



**HAL**  
open science

# Structure and Conversion Kinetics of a Bi-stable DNA i-motif: Broken Symmetry in the [d(5mCCTCC)]<sub>4</sub> Tetramer

Sylvie Nonin, Jean-Louis Leroy

► **To cite this version:**

Sylvie Nonin, Jean-Louis Leroy. Structure and Conversion Kinetics of a Bi-stable DNA i-motif: Broken Symmetry in the [d(5mCCTCC)]<sub>4</sub> Tetramer. *Journal of Molecular Biology*, 1996, 261 (3), pp.399-414. 10.1006/jmbi.1996.0472 . hal-02344764

**HAL Id: hal-02344764**

**<https://hal.science/hal-02344764>**

Submitted on 20 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Structure and Conversion Kinetics of a Bi-stable DNA i-motif: Broken Symmetry in the [d(5mCCTCC)]<sub>4</sub> Tetramer

Sylvie Nonin and Jean-Louis Leroy\*

*Groupe de Biophysique de l'Ecole Polytechnique et de l'URA D1254 du CNRS 91128 Palaiseau, France*

At slightly acidic pH, protonation of C-rich oligomers results in the formation of a four-stranded structure composed of two parallel duplexes in a head to tail orientation with their hemi-protonated C·C<sup>+</sup> pairs intercalated in a so-called i-motif. In all cases reported previously the duplexes are identical.

The tetramer formed by the d(5mCCTCC) oligomer is different. The structure is computed on the basis of 55 inter-residue distances derived from NOESY cross-peaks measured at short mixing times. It consists of two intercalated non-equivalent symmetrical duplexes. The base stacking order is C5\* C1 C4\* C2 (T3\*) T3 C2\* C4 C1\* C5, but the thymidine bases (T3\*) of one duplex are looped out and lie in the wide grooves of the tetramer. The thymidine bases T3 stack as a symmetrical T·T pair between the sequentially adjacent C2·C2\* pair and the C2\*·C2\*\* pair of the other duplex.

Numerous exchange cross-peaks provide evidence for duplex interconversion. The interconversion rate is 1.4 s<sup>-1</sup> at 0°C and the activation energy is 94 kJ/mol.

The opening of the T3·T3 pair, the closing of the T3\*·T3 pair, and the opening of the C2\*·C2\*\* pair occur simultaneously with the duplex interconversion. This suggests that the concerted opening and closing of the thymidine bases drive the duplex interconversion. Opening of the C4·C4\* and C4\*·C4\*\* pairs, and dissociation of the tetramer are not part of the interconversion since they occur at much slower rates. Duplex interconversion within the [d(5mCCTCC)]<sub>4</sub> tetramer provides the first structural and kinetics characterization of broken symmetry in a biopolymer.

The tetramer formed by d(5mCCUCC) adopts a similar structure, but the rate of duplex interconversion is faster: 40 s<sup>-1</sup> at 0°C. At 32°C, interconversion is fast on the NMR time scale.

**Keywords:** i-motif; proton exchange; DNA solution structure; telomer; tetramer

\*Corresponding author

## Introduction

Parallel duplexes of cytidine-rich oligomers held together by hemi-protonated C·C<sup>+</sup> pairs were

described more than 30 years ago (Langridge & Rich, 1963; Inman, 1964). Recently, NMR experiments established that two parallel deoxy-oligonucleotide duplexes associate head-to-tail to form a four-stranded structure with C·C<sup>+</sup> pairs intercalated face-to-face in a so called i-motif (Gehring *et al.*, 1993).

The i-motif structure can be formed by four strands carrying one oligo(dC) stretch each (Gehring *et al.*, 1993; Chen *et al.*, 1994; Kang *et al.*, 1994, 1995; Leroy & Guéron, 1995; Berger *et al.*, 1995), by two strands carrying two oligo(dC) stretches each (Rohozinski *et al.*, 1994), or by one

---

Abbreviations used: 1D, one-dimensional; 2D, two-dimensional; 5mC, 5-methylcytidine; H4<sub>ext</sub>, cytidine external amino proton; H4<sub>int</sub>, cytidine internal (H-bonded) amino proton; JR, jump-and-return; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; RMSD, root mean square deviation; TOCSY, total correlation spectroscopy; ppm, parts per million.

strand carrying four oligo(C) stretches (Leroy *et al.*, 1994; Mergny *et al.*, 1995).

The formation of a C-C<sup>+</sup> pair shifts the protonation equilibrium of the cytidine N3 nitrogen ( $pK_{N3} = 4.3$ ). It is of particular interest that the telomeric oligo(C) repeat of vertebrates folds into an i-motif at pH 7 (Leroy *et al.*, 1994).

High definition structures of the i-motif were obtained recently for the tetramers of d(TCC) and d(5mCCT), and it was shown that the i-motif intercalation topology may be derived directly from characteristic features of the NOESY spectrum, namely the inter-residue connectivities of the H1'-H1' protons and the cross-peaks between amino protons and H2'/H2'' protons (Leroy & Guéron, 1995).

Here, we use these determinants to solve the structure formed by d(CCTCC), a sequence chosen to determine how the i-motif may incorporate a T-T pair, and by the related sequence d(5mCCTCC). The stoichiometry, the stability, and the NMR characteristics of both multimers are similar but the larger dispersion of the amino proton chemical shifts favours the methylated sequence for the structural study.

Gel filtration experiments demonstrate the tetrameric stoichiometry of the structure formed by the d(5mCCTCC) oligomer and the connectivities of the NOESY spectra are characteristic of the i-motif. The presence of two spin systems per residue indicates a non-symmetrical structure, which differs from the fully symmetrical ones reported in all previous solution studies.

The structure computed on the basis of the distance restraints corresponding to NOESY cross-peaks measured at short mixing times is a tetramer formed by intercalation of two different symmetrical duplexes. NOESY spectra recorded at high temperature reveal an interconversion process between both duplexes which we further characterize by kinetics experiments involving gel filtration chromatography, saturation transfer and proton exchange. Finally, we compare the interconversion kinetics of [d(5mCCTCC)]<sub>4</sub> and of its uridine homologue [d(5mCCUCC)]<sub>4</sub>.

## Results

### Stoichiometry, stability and dissociation kinetics of the multimer

The elution profile of the d(CCTCC) oligomer on a gel filtration column shows two components in proportions dependent on the oligomer concentration. Using a calibrated column, we ascribed one component to the monomer ( $6(\pm 2)$  nucleosides) and the other to a multimer of  $17(\pm 4)$  nucleosides.

Samples of d(CCTCC) at different concentrations were prepared by successive dilutions of a concentrated sample. Gel filtration profiles obtained after waiting different time intervals showed very slow dissociation kinetics of the multimer

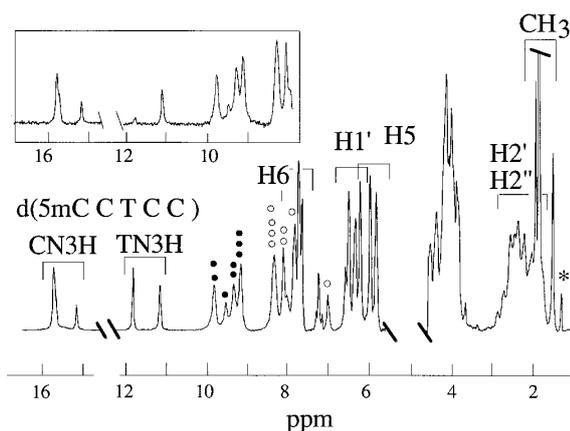


Figure 1. Proton spectrum of [d(5mCCTCC)]<sub>4</sub> in H<sub>2</sub>O solution. The observation of two thymidine imino proton peaks of equal area demonstrates the presence of two non-equivalent strands. The internal and the external amino protons are labelled by black and open circles, respectively. Insert: Imino protons and internal amino protons region of a spectrum recorded 0.25 second after selective inversion of the water magnetization. This saturation transfer experiment establishes the faster exchange rate of the T imino proton at 11.8 ppm. Experimental conditions:  $t = -7^{\circ}\text{C}$ , pH 4.6.

(hours) as compared with the time required for the chromatography (seven minutes). The multimer and the monomer concentrations at equilibrium were measured on the elution profile of solutions prepared by dilution of an aliquot of the NMR sample. The multimer fraction at equilibrium varies as the fourth power of the monomer concentration, hence the multimer species is a tetramer. The dissociation constant is  $(1.1 \times 10^{-4} \text{ M})^3$ . The chromatography profile does not disclose any dimeric component.

### Proton spectra

The NMR studies were conducted on the d(5mCCTCC) methylated derivative whose spectrum exhibits a better resolution than that of d(CCTCC).

The 1D spectrum of [d(5mCCTCC)]<sub>4</sub> in water at  $-7^{\circ}\text{C}$  is displayed in Figure 1. The spectrum of cytidine imino and amino protons is characteristic of hemi-protonated C-C<sup>+</sup> pairs (Leroy *et al.*, 1993). At the strand concentration of the NMR samples (8 mM), free strand is negligible.

The presence of two thymidine imino proton peaks with identical areas shows immediately that [d(5mCCTCC)]<sub>4</sub> does not form a fully symmetrical structure with four identical strands. When the integrated area of each thymidine imino proton is set to 1, the 15-16 ppm region corresponds to  $2.5(\pm 0.5)$  H-bonded cytidine imino protons. The thymidine imino proton at 11.84 ppm exchanges faster than that at 11.18 ppm as shown by the spectrum recorded 0.24 second after selective inversion of water (insert in Figure 1).

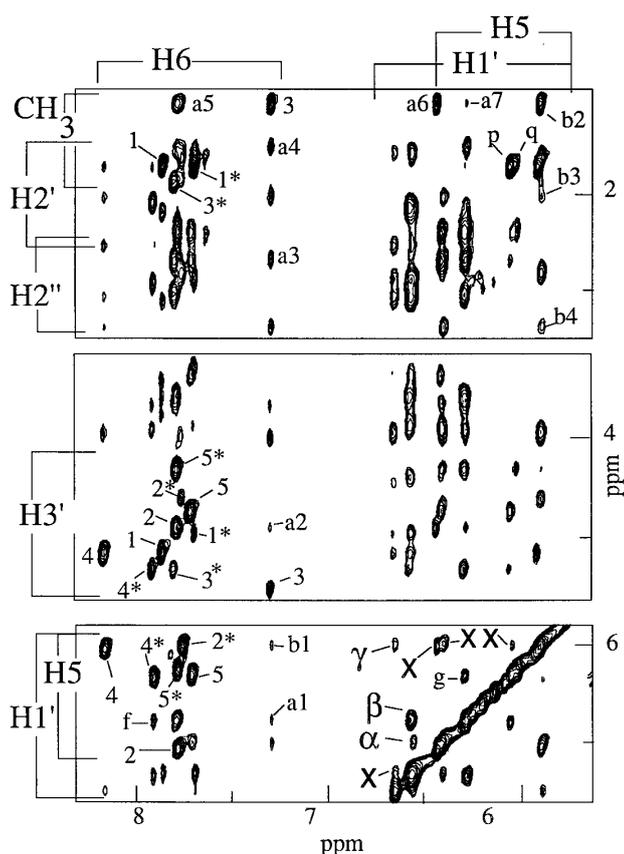


Figure 2. Expanded NOESY contour plot in  $^2\text{H}_2\text{O}$  solution (250 ms mixing time) of  $[\text{d}(5\text{mCCTCC})]_4$ . Intra-residue cross-peaks are labelled by the residue number. H5-H6, H3'-H6 and methyl-H6 cross-peaks are displayed on the lower, central and upper panels, respectively. Cross-peaks between T3 and C2 are labelled a1 (T3 H6-C2 H1'), a2 (T3 H6-C2 H3'), a3 (T3 H6-C2 H2'), a4 (T3 H6-C2 H2'), a5 (T3 methyl-C2 H6), a6 (T3 methyl-C2 H5), and a7 (T3 methyl-C2 H1'). Cross-peaks between T3 and C2\* are labelled b1 (T3 H6-C2\* H5), b2 (T3 methyl-C2\* H5), b3 (T3 H2'-C2\* H1'), and b4 (T3 H2''-C2\* H1'). Cross-peaks f and g are assigned to T3\*H1'-C4\* H6 and to T3\*H1-C4\* H5.

