The effect of 2′-fluorine substitutions on DNA i-motif conformation and stability†

Clare P. Fenna,[a] Victoria J. Wilkinson,[a] John R. P. Arnold,[a] Richard Cosstick*[b] and Julie Fisher*[a]

Received (in Cambridge, UK) 20th March 2008, Accepted 29th April 2008
First published as an Advance Article on the web 5th June 2008
DOI: 10.1039/b804833a

^{1}H and ^{19}F NMR, and UV thermal melting studies have established that the stability of d(TCCCCC) is enhanced by the inclusion of a single 2′-fluorine-modified deoxycytidine residue; the results support the notion of the importance of sugar–sugar contacts in stabilising i-motifs in general and reveal that solvation is the cause of the instability of RNA equivalents.

Almost fifteen years ago, the first intermolecular i-motif solution phase structure was reported for the sequence d(TCCCCC) (**1**).[1] This sequence has subsequently been used as a model in the analysis of features that are important in determining the overall structure and stability of i-motifs in general;[2] information that is increasingly valuable as the biological significance of i-motifs is still being revealed,[3] as is their potential for exploitation in, for example, nanotechnological applications.[4]

Sequence **1** forms a structure composed of two parallel hemiprotonated duplexes, intercalated into each other in a head-to-tail orientation (Fig. 1). It is noted that the deoxyribose rings of most of the residues adopt a predominantly C3′-*endo* (or north) conformation, more reminiscent of ribose sugars in RNA sequences. Despite this, early studies on cytidine-rich RNA sequences indicated that these have a low to no propensity to form i-motifs.[2a,5,6] Here, we report studies on **1** containing 2′-fluorocytidine, the results of which lead to the conclusion that the energetic penalty for desolvating the 2′-hydroxyl group of ribose sugars is what precludes i-motif formation.

Inclusion of a single RNA residue in the hexamer has been shown to cause a small reduction in its thermal stability (Table 1, sequence **3**).[2a] While some thermodynamic consequences of the sequence variation in this substitution were noted, these were uniformly destabilising.[2a] Multiple substitutions were investigated by the same researchers, with more dramatic sequence variation being revealed.[2a] The sequence d(TCC**CrCrC**) (**Cr** is a ribose C) displayed a melting temperature (T_m) of 41 °C, whilst the alternative arrangement, d(TC**CrCr**CC) (**5**), displayed a T_m more than 10 °C lower (Table 1, sequence **5**). This large drop was thought to be a result of either steric clashes involving the hydroxyls, or of the heavy solvation of these groups.[2a] The former effect would be anticipated to be more significant for the second sequence (with ribose residues at positions 3 and 4), as in this case, the hydroxyls share two *back-to-back*[2a] relations between intercalated strands, and are therefore in very close proximity; the first sequence (with substituents at positions 4 and 5) displays no such relation. The phrase *back-to-back* describes the situation in which the 2′-substituent of cytidines 1–6, 2–5 and 3–4 come close to each other in antiparallel strands. The alternate, *face-to-face*, arrangement is displayed between residues 6–2, 5–3 and 4–4 in antiparallel strands.[2a]

Recently, so-called locked nucleic acids (LNAs) have been introduced to **1**,[2b] with two and four LNA units being incorporated both sequentially and alternately. This modification, which includes a 2′-O (in a 2′-O–CH_2-4′ link or lock to constrain the sugar to the desired C3′-*endo* pucker), is also found to destabilise the i-motif, with the exception of when it is located at the 4 and 5 positions,[2b] where an increased T_m is noted. Unfortunately, an LNA substitution pattern similar to that for the ribocytidines has not been reported for a direct comparison to be made.

