Article:
Evans, K, Bhamra, I, Wheelhouse, RT et al. (3 more authors) (2015) Stabilization of a bimolecular triplex by 3′-S-phosphorothiolate modifications: An NMR and UV thermal melting investigation. Chemistry-A European Journal, 21 (19). 7278 - 7284. ISSN 0947-6539

https://doi.org/10.1002/chem.201500369

Stabilization of a bimolecular triplex by 3'-S-phosphorothiolate modifications: an NMR and UV thermal melting investigation

Kathryn Evans,[a] Inder Bhamra,[b] Richard T. Wheelhouse,[c] John R.P. Arnold, [d] Richard Cosstick,[b] and Julie Fisher*[a]

Abstract: Triplexes formed from oligonucleic acids are key to a number of biological processes. They have attracted attention as molecular biology tools and as a result of their relevance in novel therapeutic strategies. The recognition properties of single stranded nucleic acids are also relevant in third strand binding. Thus there has been considerable activity in generating such moieties, referred to as triplex forming oligonucleotides (TFOs). Triplexes composed of Watson-Crick (W-C) base-paired DNA duplexes and a Hoogsteen base-paired RNA strand are reported to be more thermodynamically stable than those in which the third strand is DNA. Consequently synthetic efforts have been focused on developing TFOs with RNA-like structural properties. Here we describe structural and stability studies of such a TFO; composed of deoxynucleic acids, but with 3'-S-phosphorothiolate (3'-SP) linkages at two sites. The modification results in an increase in triplex melting temperature as determined by UV absorption measurements. ^{1}H NMR analysis and structure generation for the (hairpin) duplex component and the native and modified triplexes revealed that the double helix is not significantly altered by the major groove binding of either TFO. However the triplex involving the 3'-SP modifications is more compact. The 3'-SP modification was previously shown to stabilise G-quadruplex and i-motif structures and therefore is now proposed as a generic solution to stabilising multi-stranded DNA structures.

Keywords: Triplex DNA• NMR spectroscopy • 3'-S-phosphorothiolate • CNS-solve • UV thermal melting.

Introduction

Triple stranded nucleic acid structures were first reported in 1957[1] following observations of a complex with a 1:2 stoichiometry for mixtures of poly r(A) and poly r(U) respectively. The triplex repertoire quickly expanded to systems containing r(G) and r(C)[2], and then the DNA equivalents[3]. Subsequently mixed DNA/RNA triplexes[4] were demonstrated. In these intermolecular systems the strands are differentiated on the basis of the hydrogen bonding network; the duplex being formed through Watson-Crick (W-C) base-pairing with the third strand interacting in the major groove through Hoogsteen hydrogen bonding[5]. It has been established that irrespective of whether deoxyribose or ribose sugars are present triplex formation is sequence specific and is restricted to polypurine-polypyrimidine tracts[6].

The occurrence of triplexes in vivo is well documented; they have been detected in the nuclei of cells from insects, nematodes, and mammals including humans,[7.8] and they have been implicated in a diverse range of biological processes such as transcription[9], RNA splicing[10,11], and chromatin organisation[12]. Additionally, triplex forming oligonucleotides (TFO) have shown promise as tools for a variety of applications; due to their ability to sequence-specifically recognise double stranded DNA targets, and triple helix target sites (i.e. stretches of homopurine-homopyrimidine tracts) have been found in both prokaryotes and eukaryotes, often being over represented in the latter.[6]

Of the various combinations possible for intermolecular triplexes it has been shown that the most stable has a W-C double helical DNA component and an RNA third strand.[13-18] Consequently such hybrid triplexes have been the focus of much interest for more than two decades.[16-18] Of course the issues associated with working with single stranded RNA have been well rehearsed[19] in other arenas and have needed to be addressed when designing TFOs.

A wide range of chemical variants is now available which mimic RNA residues in functional or structural terms; these include 2'-O-methyl analogues[20] and the so-called locked nucleic acids (LNA).[21,22] These modifications impose an RNA-like conformation (i.e. with C3'-endo or 'north' sugar puckers) on the modified residue and adjacent nucleotides whilst reducing or removing susceptibility to ribonucleases. NMR studies have shown[23a,b] that the 3'-SP linkage also promotes the

[a] Dr. K. Evans, Dr. J. Fisher
School of Chemistry
University of Leeds
Leeds, LS2 9JT (UK)
Fax: (+44) 113 3436577
E-mail: j.fisher@chem.leeds.ac.uk

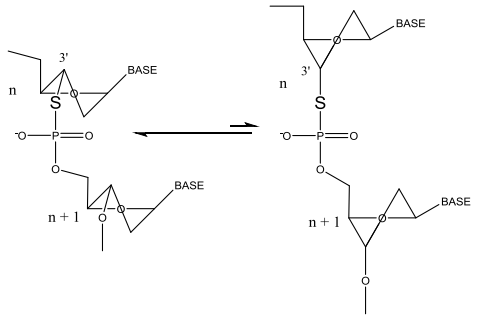
[b] Dr. I Bhamra, Prof. R. Cosstick
Department of Chemistry
University of Liverpool
Liverpool, L69 7ZD (UK)

[c] Dr. R.T. Wheelhouse
School of Pharmacy
University of Bradford
Bradford, BD7 1DP (UK)

[d] Dr. J.R.P. Arnold
Selby College
Selby, YO8 8AT (UK)

Supporting information for this article is available on the WWW under http://www.chemeurj.org/ or from the author.

