DNA Structures

DOI: 10.1002/ange.200503110

Stabilization of the DNA I-Motif Structure by Incorporation of 3′-*S*-Phosphorothiolate Linkages**

John A. Brazier, Julie Fisher, and Richard Cosstick**

The telomeric regions of human chromosomes contain a cytosine-rich strand that can associate into a unique four-

[*] Dr. J. Fisher
Department of Chemistry
University of Leeds
Leeds, LS2 9JT (UK)
Fax: (+44) 113-343-656
E-mail: J.Fisher@chem.leeds.ac.uk
Dr. J. A. Brazier, Dr. R. Cosstick
Department of Chemistry
University of Liverpool
Crown Street, Liverpool, L69 7ZD (UK)
Fax: (+44) 151-794-3588
E-mail: rcosstic@liv.ac.uk

[**] This work was financially supported by the Biotechnology and Biological Sciences Research Council. We would like to thank Allan Mills (Liverpool) for obtaining mass spectra.

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

stranded and intercalated structure, commonly known as an i-motif.[1–3] This structure is formed from two parallel-stranded duplexes containing hemiprotonated cytosine–cytosine ($C{\cdot}CH^+$) base pairs (C = cyt = cytosine). The two duplexes run antiparallel to each other and are intimately associated by base-pair intercalation.[4–6] Stoichiometrically, the i-motif can be assembled through the association of four strands,[4] by two hairpins each containing two cytosine-rich regions[7] or by the folding of a single strand containing four cytosine stretches[8] (as is the case in the human telomeric sequence d-[$(CCCTA_2)_3CCCT$]).[1,3] Consistent with the hemiprotonated base pairs, the stability of the i-motif is pH dependent and is at a maximum at a pH value close to the pK_a value of cytosine.[9] At a neutral pH value, the i-motif is much less stable although it can still be observed in intramolecular systems.[1,9] The recent discovery of proteins that specifically bind the cytosine-rich strand of human telomeric sequences, combined with the stability of an intramolecular i-motif at physiological pH values, implies a biological role.[10] Detailed structural features of the i-motif have been revealed by both NMR and X-ray crystallographic techniques.[9,11] Viewed along the helical axis, the i-motif displays two wide grooves and two narrow grooves with close sugar–sugar contacts in the narrow groove.[8] The deoxyribose sugars adopt an *endo* conformation at C3′, which is typical for RNA A-type duplexes.[1,9]

To probe the importance of some of the structural features, a number of 2′-deoxycytidine analogues have been substituted into the i-motif. Surprisingly, despite the C3′-*endo*-conformational preference of ribonucleosides, melting studies have shown that the RNA i-motif is much less stable than the analogous DNA structure.[12,13] Destabilization by the ribose sugar appears to result from a steric clash between 2′-hydroxy groups in the narrow groove. In support of this steric effect, replacement of a single 2′-hydroxy group by the bulkier 2′-*O*-methyl substituent abolishes formation of the i-motif,[12] whereas the incorporation of D-arabinocytidine (2′-hydroxy group pointing away from the narrow groove) has a neutral effect.[14] Other analogues of deoxycytidine that have been studied in this context include 5-methyldeoxycytidine [9,15] and the phosphorothioate,[15,16] methylphosphonate,[15] and 3′-*N*-phosphoramidate[15] backbone modifications. However, none of these modifications stabilize the i-motif relative to the natural sequences. In one report, an intramolecular i-motif (derived from TCCTCCTTTTCCTCCT) was stabilized through replacement of thymidine by 5-propynyl deoxyuridine.[15] This modification presumably acts by stabilization of the loop regions of the structure through hydrophobic interactions rather than stabilizing the unique structural features of the i-motif.

Our previous studies have shown that incorporation of a 3′-*S*-phosphorothiolate linkage into DNA shifts the conformational preference of the sugar to which the sulfur atom is attached to the C3′-*endo* (north) pucker (Figure 1a). This alteration thus mimics the RNA conformation without the introduction of a substituent at the C2′.[17] In this study we have introduced either one or two phosphorothiolate linkages (Figure 1b) into d(TCCCCC) and report the first example of a deoxycytidine analogue that stabilizes an i-motif.

