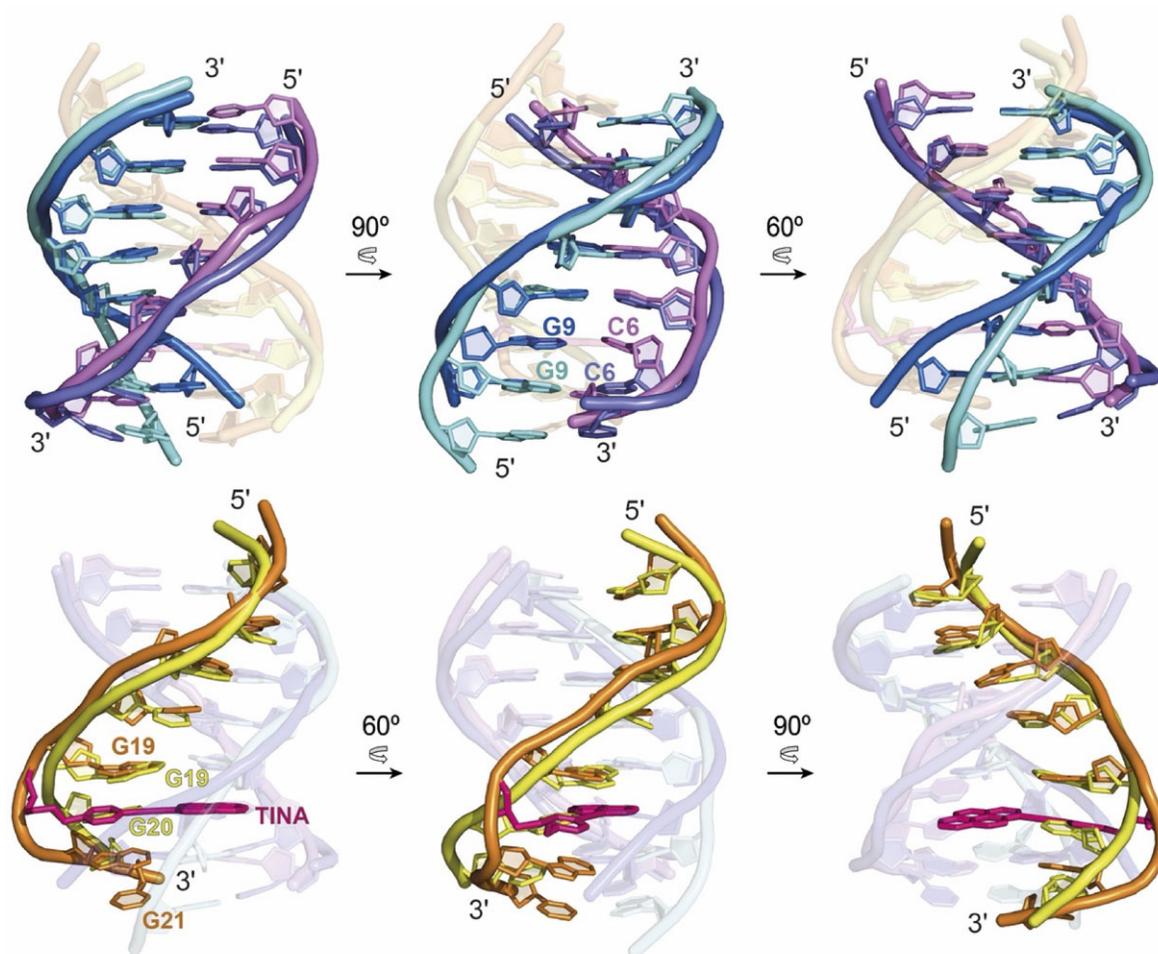
Figure 5. Structural alignment of TTa and the homologous region of the unmodified triplex PDB ID 134D (45) (UT134D). Top. The Watson–Crick strands (*slate blue*: TTa; *violet*: UT134D) and the purine strands (*light blue*: TTa; *dark blue*: UT134D) are mostly overlapped along the segment from T1:A14 to C5:G10. C6 and G9 of TTa are moved and tilted to make a wider gap between C5:G10 and C6:G9 base pairs, facilitating the intercalation of the pyrene moiety of TINA. Bottom. The Hoogsteen strands of TTa (*orange*) and UT134D (*yellow*) show a poor alignment. The bases from both strands show different tilt and the backbone of TTa is placed at longer distance to the purine strand. Bases flanking TINA (G19 and G20) are more distant to each other with respect to the homologous G19 and G20 in UT134D. G20 is tilted towards 3′-end to increasing the space for TINA intercalation. The conjugation of TINA results in a distortion of the backbone of TTa which shows as a bulge between G19 and G20.