C3'-endo conformation (Scheme **1**) and has been used to investigate the conformation and thermodynamic stability of nucleic acid duplexes[23], G-tetraplexes[24], and i-motifs[25].



Scheme 1. Illustration of the effect of a 3'-S-phosphorothiolate on the conformational equilibrium of deoxynucleic acids. The equilibrium lies significantly to the left hand side with both the 'n' and 'n + 1' sugar pucker being C3'-endo. The right hand side depicts the C2'-endo conformation typical of deoxynucleic acids.

Herein we report the first use of the 3'-S-phosphorothiolate moiety in (the Hoogsteen base paired strand of) a triplex (Scheme **2**); composed of a triethylene glycol (TEG)-linked hairpin duplex and a TFO. Inclusion of the 3'-SP linkage in the TFO strand results in enhanced thermal stability and also induces the same local changes in sugar conformation as those previously observed in duplex structures. It is also noted that the modified triplex is a more intimate complex than the non-modified system.

TFO 5'- $C_{25}T_{26}C_{27}C_{28}T_{29}T_{30}T_{31}C_{32}T_{33}T_{34}C_{35}C_{36}$

Hairpin dsDNA 5'- $G_1\ A_2\ G_3\ G_4\ A_5\ A_6\ \ A_7\ G_8\ A_9\ A_{10}\ G_{11}G_{12}$ — TEG
3'- $C_{24}T_{23}C_{22}C_{21}T_{20}T_{19}dU_{18}C_{17}T_{16}T_{15}C_{14}C_{13}$

TEG = triethylene glycol linker

Scheme 2. Sequence of the hairpin (residues 1-24) showing the TEG loop, and the parallel homopurine TFO sequence. In the modified system the 3'-S-modification is placed between T_{30} and T_{31}, and between T_{33} and T_{34} (indicated in green). The single deoxyuridine residue was introduced into the hairpin so as to simplyfy the resonance assignment process.

Results and Discussion

UV thermal melting: The UV absorbance at 260 nm was measured as a function of temperature for the hairpin duplex, the natural triplex, and the 3SP-modified triplex. These measurements were made for samples at pH 5-7.

For the hairpin a single phase sigmoidal melting curve was observed which was invariant to pH (Figure **S1** in the Supporting Information) revealing a melting temperature of 57.7 +/- 0.1 °C. It should be noted that this is slightly lower than the values determined for the hairpin in the presence of the TFO (Table 1). (We hypothesise this may be due to changes in ionic strengths due to differing counter ion concentrations as a result of increased total DNA concentrations between hairpin and triplex solutions.) The melting transition occurred over a relatively wide temperature range (ca. 30 °C) which suggests that the Gibbs free energy for hairpin unfolding has a low temperature dependence.[26]Assuming a two state equilibrium (folded – unfolded) it was possible to determine first an affinity constant (K_a) and then, from a plot of $\ln(K_a)$ against reciprocal temperature to derive standard enthalpy and entropy changes for folding and the Gibbs free energy change associated with this transition (Table **S1** in the Supporting Information). Folding is clearly enthalpically driven, a small entropy contribution resulting in a favourable negative $\Delta G°$.

The melting profiles for the two triplexes appear essentially the same at pH 7, dominated by the significant absorbance change at approximately 60 °C. However on closer inspection a small absorbance change is detected in the melting curve at approximately 21 °C for the modified triplex (Figure **S2(b)** in the Supporting Information). Progressive reduction in the pH resulted in this biphasic behaviour becoming increasingly apparent (Figure **1** and Figure **S3** in Supporting Information); the first transition resulting from dissociation of the single strand from hairpin (the melting temperature referred to as $T_{m,t}$), the second the unfolding of the hairpin duplex ($T_{m,d}$).[27.28] It is apparent that the modified triplex is more stable than the natural system, with a $\Delta T_{m,t}$ of 5.9 +/- 0.1 °C (i.e. 2.9 °C per modification) (Table 1). Assuming two separate, two-state equilibria (triplex – duplex and folded hairpin duplex – unfolded hairpin) thermodynamic parameters were again determined. If we assume that the enthalpies are additive [29] then the overall enthalpy change for the natural triplex ranges from -219 to -246 kJ mol^{-1} (depending on pH) whereas that for the modified triplex is -221 to -281 kJ mol^{-1} .

