J|A|C|S COMMUNICATIONS

Published on Web 08/24/2009

$[5mCCTCTCTCC]_4$: An i-Motif Tetramer with Intercalated T•T Pairs

Muriel Canalia and Jean-Louis Leroy*

Laboratoire de Chimie et Biologie Structurale, Institut de Chimie des Substances Naturelles, Gif-sur-Yvette, France 91190

Received April 30, 2009; E-mail: Jean-Louis.Leroy@icsn.cnrs-gif.fr

At slightly acid pH, C rich sequences form tetrameric structures in which two parallel duplexes held together by hemiprotonated $C•C^+$ pairs are intercalated in each other.[1] This structure, called an **i**-motif, is exceptional in that it involves systematic base-pair **i**ntercalation. Thymidines can form symmetrical T•T base pairs that are nearly isomorphic of $C•C^+$ pairs; however the structural investigations of the $[5mCCTCC]_4$[2] and $[5mCCTC_nTCC]_4$ ($n = 1$, 2, 3) tetramers[3] and of the $[5mCCTCACTCC]_2$ dimer[4] show that incorporation of two T•T base pairs in a face-to-face orientation in the i-motif is hindered. All these structures exhibit a common structural feature including the T•T pair of one duplex, the unstacked thymidines of the other, and the adjacent $C•C^+$ pairs of each thymidine. This assembly has been designated as an "open/paired T/T motif" (Figure 1).

We have investigated the structure and the base-pair opening kinetics of the i-motif tetramer of 5mCCTCTCTCC. The sequence was synthesized with a 5mC that provided a marker readily identifiable on the NMR spectrum. $[5mCCTCTCTCC]_4$ is fully symmetrical. It includes two open/closed T3/T7 motifs, but the central thymidines form two long-lived T5•T5 pairs that are intercalated into the i-motif core. The analysis of several i-motif structures containing unstacked/closed T/T motifs suggests that the backbone stretching induced by the larger C1′−C1′ intrastrand distance hinders thymidine intercalation and that the cumulative effect of T3 opening on each side of $[5mCCTCTCTCC]_4$ results in a reduction of the backbone stretching allowing face-to-face intercalation of T5•T5 pairs.

$[5mCCTCTCTCC]_4$ Formation. Gel exclusion chromatograms recorded as a function of the time after melting of 5mCCTCTCTCC reveal the formation of two species: a dimer and tetramer (Figure S1). The NMR spectra collected at different times show that the dimer faction is maximal in the early stage of the association process and decreases slowly in favor of the tetramer. The tetramer half formation time is ∼8 h at 25 °C in a 2.3 mM oligonucleotide solution, pH 4.8. At 0 °C it exceeds 1 month. The dimer fraction at equilibrium is negligible (Figure S2). This behavior reflects the fast formation of a short-lived dimer in competition with a long-lived tetramer whose formation rate is slower. The kinetic trapping of a dimer was also reported during $[CCTCTCC]_4$ formation.[3]

The ^1H and ^{31}P NMR spectra of $[5mCCTCTCTCC]_2$ are nearly the same as those of the i-motif dimer of $[5mCCTCACTCC]_2$ (Table S1). This, together with the quasi identity of the exchange times of the equivalent imino protons of both dimers and the similarities of their NOESY spectra, indicates comparable structures formed by association of two identical hairpins whose loops with a single A5 or T5 residue are on the same side of the i-motif core.[4] It is noteworthy that the tetramer of [5mCCTCACTCC] was not observed.

$[5mCCTCTCTCC]_4$ Structure. The TOCSY and NOESY spectra of the tetramer show a single set of NMR peaks for each residue and therefore establish that the tetramer is the symmetrical assembly of two identical duplexes. The rules allowing the assignment of the spin systems and the identification of the intercalation topology of i-motif structures are well established.[5]

The intercalation topology of $[5mCCTCTCTCC]_4$ was determined by the NOE connectivities displayed in Figure S3. The structure includes two identical open/close T3/T7 motifs (Figure 1). The NOESY cross-peaks connecting C4(H1′) to T5(H1′) and C4(amino proton) to T5 (H2′/H2″) that are characteristic of residues stacked by contacting faces oriented in the 3′ direction indicate T5•T5 pairing and stacking between the two $C4•C4^+$ pairs (Figure S3).