rings of TINA are slightly twisted to each other. The torsion angle between both aromatic planes ranges from 10° to 20° depending on the refined structure analysed being the average torsion angle 16.5°. This data is in good agreement with the molecular modelling results obtained for both antiparallel (36) and parallel (38) triplexes that reported torsion angles of 1–10° and 15.3°, respectively. A greater torsion angle might be deleterious for triplex stabilisation as previously observed for the triazole-TINA in which 1,2,3-triazole was used instead of the triple bond in *para*-TINA imposing a torsion angle of 35–40° and a significantly lower level of parallel triplex stability (67).

The analysis of melting curves for TTa and UT triplexes clearly indicates that TINA stabilization is of enthalpic origin. The insertion of TINA into the triplex introduces not only the pyrene intercalator but also a bulge in the structure and an additional negatively charged phosphate. These structural features must be accommodated within the triplex, introducing an increased unfavorable entropy component to the free Gibbs energy of triplex formation. TINA intercalation within the triplex structure suggests that the stabilizing effect is pri-

marily due to stacking interactions. The energy decomposition of our MD data enables the quantification of the TINA effect in the triplex, yielding values of −181.6 kJ/mol and −258.3 kJ/mol for the coulombic and Lennard-Jones terms, respectively. These values are sufficiently large to compensate for the distortions in the triplex structures required for the formation of the intercalation site and any associated negative entropic effect.

A comparison between the structure of TTa and the structure of the unmodified triplex reported by Radhakrishnan and Patel (45) (PDB ID: 134D) is shown in Figure 5. Although TINA intercalation provokes large local distortions in the surrounding nucleobases, the overall triplex structure is mainly unaffected. The pyrene ring is sandwiched between the Watson–Crick pairs, with a large stacking surface with the guanine nucleobases. In addition, the phenyl ring stacks on top of the 5′-neighbouring guanine within its own strand. These stabilizing interactions, primarily resulting from highly favourable stacking, compensate for the local distortions at the intercalation site and are responsible for the significant stabilization caused by TINA. This explains the enhanced

stability of antiparallel triplexes formed by TFOs containing TINA monomers between two guanines (purines), instead of between a purine and a pyrimidine (35). This sequence design also reduces the ability of G-rich strands to form G-quadruplexes (35,36) which contributes to the shift of triplex–G-quadruplex equilibrium towards triplex formation. It is also interesting to notice that TINA fits very well with the overall size of Pyrimidine:Purine*Purine (Y:R*R) triad. Although a significant bulge is observed in the Hoogsteen strand, no large distortions occur in the purine and pyrimidine strands backbone. Since TINA is probably more difficult to accommodate between the smaller Y:R*Y triads, this may explain the enhanced stabilization of TINA in antiparallel vs parallel triplex (35). The good fitting of TINA between Y:R*R triads suggests that only a few modifications in its chemical design such as using polyaromatic platforms slightly larger than phenyl (68) and pyrene may improve TINA-induced antiparallel triplex stability.