In a similar manner to above, the effect of salt on the thermal stability of the hairpin and natural triplex was analysed (Table **S2** and figure **S4** in Supporting Information). The presence of 100 mM NaCl resulted in a significant stabilisation of the hairpin structure with concomitant destabilization of the triplex, in line with previous observations. [27,28]

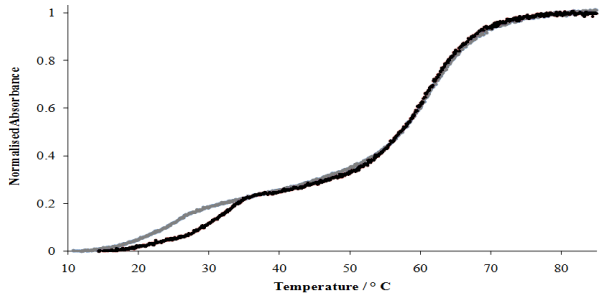


Figure 1. Normalised melting profiles for the non-modified triplex (blue) and modified triplex (red) at pH 6. Absorbance was measured at 260 nm and samples were heated from 10 - 85 °C at a rate of 1 °C /min. Sample concentrations were 10 µM in 20 mM sodium phosphate buffer.

Assignment of NMR spectrum of hairpin duplex: A combination of DQF-COSY, TOCSY and NOESY spectra and the assumption of a B-type double helix model was used to assign the ^{1}H NMR spectrum of the hairpin.[30] Almost all of the protons were assigned; the exceptions being the H5'/H5" of A_{10}, T_{16}, C_{17}, and dU_{18} and the exchangeable proton resonances for residues at the extremities (resonance assignments are provided in Table **S3** together with some sample spectra (Figure **S5)** in Supplementary Information). We had previously found[31], in agreement with literature reports[32], that the presence of the TEG linker has no significant effect on the structure of the duplex. The same was true in this study with 'normal' nOe constraints being observed along the length of the hairpin including involving the residues adjacent to the TEG tether. No nOe connections were observed between nucleotide residues and the TEG protons.

Prior to quantitative structural analysis a qualitative view was taken by comparing key nOe intensities; specifically between H2'-H6/H8[33] and H1' – H2'/H2"[34] sets of protons. In this way all sugar puckers were determined to be C2'-endo (south). Additionally almost all glycosidic bond orientations were found to be anti; around C_{13} the orientation appears to be syn, and the geometry for T_{16} could not be determined due to resonance overlap.

Table 1. Melting temperatures and van't Hoff enthalpies for non-modified and modified triplex formation calculated over a range of pH values. Melting temperatures are given in °C (± 0.5°C), ΔH_{vH} are given in kJ mol^{-1} (± 10 %). (1) denotes triplex to duplex unfolding, (2) denotes duplex to single strand melting. * While biphasic behaviour is noted at pH 7 for the modified triplex, the absorbance change was very small making calculation of thermodynamic parameters unreliable.

Triplex	pH	T_m (1)	ΔH_{vH} (1)	T_m (2)	ΔH_{vH} (2)
Non-modified	7	-	-	60.1	-203
	6.5	-	-	59.2	-180
	6	27.5	– 65	60.3	-154
	5	34.5	– 108	58.6	-138
Modified	7*	20.7*	-	59.8	-204
	6.5	26.9	– 57	60.3	-164
	6	30.7	– 88	59.2	-158
	5	41.3	– 151	58.8	-130

Assignment of NMR spectrum for natural and 3'-SP-modified triplexes: Resonance assignment followed a similar procedure to that adopted for the hairpin duplex; data recorded at pH 5 were utilised as there was a greater number of, and

more intense, nOe's as opposed to more significant chemical shift dispersion. The formation of a triplex was immediately apparent through the observation of C^+ imino (14-16 ppm) and amino (9.5 – 10.5 ppm) protons.[35] Four of the six C^+ imino resonances could be unambiguously assigned; the two iminos at 15.4 and 15.7 ppm are broad and did not display nOe connections, both of these observations are consistent with these belonging to the terminal cytidines C25 and C36.

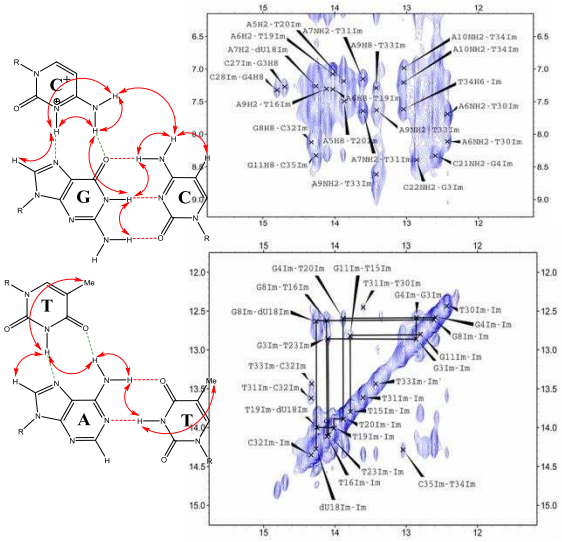


Figure 2. Expected through-space connections for the triplex (left hand side) and sections of the NOESY spectrum for the non-modified triplex indicating nOe connections involving imino and amino protons.