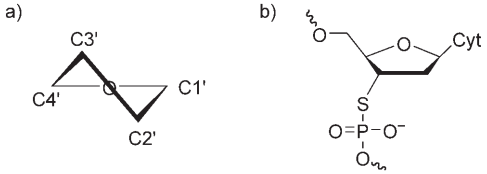

Figure 1. a) North (C3′-*endo*) sugar pucker. b) 3′-S-phosphorothiolate linkage. Cyt = cytosin-1-yl.

Oligonucleotides that contain phosphorothiolate linkages were synthesized by using a 3′-*S*-phosphorothioamidite derived from deoxycytidine and prepared as previously described.[18] The sequences studied are presented in Table 1. i-Motif structures were assembled in citrate buffer at pH 4.6 and the samples were stored, as previously described, at 4 °C for one week to give a single i-motif tetrad.[12]

Table 1: ESI mass spectrometry data and UV T_m data.

Sequence[a]	Mass calcd	Mass obs.[b]	T_m [°C][c]
d(TCCCCC)	1687.3221	1687.3330	46.0
d(TCC**Cs**CC)	1703.2993	1703.2780	48.5
d(TCC**Cs**C**Cs**C)	1719.2765	1719.2950	53.0
d(TC**Cs**C**Cs**CC)	1719.2765	1719.2991	54.5

[a] **Cs** indicates a cytidine phosphorothiolate analogue. [b] The measurement was carried out in negative mode and the sample prepared in aqueous methanol + diethylamine (0.1%). [c] Sodium citrate buffer (50 mM, pH 4.6), 5 μM strand concentration.

Both the natural and modified sequences exhibited melting behavior characteristic of an i-motif with an increase in absorption at 260 nm and a hypochromic transition at 295 nm. The melting temperature (T_m) of the unmodified i-motif is consistent with that previously reported for this sequence (48 °C) at a higher concentration (11.5 μM).[12] Incorporation of a single 3′-*S*-phosphorothiolate linkage at C4 [d(TCC**Cs**CC)] gave a modest increase in T_m (ΔT_m) of 2.5 °C but a much greater effect was seen when two modifications were introduced. The modifications of d(TCC**Cs**C**Cs**C) and d(TC**Cs**C**Cs**CC) resulted in ΔT_m values of 7.0 and 8.5 °C, respectively. The moderate increase of the singly modified i-motif may be related to the twofold symmetry axis of the tetrad present at C4. These data demonstrate the considerable stabilizing effect of the phosphorothiolate modification on the i-motif structure.

Although the hypochromic transition at 295 nm is supportive of an i-motif structure, further evidence was sought from solution NMR spectroscopic investigations. 1D ^{1}H NMR spectra recorded for d(TCC**Cs**CC) (Figure 2a) displayed four distinct imino proton resonances at chemical shifts greater than $\delta = 15$ ppm; the imino group at C6 was exchanged/broadened beyond observation under the conditions adopted. These signals arise from imino hydrogen atoms on protonated cytosines that are involved in hydrogen bonding and are thus characteristic of an i-motif. Non-hydrogen bonded imino protons of this type would be shifted

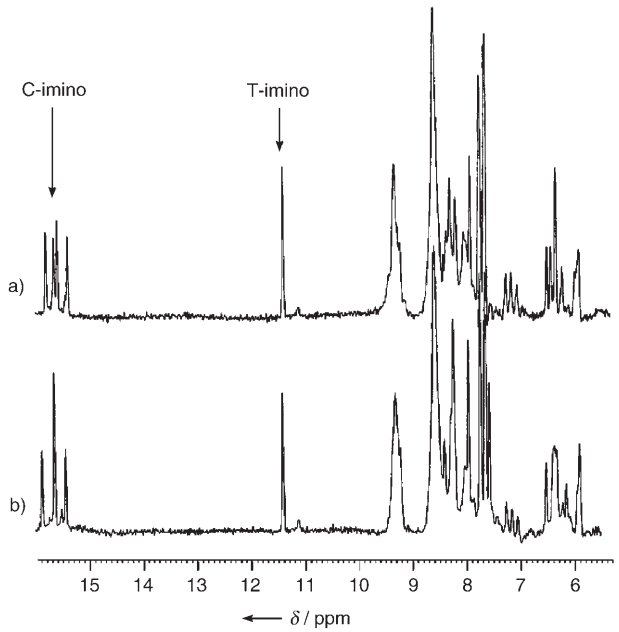

Figure 2. Downfield region of the 500 MHz 1D ^{1}H NMR spectra recorded for i-motifs a) d(TCC**Cs**CC) and b) d(TCCCCC). Conditions: 1 mM oligonucleotide in 50 mM citrate buffer, H_2O/D_2O solution (9:1), pH 4.6, 10 °C. The signal for water was suppressed by using the WATERGATE sequence.