H1'-H1' cross-peaks, a characteristic feature of the i-motif, are indicated by Greek letters:  $\alpha$  is assigned to the mC1\* H1'-C5 H1' cross-peak,  $\beta$  to the overlapping 5mC1 H1'-C5\* H1', C2 H1'-T3\* H1', and C4\* H1'-C2 H1' cross-peaks, and  $\gamma$  to the C2\* H1'-C4\* H1' cross-peaks. Exchange cross-peaks are labelled by crosses. Cross-peaks p and q are ascribed to the non-resolved interactions between the methyl groups of 5mC1 and 5mC1\*, and the H5 protons of C5 and C5\*. Experimental conditions:  $t = -7^\circ\text{C}$ , pH 4.2. Spectrometer frequency 600 MHz.

The NOESY spectra of  $[\text{d}(5\text{mCCTCC})]_4$  in  $^2\text{H}_2\text{O}$  (Figure 2) and in  $\text{H}_2\text{O}$  solution (Figure 3) display intra-residue connectivities of comparable intensities corresponding to ten spin systems, establishing the presence of two distinct strand species.

We examined the effect of temperature and pH on the spectrum of  $[\text{d}(5\text{mCCTCC})]_4$  (data not shown). At a strand concentration of 8 mM, dissociation occurs between 35 and 43°C. At  $-7^\circ\text{C}$ , the

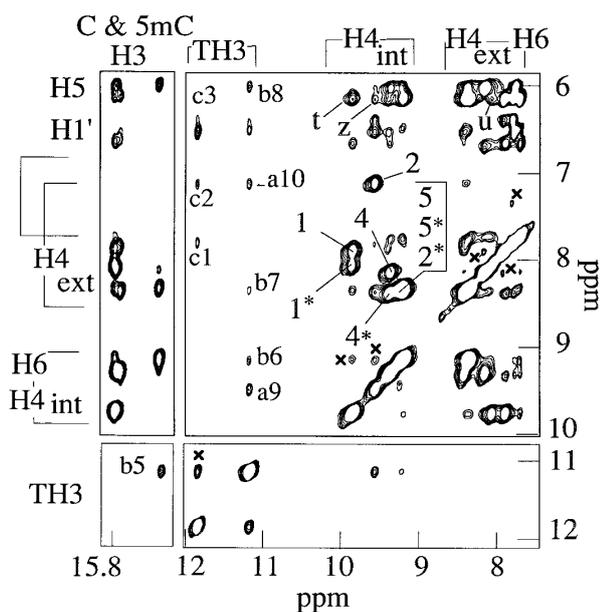


Figure 3. Exchangeable proton regions of the NOESY spectrum of  $[\text{d}(5\text{mCCTCC})]_4$ . Intra-residue cross-peaks between internal and external amino protons are labelled by the residue number. Amino protons of the 5mC1-5mC1\*, C4-C4\*, 5mC1\*5mC1\*, C2\*C2\* and C4\*C4\* base-pairs are connected to imino protons. Cross-peaks between thymidine imino protons and C2 protons are numbered a9 (T3 H3-C2 H4<sub>int</sub>), a10 (T3 H3-C2 H4<sub>ext</sub>), c1 (T3\* H3-C2 H6), c2 (T3\* H3-C2 H4<sub>ext</sub>), and c3 (T3 H3-C2 H5). Cross-peaks between T3 and C2\* are labelled b5 (T3 H3-C2\*H3), b6 (T3 H3-C2\* H4<sub>int</sub>), b7 (T3 H3-C2\* H4<sub>ext</sub>), and b8 (T3 H3-C2\*). The cross-peaks between the non-resolved H5 protons of the C5\*C5 couple and the non-resolved internal amino protons of the 5mC1-5mC1\* couple are labelled t. The cross-peak between 5mC1\* H4<sub>ext</sub>-C5 H5 is labelled u. C2 H4<sub>int</sub> and C4\*H5 are connected by cross-peak z. Experimental conditions  $t = -7^\circ\text{C}$ , pH 4.7, 250 ms mixing time, spectrometer frequency 360 MHz.

tetramer is stable between pH 3.5 and 5.8, and the proton chemical shifts are pH-independent. At all pH and temperatures, the concentration of the two strand species are equal.

Combined NOESY-ROESY experiments in  $^2\text{H}_2\text{O}$  (Figure 4) and in  $\text{H}_2\text{O}$ , show cross-peaks due to structural exchange without any contribution from the dipolar interaction. The intensity of all the exchange cross-peaks increases similarly with the temperature. The exchange process corresponds to interconversion between two strand species.

Consider for example the H5-H6 intra-residue NOESY cross-peak, and the corresponding H5\*-H6\* cross-peak on the homologous residue. At  $-7^\circ\text{C}$ , with a mixing time of 70 ms the exchange cross-peaks H6-H6\* (left panel of Figure 5) and H5-H5\* are weak. At  $18^\circ\text{C}$  with a mixing time of 250 ms (central panel of Figure 5), both dipolar transfer and chemical exchange occur. In that case, we also detect (H6\*-H5) and (H6-H5\*) exchange cross-peaks (Figure 5, central panel). From the pattern of the exchange cross-peaks, we defined

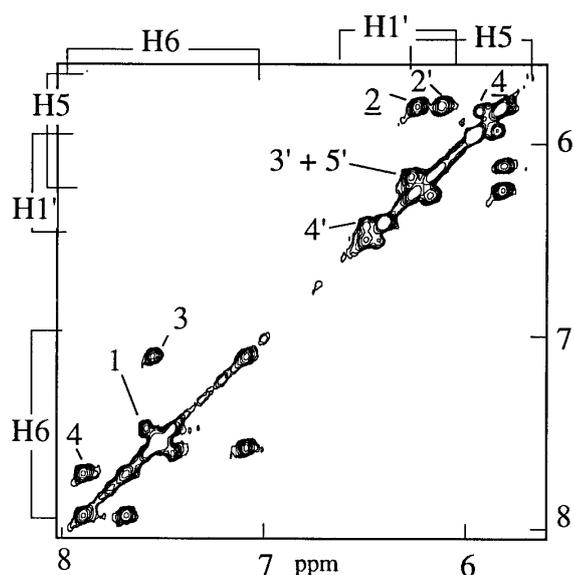


Figure 4. Pure chemical exchange experiment on the [d(5mCCTCC)]<sub>4</sub> tetramer at 20°C (mixing time 150 ms) in <sup>2</sup>H<sub>2</sub>O solution. The cross-relaxation is suppressed by a combination of interleaved NOESY and ROESY pulses (Fejzo *et al.* 1991). The exchange cross-peaks are labelled by the residue number of the exchanging couple. Primed numbers indicate H1'-H1' exchange cross-peaks. The number of the H5-H5 exchange cross-peaks is underlined. The exchange cross-peaks C1 H6-C1\* H6, C2 H6-C2\* H6, C1 H1'-C1\* H1', and C1 H5-C1\* H5 corresponding to couples of exchanging protons insufficiently resolved are too close of the diagonal to be detected.

five couples of homologue residues which occupy the same sequential position on each d(5mCCTCC) strand. From the comparison of NOESY and of exchange ROESY-NOESY spectra, we could identify the cross-peaks which have no exchange contribution.

### Tetramer symmetry

The characteristic amino-H2'/H2'' cross-peaks (labelled 1 to 13 in Figure 6), the inter-residue connectivities between H1'-H1' protons (Greek letters in Figure 2), and the strong intra-residue H3'-H6 cross-peaks (Figure 2) observed on the NOESY spectra of [d(5mCCTCC)]<sub>4</sub> provide evidence for an i-motif arrangement.

Figure 7 shows three hypothetical arrangements which could account for the formation of an i-motif tetramer containing two distinct strands in identical proportions: a single symmetrical tetramer consisting of two distinct symmetrical duplexes (Figure 7a), or consisting of two identical asymmetrical duplexes (Figure 7b); an equimolar mixture of two distinct symmetrical tetramers (Figure 7c).

The observation of NOE connectivities between both strand species eliminates the strand arrangement of Figure 7c. For example, both thymidine imino protons are connected to the same external

amino proton (cross-peaks c2 and a10, Figure 3). The same NOESY spectrum shows that the cytidine imino protons are connected to a single amino proton group. This establishes that the hemi-protonated C-C<sup>+</sup> pairs are symmetrical on the NMR time scale and excludes the non-symmetrical base-pairing of Figure 7b. Thus we are left with the strand arrangement of Figure 7a, corresponding to the association of two distinct symmetrical duplexes anti-parallel to one another. Henceforth, one duplex made of two identical strands a and c will be designated {ac}, and the second duplex will be similarly indicated {bd}. The residues of the {bd} duplex will be identified by a star.

### Proton assignment and intercalation topology

In all i-motif structures investigated so far by NMR (Gehring *et al.*, 1993; Leroy & Guéron, 1995), and more generally in most DNA structures, the H3' proton of 3'-end residues is shifted upfield from those connected to a phosphate group. We thus assign the peaks at 4.39 and 4.6 ppm to the H3' protons of cytidine C5 and C5\* (middle panel of Figure 2 and Table 1). We arbitrarily assign the H3' proton at 4.61 ppm to C5. The 5mC1-5mC1\* couple is readily identified by the intra-residue connectivities with the methyl protons (upper panel, Figure 2).

The NOE connectivities observed between the couples 5mC1-5mC1\* and C5-C5\* are consistent with the intercalation of {ac} and {bd} duplexes in a head-to-tail orientation, resulting in the stacking of 5mC1 with C5\* at one end of the tetramer, and of 5mC1\* with C5 at the other end. They include: (1) cross-peaks between the non-resolved 5mC1 or 5mC1\* methyl protons and the H5 protons of C5 and C5\* (peaks labelled p and g on Figure 2); (2) cross-peaks between the cytidine methyl protons and the amino protons of the C5 C5\* couple (cross-peaks r and s, Figure 6); (3) cross-peaks between C5 H5 and C5\* H5 protons and both internal and external amino protons of the methylated cytidine residues (t and u, Figure 3).

The protons of the C5 and C5\* residues are NOE-connected exclusively to the 5mC1\*-5mC1. The latter couple is also connected to a second couple of exchanging cytidine residues. This shows that C5 and C5\* terminate the tetramer, and leads to the intercalation scheme displayed in Figure 8. The scheme identifies the second couple of cytidine residues as C4 and C4\*. Characteristic reciprocal amino-H2'/H2'' cross-peaks connect 5mC1 and C4\* (cross-peaks 5, 6, 9, and 10 between 5mC1 H2'/H2'' and C4\* amino protons, and cross-peaks 1 and 3 between the internal amino proton of 5mC1 and C4\* H2'/H2'' protons, Figure 6). Similar connections are seen between 5mC1\* and C4 (cross-peaks 2, 4, and 13 in Figure 6). The amino protons of the C4-C4\* couple are also connected to the non-resolved methyl groups of the 5mC1-5mC1\* couples.