The location and orientation of the third strand was established through the observation of nOes involving both WC and Hoogsteen hydrogen bonded imino and amino protons together with a number of connections from these to H6/H5/Me protons (Figure **2**).

Tables **S4** and **S5** in the Supporting Information contain all the completed resonance assignments. Figures **S6** and **S7** provide some sample spectra. As found previously[36] the H3' protons for the two residues with the 3'-SP modification were approximately 1 ppm upfield of the those for the non-sulfur modified system. A comparison of chemical shifts for the natural and modified triplex indicated that the effect of the third strand on the hairpin was similar for both sequences. It was however noted that the chemical shifts of a number of H6/H8 protons of the third strand differed for residues adjacent to the sites of modification suggestive of a slightly different base-stacking arrangement[36] (see below and also Figure **S6** in Supporting Information). Again a qualitative assessment of the conformation of the complexes was made. In the natural triplex all sugar puckers were found to be C2'-endo. In the modified system a change in the sugar pucker from C2'-endo to C3'-endo was observed for the modified T_{30} and T_{33} residues and the T_{31} and T_{34} (n + 1) residues, a very similar effect to that previously observed in duplex systems containing the 3'-SP linkage (Scheme **1**).[36]

Structure generation for the hairpin duplex: Structures were generated for the hairpin following two approaches; based on nOe cross-peak volumes measured from a NOESY spectrum recorded with a 200 ms mixing time, and then based on nOe build-up curves generated from spectra recorded with mixing times ranging from 50-250 ms (an illustrative build-up curve is provided in Figure **S9** in Supporting Information). These data sets provided 13 and 12 nOe constraints per-residue, respectively. Average structures were calculated from the 10 lowest energy structures generated and both of the averaged structures were found to adopt a right-handed helical structure with B-type characteristics; the helix diameter being 19.2 and 18.9 Å for the 200 ms and build-up structures respectively (Figure **S10** in Supporting Information). The energy associated with the various constraints for the hairpin duplex compare favourably with literature values [37]; and is provided in Table **S6** in Supporting Information. It is notable that the two approaches adopted here generate structures with essentially the same energy terms. Pairwise atomic RMSD values describe the precision of a set of structures and the low values (averaged over 24 residues to less than 0.2 and 0.5 Å for the 200 ms and build-up structures respectively) produced here suggest well defined structures have been generated[38].

Structure generation for the non-modified and modified triplex: The structure generation process adopted for the triplexes followed essentially the same format as that described for the hairpin, using nOe constraints measured from spectra recorded with a 200 ms mixing time. An average of 12 nOe restraints was available for both triplexes, covering all residues (but again excluding the TEG link). Both systems adopted a right-handed helical form with triplex diameter of 19.8 and 19.1 Å for the non-modified and modified triplex respectively (Figure **S11** in Supporting Information). The atomic RMSD values calculated from ensembles of the 10 lowest energy structures, for each system, were significantly lower than 1 Å (Table **S7** in Supporting Information) indicating excellent convergence characteristics. Amongst other parameters adjacent phosphate-phosphate distances were measured. A typical value for this separation in B-type helices (with south sugar puckers) is 7.0 Å, while for an A-type helix (with north sugar pucker) is 5.9 Å. For the non-modified triplex P-P distances ranged from 6.7 – 7.2 Å (excluding the first link) and from 5.7 – 7.3 Å for the modified triplex with the lower (A-helix-type) values arising at the site of modification and at the n + 1 site (Table **S8** in Supporting Information) (Figure **3**).

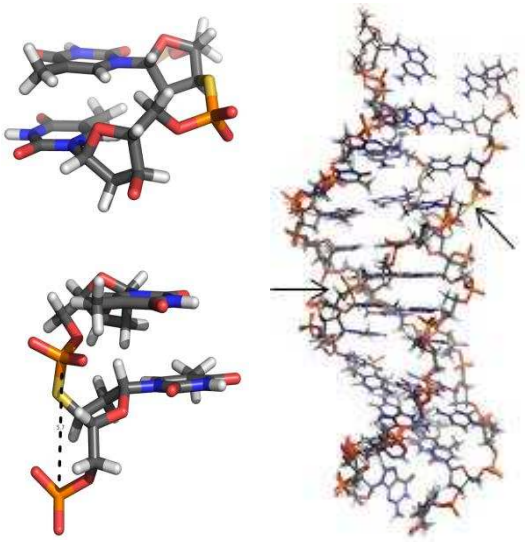


Figure 3. Local geometries around one of the *3'*-S phosphorothiolate links in the TFO (left hand side). The dotted line indicates a 5.7 Å separation between the phosphorus atoms of residue 6 and 7. The 10 lowest energy structures for the modified triplex (right hand side); the arrows indicate the location of the phosphorothiolate groups.