several ppm upfield.[4,5] The single resonance at $\delta = 11.4$ ppm is due to the imino proton of the thymine residue and is consistent with the presence of a single i-motif structure. Overall, the spectrum is very similar to that obtained for the nonmodified sequence d(TCCCCC) (Figure 2b),[5] with some variation apparent in the midrange imino proton and the H1′ chemical shifts.

A comparison of NOESY spectra recorded for d(TCC**Cs**CC) with those obtained for d(TCCCCC)[4] enabled resonance assignments to be made, some of which are shown in Figure 3. It is interesting that the signal for the imino hydrogen atom at C3 is shifted downfield in d(TCC**Cs**CC) and likely reflects a change in conformation of the tetraplex in this region. The H1′ of C6 shifts downfield upon introduction of the phosphorothiolate linkage, which may reflect a tightening of the structure in this region. The NOESY data for d(TCC**Cs**CC) also exhibits nonsequential imino–imino and H1′–H1′ (Figure 3b) close-spatial relationships that are another characteristic of i-motif structures. These data are consistent with a fully intercalated structure with a base-pair stacking order of Thy1-Cyt6-Cyt2-Cyt5-Cyt3-Cyt4-Cyt4-Cyt3-Cyt5-Cyt2-Cyt6-Thy1 (bases from one duplex are underlined), which is the same base-pair order as that observed for the unmodified sequence.

In conclusion, the UV melting and NMR spectroscopic studies clearly demonstrate that phosphorothiolate incorporation stabilizes the i-motif with a minimum perturbation of the overall structure and is the first example of a deoxycytidine analogue that stabilizes the i-motif. This stabilization almost certainly arises from the preference of the phosphorothiolate residues for the C3′-*endo* sugar pucker, which is universally observed in solution for the cytidine sugars in the i-motif. These results also strongly suggest that the destabilization of the i-motif by ribose sugars is due to the 2′-substituent and not the C3′-*endo* sugar pucker.

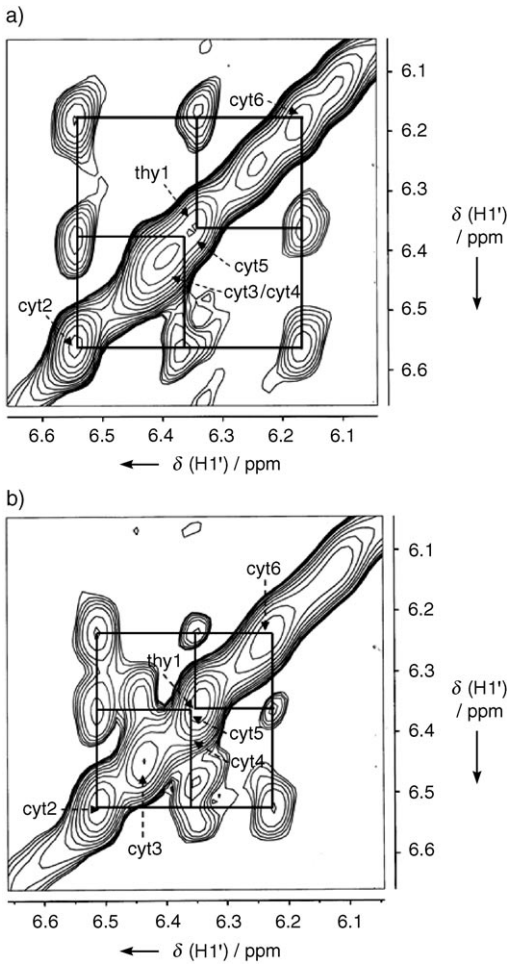

Figure 3. 500 MHz NOESY spectrum of the H1′–H1′ region, recorded for a) d(TCCCCC) and b) d(TCC**Cs**CC). Conditions are as described in Figure 2. NOESY spectra were recorded with $\tau_M = 200$ ms, 2048 points were collected in t_2 and 512 in t_1 by using TPPI phase cycling. Prior to Fourier transformation, a Gaussian function with gb = 0.02 and lb = −10 Hz was applied in both dimensions. The assignments shown for d(TCCCCC) were taken directly from reference [4], and those for d(TCC**Cs**CC) were made through comparison. Cyt = cytosine, Thy = thymine.