Next, C2\* is identified by the H1'-H1' cross-peak with C4 (cross-peak  $\gamma$ , Figure 2, lower panel). The

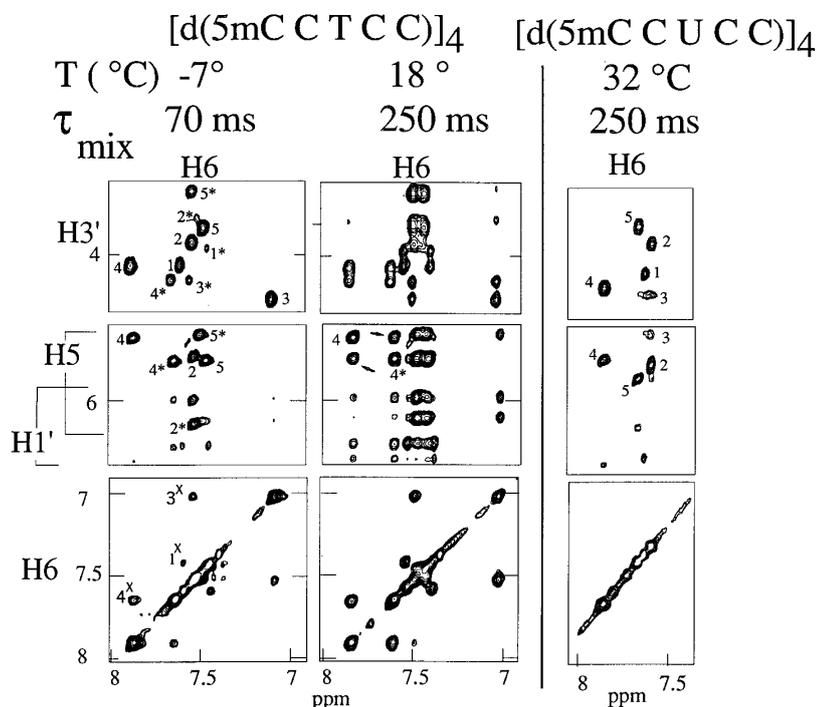


Figure 5. Comparison of the NOESY spectra of  $[d(5mCCTCC)]_4$  and  $[d(5mCCUCC)]_4$  in  $^2H_2O$  solution. Left panel: the NOESY spectrum of  $[d(5mCCTCC)]_4$  at  $-7^\circ C$  (70 ms mixing time) shows ten spin systems corresponding to two non-equivalent duplexes. The intra-residue cross-peaks are labelled by the residue number. Upper row:  $H3'$ - $H6$  cross-peaks. Central row:  $H5$ - $H6$  cross-peaks. The  $H6$ - $H6$  exchange cross-peaks (crosses) reflect interconversion of the duplexes.

Central panel: At  $18^\circ C$  (250 ms mixing time) the interconversion rate between the duplexes is fast enough to give rise to cross-peaks involving a combination of dipolar and exchange transfers during the mixing time. The  $C4$   $H6$ - $C4^*$   $H5$  and the  $C4$   $H5$ - $C4^*$   $H6$  cross-peaks are representative of this phenomenon (arrows).

Right panel: The NOESY spectrum of  $[d(5mCCUCC)]_4$  at  $32^\circ C$  (250 ms mixing time) shows only five spin systems, due to equivalence of the four  $d(5mCCUCC)$  strands on the NMR time-scale at this temperature.

homologous residues,  $C2$  and  $C4^*$ , are connected by a NOE between the internal amino proton of  $C2$  and  $C4^*$   $H5$  (cross-peak z in Figure 3).

The characteristic NOEs between  $C2^*$  amino protons and  $T3$   $H2'/H2''$  protons (cross-peaks 7, 8, 11, and 12 on Figure 6) show that  $T3$  is intercalated

into the i-motif in position adjacent of  $C2^*$ .  $T3$  exhibits several other connectivities with  $C2^*$  ring and sugar protons (cross-peaks b1-b4 in Figure 2, and b5 to b8 in Figure 3). Surprisingly,  $T3$  is also connected by sequential NOEs with  $C2$  (cross-peaks a1 to a7, Figure 2, and cross-peaks a9 and a10, Figure 3) suggesting that it is stacked with  $C2$ .

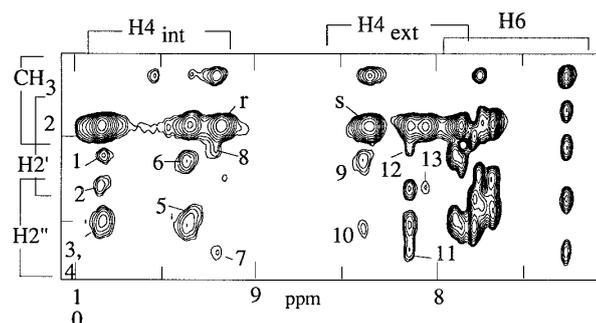


Figure 6. Expanded NOESY contour plot of  $[d(5mCCTCC)]_4$  in  $H_2O$  solution at  $-5^\circ C$  (150 ms mixing time). Cross-peaks 1 to 13 correspond to the NOEs between amino protons and  $H2'/H2''$  protons which determine the intercalation topology: 1 ( $5mC1$   $H4_{int}$ - $C4^*$   $H2'$ ); 2 ( $5mC1^*$   $H4_{int}$ - $C4$   $H2'$ ); 3 ( $5mC1$   $H4_{int}$ - $C4^*$   $H2''$ ); 4 ( $5mC1^*$   $H4_{int}$ - $C4$   $H2''$ ); 5 ( $C4^*$   $H4_{int}$ - $5mC1$   $H2''$ ); 6 ( $C4^*$   $H4_{int}$ - $5mC1$   $H2'$ ); 7 ( $C2^*$   $H4_{int}$ - $T3$   $H2''$ ); 8 ( $C2^*$   $H4_{int}$ - $T3$   $H2'$ ); 9 ( $C4^*$   $H4_{ext}$ - $5mC1$   $H2''$ ); 10 ( $C4^*$   $H4_{ext}$ - $5mC1$   $H2'$ ); 11 ( $C4^*$   $H4_{ext}$ - $5mC1$   $H2''$ ); 12 ( $C2^*$   $H4_{ext}$ - $T3$   $H2''$ ); 13 ( $5mC1^*$   $H4_{ext}$ - $C4$   $H2'$ ). Cross-peaks between the non-resolved methyl protons of the  $5mC1$ - $5mC1^*$  couple and the non-resolved internal and external amino protons of the  $C5$ - $C5^*$  couple are labelled r and s, respectively.

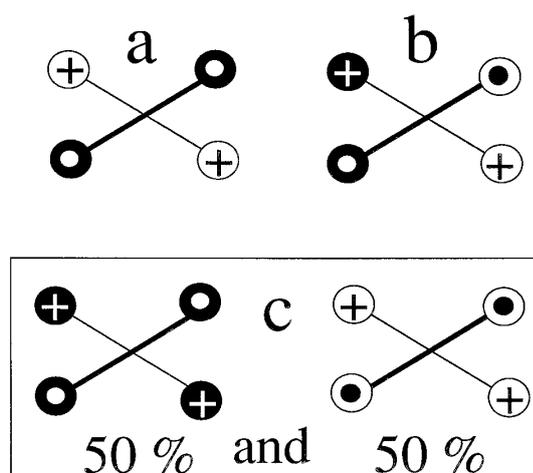


Figure 7. End-on-view of three hypothetical strand arrangements in an i-motif tetramer containing two kind of strands (open and closed circle) in identical proportions. The strand orientation is indicated by the symbols  $\circ$  and  $\bullet$ . Base-pairing between duplexes is indicated by thin lines.

Table 1. Proton chemical shifts (ppm) in the [d(mCCTCC)]<sub>4</sub> tetramer at -7°C

	H3	H4 <sub>int</sub>	H4 <sub>ext</sub>	H6	H5/CH3	H1'	H2'	H2''	H3'	H4'
mC1	15.711	9.86	8.05	7.85	1.87	6.54	2.12	2.59	4.83	4.10
C2	<sup>a</sup>	9.55	7.03	7.78	6.41	6.25	1.77	2.38	4.69	4.02
T3	11.18	—	—	7.28	1.54	6.39	2.04	2.74	5.04	4.22
C4	15.71	9.35	8.13	8.16	5.89	6.63	2.31	2.58	4.84	4.18
C5	<sup>a</sup>	9.19	8.34	7.71	6.02	6.38	2.23	2.39	4.61	3.90
mC1*	15.711	9.84	7.86	7.69	1.88	6.54	1.80	2.48	4.73	4.14
C2*	15.15	9.56	9.22	7.74	5.86	5.86	1.80	2.46	4.54	4.20
T3*	11.84	—	—	7.79	1.96	6.27	2.39	2.52	4.92	4.04
C4*	15.6	9.36	8.39	7.91	6.03	6.55	2.08	2.50	4.92	4.16
C5*	<sup>a</sup>	9.19	8.34	7.78	6.00	6.26	2.23	2.23	4.39	3.98

<sup>a</sup> Imino proton line too broad to be detected.

Thymidine T3\* shows NOEs to C2 (cross-peaks c1 to c3, Figure 3), to C2\* (cross-peaks T3 CH<sub>3</sub>-C2 H3'), and to C4\* (cross-peaks f and g between T3\* H1' and C4\* H6/H5, Figure 2, lower panel).

With the remarkable exception of the C2 external amino proton, which is shifted by 1.34 ppm upfield from that of C2\*, chemical shift differences between protons belonging to homologous residues are maximum for the T3-T3\* couple, and decrease in both 3' and 5' directions (Table 1).

### Structural analysis

The structure of [d(5mCCTCC)]<sub>4</sub> was computed using the simulated annealing method (Nilges, 1993) with the inter-residue distance restraints displayed in Figure 8. The distances were computed from the build-up of 55 NOESY cross-peaks as described in Materials and Methods. All the restraints used in the structure calculation are shown in the histogram displayed in Figure 9.