Essentially identical average backbone torsion angles were observed for the W-C components of each triplex, notable exceptions being the χ and γ angles relating to the glycosidic and 5'C- 4'C sugar link; the former being almost 20° different for the hairpin component of the non-modified triplex (30° when compared to the hairpin alone) . Differences were more apparent when comparing torsion angles in the TFOs; almost all of the torsion angles (but most notably α and γ) of the modified TFO, when compared to the non-modified version, were more A-helix-like than B-helix-like (see Tables **S9(a)** and **S9(b)** in Supporting Information).

Discussion

The aim of these studies was to determine the consequences for triplex structure and stability of the introduction of *3'*-S phosphorothiolate linkages to the TFO. This modification has previously been incorporated in a number of DNA motifs in which it has been observed to enhance their thermodynamic stability[23]. The (hairpin) duplex investigated has previously been shown to form a parallel triplex[13] and ethylene glycol-based linkers have been employed to reduce the molecularity of the triple helix system[32]. The single deoxyuridine residue was introduced so as to ease the resonance assignment process.

Initial insight to the thermodynamic stability of the hairpin and the two triplexes was obtained through UV thermal melting studies; as a function of pH and then salt concentration. In the absence of a TFO the hairpin melting transition was observed to be independent of pH but moved to a higher temperature (ΔT + 8 °C) in the presence of 100 mM NaCl, in line with previous observations[39]. In contrast the melting temperature for the triplexes was reduced by the presence of salt at each pH

tested, while reduction in pH in the absence of salt resulted in an increase in the temperature for the triplex melting transition; this is consistent with increased protonation of the (TFO) cytidines (pKa for N3 is ca. 4.5-4.8, depending on temperature and ionic strength [40]). At each pH the modified triplex was more thermodynamically stable than the natural system, and thus the 3'-SP modification extends the pH range for triplex formation involving cytidines. The level of stabilisation induced by this modification in the triplex (ca. 2.5 °C per 3'-SP modification) is greater than that previously observed in DNA-RNA duplexes (ca. 1.5 per 3'-SP modification) and very similar to that observed for the i-motif . [23-25] We have previously demonstrated that increasing the number of 3'-SP modifications enhances in an additive manner the thermodynamic stability of thymine rich DNA-RNA hybrid duplexes[23c], and also an i-motif structure which shows a very similar pH dependence to the triplex studied here[25]. On the basis of the observations made here it would seem reasonable to suggest that a similar effect would be apparent for triplexes involving longer thymine sequences than in the current system. It is possible that incorporation of T-SP-C or C-SP-C[25] modifications could extend the pH stability of the current triplex sequence in the same manner, due to further favourable conformational changes in the TFO and this is clearly something that should be explored if C-containing 3'-S modified TFOs are to be used for therapeutic purposes.

Our NMR studies revealed firstly that the TEG-linked hairpin adopted a standard B-helical form up to and including the point at which the TEG group was linked. Subsequently the parallel orientation of the TFO with respect to the purine tract of the hairpin was confirmed through the observation of a series of TFO-hairpin nOes. Structure generation based on nOe and other conformational restraints enabled the geometry around the modification sites to be established and compared with the non-modified system. The expected sugar pucker change (from C2'-endo to C3'-endo') was observed for the modified T_{30} and T_{33} residues and the T_{31} and T_{34} (n + 1) residues, with the concomitant reduction in the P-P backbone distance (from ca. 6.8 to 5.7 Å). This conformational shift appears to result in enhanced base stacking interactions in the (modified) single strand; this has been reported in previous studies of short *3'*-S-modified sequences[36] suggesting some pre-ordering of the single strand[36c]. Thus the enhanced stability as judged by UV thermal melting measurements could be due (at least in part) to the additional stability of the modified TFO (as compared to the natural TFO). Nevertheless the ultimate modified triplex structure has features which suggest that the folded form of the molecule is also important. For example, the width of the major groove of the hairpin component increased slightly on changing from the natural to modified TFO (17.8 and 18.2 Å respectively) the diameter of the resultant triplexes was found to be narrower when the modified TFO was utilised (the natural triplex has a diameter 0.7 Å greater).