Received: September 1, 2005
Published online: November 21, 2005

Keywords: DNA structures · i-motif · NMR spectroscopy · nucleic acids · phosphorothiolates

[1] J. L. Leroy, M. Gueron, J. L. Mergny, C. Helene, *Nucleic Acids Res.* **1994**, *22*, 1600.
[2] S. Ahmed, A. Kintanar, E. Henderson, *Nat. Struct. Biol.* **1994**, *1*, 83.
[3] A. T. Phan, M. Gueron, J. L. Leroy, *J. Mol. Biol.* **2000**, *299*, 123.
[4] K. Gehring, J. L. Leroy, M. Gueron, *Nature* **1993**, *363*, 561.

[5] J. L. Leroy, K. Gehring, A. Kettani, M. Gueron, *Biochemistry* **1993**, *32*, 6019.
[6] M. Gueron, J. L. Leroy, *Curr. Opin. Struct. Biol.* **2000**, *10*, 326.
[7] P. Catasti, X. Chen, L. L. Deaven, R. K. Moyzis, E. M. Bradbury, G. Gupta, *J. Mol. Biol.* **1997**, *272*, 369.
[8] X. G. Han, J. L. Leroy, M. Gueron, *J. Mol. Biol.* **1998**, *278*, 949.
[9] J. L. Mergny, L. Lacroix, X. G. Han, J. L. Leroy, C. Helene, *J. Am. Chem. Soc.* **1995**, *117*, 8887.
[10] a) E. Marsich, A. Piccini, L. E. Xodo, G. Manzini, *Nucleic Acids Res.* **1996**, *24*, 4029; b) L. Lacroix, H. Lienard, E. Labourier, M. Djavaheri-Mergny, J. Lacoste, H. Leffers, J. Tazi, C. Helene, J. L. Mergny, *Nucleic Acids Res.* **2000**, *28*, 1564.
[11] a) L. Q. Chen, L. Cai, X. H. Zhang, A. Rich, *Biochemistry* **1994**, *33*, 13540; b) C. H. Kang, I. Berger, C. Lockshin, R. Ratliff, R. Moyzis, A. Rich, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 3874.
[12] D. Collin, K. Gehring, *J. Am. Chem. Soc.* **1998**, *120*, 4069.
[13] K. Snoussi, S. Nonin-Lecomte, J. L. Leroy, *J. Mol. Biol.* **2001**, *309*, 139.
[14] S. Robidoux, M. J. Damha, *J. Biomol. Struct. Dyn.* **1997**, *15*, 529.
[15] L. Lacroix, J. L. Mergny, *Arch. Biochem. Biophys.* **2000**, *381*, 153.
[16] a) H. Kanehara, M. Mizuguchi, K. Tajima, K. Kanaori, K. Makino, *Biochemistry* **1997**, *36*, 1790; b) K. Kanaori, S. Sakamoto, H. Yoshida, P. Guga, W. Stec, K. Tajima, K. Makino, *Biochemistry* **2004**, *43*, 5672.
[17] a) A. P. G. Beevers, K. J. Fettes, I. A. O'Neil, S. M. Roberts, J. R. P. Arnold, R. Cosstick, J. Fisher, *Chem. Commun.* **2002**, 1458; b) A. P. G. Beevers, K. J. Fettes, G. Sabbagh, F. K. Murad, J. R. P. Arnold, R. Cosstick, J. Fisher, *Org. Biomol. Chem.* **2004**, *2*, 114.
[18] G. Sabbagh, K. J. Fettes, R. Gosain, I. A. O'Neil, R. Cosstick, *Nucleic Acids Res.* **2004**, *32*, 495.