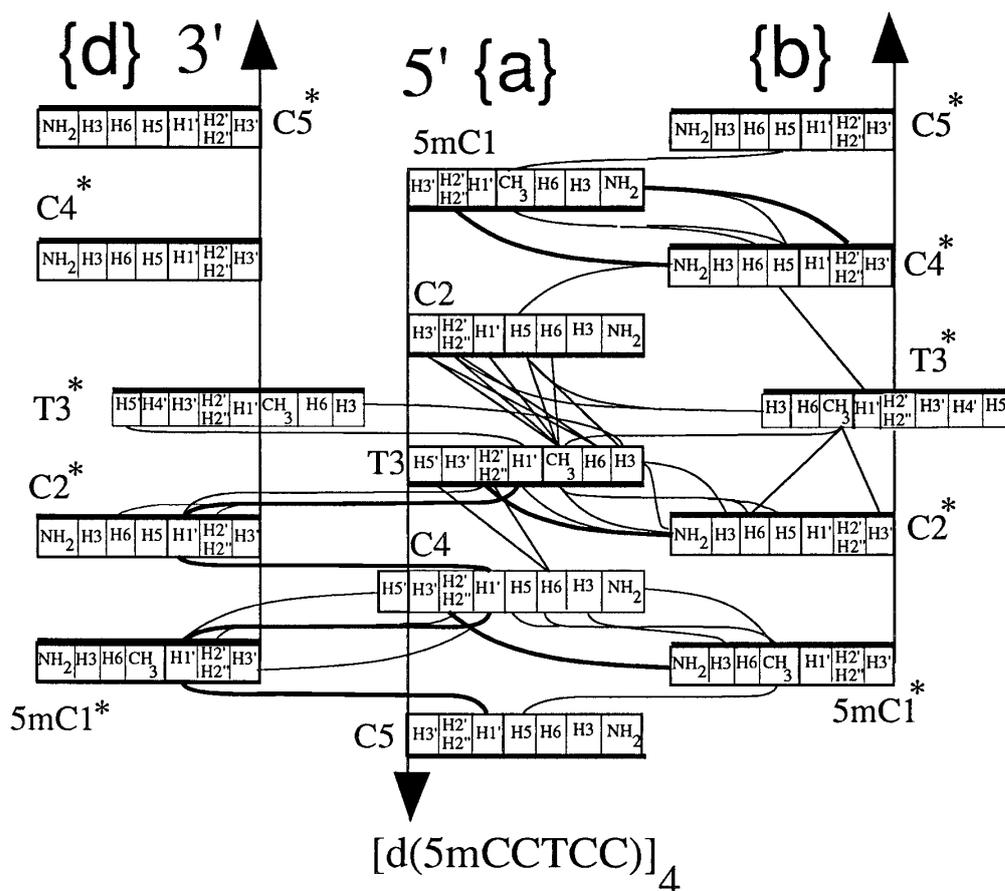


Figure 8. Schematic representation of the inter-residue distance restraints used to compute the [d(5mCCTCC)]<sub>4</sub> structure. The distance restraints between strand {a} and strand {b}, and between strand {a} and strand {d} define the inter-strand relations across the wide groove and the narrow groove, respectively. The amino-H2'/H2'' and the H1'-H1' inter-strand connectivities indicated by heavy lines are characteristic of the i-motif structure and determine the intercalation topology.

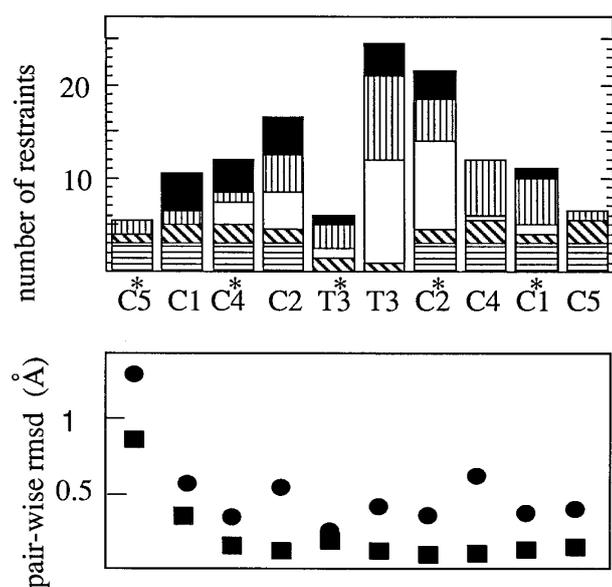


Figure 9. Upper panel: Distribution of the NMR-derived distance restraints used in the structure calculation of  $[d(5mCCTCC)]_4$ . The intra-residue cross-peaks are counted as one, and the inter-residue restraints as one half for each partner. Black area, distances involving an exchangeable proton; vertically hatched area, inter-residue distances involving two non-exchangeable protons. Blank area, the repulsive restraints reflecting the absence of direct NOE cross-peaks at long mixing times. The diagonally hatched area represents restraints on H-bond lengths used to define the C-C<sup>+</sup> pairing and the horizontally hatched area the intra-residue distance restraints.

Lower panel: pair-wise RMSD of the heavy atom positions for the bases (squares) and the sugar-phosphate backbones (circles). The RMSD are computed from the pair-wise comparison of 13 conformers aligned on the C1' positions.

We computed a set of 24 conformers. All had the same topology. The structural parameters were derived from the analysis of 13 conformers selected for their low energy associated with NOE violations. Violations from ideal geometry are displayed in Table 2.

The tetramer is formed by two parallel duplexes intercalated head-to-tail, differing by the arrangement of the thymidine residues (Figure 10).

In the {ac} duplex, the short inter-strand distance found in the computed structure between T3, H3

Table 2. Violations and deviations from ideal geometry in structural computations

Number of NOE violations larger than 0.2 Å	6
Largest NOE violation (Å)	0.29
RMSD of input distance restraints (Å)	0.065
Number of ideal bond angle violations larger than 6°	2
Largest ideal bond angle violation (°)	6.2
RMSD from ideal bond angles	1.3
Number of improper violations larger than 5°	0
RMSD from ideal impropers (°)	0.59
Bond violation larger than 0.05 (Å)	0
RMSD from ideal bonds (Å)	0.007

and T3 O2,  $2.15(\pm 0.15)$  Å, shows that the thymidine residues are base-paired. This pair is symmetrical (a single spin system) and it is stacked into the i-motif between the C2·C2<sup>+</sup> pair of the same duplex and the C2\*·C2\* pair of the other duplex.

The thymidines T3\* of the {bd} duplex are unpaired and not stacked. They lie symmetrically in the wide grooves along the C2·C2<sup>+</sup> base-pair. This position is imposed by the strong T3\* H3-C2 H5 cross-peak (labelled c3 in Figure 3).

Except for the C5\* residue which is poorly defined by the NMR data, the RMSD plotted as a function of the stacking order in the lower panel of Figure 9 shows a uniform definition of the structure. The overall pair-wise RMSD is 0.9 Å. The geometrical parameters are presented in Table 3.

For all residues, the glycosidic conformation is *anti* with glycosidic torsion angles ranging from  $-107^\circ$  to  $-154^\circ$ . Most sugar puckers are N-type. The tetramer has two identical narrow grooves and two identical wide grooves. The inter-strand phosphate-phosphate distances vary from 9 to 12.7 Å across the narrow grooves and from 13.4 to 16.1 Å across the wide grooves (Table 4). The flatness of the bases may be related to the small stacking distance (3.06 Å on the average). The duplexes are right-handed. The C1'-C1' distance between the paired bases ranges from 9.2 to 10 Å for the cytidines and is equal to 10.4 Å for the T3·T3 pair. This distance is larger in the case of the non-stacked T3\* residues, in relation with the backbone distortions at this point.

### Duplex interconversion kinetics

Due to their large chemical shift difference, thymidine H6 protons are good markers of duplex interconversion. Figure 11 shows magnetization transfer from T3 H6 to T3\* H6 as a function of the time during which T3 H6 is saturated. The interconversion time was computed from the fits displayed in the right panel using equation (1). Magnetization transfer experiments performed on the H6 protons of the C4-C4\* couple gave identical interconversion times. The temperature dependence of the duplex interconversion time (Figure 11) corresponds to an activation energy of 94 kJ/mol. The interconversion time is unchanged when the sample is diluted by a factor of two. It decreases twofold in 0.18 M potassium phosphate (Figure 12, central panel), and threefold in 0.7 M ammonium chloride.

### Proton exchange with water

At  $-70^\circ\text{C}$ , imino protons of C5, C5\* and C2 exchange too rapidly to give rise to intra-residue NOEs with the amino protons (Figure 3). The exchange times of T3\*, T3, and C2\* imino protons are plotted *versus* temperature on Figure 12. The exchange time of T3\* imino proton is close to that measured on the monomer, as expected for a non-paired residue.

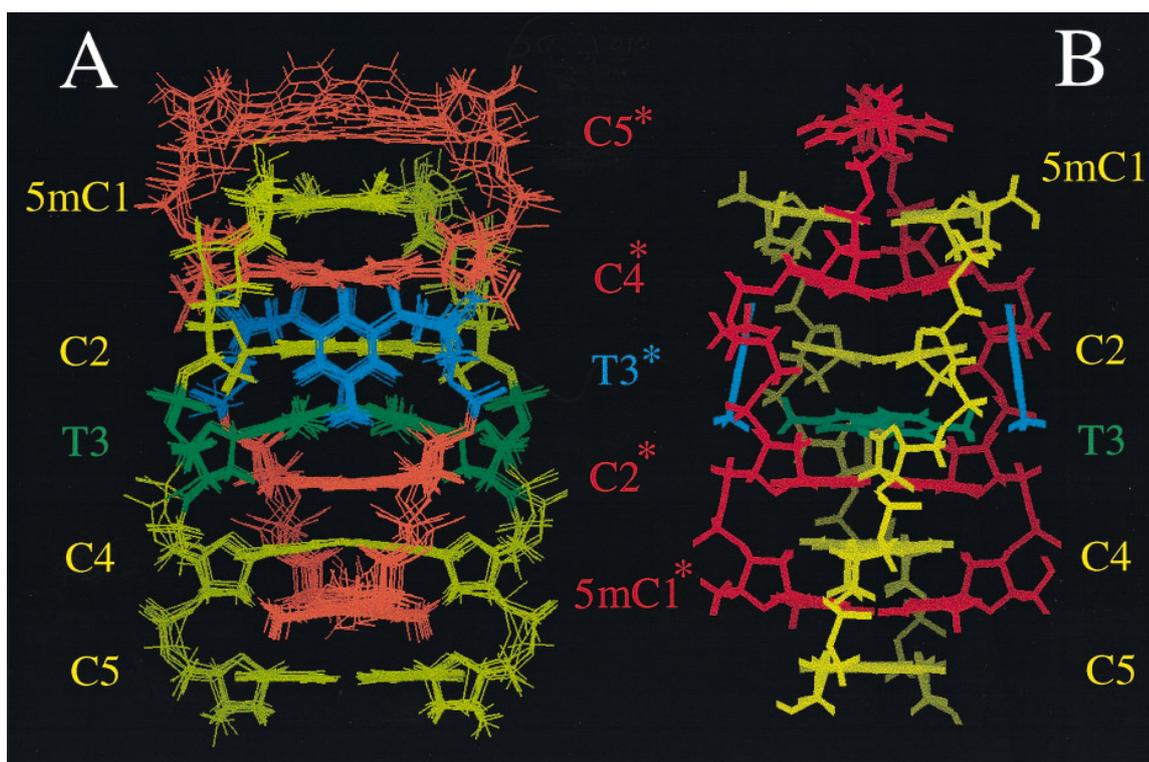


Figure 10. Computed structure of [d(5mCCTCC)4]. A, Wide groove view of the eight lowest-energy conformers. B, View normal to the minor groove of the lowest energy conformer. The yellow cytidine residues belong to the duplex whose thymidine residues are base-paired (green T3·T3 pair). In the second duplex, the cytidine residues are red and the non-base-paired thymidine residues T3\* are blue.

The exchange time of the imino proton of the C2\*·C2\*\* pair is equal to the duplex interconversion time. This corresponds to an exchange time in the C2\*·C2\*\* pair which is longer than the duplex lifetime. Interconversion of the duplex changes the C2\*·C2\*\* pair into a C2·C2+ pair from which imino proton exchange is very fast, as mentioned above.

Hydroxyl and phosphate catalysis of imino proton of the non-base-paired T3\* occur at rates similar to those of the monomeric thymidine.