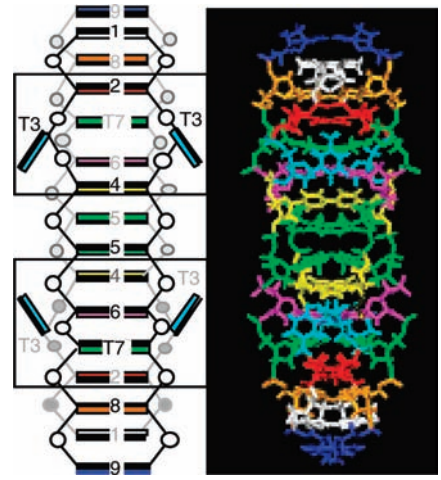

Figure 1. Schematic representation of the arrangement of the fully symmetrical and computed structure of $[5mCCTCTCTCC]_4$. The tetramer is formed by intercalation of two identical symmetrical duplexes (gray and black backbones) whose T3 bases are flipped out. T3 opening allows intercalation of the sequentially adjacent T7 and C6 pairs of the other duplex. The open/paired T3/T7 motifs are indicated by boxes. Two T5•T5 pairs are intercalated in a face-to-face orientation.

All the T and C imino protons are detected on the NMR spectrum at −5 °C, except those of the outer $C9•C9^+$ pair. The lifetime of $C•C^+$ pairs was identified with the exchange time of their imino proton,[6] and that of T•T pairs was obtained from the extrapolation at infinite phosphate concentration of the T(H3) exchange times.[7]

The lifetimes of T5•T5 and T7•T7, respectively, 0.2 and 0.11 s at 20 °C (Figure 2), are 3−4 orders of magnitude longer than that of wobble G•T pairs.[8] The fast exchange rate of T3(H3) shows that T3 is fully exposed.

The $[5mCCTCTCTCC]_4$ structure was computed using XPLOR[9] on the basis of interproton distances obtained from the buildup of NOE cross-peaks measured at a short mixing time in ^2H_2O and H_2O (Figure S3). On each duplex, the T3 bases are flipped out in the wide grooves. The intercalation topology of $[5mCCTCTCTCC]_4$ is the same as that reported previously for $[5mCCTCCCTCC]_4$, and the open/closed T3/T7 motifs of each structure are superim-

10.1021/ja903210t CCC: $40.75

posable.[3] This, together with the similarities of the lifetimes of the equivalent base pairs of both tetramers (Figure S4), indicates closely related structures. The T(H3)−T(O4) distances between the opposite thymidines of pairs T7•T7 and T5•T5 of the computed structure, 1.8 ± 0.2 Å, are consistent with H-bonding. Simulations enforcing T(O2)−T(H3) bonds resulted in structures exhibiting interproton distances inconsistent with the NOESY spectra. The deviation from planarity of the T5•T5 pairs is less than 15°. The interstrand C1′−C1′ distance in T•T pairs, 11.1 ± 0.2 Å, is significantly larger than that measured for the C•C$^+$ pairs: 9.5 ± 0.2 Å. The pseudo rotation angle of T5, 0 ± 8°, and T7, 40 ± 10°, indicate a sugar pucker in the same conformational range as that of intercalated C•C$^+$ pairs.

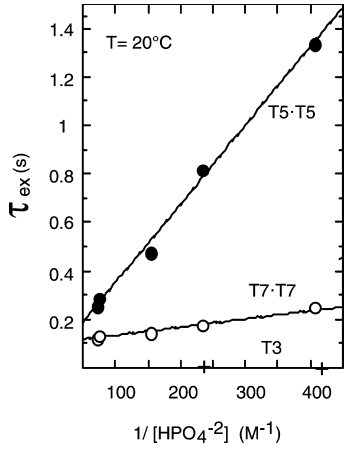

Figure 2. Exchange catalysis by phosphate of the T imino protons of [5mCCTCTCTCC]$_4$ at 20 °C. The exchange times of T5(H3) and T7(H3) tend toward limit values at infinite phosphate concentration corresponding to the lifetimes of the corresponding T•T pairs. T3 is not paired.

It is particularly interesting to show that the outer thymidines of [5mCCTCTCTCC]$_4$ form two symmetry related T3/T7 motifs but that the central thymidines form two T5•T5 pairs face-to-face intercalated in the i-motif. This observation provides indications of the origin of the conflict that usually hinder T•T intercalation into i-motif structures and more generally of the constraints influencing i-motif formation.

It has been suggested that unfavorable contacts between atoms with the same charge could prevent thymidine intercalation.[2] It is clear that this argument is inconsistent with the stability of the T5•T5 pairs of [5mCCTCTCTCC]$_4$ and that another explanation should be considered.