The use of modified TFOs with improved stability inside cells and higher affinity for the target duplex (24,69–73) is further expanding the potential of triplex formation as a tool for therapeutic and biotechnological applications. Particularly, *para*-TINA modification favours the formation of more stable triplexes without compromising the specificity. Indeed, the thermal stability of parallel triplexes formed by TINA-modified TFOs decreases more than 11.5°C when the resulting triplex has a single mismatch (38). Similarly, for TINA-modified TFOs forming antiparallel triplexes, the presence of a mismatch next to TINA completely disrupts triplex formation (36,37). Moreover, the fluorescence properties of TINA can be used in probes for detection of DNA and RNA specific sequences (74–79). The structural details unveiled in this study help rationalize the different effects of TINA insertion when it occurs in the middle of purine-rich strands of TFOs and in G4-forming sequences. Insertion in quadruplex G-tracts usually results in lower thermal stability or the formation of higher-order structures through TINA-TINA interactions (64,80), making this modification a very useful tool to modulate G-rich sequence folding and to overcome the main limitations of triplex formation and visualization *in vitro* and *in vivo*.

The high-resolution structural data presented here will greatly contribute to a more rational and efficient design of modified oligonucleotides aimed to be used as antigenic agents and DNA-based biotechnological devices. These results are of interest for further design of more efficient modulators of hybridization properties of oligonucleotides allowing the formation and stabilization of complexes with genomic DNA for therapeutic and diagnostic applications.

Data availability

The data underlying this article are available in the Protein Data Bank (PDB) at https://www.rcsb.org/, and can be accessed with ID 8PWR. All other data are available in the article or Supplementary information. Raw data will be shared on request to the corresponding/first authors.

Supplementary data

[Supplementary Data](#) are available at NAR Online.

Acknowledgements

Funding

Conflict of interest statement

None declared.

References

1. Chan,P.P. and Glazer,P.M. (1997) Triplex DNA: fundamentals, advances, and potential applications for gene therapy. *J. Mol. Med.*, **75**, 267–282.
2. Dalla Pozza,M., Abdullrahman,A., Cardin,C.J., Gasser,G. and Hall,J.P. (2022) Three's a crowd – stabilisation, structure, and applications of DNA triplexes. *Chem. Sci.*, **13**, 10193–10215.
3. Frank-Kamenetskii,M.D. and Mirkin,S.M. (1995) Triplex DNA structures. *Annu. Rev. Biochem.*, **64**, 65–95.
4. Vasquez,K.M. and Glazer,P.M. (2002) Triplex-forming oligonucleotides: principles and applications. *Q. Rev. Biophys.*, **35**, 89–107.
5. Sakamoto,T., Yamaoki,Y., Nagata,T. and Katahira,M. (2021) Detection of parallel and antiparallel DNA triplex structures in living human cells using in-cell NMR. *Chem. Commun.*, **57**, 6364–6367.
6. Gorab,E., Amabis,J.M., Stocker,A.J., Drummond,L. and Stollar,B.D. (2009) Potential sites of triple-helical nucleic acid formation in chromosomes of Rhynchosciara (Diptera: sciaridae) and *Drosophila melanogaster*. *Chromosome Res.*, **17**, 821.
7. Piergentili,R. and Mencarelli,C. (2008) *Drosophila melanogaster* kl-3 and kl-5 Y-loops harbor triple-stranded nucleic acids. *J. Cell Sci.*, **121**, 1605–1612.
8. Černá,A., López-Fernández,C., Fernández,J.L., de la Espina,S.M.D., De la Torre,C. and Gosálvez,J. (2008) Triplex configuration in the nick-free DNAs that constitute the chromosomal scaffolds in grasshopper spermatids. *Chromosoma*, **117**, 15–24.
9. Ohno,M., Fukagawa,T., Lee,J.S. and Ikemura,T. (2002) Triplex-forming DNAs in the human interphase nucleus visualized in situ by polypurine/polypyrimidine DNA probes and antitriplex antibodies. *Chromosoma*, **111**, 201–213.
10. Brázdová,M., Tichý,V., Helma,R., Bažantová,P., Polášková,A., Krejčí,A., Petr,M., Navrátilová,L., Tichá,O. and Nejedlý,K. (2016) p53 specifically binds triplex DNA in vitro and in cells. *PLoS One*, **11**, e0167439.
11. Guo,M., Hundseth,K., Ding,H., Vidhyasagar,V., Inoue,A., Nguyen,C.-H., Zain,R., Lee,J.S. and Wu,Y. (2015) A distinct triplex DNA unwinding activity of ChlR1 helicase. *J. Biol. Chem.*, **M114**, 634923.
12. Jiménez-García,E., Vaquero,A., Espinás,M.L.s., Soliva,R., Orozco,M., Bernués,J. and Azorín,F. (1998) The GAGA factor of Drosophila binds triple-stranded DNA. *J. Biol. Chem.*, **273**, 24640–24648.
13. Gorab,E. (2018) Triple-helical DNA in Drosophila heterochromatin. *Cells*, **7**, 227.