In the presence of 0.18 M phosphate (pH 5.6), the exchange rate of the T3\* imino proton is faster than the duplex interconversion rate, and the exchange rate of the T3 imino proton is equal to the duplex interconversion rate. In this case, T3 imino proton exchange is controlled by duplex interconversion, as in the case of C2\*/C2 just described (Figure 12, central panel).

Upon addition of 0.7 M ammonia at pH 5.6, i.e. in the presence of  $12.5 \times 10^{-4}$  M of the acceptor species  $\text{NH}_3$ , the exchange time of T3 imino proton remains identical to the duplex interconversion time. In such drastic conditions of catalysis, the exchange rate of the imino proton of the thymidine monomer is  $3 \times 10^4 \text{ s}^{-1}$  (Guéron *et al.*, 1989). The lack of effect of  $\text{NH}_3$  on proton exchange ( $<3 \text{ s}^{-1}$ ), gives an upper bound of  $10^{-4}$  for the dissociation constant of the T3·T3 pair at 15°C.

At pH 4.7 in the absence of added exchange catalyst, T3 imino proton exchange is controlled by

the rate of duplex interconversion (below 20°C) or by the exchange rate of T3\* imino proton (above 20°C).

Imino and amino proton exchange times measured in real time at 0°C from a series of spectra recorded after dilution of a protonated sample in  $^2\text{H}_2\text{O}$  are displayed in Table 5. Two imino protons exchange in 730 seconds. The resolved low field proton belongs to the C4·C4+ pair. Since the exchange time is much larger than the duplex interconversion time, the second slow-exchanging imino proton is necessarily that of the C4\*·C4\*\* pair.

Internal and external amino protons exchange at distinct rates. The exchange rates of the amino protons of the C4·C4\* and C1·C1\* couples are particularly slow (Table 5).

### The 5mCCUCC tetramer

Gel filtration chromatography shows that the deoxy-uridine derivative, d(5mCCUCC), forms a tetramer whose stability is similar to that of d(5mCCTCC).

### Tetramer symmetry and interconversion kinetics

The 1D NMR spectrum of [d(5mCCUCC)]4 is quite comparable to that of [d(5mCCTCC)]4. At 0°C, the NOESY spectra exhibit also two exchanging spin systems for each residue, but the intensities

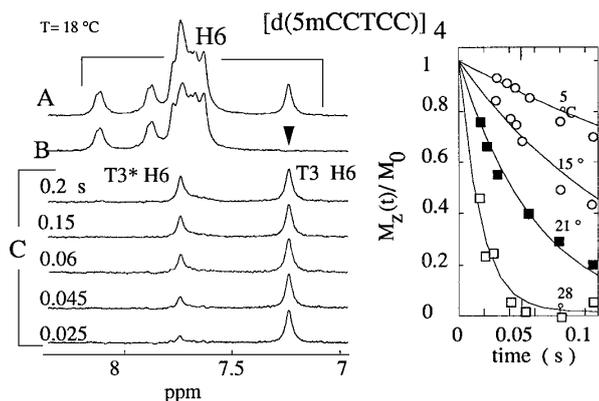


Figure 11. Magnetization transfer experiment between the T3 H6 proton of the {ac} duplex and its homologue (T3\* H6) on the {bd} duplex. Left side: H6 proton region of the [d(5mCCTCC)]4 spectrum. A, Reference spectrum; B, T3 H6 (arrow) was selectively saturated during 0.15 second prior to detection. C, Difference spectra between the spectrum in A and spectra obtained after saturation of T3 H6 during the delay indicated on the Figure. Right side: Magnetisation of T3\* H6 as a function of the time during which T3 H6 was maintained saturated. The duplex interconversion time was computed using expression (1). The exchange times displayed in Figure 12 are derived from the fits.

of the exchange cross-peaks indicate a faster duplex interconversion rate in the case of the uridine oligomer. Indeed, the NMR lines of homologous protons belonging to most of the exchanging residues coalesce between 10°C and 30°C. The NOESY spectrum recorded just below the melting transition at 32°C exhibits only five spin systems (Figure 5, right panel). The narrow cytidine intra-residue H6-H5 and H3'-H6 cross-peaks show that the duplex interconversion rate is fast enough to average the environment of these protons. The environment of the homologous protons which have a large chemical shift difference is incompletely averaged on the 32°C spectrum, as shown by the large linewidth of the uridine H6 proton and the absence of the C2 external amino proton on the NOESY spectrum in H<sub>2</sub>O at 32°C (data not shown).

The interconversion time of the [d(5mCCUCC)]2 duplexes measured by magnetization transfer, 25 ms at 0°C, is 30 times shorter than that measured on the tetramer of d(5mCCTCC) (Figure 12, right panel). The activation energy of the duplex interconversion rate is 55 kJ/mol.

#### Proton assignment and intercalation topology of the motion-averaged structure of [d(5mCCUCC)]4

We design by the subscript av the motion averaged NMR peaks. The spectrum is assigned as explained above in the case of [d(5mCCTCC)]4, using the characteristic H1'-H1' and amino-H2'-H2'' cross-peaks detected on the NOESY spectrum at 32°C. The C5<sub>av</sub> residue, identified by its upfield shifted H3' proton, is connected to the methylated

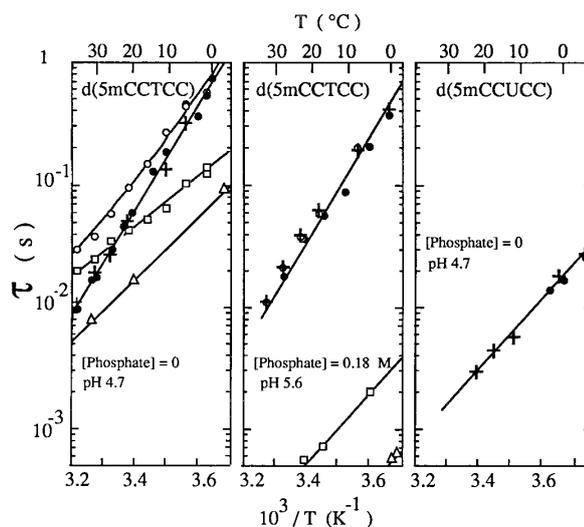


Figure 12. Duplex interconversion times (●) and imino proton exchange times versus temperature in [d(5mCCTCC)]4 and in [d(5mCCUCC)]4: T3, (○); T3\* (□); and C2\* (+). Imino proton exchange and interconversion times in [d(5mCCTCC)]4 were measured at pH 4.7 in the absence (left panel) and in the presence (central panel) of a proton exchange catalyst (0.18 M phosphate) at pH 5.6. The exchange time of the T3\* imino proton is comparable to that of the thymidine monomer (△). The exchange time of the C2\* imino proton is identical to the duplex interconversion time. The T3 imino proton does not exchange from the {ac} duplex. In 0.18 M phosphate its exchange time is equal to the duplex interconversion time. In the absence of added proton exchange catalyst, T3 imino proton exchange is controlled either by the duplex interconversion ( $t < 15^\circ\text{C}$ ), or by the exchange rate from T3\*, the homologous residue on the {bd} duplex ( $t > 30^\circ\text{C}$ ). Right panel: Duplex interconversion time (l) and C2\* imino proton exchange time (+) in [d(5mCCUCC)]4.

cytidine by strong H1'-H1' and C1<sub>av</sub> CH3-C5<sub>av</sub> H5 cross-peaks. C4<sub>av</sub> was identified by reciprocal amino-H2'/H2'' cross-peaks to C5<sub>av</sub>. The last cytidine spin system was assigned to C2<sub>av</sub> by the strong H1'-H1' cross-peak with C4<sub>av</sub>. The intercalation topology of the cytidine residues in the motion-averaged structure of [d(5mCCUCC)]4 is the same as that determined for [d(5mCCTCC)]4.

#### Proton exchange in [d(5mCCUCC)]4

The proton exchange characteristics are similar to those described above for [d(5mCCTCC)]4. At low temperature, the duplex interconversion is slow on the NMR time scale. In this case, exchange of the C2\* imino proton is controlled by duplex interconversion as for the [d(5mCCTCC)]4 tetramer (Figure 12). At 32°C, when the duplex interconversion is fast on the NMR time scale, all cytidine imino proton peaks except that of the (C4-C4<sub>av</sub>)<sub>av</sub> pair are broadened by exchange with water. This establishes that the exchange rate of the imino proton of the (C4-C4<sub>av</sub>)<sub>av</sub> pairs is much slower than the duplex interconversion rate (about 1000 s<sup>-1</sup> at

Table 3. Geometrical parameters of the [d(mCCTCC)]<sub>4</sub> tetramer<sup>a</sup>

	mC1-C2-T3-C4-C5 {ac} duplex				mC1*-C2*-T3*-C4*-C5* {bd} duplex				
	Glycosidic torsion (°)	Pseudo-rotation angle (°)	Rise <sup>b</sup> (Å)	Buckle <sup>d</sup> (°)	Glycosidic torsion (°)	Pseudo-rotation angle (°)	Rise <sup>b</sup> (Å)	Buckle <sup>d</sup> (°)	C1'-C1' distance (Å)
Base	C1	-130 ± 5	9 ± 13						
	C2	-110 ± 3	66 ± 64		C1*	-143 ± 3	52 ± 40		
	T3	-141 ± 3	17 ± 28		C2*	-154 ± 3	40 ± 40		
	C4	-132 ± 3	35 ± 20		T3*	-107 ± 2	-215 ± 10		
	C5	-143 ± 4	21 ± 10		C4*	-133 ± 6	-205 ± 50		
Duplex step		Helical twist (°)				Helical twist (°)			
	C1-C2	16 ± 4	6.5 ± 0.1		C1*-C2*	22 ± 3.2	5.7 ± 0.6		
	C2-T3	15 ± 2	3.3 ± 0.2		C2*-T3*	35 ± 2	—		
	T3-C4	31 ± 5	5.8 ± 0.3		T3*-C4*	39 ± 3	—		
	C4-C5	19 ± 5	5.7 ± 1		C4*-C5*	27 ± 10	7.1 ± 0.4		
Base-pair		Propeller twist (°)				Propeller <sup>e</sup> twist (°)			
	C1.C1	5 ± 2	-2 ± 4		C1*.C1*	-8 ± 2	-6 ± 6		
	C2.C2	5 ± 2	-6 ± 4		C2*.C2*	-3 ± 2	3 ± 1		
	T3.T3	7 ± 2	17 ± 4		T3*.T3*	—	—		
	C4.C4	2.5 ± 2	6 ± 3		C4*.C4*	4 ± 3	-12 ± 3		
C5-C5	-2.4 ± 3	0 ± 3		C5*.C5*	-10 ± 7	16 ± 15			

<sup>a</sup> Average and RMSD are computed for 14 conformers.

<sup>b</sup> The rise is defined as the distance between the average of the base planes.

<sup>c</sup> The propeller twist is the complement of the angle between the C4-C2 vector and the longitudinal symmetry axis. It is negative when the vector points in the 3' direction.

<sup>d</sup> The buckle is computed as the complement of the angle between the C6-N3 vector and the longitudinal symmetry axis. It is negative when the vector points in the 3' direction.