The structural differences observed in the organization of the thymidine groups of [5mCCTCTCTCC]$_4$ suggest that T5 intercalation may be a consequence of the dual opening of both T3 bases. Mutual intercalation of the C•C$^+$ pair requires a helical rise of ∼6.3 Å, a value approximately twice that of B DNA. This is obtained by a reduction of the helical twist of 15° to 30° by comparison with B-DNA and by backbone stretching enforcing the sugar pucker in the C3′*endo* to C4′*exo* conformational range instead of C2′ *endo* as in B DNA. Molecular modeling indicates that further backbone stretching is energetically prohibitive and results in sugar distortion leading to base-pair disruption.[10]

The interstrand C1′−C1′ distance in T•T pairs is ∼1.6 ± 0.4 Å larger than that in C•C$^+$ pairs. If this effect cannot be counterbalanced by backbone extension, this should entail a reduction of the rise at the C−T−C steps hindering intercalation. The open/closed T/T motif appears therefore as the best compromise with respect to the energetic balance related to backbone stretching, base stacking/unstacking, and base pairing. T3 opening allows intercalation of the sequentially adjacent C6•C6$^+$ and T7•T7 pairs. For this reason, the helical rise measured on the computed structure of [5mCCTCTCTCC]$_4$ at step C6−C7 is only 3.5 ± 0.15 Å. In addition the measured T3(C1′)−C4(C1′) intrastrand distance, 5.5 ± 0.2 Å, shows that T3 opening slackens the backbone stretch at step T3−C4. Considering that T3 opening releases the backbone stretch at the T3−C4 and T7−C6 steps, we suggest that the concurrent effect induced by T3 unstacking at both sides of the tetramer loosen the compactness of the central region and thus allows face-to-face intercalation of the central T5•T5 pair.

Two types of the symmetrical A•A pair could be incorporated in the i-motif. The observation that the thermodynamically stable structure formed by [5mCCTCACTCC] is an i-motif dimer with two A5 loops[3] indicates that T3 opening cannot compensate for the backbone distortion that should be induced by A•A intercalation in a tetramer. For the A•A pair with H6 *cis*-N1 bonds this may be due to the large C1′−C1′ interstrand distances of 13.5 Å.[12] However, the similarity of the C1′−C1′ interstrand distance of the A•A pairs with H6 *trans*-N7 and of T•T pairs suggests that it is more likely the size of the purine base that hinders adenine intercalation.

Alternated [C−T]$_n$ sequences are known to form nonelucidated pH dependent structures containing protonated cytidines whose half titration pH are close to that of the corresponding [C]$_n$ homopolymer.[11] It is conceivable that these sequences could form i-motif structures combining open/closed T/T motifs and stacked T•T pairs similar to those observed in the present study of [5mC(C−T)$_3$CC]$_4$.

Coordinate Deposition. The coordinates of [d(5mCCTCTCTCC)]$_4$ and the distance restraints used in the molecular dynamics have been submitted to the protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973, USA (RCSB code 101235 and PDB ID code 2kkk.).

Supporting Information Available: The 1d proton spectra and chromatograms of [5mCCTCTCTCC]$_4$ and [5mCCTCTCTCC]$_4$; distance restraints used to compute the tetramer structure; base-pair lifetimes in [5mCCTCXCTCC]$_4$ (X = T or C); table of geometrical parameters; table of ^{1}H and ^{31}P chemical shifts in [5mCCTCTCTCC]$_2$ and [5mCCTCTCTCC]$_2$. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Gehring, K.; Leroy, J.-L.; Guéron, M. *Nature* **1993**, *363*, 561–565.
(2) Nonin, S.; Leroy, J.-L. *J. Mol. Biol.* **1996**, *261*, 399–414.
(3) Leroy, J.-L. *J. Mol. Biol.* **2004**, *333*, 125–139.
(4) Canalia, M.; Leroy, J.-L. *Nucleic Acids Res.* **2005**, *33*, 5471–5481.
(5) Leroy, J.-L.; Guéron, M. *Structure* **1995**, *3*, 101–120.
(6) Leroy, J.-L.; Gehring, K.; Kettani, A.; Guéron, M. *Biochemistry* **1993**, *32*, 6019–6031.
(7) Leroy, J.-L.; Kochoyan, M.; Huyn-Dinh, T.; Guéron, M. *J. Mol. Biol.* **1988**, *200*, 223–238.
(8) Varnai, P.; Canalia, M.; Leroy, J.-L. *J. Am. Chem. Soc.* **2004**, *126*, 14659–14667.
(9) Brünger, A. T. *X-PLOR Version 3, a system for X-ray crystallography and NMR*; Yale University: New Haven, CT, 1990.
(10) Lebrun, A.; Lavery, R. *Nucleic Acids Res.* **1996**, *24*, 2260–2267.
(11) Jaishree, T. N.; Wang, A. H. *Nucleic Acids Res.* **1993**, *21*, 3839–3844.
(12) Leontis, N. B.; Jesse Stombaugh, J.; Westhof, E. *Nucleic Acids Res.* **2002**, *30*, 3497–3531.

JA903210T