14. Bacolla,A., Collins,J.R., Gold,B., Chuzhanova,N., Yi,M., Stephens,R.M., Stefanov,S., Olsh,A., Jakupciak,J.P. and Dean,M. (2006) Long homopurine• homopyrimidine sequences are characteristic of genes expressed in brain and the pseudoautosomal region. *Nucleic Acids Res.*, **34**, 2663–2675.

15. Schroth,G.P. and Ho,P.S. (1995) Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. *Nucleic Acids Res.*, **23**, 1977–1983.

16. Wang,G., Seidman,M.M. and Glazer,P.M. (1996) Mutagenesis in mammalian cells induced by triple helix formation and transcription-coupled repair. *Science*, **271**, 802–805.

17. Tiwari,M.K., Adaku,N., Peart,N. and Rogers,F.A. (2016) Triplex structures induce DNA double strand breaks via replication fork collapse in NER deficient cells. *Nucleic Acids Res.*, **44**, 7742–7754.

18. Tiwari,K. and Rogers,F.A. (2013) XPD-dependent activation of apoptosis in response to triplex-induced DNA damage. *Nucleic Acids Res.*, **41**, 8979–8994.

19. Wang,G. and Vasquez,K.M. (2004) Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 13448–13453.

20. Guntaka,R.V., Varma,B.R. and Weber,K.T. (2003) Triplex-forming oligonucleotides as modulators of gene expression. *Int. J. Biochem. Cell Biol.*, **35**, 22–31.

21. Jain,A., Wang,G. and Vasquez,K.M. (2008) DNA triple helices: biological consequences and therapeutic potential. *Biochimie*, **90**, 1117–1130.

22. Praseuth,D., Guieysse,A.L. and Helene,C. (1999) Triple helix formation and the antigene strategy for sequence-specific control of gene expression. *Biochim. Biophys. Acta*, **1489**, 181–206.

23. Rogers,F.A., Lloyd,J.A. and Glazer,P.M. (2005) Triplex-forming oligonucleotides as potential tools for modulation of gene expression. *Curr. Med. Chem. Anticancer*, **5**, 319–326.

24. Taniguchi,Y. and Sasaki,S. (2012) An efficient antigene activity and antiproliferative effect by targeting the Bcl-2 or survivin gene with triplex forming oligonucleotides containing a W-shaped nucleoside analogue (WNA-βT). *Org. Biomol. Chem.*, **10**, 8336–8341.

25. Aviñó,A., Eritja,R., Ciudad,C.J. and Noé,V. (2019) Parallel clamps and polypurine Hairpins (PPRH) for gene silencing and triplex-affinity capture: design, synthesis, and use. *Curr. Protoc. Nucleic Acid Chem.*, **77**, e78.

26. Wang,G. and Vasquez,K.M. (2014) Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. *DNA Repair (Amst.)*, **19**, 143–151.

27. Mukherjee,A. and Vasquez,K.M. (2011) Triplex technology in studies of DNA damage, DNA repair, and mutagenesis. *Biochimie*, **93**, 1197–1208.

28. Simon,P., Cannata,F., Concordet,J.-P. and Giovannangeli,C. (2008) Targeting DNA with triplex-forming oligonucleotides to modify gene sequence. *Biochimie*, **90**, 1109–1116.

29. Wu,Q., Gaddis,S.S., MacLeod,M.C., Walborg,E.F., Thames,H.D., DiGiovanni,J. and Vasquez,K.M. (2007) High-affinity triplex-forming oligonucleotide target sequences in mammalian genomes. *Mol. Carcinog.*, **46**, 15–23.