Table 4. Inter-duplex phosphorus-phosphorus distances and base stacking distances in [d(mCCTCC)]<sub>4</sub>

Inter-strand phosphorus-phosphorus distances (Å)				Paired bases	Stacking distances (Å)
Across the wide grooves		Across the narrow grooves		C5*C5*-C1.C1	3.7 ± 0.5
C2(P)-C5*(P)	14.9 ± 0.7	C2(P)-C5*(P)	9 ± 2	C1.C1-C4*.C4*	3 ± 0.4
C2(P)-C4*(P)	13.4 ± 2	C2(P)-C4*(P)	12.7 ± 0.9	C4*.C4*-C2.C2	3.1 ± 0.4
T3(P)-C4*(P)	15.5 ± 0.6	T3(P)-C4*(P)	12.6 ± 0.6	C2.C2-T3.T3	3.1 ± 0.3
T3(P)-T3*(P)	13.7 ± 0.3	T3(P)-T3*(P)	11.1 ± 0.6	T3.T3-C2*C2*	2.8 ± 0.3
C4(P)-T3*(P)	16.1 ± 0.6	C4(P)-T3*(P)	9.4 ± 1.1	C2*.C2*-C4.C4	3 ± 0.1
C4(P)-C2*(P)	13.8 ± 1.2	C4(P)-C2*(P)	10.7 ± 0.7	C4.C4-C1*.C1*	2.7 ± 0.3
C5(P)-C2*(P)	16 ± 0.4	C5(P)-C2*(P)	9.5 ± 0.8	C1*.C1*-C5.C5	3 ± 0.4

32°C). Due to line overlap, exchange of the U imino proton is not accessible to quantitative measurements.

## Discussion

We have previously reported the structures of the i-motif tetramers formed by d(TCCCC), (Gehring *et al.*, 1993), d(TCC), d(5mCCT) and d(T5mCC) (Leroy & Guéron, 1995). Each tetramer structure consists of two intercalating duplexes of parallel strands and the four strands are identical on the NMR time-scale.

The first three oligomers form a single tetrameric structure, which maximizes the degree of intercalation of the duplexes. For (d(T5mCC)), two tetramers co-exist. One maximizes intercalations, the second avoids steric hindrance between the cytidine methyl groups by a lesser degree of intercalation. In all the cases, each duplex is made of two identical strands, and the tetramers of four identical strands.

In common with these structures, the tetramer of d(5mCCTCC) consists of two intercalated duplexes, each made of two parallel and identical strands. However, the duplexes are not identical: one has its thymidine residues turned inward (designated as {ac} or A) and the other outwards (designated as {bd} or B). The tetramer is an asymmetric assembly AB and it fluctuates spontaneously between the strictly equivalent AB and BA forms. The transition can be followed by the kinetics of the exchange cross-peaks. It is fully concerted: the spectrum shows no indication of other forms, neither AA nor BB, nor of any forms which include a duplex built of non-equivalent strands.

### Intercalation of thymidine and tetramer stability

Duplexes A and B differ essentially by the arrangement of the thymidine residues. In the A duplex, the NOEs observed between the thymidine residues and the adjacent C2·C2<sup>+</sup> and C2\*·C2\*<sup>+</sup> pairs indicate a symmetrical T3·T3 base-pair. The C1'-C1' distance between the paired thymidine residues (10.4 Å) is about 1 Å larger than the corresponding distance in C·C<sup>+</sup> pairs. This allows the incorporation of the T3·T3 pair within the i-motif without marked distortions of the ribose backbone.

In the B duplex, the T3\* residues are symmetrically extruded in the wide grooves.

The similarity of the structures adopted by d(5mCCTCC) and by d(5mCCUCC) excludes a steric effect in relation with the methyl groups of the thymidine residues, but the difference of the activation energy for the motion of uridine and thymidine residues shows that swinging of the latter is impeded by the methyl group.

One can model an AA tetramer, without any particular van der Waals strain. However, the disposition of the two stacked thymidine pairs which juxtapose negatively charged atoms (N1-02, 02-N1, N3-N3, 04-04) on one hand, and positive carbons (C2-C2) on the other, is electrostatically unfavourable. A BB tetramer can also be modelled but it would have less stacking interactions than an AB.

The comparison of the dissociation constant of [d(5mCCTCC)]<sub>4</sub>,  $K_d = (1.1 \times 10^{-4})^3 \text{ M}^3$ , with that of tetramers formed by a continuous stretch of four cytidine residues such as [d(TCCCC)]<sub>4</sub>,  $K_d = (10^{-5})^3$  (unpublished result) and [d(CCCCAA)]<sub>4</sub>,  $K_d = (1.2 \times 10^{-5})^3 \text{ M}^3$  (Leroy *et al.*, 1994), suggests that incorporation of a T·T pair in the i-motif is somewhat destabilizing.

### Chemical shift differences between homologous residues

The chemical shifts in nucleic acids are affected by ring currents and by base protonation. H-bonding also contributes to the chemical shifts of exchangeable protons. The chemical shifts induced by the weak ring current of thymidine (Giessner-Prettre *et al.*, 1976) cannot account for the large shift differences between homologue cytidine residues. The chemical shift differences between the homologous pairs of each duplex are maximal between the C2·C2<sup>+</sup> and C2\*·C2\*<sup>+</sup> pairs, and decrease towards both ends of the tetramer. They reflect the difference of structure between the two duplexes. The NMR peak of the C2 external amino proton has an intriguing upfield shift of 1.34 ppm by comparison to its C2\* homologue. By contrast, the internal amino protons of C2 and C2\* are both in the frequency range expected for H-bonded amino protons in C·C<sup>+</sup> pairs. In the computed structure of [d(5mCCTCC)]<sub>4</sub>, the thymidine residues T3\* are

Table 5. Imino and amino proton exchange times (seconds) in the [d(5mCCTCC)]<sub>4</sub> tetramer

Residue	C5*	C1	C4*	C2	T3*	T3	C2*	C4	C1*	C5
H3	<sup>a</sup>	<sup>b</sup>	730	<sup>a</sup>	0.15	1	0.7	730	<sup>b</sup>	<sup>a</sup>
H <sub>4int</sub>	450 ≤ <sup>c</sup>	5300	5300	570	—	—	570	5300	5300	450 ≤ <sup>c</sup>
H <sub>4ext</sub>	450 ≤ <sup>c</sup>	450 ≤ <sup>c</sup>	800	≈600	—	—	800	800	450 ≤ <sup>c</sup>	450 ≤ <sup>c</sup>

<sup>a</sup> The absence of the imino proton peak indicates exchange times shorter than 5 ms.

<sup>b</sup> Unresolved proton. Magnetization transfer experiments from water indicate exchange times longer than 0.1 second but too short to be measured in real time experiments:  $0.1 < \tau_{\text{ex}} < 60$  seconds.

<sup>c</sup> Inaccurate estimation due to insufficient resolution.

symmetrically embedded in the large grooves along the C2·C2<sup>+</sup> pair. The external amino proton of C2 is located 3.5 Å under the N1 atom of T3\* in the direction perpendicular to the T3\* aromatic ring towards the axis of the helix (Figure 10). In this position, the ring currents of T3\* cannot contribute for a chemical shift larger than 0.15 ppm (Giessner-Prettre *et al.*, 1976).

The chemical shifts of C2 amino protons are unchanged between pH 6 and pH 3, in a pH range where titration of the N3 position would induce a 1.5 ppm lowfield shift to the amino protons of the cytidine monomer (Razka, 1974). This shows that the anomalous shift is not due to incomplete N3 protonation.

Imino and amino proton chemical shifts are greatly affected by H-bonding to the oxygen or to the nitrogen of the complementary base, and to the oxygen of hydration water. The familiar 11 ppm position of the imino protons of T and G monomers corresponds to imino protons H-bonded to water. When dissolved in a solvent such as chloroform, the imino proton of uridine is shifted 2.5 ppm upfield from its position in H<sub>2</sub>O (Guéron *et al.*, 1983). Similarly, the amino protons of the neutral and of the protonated N1-methyl cytidine derivative dissolved in liquid SO<sub>2</sub> are shifted upfield by about 2 ppm by reference to their position in H<sub>2</sub>O (Becker *et al.*, 1965). We note that such effects are not observed in dry DMSO solution, due most probably to H-bonding with the polar DMSO oxygen.

In the i-motif, the external amino protons point toward the wide grooves and are usually accessible to water. C2 external amino proton has the particularity to be buried beneath thymidine T3\*. Using the surface generation routine of the Quanta program, we observed that it is the sole external amino proton to be under the surface accessible to a molecule with a radius of 1.3 Å. This suggests that the upfield shift of this proton is due to the lack of hydration. According to this interpretation, T3\* would form an hydrophobic pocket around the C2 residue, impeding hydration of its external amino proton. The internal amino proton of C2, which is H-bonded to the oxygen O2 of the complementary cytidine, is not affected by this effect. This interpretation is consistent with the absence of cross-peaks between water and C2 amino protons, and with the slow exchange with water of the C2 external amino proton (Table 5).

## Duplex interconversion

Tetramer dissociation observed by gel filtration chromatography is extremely slow in comparison with duplex interconversion. This shows that the motion of the thymidine residues does not require disruption of the tetramer.

The imino protons of [d(5mCCTCC)]<sub>4</sub> provide excellent markers to follow the internal motions related to the concerted opening-closing of the thymidine residues. The interpretation of the proton exchange kinetics is particularly straightforward for the imino proton of C·C<sup>+</sup> pairs whose exchange time may be identified with the base-pair lifetime (Leroy *et al.*, 1994).

The amino proton spectrum of the C5·C5<sup>+</sup> and C5\*·C5\*\* pairs shows that these pairs are hemi-protonated, but the absence of imino proton signals on the NMR spectrum at -7°C indicates fast exchange with water. This reflects the fast opening motion of the terminal pairs by reference to the case of terminal Watson-Crick base-pairs (Nonin *et al.*, 1995). The fast exchange rate of the C2·C2<sup>+</sup> pair is also deduced from the lack of imino proton signal for this pair (Figure 3).

There is no indication for T3 imino proton exchange from the {ac} duplex. The absence of effect of ammonia provides an upper limit of 10<sup>-4</sup> for the dissociation constant of the T3·T3 pair. The low pH conditions required for the i-motif formation impede the study of T imino proton exchange, and we cannot exclude an opening motion faster than the duplex interconversion for this pair. The local distortions caused by the switch of the thymidine residues entail opening of the C2\*·C2<sup>+</sup> pair which is found to be synchronized with duplex interconversion. The fast exchange rate of the imino proton of the C2·C2 pair indicates a lifetime shorter than the duplex interconversion time. It may result from unfavourable stacking with the sequentially adjacent T3·T3 pair, or from backbone distortions induced by the open T3\* residues.

The exchange rate of C4 and C4\* imino protons is slower than the duplex interconversion rate. According to equation (4) it is equal to the average of the opening rate of C4·C4<sup>+</sup> and C4\*·C4\*\* base-pair. Hence, the thymidines switch does not require opening of the C4·C4<sup>+</sup> and C4\*·C4\*\* pairs whose lifetimes are much longer than the duplex interconversion time.

## [d(5mCCTCC)]<sub>4</sub> and other i-motif structures

The similarity of the [d(5mCCTCC)]<sub>4</sub> structure with those previously described for other i-motif tetramers, shows that the main features of the i-motif are maintained when a T·T pair is incorporated.