Conclusions

Triple helical nucleic acids have been implicated in a diverse range of biological processes and functions. Homopurine-homopyrimidine tracts of DNA, which are required for triplex formation, are overexpressed in genes associated with several diseases such as HIV[41], diabetes[7] and cancer[42]. It is therefore of interest to design TFOs which form highly stable triplexes when bound to a DNA duplex; these could potentially prevent DNA replication and, ultimately , gene expression when utilised as therapeutic agents. The 3'-S-phosphorothiolate linkage is unusual in that it has been shown to stabilise G-quadruplex, i-motif and now triplex structures and therefore represents a generic solution to stabilising multi-stranded DNA structures. Additionally, it is able to achieve this with a single atom substitution and thus may prove particularly useful where there is a need to stabilise these structures without introducing steric bulk which might interfere with interactions with biological systems. The 3'-S-P linkage also introduces beneficial features that are not available from other stabilising modifications such as LNA or 2'-O-methyl RNA and enable the DNA structures to be interrogated and manipulated by additional techniques. For example, ^{31}P NMR provides information on the backbone torsion angles and the characteristic downfield shift in the ^{31}P NMR signals from the introduction of sulfur considerably simplifies assignment of resonances and the extraction of this data[36]. In addition, the very specific hydrolysis of S-P bonds with aqueous silver ions, [43] provides distinctive cleavage sites that liberate a 5'-terminus which, for example, is ready for ligation, should further manipulation of these DNA structures be required.

Experimental Section

Materials: The non-modified and TEG-modified oligomers were synthesised using standard phosphoramidite chemistry. The TEG linker (Spacer-CE Phosphoramidite 9) was purchased from Link Technologies Ltd. The 3'-S-phosphorothiolate modified (third) strand was prepared following our literature protocol[43]. All oligomers were purified by reverse phase HPLC, lyophilised and re-dissolved in 20 mM sodium phosphate buffer (in either D_2O or 9:1 H_2O:D_2O); the pH adjusted as required by the addition of deuterium chloride or sodium deuteroxide. Sample concentrations were measured using a Hewlett-Packard 8452A diode array spectrophotometer; the molar extinction coefficients (ε) were calculated[44] ; for the hairpin ε = 235970 cm^{-1} M^{-1}, and for single strands ε = 94000 cm^{-1} M^{-1}.

Spectroscopy: Prior to UV and NMR experiments samples were annealed by heating to 95 °C for 5 min and then leaving to cool slowly to room temperature. UV thermal melting spectra were recorded using a Cary 400 Bio spectrophotometer. A

20mM phosphate buffer sample was heated alongside the DNA samples for blanking at each temperature. Absorbance was measured at 260 nm over the temperature range 10 – 85 °C at a rate of 1 °C per min, and readings taken every 0.1 °C. At least two runs were conducted for each sample, with melting temperatures differing between repetitions by < 0.2 °C. Melting temperatures were determined from fraction-folded versus temperature plots. NMR spectra were recorded using either a Varian Unity INOVA 500 or a 4 channel Varian INOVA 750 MHz instrument equipped with 5 mm z-gradient triple resonance cryoprobe. Spectra were recorded at 20 °C for samples in D_2O-buffer and at 2 °C for H_2O:D_2O-buffered samples. 1-d presat spectra (at 500 MHz) were recorded with a spectral width of 6000 Hz, a 3 µs high power pulse (equivalent to a 30 ° tilt angle) and a 1 s recycle delay, in 8k pairs of data points. DQF-COSY, TOCSY and NOESY spectra were acquired using 2048 pairs of data points in t_2 and 512 t_1 increments. Mixing times of 50 to 400 ms were used for the NOESY experiments and 80 ms for the TOCSY. A 1 s recycle delay was employed during which the residual water signal was saturated. Similar parameters were employed for the 750 MHz spectra. For samples in 9:1 H_2O:D_2O-buffer spectral widths from 15 000 to 18 000 Hz were used. All data were processed using Varian VNMR 6.1C software using a Gaussian function, and visualised, assigned and quantified using SPARKY 3.111 [45].