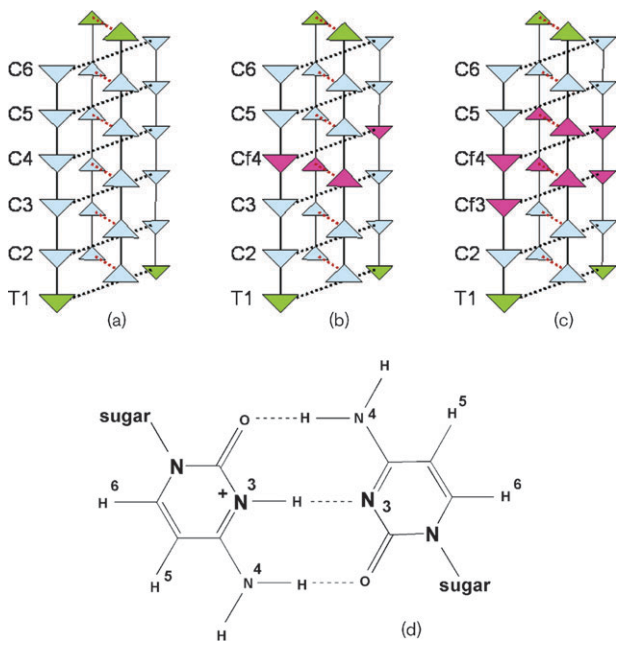

Fig. 1 The topology of the i-motif structure formed by (a) **1**, (b) d(TCC**CfC**C) (**2**), (c) d(TC**CfCf**CC) (**4**) and (d) a C^{+}·C base pair.

[a] *School of Chemistry, University of Leeds, Leeds, UK LS2 9JT. E-mail: j.fisher@chem.leeds.ac.uk; Fax: +44 (0)113 343 6565; Tel: +44 (0)113 343 6577*
[b] *Department of Chemistry, University of Liverpool, Crown Street, Liverpool, UK L69 7ZD. E-mail: rcosstic@liverpool.ac.uk; Fax: +44 (0)151 794 3588; Tel: +44 (0)151 794 3514*

† Electronic supplementary information (ESI) available: Sample preparation details, imino proton chemical shifts, NOESY data and pK_a information. See DOI: 10.1039/b804833a

Table 1 UV T_m data

	Sequence[a]	T_m/°C[b]
(**1**)	d(TCCCCC)	44.8[c], 46.0[2cb, d] 48.0[2ae]
(**2**)	d(TCC**Cf**CC)	48.8[c]
(**3**)	d(TCC**Cr**CC)	44.5[2a]
(**4**)	d(TC**Cf**C**f**CC)	44.2[c]
(**5**)	d(TC**Cr**C**r**CC)	26.5[2a]

[a] **Cf** indicates a 2′-fluoro-2′-deoxycytidine residue; **Cr** indicates ribocytidine. [b] Sodium citrate buffer (50 mM, pH 4.2). [c] Sodium citrate buffer (50 mM, pH 4.6). [d] 5 μM strand concentration. [e] Sodium citrate buffer (50 mM, pH 4.6, 11.5 ± 0.5 μM).

Given the apparent sensitivity of the i-motif structure to modification at the 2′-position, it was intriguing to explore the effect of 2′-fluoro substitution. Fluorine is of particular interest, as it has found many uses as an NMR label in studies of proteins and nucleic acids, suggesting an alternative spectroscopic approach for probing the i-motif structure.[7] A fluorine atom at the 2′-position in a deoxyribose ring results in a preferential north or C3′-*endo* pucker;[8] hence, this should stabilise the i-motif (Fig. 2). Fluorine is sterically equivalent[9] to a hydroxyl group but has lower hydrogen-bonding capabilities.[10] Through appropriate positioning of these residues in the hexamer sequence, we have been able to explore steric and solvation contributions to overall stability.

The presence of an i-motif for the natural sequence and two modified sequences (Table 1, sequences **2** and **4**) was confirmed by ^{1}H NMR measurements. 1D ^{1}H NMR spectra recorded for **1** and **2** each display four distinct imino proton resonances at chemical shifts greater than 15 ppm (Fig. 3). These signals are a result of imino hydrogen atoms on protonated cytosines that are shifted several ppm downfield of the standard imino region due to hydrogen-bonding; this being characteristic of i-motifs.[1] A similar shift range is observed for **4**; however, two signals are observed for the thymine imino group, which is suggestive of more than one i-motif conformation being present under the conditions used;[2b] only one thymine imino is observed for the natural and singly-modified sequence.[1,2c] The alternative conformer(s) is (are) not highly populated and could not be characterised. An additional set of peaks appears lowfield shifted 7 ppm in the ^{19}F spectrum of **3**. These low level peaks may be due to a shifted i-motif or a duplex structure, but this could not be determined due to the low population of this (these) conformer(s). Further support for the presence of an i-motif was gained through the observation of characteristic H1′–H1′ and imino–imino proton connectivities in NOESY spectra.[1] The similarity of the NOESY data set obtained and assigned in-house for **1** with those of **2** and **4** are supportive of the sequences having the same topology (Fig. 1).[1]