The helical twist at the C2-T3 step (16°) is typical of values found in i-motif structures but the twists at the T3-C4, C2\*-T3\* and T3\*-C2\* steps are larger. As indicated by the strong NOEs between H3' and H6 protons, all the sugars are N-type except for the open thymidine (T3\*) and for the adjacent C4\* residue which are S-type. This is unexpected for C2 and T3 which are stacked together, and for which a 2' *endo* sugar-puckering might be expected.

The inter-strand phosphorus-phosphorus distances across the wide grooves are similar to values commonly found in other i-motif structures, but the phosphorus-phosphorus distances across the narrow grooves, in particular those involving the C2-T3\*-T3 segment, are larger (Table 3). Examination of the structure shows that the rotation of the T3\* sugar backbone towards the major groove may contribute to increase the narrow groove width. The potential H1'-H1' NOE expected between couples 5mC1-C5\*, C2-C4\* and C2-T3\* (see Figure 8) falls in the strong non-resolved  $\beta$  cross-peak (Figure 2). The absence of distance constraints across this section of the narrow groove may result in an over-evaluation of the groove width.

Each hemi-protonated duplex of [d(5mCCTCC)]<sub>4</sub> is symmetrical on the NMR time-scale. Common to all solution structures, this feature contrasts with the duplex asymmetry observed in the crystal structures of i-motif tetramers (Chen *et al.*, 1994; Kang *et al.*, 1995; Berger *et al.*, 1995). In the case of d[d(CCCT)]<sub>4</sub>, the solution (J.-L. Leroy, unpublished result) and crystal structures (Kang *et al.*, 1994) are available. In the crystal, the tetramer is formed by the intercalation of two identical duplexes, each of which is non-symmetrical, whereas in solution the duplexes are symmetrical even at very low temperature: -30°C for a spectrum obtained in 33% deuterated methanol 67% H<sub>2</sub>O.

Different phenomena can contribute to this difference. The NMR structure may correspond to fast exchange in solution between different strand structures. The crystal structure may be due in part or entirely to crystal packing.

The asymmetric [d(5mCCTCC)]<sub>4</sub> tetramer is an example of an association of identical units in which the equivalence is lost. It is a case of broken symmetry in a biological macromolecular assembly, the first in which the structure and switching kinetics are fully characterized. The switching properties of the asymmetrical "AB" tetramer could be investigated in other sequences derived from 5mCCTCC. The number of thymidine or of cytidine residues may be increased.

## Materials and Methods

### Oligomer preparation and NMR samples

The oligomers were synthesized and purified according to procedures already described (Leroy & Guéron, 1995). The strand concentration of the NMR samples, 6 mM for d(5mCCUCC) and 8 mM for d(5mCCTCC), was determined from the absorbance measured at neutral pH using the  $A_{260}$  value of 39,250 M<sup>-1</sup> computed according to a nearest neighbour model (Cantor & Warshaw, 1970).

### Gel filtration chromatography

The oligomer stoichiometry was determined from the concentration dependence of the multimer-monomer equilibrium, as measured by high pressure chromatography on a Synchropack GPC 100 Gel filtration column. The column was calibrated with mononucleosides (C and T), with single-stranded oligomers, with DNA duplexes, and with oligo(C) tetramers such as [d(TCCCC)]<sub>4</sub>, [d(CCCCAA)]<sub>4</sub> and [d(TTCCCCCCTT)]<sub>4</sub>. The samples injected in the chromatography column were prepared by dilution of small aliquots obtained from the NMR samples into a 0.3 M NaCl, 20 mM acetate buffer (pH 4.7). They were incubated four days at room temperature in order to reach the monomer-tetramer equilibrium. The sample were eluted in less eight minutes (flow rate 0.4 ml/min).

### NMR methods

NMR experiments were performed on a home-built 360 MHz spectrometer and on a Bruker AMX600 spectrometer. The "jump and return" sequence with maximum sensitivity at 13.5 ppm was used for water suppression (Plateau & Guéron, 1982).

All NOESY experiments used 90° pulses for preparation, mixing and detection.

Pure exchange spectra were obtained using a mixed "NOESY-ROESY" sequence designed to compensate the effects of longitudinal and transverse relaxation in order to suppress the cross-relaxation. This is achieved by forcing the magnetization to spend twice as much time along the main field axis than in the perpendicular direction. The sequence was derived from that described by Fejzo *et al.* (1991), by addition of z-gradient pulses before and after the NOESY-ROESY loop for the suppression of extraneous transverse magnetization. The elimination of the dipolar cross-peaks was checked by the absence of the cytidine intra-residue H5-H6 NOE cross-peaks (Figure 4).

NOESY and NOESY-ROESY experiments in H<sub>2</sub>O were read with the JR sequence. The spectral width was 7.2 kHz, the acquisition time 0.23 second, and the repetition rate one second. The  $t_1$  delay (240 increments) was incremented from 2  $\mu$ s to 60 ms.

2D experiments in <sup>2</sup>H<sub>2</sub>O were recorded with a spectral width of 3 kHz and an acquisition time of 0.34 second. The repetition rate was two seconds. The residual H<sup>2</sup>O signal was generally not saturated in order to avoid saturation of the H3' protons. The  $t_1$  delay was incremented from 2  $\mu$ s to 84 ms (256 increments). TOCSY experiments used ten MLEV-17 repetitions (total time 15 ms) without trim pulses (Bax *et al.*, 1985).

The 2D data were processed with Felix 1.1 (courtesy of Hare Research). Apodization in both dimensions

included exponential broadening (1 Hz in  $^2\text{H}_2\text{O}$ , 3 Hz in  $\text{H}_2\text{O}$ ), a sine-bell with a phase shift of either  $40^\circ$  for peak assignment, or  $60^\circ$  for measurement of NOESY cross-peak volumes and for observation of exchange cross-peaks near the diagonal, and with a Felix skew factor of 0.8. The remaining water signal was reduced on 1D and 2D experiments by post-acquisitional processing. 1D spectra were corrected for the frequency response to the JR excitation (Guéron *et al.*, 1991).

### Rate of duplex interconversion

The rate of duplex interconversion was determined in  $^2\text{H}_2\text{O}$  solution from the rate of magnetization transfer between the H6 protons of the thymidine residues of each duplex. Let {ac} and {bd} be the duplexes, and T3 and T3\* their respective thymidine residues. The magnetization of T3 H6 was selectively saturated using a  $90^\circ$  DANTE pulse followed by a 1 ms z-gradient pulse. The peak was saturated by a succession of DANTE pulses followed by a 0.5 ms z-gradient pulse. The longitudinal magnetization,  $M_z(t)$ , of T3\* H6 was measured as a function of the time during which T3 H6 was maintained saturated. According to Forsèn & Hoffman (1963):

$$M_z(t) = M_0[(T_{1d}/\tau_d)\exp(-t/T_{1d}) + (T_{1d}/T_1)] \quad (1)$$

where  $\tau_d$  is the duplex interconversion time, and  $T_1$  is the longitudinal relaxation time of T3\* H6.  $M_0$  is the initial magnetization of T3\* H6, and  $T_{1d} = (1/\tau_d + 1/T_1)^{-1}$ .  $T_1$  was determined in a separate set of inversion-recovery experiments.

### Imino proton exchange

Exchange of the imino proton in a base-pair requires disruption of the pair, followed by proton transfer to an acceptor such as  $\text{NH}_3$  or phosphate acting as catalyst. Imino proton exchange from the open pair is also catalysed by the cyclic nitrogen of the paired nucleotide (A N1 in A:T pairs, C N3 in C:C+ pairs) which acts as an intrinsic catalyst in a concerted transfer involving a water molecule (Guéron *et al.*, 1987). The exchange time,  $\tau_{\text{ex}}$ , is given by:

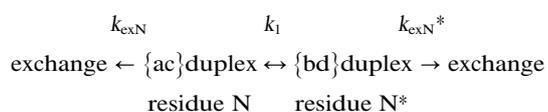
$$\tau_{\text{ex}} = \tau_0[1 + 1/(K_{\text{diss}} \times k_{\text{ex,acc open}})] \quad (2)$$

where  $\tau_0$  is the base-pair lifetime,  $K_{\text{diss}}$  the dissociation constant, and  $k_{\text{ex,acc open}}$ , the rate of transfer to the proton acceptor from the open pair.

In the case of a C:C+ base-pair, the intrinsic catalysis is so efficient that the imino proton exchanges with water at each opening event, so that the exchange time is equal to the base-pair lifetime (Leroy *et al.*, 1994).

### Imino proton exchange in the case of two exchanging structures

The exchange rate of several imino protons of the [d(5mCCTCC)]4 tetramer depends on the rate of duplex interconversion. Let N be a residue of the {ac} duplex, and N\* the corresponding residue in the {bd} duplex.



The expression of imino proton exchange is simple in three extreme cases.

(1) When the duplex interconversion rate  $K_1$  is slower

than both imino proton exchange rates, ( $k_{\text{exN}}, k_{\text{exN}^*} > k_1$ ), exchange occurs independently from N and from N\* with the rate  $k_{\text{exN}}$  and  $k_{\text{exN}^*}$ .

(2) When the duplex interconversion rate is faster than the exchange rate of the imino proton of one species ( $k_{\text{exN}^*} > k_1 > k_{\text{exN}}$  for instance), the measured exchange rate of the slow exchanging imino proton,  $k_{\text{mesN}}$ , depends on the interconversion rate and on the exchange rate from N\* according to:

$$k_{\text{mesN}} = k_1 \cdot k_{\text{exN}^*} / (k_1 + k_{\text{exN}^*}) \quad (3)$$

whereas the exchange rate of the fast exchange rate is not affected.

(3) When the interconversion rate is faster than both imino proton exchange rates ( $k_1 > k_{\text{exN}^*}, k_{\text{exN}}$ ), the imino protons exchange at the same rate:

$$k_{\text{mesN}} = k_{\text{mesN}^*} = (k_{\text{exN}^*} + k_{\text{exN}}) / 2 \quad (4)$$

The rates of proton exchange were measured as described previously (Guéron & Leroy, 1996).

### Structure determination

#### NMR determinants of the i-motif

In the intercalated duplexes, the large distances between sequentially adjacent residues result in the quasi-absence of sequential NOE connectivities at short mixing time. The intercalation topology of the i-motif is read off the patterns of H1'-H1' cross-peaks which connect the stacked residues across the narrow grooves (Gehring *et al.*, 1993), and of the amino-H2'/H2'' cross-peaks connecting across the wide grooves the nucleotides stacked *via* contacting 3'-faces. Stacked C'C+ pairs are also connected by strong H5-amino proton cross-peaks (Leroy & Guéron, 1995). The intercalation stretches the sugar-phosphate backbones, leading to N-type puckers and strong intra-residue H3'-H6 cross-peaks.

#### NOESY cross-peaks and distance restraints

The structure of [d(5mCCTCC)]4 was determined on the basis of the inter-proton distances obtained from the build-up of NOE cross-peaks measured at mixing times of 40, 60, 85, 110, 135, 150 and 230 ms in  $\text{H}_2\text{O}$  at  $-7^\circ\text{C}$ , and of 35, 70, 90, 120 and 250 ms in  $^2\text{H}_2\text{O}$  at  $-5^\circ\text{C}$ . We distinguished the cross-peaks due to cross-relaxation from those due to chemical exchange using pure chemical exchange spectra.

The NOESY cross-peak volumes were measured manually as the product of the number of contour plots by the cross-peak section at half-height. Inter-proton distances were scaled using the intra-residue H5-H6 cross-peak of C4, C4\*, C5, C5\* and C2\* and the corresponding inter-residue distance of 2.45 Å.