Structure calculations: Structures were generated for the hairpin duplex, non-modified triplex and modified triplex using CNS_solve 1.1[46] . New topology and parameter entries were included for the TEG link (using standard covalent bond lengths, angles and partial charges), the *3'*-S modification, and the protonated cytidine[47]. NOe build-up curves were linear to a mixing time of 250 ms hence distance constraints were extracted from NOESY spectra recorded with a 200 ms mixing time. The cytidine H5-H6 distance (2.4 Å) was used to normalise cross-peak volumes. Loose errors of +/- 30 % were applied to distances between non-exchangeable protons, increasing to 50 % for terminal base-pairs. Error boundaries of +/- 40 % were applied to distance constraints involving exchangeable protons. All hydrogen bonding distances were taken from standard values with error boundaries of +/- 35 %. Planarity constraints were applied to all except the terminal base-pairs [48]. Backbone torsion angles were loosely constrained to a range which encompassed typical values for both A- and B-type helices, in accordance with the literature. [48,49]. For the modified triplex, torsion angles were not additionally constrained for the Hoogsteen bonded strand. [50] On average 12 nOe constraints were available per residue. Simulated annealing was employed with an initial extended starting structure for both hairpin and modified and non-modified single strands, using torsion angle dynamics followed by Cartesian molecular dynamics. Parameters used for simulated annealing were as follows; heating to 20000 K (phase 1 torsion), cooling 20000 to 1000 K (phase 2 torsion), slow cooling 1000 to 300 K (phase 3 Cartesian) followed by 1000 steps of conjugate gradient minimisation. In phase 1 and 2 the time step was 0.015 and dihedral scale factor of 5 were employed. In phase 3 and 4 the dihedral scale factor was 200 and 400 respectively. The nOe scale factor was 150 throughout. From phase 1 to 4 the scale factor for van der Waals interactions was varied from 0.1 to 4. The annealing procedure was performed ten times using different initial velocities and was terminated once 100 acceptable structures had been generated. Acceptable structures were generated having no nOe constraint great than 0.5 Å, no torsion angle violations greater than 5 °, RMSD of bonds from ideal values of less than 0.02 Å and of angles not exceeding 2.0 ° [51]. The lowest energy acceptable structures were submitted to a further round of Cartesian molecular dynamics. Resultant structures were visualised with PyMOL[52] and helical parameters extracted using CURVES$^+$ [53] and groove widths using 3DNA [54]. The coordinates for the structures generated are included in the Supplementary Information section (Table S10-S12).

Acknowledgements

We would like to thank Dr. Arnout Kalverda, the Astbury Centre for Structural Molecular Biology, University of Leeds, UK, for recording 750 MHz NMR spectra.

[1] G. Felsenfeld, D. R. Davies and A. Rich, J. Am. Chem. Soc. **1957**, 79, 2023-2024.

[2] M.N. Lipsett, Biochem. Biophys.Res. Commun. **1963**, 11, 224-228.

[3] M. J. Chamberlain and D. L. Patterson, J. Mol. Biol. **1965**, 12, 410-428.

[4] A. Rich, Proc. Natl. Acad. Sci. USA, **1960**, 46, 1044-1053.

[5] K. Hoogsteen, Acta Crystallogr., **1959**, 12, 822-823.

[6] S. Neidle, Nucleic Acid Structure, Oxford University Press, Oxford. **1999**.

[7] F.A. Buske, J.S. Mattick and T.L. Bailey, RNA Biology, **2011**, 8, 427-439.

[8] M. Ohno, T. Fukagawa, J. Lee and T. Ikemura, Chromosoma, **2002**, 111, 201-213.

[9] I. Martianov, A. Ramadass, A.S. Barros, N. Chow and A. Akoulitchev, Nature, **2007**, 445, 666-670.

[10] L.J. Collins, C.g. Kurland, P.Biggs and D. Penny, J. Heredity, **2009**, 100, 507-604.

[11] B.M. Lunde, C. Moore and G. Varani, Nat. Rev. Mol. Cell Biol., **2007**, 8, 479-490.

[12] K.J. Hampel, G.D. Burkholder and J.S. Lee, Biochemistry, **1993**, 32, 1072-1077.

[13] A.R. Srinivasan and W.K. Olson, J. Am. Chem. Soc. **1998**, 120, 484-491

[14] R.W. Roberts and D.M. Crothers, Science, **1992**, 258, 1463-1466.

[15] G.M. Hashem, J.-D. Wen, Q. Do and D.M. Gray, Nucleic Acids Res., **1999**, 27, 3371-3379.

[16] C. Escude, J.C. Francois, J.S. Sun, G. Ott, M. Sprinzl, T. Garestier and C. Helene, Nucleic Acids Res., **1993**, 21, 5547-5553.

[17] M.J.J. Bloomers, F. Natt, W. Jahnke and B. Cuenoud, Biochemistry, **1998**, 37, 17714-17725.

[18] J.L. Asensio, R. Carr, T. Brown and A.N. Lane, J. Am. Chem. Soc., **1999**, 121, 11063-11070.

[19] F. Eckstein, Oligonucleotides and Analogues, Oxford University Press, Oxford. **1991**.

[20] H. Torigoe, R. Shimizume, A. Sarai and H. Shindo, Biochemistry, **1999**, 38, 14653-14659.

[21] S. Karkare and D. Bhatnagar, Appl. Microbiol. Biotechnol., **2006**, 71, 575-586.

[22] J.J. Sorensen, J.T. Nielsen and M. Petersen, Nucleic Acids Res., **2004**, 32, 6078-6085.

[23] (a) A.P.G. Beevers, K.J. Fettes, I.A. O'Neil, S.M. Roberts, J.R.P. Arnold, R. Cosstick and J. Fisher, Chem. Commun., **2002**, 1458-1459. (b) A.P.G Beevers, E.M. Witch, B.C.N.M. Jones, R. Cosstick, J.R.P. Arnold and J. Fisher, Magn. Reson. Chem. **1999,** 37, 814-820. (c) J. Bentley, J.A. Brazier, J. Fisher and R. Cosstick, Org. Biomol. Chem., **2007**, 5, 3698-3702.