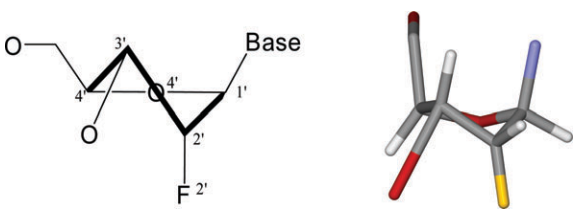

Fig. 2 The chemical structure of a fluorine-modified nucleotide, showing a 3′-*endo* sugar conformation.

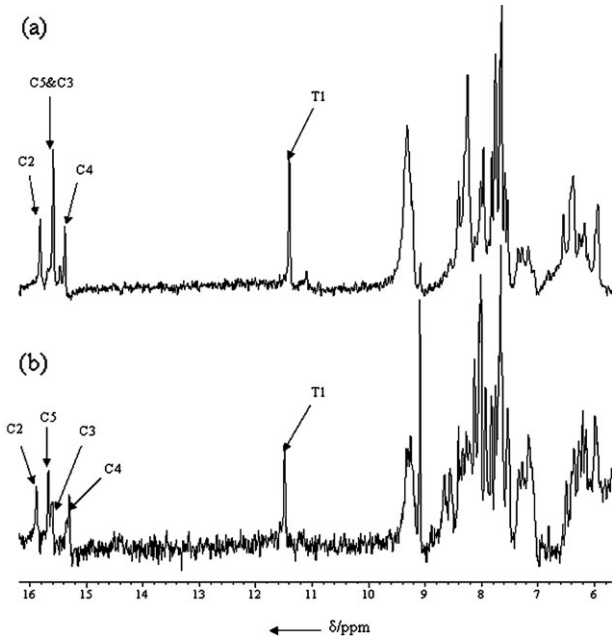

Fig. 3 Downfield region of the 500 MHz 1D NMR spectra recorded for i-motifs (a) **1** and (b) **2**. Conditions: 0.4 mM oligonucleotide in 50 mM citrate buffer, H_2O/D_2O solution (9 : 1), pH 4.2, 10 °C.

Melting temperatures were recorded for all three sequences by monitoring their UV absorbance at 260 nm. The sequence containing a single fluorine had a higher T_m than the natural strand. Conversely, a small reduction (0.6 °C) was seen when two sequential residues were modified, compared with the almost 20 °C loss when ribose residues were placed in identical positions.[2a] Given the similarity in size of the C–F group compared to the C–O–H group,[9] these data suggest that there is sufficient space to accommodate such modifications at the 2′-position. It should be noted that these observations were made at pH 4.2, which is at the pK_a for the N3 of the fluorinated cytidine (see ESI†) and within the range that has been quoted for the structural stability of the natural hexamer.[11]

The ^{19}F NMR chemical shift[12a] is highly sensitive to a fluorine atom's environment, such that when the solvent is changed to a heavier isotope form, as in H_2O to D_2O, a solvent-induced isotope chemical shift (SIIS) is observed.[7,12] ^{19}F NMR measurements for the i-motif in H_2O and D_2O solution therefore reveal the solvent exposure of the 2′-fluorine group (Table 2).

^{19}F NMR spectra recorded at ambient temperature for the i-motif sequences displayed a very small upfield shift upon solvent change. The somewhat larger change observed for the

Table 2 Solvent-induced isotope shifts

	System	Temperature/°C	H_2O/D_2O shift/ppm[a]
	2′-Fluoro-2′-deoxycytidine	20	−0.188
2	d(TCC**Cf**CC)	20	−0.030
		50	−0.044
4	d(TC**Cf**C**f**CC)	20	−0.022, −0.044
		50	−0.198, −0.202

[a] Negative values indicate an upfield shift for heavier isotope solvents. The estimated error margins are ±0.0005 ppm.

nucleotide monomer is in line with what would be expected for a fully solvent-exposed nucleus.[12c] Data was not available for duplex DNA; this may have been another useful comparator.