Distances involving a methyl group were scaled by reference to the H6-CH3 pseudo-atom distance of 2.9 Å.

The distances were sorted into three categories corresponding to lower and upper bounds of 1.8 to 2.9, 1.8 to 3.7 and 2.9 to 4.2 Å. In some cases, the absence of NOE or the observation of weak cross-peaks at long mixing time was converted into repulsive restraints with a lower bond of 4.2 Å. For cross-peaks involving geminal protons, we used the distance corresponding to the strongest cross-peak as a restraint. The distances derived from the other cross-peaks were considered as lower

bonds. The distances involving methyl protons or C-C<sup>+</sup> imino protons were loosened by  $\pm 0.5$  Å.

Due to insufficient spectral resolution, several cross-peaks, in particular many of those which connect C5\* to 5mC1 at one end of the tetramer, and C5 to 5mC1\* at the other end, could not be translated into distance restraints. In such cases, we verified on the final computed structure that at least one of the corresponding inter-proton distances can account for the intensity of these cross-peaks.

The sugar conformation and the glycosidic torsion angles were constrained by distances derived from intra-residue H3'-H6, H1'-H6, and H3'-H2'/H2'' NOE cross-peaks. The structures computed with and without the intra-residue restraints show the same intercalation topology.

#### Base-pairing restraints

Based-paired strands were designated {a} and {c} for one duplex, and {b} and {d} for the other. The chemical shifts and the exchange properties of imino and amino protons indicated the base-pairing of the cytidine residues. The C-C<sup>+</sup> pairs were defined by the O2-N4 and N3-N3 distances which were both set to  $2.75(\pm 0.1)$  Å. No base-pairing was imposed for the thymidine residues. We did not impose planarity of the base-pairs.

#### Strand equivalence and symmetry restraints

The NMR spectra of [d(5mCCTCC)]<sub>4</sub> show the intercalation of two distinct symmetrical duplexes. The strategy used to overcome the ambiguities resulting from the duplex symmetry is derived from that previously used to solve the structure formed by four identical d(TCC) strands (Leroy & Guéron, 1995). The modelling started with four strands ({a}, {b}, {c}, and {d}) in arbitrary extended conformations with ideal geometry. Each parallel duplex, {ac} and {bd}, was defined by the base-pairing distance restraints declared between the equivalent cytidine. In order to enforce the identity of the base-paired strands, several intra-residue distances (O2-O3', N4/O4-C4', N1-C5', and N3-C2') were declared identical (within 0.1 Å) for each base-paired residue. The equivalence of the {a}-{b} and {c}-{d} inter-strand relations was enforced by six inter-strand segments, which were defined between a selected set of atoms (N3, C4, C5, N1, O4', O3') of the residues of strand {a}, and the same set of atoms of the corresponding residues of strand {b}. These segments were declared equivalent (within  $\pm 0.1$  Å) to the equivalent segments defined between strands {c} and {d}. Similarly the corresponding segments defined between strands {a} and {d}, and between strands {c} and {b} were declared equivalent. Using this procedure we obtained an RMSD of about 0.1 Å between the equivalent strands of each computed conformer, a value much smaller than the RMSD between two computed conformers.

Due to the symmetry of the duplexes, the strand appartenance of the NOE-connected protons is ambiguous. To solve these ambiguities, a preliminary model was built using only the distance restraints derived from the characteristic H1'-H1' and amino-H2'/H2'' cross-peaks. By reference to the conventions used in the structural study of [d(TCC)]<sub>4</sub> and [d(5mCCT)]<sub>4</sub>, (Leroy & Guéron, 1995), we designate {a}-{b} the couples of strands connected by amino-H2'/H2'' NOE cross-peaks, and {c}-{d} the couple of strands connected by the short

H1'-H1' distances. Using this preliminary model, we could assign all the other inter-duplex connectivities to either the {a}-{b} or to the {a}-{d} couples of strands. Similarly, the intra-duplex NOE cross-peaks could be assigned either to intra-strand distance connectivities or to distance connectivities between equivalent base-paired strands.

The NMR spectrum implies an identical environment for the two strands of a duplex. To this end, identical sets of distance restraints were declared between the couples of strands {a}-{d}, and {c}-{b}, and between the couples {a}-{b} and {c}-{d}. Identical sets of intra-strand distances were declared in strands {a} and {c}, and in strands {b} and {d}.

#### Distance restrained molecular dynamics

The structures were computed using the simulated annealing method (Brünger, 1990) of the X-PLOR program. Potential energy terms related to electrostatics and "empirical dihedral" were omitted. The force constant expressing distance restraints was set to  $50 \text{ kcal mol}^{-1} \text{ Å}^{-2}$ . The force constant related to H-bonding restraints was set to  $500 \text{ kcal mol}^{-1} \text{ Å}^{-2}$ . The standard harmonic potential for covalent geometry was used. The simulated annealing proceeded as follows. In a preliminary phase, the energy of the starting structure was minimized by five Powell cycles. The molecular dynamics computation was run for 2000 steps of 2 ps, with an initial velocity corresponding to 2000 K. H-bonding energy was introduced during the cooling procedure to 300 K in steps of 25 K, each step being followed by a 0.1 ps dynamics computation. Lastly, the energy was minimized by 400 Powell cycles. The computation was repeated to yield 24 conformers.

Thirteen conformers were selected for their low NOESY-related energy. The RMSD were determined by pairwise comparison of the conformers aligned on their C1' positions.

#### Coordinates deposition

The coordinates of the lowest NOESY-related energy conformer [d(5mCCTCC)]<sub>4</sub> have been submitted to the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973, USA, and are available from the authors on request until released.

#### Acknowledgements

We thank M. Guéron for stimulating discussions and help in the preparation of the manuscript. We thank M. Leblond for synthesis and purification of DNA. X. Han is acknowledged for assistance with gel filtration chromatography. This work was supported by grant 6520/6 July 1994 from The Association pour la Recherche contre le Cancer.

#### References

- Bax, A. & Davis, D. G. (1985). MLEV-17 based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* 65, 355-360.
- Becker, E. D., Todd Miles, H. & Bradley, R. B. (1965). Nuclear magnetic resonance studies of methyl

- derivatives of cytosine. *J. Am. Chem. Soc.* 87, 5575–5582.
- Berger, I., Kang, C., Fredian, A., Ratliff, R., Moyzis, R. & Rich, A. (1995). Extension of the four-stranded intercalated cytosine motif by adenine-adenine base-pairing in the crystal structure of d(CCCAAT). *Struct. Biol.* 2, 416–425.
- Brünger, A. T. (1990). X-PLOR Version 3.1, a system for X-ray crystallography.
- Cantor, C. R. & Warshaw, M. M. (1970). Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers*, 9, 1059–1077.
- Chen, L., Cai, L., Zhang, X. & Rich, A. (1994). Crystal structure of a four-stranded intercalated DNA: d(C4). *Biochemistry*, 33, 13540–12546.
- Fejzo, J., Westler, W. M., Macura, S. & Markley, J. L. (1991). Strategies for eliminating unwanted cross-relaxation and coherence-transfer effects from two-dimensional chemical-exchange spectra. *J. Magn. Reson.* 92, 20–29.
- Forsén, S. & Hoffman, R. A. (1963). Study of moderately rapid chemical exchange reaction by means of nuclear magnetic double resonance. *J. Chem. Phys.* 39, 2892–2901.
- Gehring, K., Leroy, J.-L. & Guéron, M. (1993). A tetrameric DNA structure with protonated cytidine-cytidine base pairs. *Nature*, 363, 561–565.
- Giessner-Prettre, C., Pullman, B., Borer, P. N., Kan, L.-S. & Ts'o P. O. P. (1976). Ring-current effects in the NMR of nucleic acids. *Biopolymers*, 15, 2277–2286.
- Guéron, M. & Leroy, J.-L. (1996). Nucleic acid dynamics. *Methods Enzymol.* 261, 383–413.
- Guéron, M., Leroy, J.-L. & Griffey, R. H. (1983). Proton nuclear magnetic relaxation of <sup>15</sup>N-labelled nucleic acids via dipolar coupling and chemical shift anisotropy. *J. Am. Chem. Soc.* 105, 7262–7266.
- Guéron, M., Kochoyan, M. & Leroy, J.-L. (1987). A single mode of DNA base-pair opening drives imino proton exchange. *Nature*, 382, 89–92.
- Guéron, M., Charretier, E., Hagerhorst, J., Kochoyan, M., Leroy, J.-L. & Moraillon, A. (1989). In *Biological Structure, Dynamics, Interactions & Expression; Proceedings of the Sixth Conversation in Biomolecular Stereodynamics* (Sarma, R. H. & Sarma, M. H., eds), vol. 2, pp. 113–137, Adenine Press, New York.
- Guéron, M., Plateau, P. & Decors, M. (1991). Solvent signal suppression in NMR. *Prog. NMR Spectrosc.* 23, 135–209.
- Inman, R. B. (1964). Transition of DNA homopolymers. *J. Mol. Biol.* 9, 624–637.
- Kang, C., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R. & Rich, A. (1984). Crystal structure of intercalated four-stranded d(C3T) at 1.4 Å. *Proc. Natl Acad. Sci. USA*, 91, 11636–11640.
- Kang, C., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R. & Rich, A. (1995). Stable loop in the crystal structure of the intercalated four-stranded cytosine-rich metazoan telomer. *Proc. Natl Acad. Sci. USA*, 92, 3874–3878.
- Langridge, R. & Rich, A. (1963). Molecular structure of helical polycytidylic acid. *Nature*, 198, 725–728.
- Leroy, J.-L. & Guéron, M. (1995). Solution structure of the i-motif tetramers of d(TCC), d(5methylCCT) and d(T5methylCC): novel NOE connection between amino protons and sugar protons. *Structure*, 3, 101–120.
- Leroy, J.-L., Gehring, K., Kettani, A. & Guéron, M. (1993). Acid multimers of oligo-cytidine strands: stoichiometry, base-pair characterization and proton exchange properties. *Biochemistry*, 3, 6019–6031.
- Leroy, J.-L., Guéron, M., Mergny, J.-L. & Hélène, C. (1994). Intra-molecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif. *Nucl. Acids Res.* 22, 1600–1606.
- Mergny, J.-L., Lacroix, L., Han, X., Leroy, J.-L. & Hélène, C. (1995). Intra-molecular folding of pyrimidine oligodeoxynucleotides into an i-DNA motif. *J. Am. Chem. Soc.* 117, 8887–8898.
- Nilges, M. (1993). A Calculation strategy for the structure determination of symmetric dimers by NMR. *Proteins: Struct. Funct. Genet.* 17, 297–309.
- Nonin, S., Leroy, J.-L. & Guéron, M. (1995). Terminal base pair of oligodeoxynucleotides: imino proton exchange and fraying. *Biochemistry*, 34, 10652–10659.
- Plateau, P. & Guéron, M. (1982). Exchangeable proton NMR without base-line distortion, using strong pulse sequences. *J. Am. Chem. Soc.* 104, 7310–7311.
- Razka, M. (1974). Mononucleotides in aqueous solution: proton magnetic resonance studies of amino groups. *Biochemistry*, 13, 4616–4622.
- Rohozinski, J., Hancock, J. M. & Keniry, M. A. (1994). Oligocytidine regions contained in DNA hairpin loops interact via a four-stranded, parallel structure similar to the i-motif. *Nucl. Acids Res.* 22, 4653–4659.