[24] M.M. Piperakis, J.W. Gaynor, J. Fisher and R. Cosstick, Org. Biomol. Chem. **2013**, 11, 966-974.

[25] J. A. Brazier, J. Fisher and R. Cosstick, Angew. Chem., Int. Ed. Engl., **2006**, 45, 340-345.

[26] J.L. Mergny and L. Lacroix, Oligonucleotides, **2003**, 13, 515-537.

[27] D. Leitner, W. Schroder and K. Weisz, Biochemistry, **2000**, 39, 5886-5892.

[28] A.M. Soto, J. Loo and L.A. Marky, J. Am. Chem. Soc., **2002**. 124, 14355-14363.

[29] H.-T. Lee, S. Arciniegas and L.A. Marky, J. Phys. Chem. B., **2008**, 112, 4833-4840.

[30] S.S. Wijmenga and B.N.M. van Buuren, Prog. Nucl. Magn. Reson. Spec., **1998**, 32, 287-387.

[31] L.A. Shaw, PhD Thesis, University of Leeds, 2011.

[32] C.H. Gotfredsen, P. Schultze and J. Feigon, J. Am. Chem. Soc. **1998**, 120, 4281-4289.

[33] C. Glemarec, A. Nyilas, C. Sund and J. Chattopadhyaya, J. Biochem. Biophys. Met., **1990**, 21, 311-332.

[34] M.R. Conte, T.C. Jenkins and A.N. Lane, J. Biochem., **1995**, 229, 433-444.

[35] J. Feigon, K.M. Koshlap and F.W. Smith, Methods Enzymol., **1995**, 261, 225-255.

[36] (a) A.P.G. Beevers, E.M. Withc, B.C.N.M. Jones, R. Cosstick, J.R.P. Arnold and J.Fisher, Magn. Reson. Chem., **1999,** 37, 814-820. (b) H.K. Jayakumar, J.L. Buckingham, J.A. Brazier, N.G. Berry, R. Cosstick and J. Fisher, Magn. Reson. Chem.**2007**, 45, 340-345. (c) A.K. Collier, J.R.P. Arnold and J. Fisher, Magn. Reson. Chem. **1996**, 34, 191-196.

[37] J. Lee, V. Guelev, S. Sorey, D.W. Hoffman and B.L. Iverson, J. Am. Chem. Soc., **2004**, 126, 14036-14042.

[38] G. Varani, F. Aboul-ela and F.H.T. Allain, Prog. Nucl. Magn. Reson. Spec., **1996**, 29, 51-127.

[39] K.J. Breslauer, G.K.A. Michael and L. Johnson, Methods Enzymol **, 1995**, 259, 221-242.

[40] J.L. Mergny, L. Lacroix, X.G. Han, J.L. Leroy, and C. Helene, J. Am. Chem. Soc., **1995**, 117, 8887-8898.

[41] C.D. Pesce et al., Antiviral Res., **2005**, 66, 13-22.

[42] G.M. Carbone, E.M. McGuffie, A. Collier and C.V. Catapano, Nucleic Acids. Res., **2003**, 31, 833-843.

[43] J.S. Vyle, B.A. Connolly, D. Kemp and R. Cosstick, Biochemistry, **1992**, 31, 3012-3018.

[44] J.W. Gaynor, J. Bentley and R. Cosstick, Nat. Protoc, **2007**, 2, 3122-3135.

[45] J.D. Puglisi and I. Tinoco, Jr., Methods Enzymol., **1989**, 180, 304-325.

[46] T.D. Goddard and D.G. Kneller, Sparky 3, **2006**, University of California, San Fransisco, USA.

[47] A.T. Brunger et al., Acta Cryst. Sec. D : Biol. Cryst. **1998**, 54, 905-921.

[48] Parameters for protonated cytidine were supplied by Dr. C. Laughton, School of Pharmacy, University of Nottingham, UK,

[49] J.W. Pham, I. Radhakrishnan and E.J. Sontheimer, Nucleic Acids Res., **2004**, 32, 3446-3455.

[50] X.Y. Yang, X.G. Han C. Cross, S. Bare, Y. Sanghvi and X.L. Goa, Biochemistry, **1999**, 38, 12586-12596.

[51] E. Wang, K.M. Koshlap, P. Gillespie, P.B. Dervan and J. Feigon, J. Mol. Biol., **1996**, 257, 1052-1069.

[52] B.M. Znosko, T.W. Barnes, III, T.R. Krugh and D.H. Turner, J. Am. Chem. Soc., **2003**, 125, 6090-6097.

[53] http://pymol.org/

[54] C. Blanchet, M. Pasi, K. Zakrzewska and R. Lavery, Nucleic Acids. Res., **2011**, 39, W68-W73.

[55] http://chem.rutgers.edu/~xiangjun/3DNA/images/bp_step_hel.gif