30. Hu,Y., Cecconello,A., Idili,A., Ricci,F. and Willner,I. (2017) Triplex DNA nanostructures: from basic properties to applications. *Angew. Chem. Int. Ed.*, **56**, 15210–15233.

31. Duca,M., Vekhoff,P., Oussedik,K., Halby,L. and Arimondo,P.B. (2008) The triple helix: 50 years later, the outcome. *Nucleic Acids Res.*, **36**, 5123–5138.

32. Cheng,A.-J., Wang,J.C. and Van Dyke,M.W. (1998) Self-Association of G-rich oligodeoxyribonucleotides under conditions promoting purine-motif triplex formation. *Antisense Nucleic Acid Drug Dev.*, **8**, 215–225.

33. Olivas,W.M. and Maher,L.J., III. (1995) Competitive triplex/quadruplex equilibria involving guanine-rich oligonucleotides. *Biochemistry*, **34**, 278–284.

34. Doluca,O., Withers,J.M. and Filichev,V.V. (2013) Molecular engineering of guanine-rich sequences: Z-DNA, DNA triplexes, and G-quadruplexes. *Chem. Rev.*, **113**, 3044–3083.

35. Doluca,O., Boutorine,A.S. and Filichev,V.V. (2011) Triplex-forming twisted intercalating nucleic acids (TINAs): design rules, stabilization of antiparallel DNA triplexes and inhibition of G-quartet-dependent self-association. *ChemBioChem*, **12**, 2365–2374.

36. Paramasivam,M., Cogoi,S., Filichev,V.V., Bomholt,N., Pedersen,E.B. and Xodo,L.E. (2008) Purine twisted-intercalating nucleic acids: a new class of anti-gene molecules resistant to potassium-induced aggregation. *Nucleic Acids Res.*, **36**, 3494–3507.

37. Doluca,O., Hale,T.K., Edwards,P.J., González,C. and Filichev,V.V. (2014) Assembly dependent fluorescence enhancing nucleic acids in sequence-specific detection of double-stranded DNA. *ChemPlusChem*, **79**, 58–66.

38. Filichev,V.V. and Pedersen,E.B. (2005) Stable and selective formation of hoogsteen-type triplexes and duplexes using twisted intercalating nucleic acids (TINA) prepared via postsynthetic Sonogashira solid-phase coupling reactions. *J. Am. Chem. Soc.*, **127**, 14849–14858.

39. Filichev,V.V., Nielsen,M.C., Bomholt,N., Jessen,C.H. and Pedersen,E.B. (2006) High thermal stability of 5'-5'-linked alternate Hoogsteen triplexes at physiological pH. *Angew. Chem., Int. Ed.*, **45**, 5311–5315.

40. Narayanan,S., Gall,J. and Richert,C. (2004) Clamping down on weak terminal base pairs: oligonucleotides with molecular caps as fidelity-enhancing elements at the 5'-and 3'-terminal residues. *Nucleic Acids Res.*, **32**, 2901–2911.

41. Sau,S.P., Kumar,T.S. and Hrdlicka,P.J. (2010) Invader LNA: efficient targeting of short double stranded DNA. *Org. Biomol. Chem.*, **8**, 2028–2036.

42. Filichev,V.V., Astakhova,I.V., Malakhov,A.D., Korshun,V.A. and Pedersen,E.B. (2008) 1-, 2-, and 4-Ethynylpyrenes in the structure of twisted intercalating nucleic acids: structure, thermal stability, and fluorescence relationship. *Chem. Eur. J.*, **14**, 9968–9980.

43. Christensen,U.B. and Pedersen,E.B. (2002) Intercalating nucleic acids containing insertions of 1-O-(1-pyrenylmethyl) glycerol: stabilisation of dsDNA and discrimination of DNA over RNA. *Nucleic Acids Res.*, **30**, 4918–4925.