These results are therefore consistent with the limited solvent accessibility of the narrow groove occupied by the fluorine atoms. When the temperature was increased to 50 °C, the shift induced by the solvent change increased slightly for the monosubstituted sequence and almost 10-fold for the disubstituted sequence; these observations are consistent with the UV results, as the NMR data were collected at temperatures slightly above and significantly above the respective T_ms. These isotope-induced shift studies provide alternative evidence for the relevance of hydrophobic sugar–sugar contacts, which have been implicated in driving i-motif formation.[11]

The results of our studies suggest that 2′-fluorocytidines may be introduced more widely into i-motifs to enhance their stability, and that multiple substitutions may be tolerated, even when this results in fluorine atoms sharing *back-to-back* relations; the dramatic drop in T_m for the two *back-to-back* hydroxyls in **5** must be due to solvation of the hydroxyl groups, a process that would not be so efficient for fluorine, even if these atoms were solvent accessible in **2** and **4**. Importantly, this, coupled with the utility of ^{19}F as an NMR probe, suggests such modifications could be used to study, for example, the kinetics of intramolecular i-motif folding by NMR. Information from such studies would, of course, provide considerable insight to the factors that influence the formation of these structures in biology.

Notes and references

1. K. Gehring, J.-L. Leroy and M. Gueron, *Nature*, 1993, **363**, 561.
2. (*a*) D. Collin and K. Gehring, *J. Am. Chem. Soc.*, 1998, **120**, 4069; (*b*) N. Kumar, J. T. Nielsen, S. Maiti and M. Petersen, *Angew. Chem., Int. Ed.*, 2007, **46**, 9220; (*c*) J. A. Brazier, J. Fisher and R. Cosstick, *Angew. Chem., Int. Ed.*, 2006, **45**, 114; (*d*) N. K. Sharma and K. N. Ganesh, *Chem. Commun.*, 2005, 4330.
3. (*a*) E. Marsich, A. Piccini, L. E. Xodo and 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 and J. L. Mergny, *Nucleic Acids Res.*, 2000, **28**, 1564; (*c*) J.-F. Cornuel, A. Moraillon and M. Gueron, *Biochimie*, 2002, **84**, 279.
4. (*a*) W. Wang, H. Liu, D. Liu, Y. Xu, Y. Yang and D. Zhou, *Langmuir*, 2007, **23**, 11956; (*b*) P. Alberti, A. Bourdoncle, B. Sacca, L. Lacroix and J.-L. Mergny, *Org. Biomol. Chem.*, 2006, **4**, 3383; (*c*) S. Chakraborty, S. Modi and Y. Krishnan, *Chem. Commun.*, 2008, 70.
5. J.-L. Leroy, K. Gehring, A. Kettani and M. Gueron, *Biochemistry*, 1993, **32**, 6019.
6. S. Robidoux and M. J. Damha, *J. Biomol. Struct. Dyn.*, 1997, **15**, 529.
7. (*a*) J. T. Gerig, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1994, **26**, 293; (*b*) M. A. Danielson and J. J. Falke, *Annu. Rev. Biophys. Biomol. Struct.*, 1996, **25**, 163.
8. M. Blandin, T. D. Son, J. C. Catlin and W. Guschlbauer, *Biochim. Biophys. Acta*, 1974, **361**, 249.
9. J. C. Biffinger, H. W. Kim and S. G. DiMagno, *ChemBioChem*, 2004, **5**, 622.
10. J. Parsch and J. W. Engels, *J. Am. Chem. Soc.*, 2002, **124**, 5664.
11. S. Modi, A. W. Wani and Y. Krishnan, *Nucleic Acids Res.*, 2006, **34**, 4354.
12. (*a*) P. E. Hansen, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1998, **20**, 207; (*b*) J. R. P. Arnold and J. Fisher, *J. Magn. Reson.*, 2000, **142**, 1; (*c*) J. R. P. Arnold and J. Fisher, *Chem. Commun.*, 1998, 1859.