44. Christensen,U.B., Wamberg,M., El-Essawy,F.A., Ismail,A.E.H., Nielsen,C.B., Filichev,V.V., Jessen,C.H., Petersen,M. and Pedersen,E.B. (2004) Intercalating nucleic acids: the influence of linker length and intercalator type on their duplex stabilities. *Nucleosides Nucleotides Nucleic Acids*, **23**, 207–225.

45. Radhakrishnan,I. and Patel,D.J. (1993) Solution structure of a purine.purine.pyrimidine DNA triplex containing G.GC and T.AT triples. *Structure*, **1**, 135–152.

46. Mergny,J.-L. and Lacroix,L. (2003) Analysis of thermal melting curves. *Oligonucleotides*, **13**, 515–537.

47. Piotto,M., Saudek,V. and Sklenář,V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661–665.

48. Lee,W., Tonelli,M. and Markley,J.L. (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, **31**, 1325–1327.

49. Güntert,P., Mumenthaler,C. and Wüthrich,K. (1997) Torsion angle dynamics for NMR structure calculation with the new program Dyana1. *J. Mol. Biol.*, **273**, 283–298.

50. Cai,L., Chen,L., Raghavan,S., Rich,A., Ratliff,R. and Moyzis,R. (1998) Intercalated cytosine motif and novel adenine clusters in the crystal structure of the Tetrahymena telomere. *Nucleic Acids Res.*, **26**, 4696–4705.

51. Case,D., Ben-Shalom,I., Brozell,S., Cerutti,D., Cheatham,T. III, Cruzeiro,V., Darden,T., Duke,R., Ghoreishi,D. and Gilson,M. (2018) In: *AMBER*. University of California, San Francisco.

52. Ivani,I., Dans,P., Noy,A., Pérez,A., Faustino,I., Hospital,A., Orozco,M., Walther,J., Andrio,P., Goñi,R., *et al.* (2016) Parmbsc1: a refined force field for DNA simulations. *Nat. Methods*, **13**, 55–58.

53. El-Khoury,R., Macaluso,V., Hennecker,C., Mittermaier,A.K., Orozco,M., González,C., Garavís,M. and Damha,M.J. (2023) i-Motif folding intermediates with zero-nucleotide loops are trapped by 2′-fluoroarabinocytidine via F··· H and O··· H hydrogen bonds. *Commun. Chem.*, **6**, 31.

54. Koradi,R., Billeter,M. and Wüthrich,K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 51–55.

55. Lu,X.-J. and Olson,W.K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.*, **3**, 1213–1227.

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

57. Schrödinger, LLC (2010) In: *The Pymol Molecular Graphics System*. Version 1.3.

58. Lindahl,E., Hess,B. and Van Der Spoel,D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Mol. Model. Annu.*, **7**, 306–317.

59. Wieczor,M. and Czub,J. (2022) Gromologist: a gromacs-oriented utility library for structure and topology manipulation. ChemRxiv doi: https://doi.org/10.26434/chemrxiv-2022-dhswc, 24 May 2022, preprint: not peer reviewed.

60. Bayly,C.I., Cieplak,P., Cornell,W. and Kollman,P.A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.*, **97**, 10269–10280.

61. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.

62. Ivani,I., Dans,P.D., Noy,A., Pérez,A., Faustino,I., Hospital,A., Walther,J., Andrio,P., Goñi,R. and Balaceanu,A. (2016) Parmbsc1: a refined force field for DNA simulations. *Nat. Methods*, **13**, 55–58.

63. Wang,J., Wolf,R.M., Caldwell,J.W., Kollman,P.A. and Case,D.A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.*, **25**, 1157–1174.

64. Kovačič,M., Podbevšek,P., Tateishi-Karimata,H., Takahashi,S., Sugimoto,N. and Plavec,J. (2020) Thrombin binding aptamer G-quadruplex stabilized by pyrene-modified nucleotides. *Nucleic Acids Res.*, **48**, 3975–3986.

65. Vasquez,K.M., Narayanan,L. and Glazer,P.M. (2000) Specific mutations induced by triplex-forming oligonucleotides in mice. *Science*, **290**, 530–533.

66. Bomholt,N., Osman,A.M.A. and Pedersen,E.B. (2008) High physiological thermal triplex stability optimization of twisted intercalating nucleic acids (TINA). *Org. Biomol. Chem.*, **6**, 3714–3722.

67. Géci,I., Filichev,V.V. and Pedersen,E.B. (2007) Stabilization of parallel triplexes by twisted intercalating nucleic acids (TINAs) incorporating 1, 2, 3-triazole units and prepared by microwave-accelerated click chemistry. *Chem. Eur. J.*, **13**, 6379–6386.

68. Bomholt,N., Osman,A.M. and Pedersen,E.B. (2008) High physiological thermal triplex stability optimization of twisted intercalating nucleic acids (TINA). *Org. Biomol. Chem.*, **6**, 3714–3722.

69. Okamura,H., Taniguchi,Y. and Sasaki,S. (2013) N-(Guanidinoethyl)-2′-deoxy-5-methylisocytidine exhibits selective recognition of a CG interrupting site for the formation of anti-parallel triplexes. *Org. Biomol. Chem.*, **11**, 3918–3924.

70. Okamura,H., Taniguchi,Y. and Sasaki,S. (2014) An isocytidine derivative with a 2-amino-6-methylpyridine unit for selective recognition of the CG interrupting site in an antiparallel triplex DNA. *ChemBioChem*, **15**, 2374–2378.

71. Okamura,H., Taniguchi,Y. and Sasaki,S. (2016) Aminopyridinyl-pseudodeoxycytidine derivatives selectively stabilize antiparallel triplex DNA with multiple CG inversion sites. *Angew. Chem. Int. Ed. Engl.*, **55**, 12445–12449.

72. Taniguchi,Y., Miyazaki,M., Matsueda,N., Wang,L., Okamura,H. and Sasaki,S. (2018) Stable and selective antiparallel type triplex DNA formation by targeting a GC base pair with the TFO containing one N2-phenyl-2′-deoxyguanosine. *Chem. Pharm. Bull. (Tokyo)*, **66**, 624–631.

73. Wang,L., Taniguchi,Y., Okamura,H. and Sasaki,S. (2017) Effect of the 3-halo substitution of the 2′-deoxy aminopyridinyl-pseudocytidine derivatives on the selectivity and stability of antiparallel triplex DNA with a CG inversion site. *Bioorg. Med. Chem.*, **25**, 3853–3860.

74. Chen,Y., Murayama,K., Kashida,H., Kamiya,Y. and Asanuma,H. (2020) A triplex-forming linear probe for sequence-specific detection of duplex DNA with high sensitivity and affinity. *Chem. Commun.*, **56**, 5358–5361.

75. Filichev,V.V. and Pedersen,E.B. (2009) DNA-conjugated organic chromophores in DNA stacking interactions. *Wiley Encycl. Chem. Biol.*, **1**, 493–524.

76. Okamoto,A., Saito,Y. and Saito,I. (2005) Design of base-discriminating fluorescent nucleosides. *J. Photochem. Photobiol. C: Photochem. Rev.*, **6**, 108–122.

77. Wilson,J.N. and Kool,E.T. (2006) Fluorescent DNA base replacements: reporters and sensors for biological systems. *Org. Biomol. Chem.*, **4**, 4265–4274.

78. Schneider,U.V., Geci,I., Johnk,N., Mikkelsen,N.D., Pedersen,E.B. and Lisby,G. (2011) Increasing the analytical sensitivity by oligonucleotides modified with para- and ortho-twisted intercalating nucleic acids–TINA. *PLoS One*, **6**, e20565.

79. Van Daele,I., Bomholt,N., Filichev,V.V., Van Calenbergh,S. and Pedersen,E.B. (2008) Triplex formation by pyrene-labelled probes for nucleic acid detection in fluorescence assays. *ChemBioChem*, **9**, 791–801.

80. Doluca,O., Withers,J.M., Loo,T.S., Edwards,P.J., González,C. and Filichev,V.V. (2015) Interdependence of pyrene interactions and tetramolecular G4-DNA assembly. *Org. Biomol. Chem.*, **13**, 3742–3748.