Structure of a C-rich strand fragment of the human centromeric satellite III: a pH-dependent intercalation topology

Sylvie Nonin-Lecomte, Jean Louis Leroy

Structure of a C-rich Strand Fragment of the Human Centromeric Satellite III: A pH-dependent Intercalation Topology

Sylvie Nonin-Lecomte* and Jean Louis Leroy

PMC Groupe de Biophysique de l'École Polytechnique et de l'UMR 7643 du CNRS
91128, Palaiseau France

*Corresponding author: sn@pmc.polytechnique.fr

Abstract

Repetitive DNA sequences may adopt unusual pairing arrangements. At acid to neutral pH, cytidine-rich DNA oligodeoxynucleotides can form the i-motif structure in which two parallel-stranded duplexes with $C{\cdot}C^+$ pairs are intercalated head-to-tail. The i-motif may be formed by multimeric associations or by intra-molecular folding, depending on the number of cytidine tracts, the nucleotide sequences between them, and the experimental conditions.
We have found that a natural fragment of the human centromeric satellite III, d(CCATTCCATTCCTTTCC), can form two monomeric i-motif structures that differ in their intercalation topology and that are favored at pH values higher (the η-form) and lower (the λ-form) than 4.6. The change in intercalation may be related to adenine protonation in the loops.
We studied the uridine derivative methylated on the first cytidine base, d(5mCCATTCCAUTCCUTTCC), whose proton spectrum is better resolved. The intercalation topologies are (C7·C17)/(5mC1· C11)/ (C6·C16)/(C2· C12) for form λ and (5mC1·C11)/(C7· C17)/(C2· C12)/ (C6·C16) for form η. We have solved the structure of the η-form, and we present a model for the λ-form. The switch from η to λ involves disruption of the i-motif. In both forms, the central AUT linker crosses the wide groove, and the first and the third linkers loop across the minor grooves. The i-motif core is extended in the η-form by the inter-loop reverse Watson-Crick A3·U13 pair, whose dissociation constant is around 10^{-2} at 0° C, and in the λ-form by the interloop T5·T15 pair.
In contrast, d(5mCCATTCCTTACCTTTCC) folds into a pH-independent structure that has the same intercalation topology as the λ-form. The i-motif core is extended below by the interloop T5·T15 pair and closed on top by the T8·A10 pair.
Thus, the C-rich strand of the human satellite III tandem repeats, like the G-rich strand, can fold into various compact structures. The relevance of these features to centromeric function remains unknown.

Keywords: i-motif; centromere; satellite III; intercalation; NMR

Introduction

Non-coding DNA sequences account for 90 % of the human genome. Many of them are repetitive, and are biologically influent: telomeres are involved in the maintenance of chromosome integrity, some genetic pathologies such as mental retar-dation (fragile X syndrome) and dystrophy (Friedrich's ataxia, Hutchinson's disease …) are linked to the expansion of triple repeats (Ashley & Warren, 1995). In vitro, and potentially *in vivo*, these repeats may adopt unusual structures such as the i-motif or the G-tetrad for C-rich and G-rich strands, respectively.

The i-motif forms *in vitro* at acid to neutral pH values. It involves head-to-tail intercalation of two duplexes with parallel strands connected by hemi-

Abbreviations used: COSY, correlated spectroscopy; HMBC, heteronuclear multiple bond correlation; JR, jump and return; NOESY, nuclear Overhauser enhancement spectroscopy; TOCSY, total COSY.

protonated C·C$^+$ pairs (Gehring et al., 1993; Leroy et al., 1993). It may be formed by multimeric associations or by intra-molecular folding, depending on the number of cytidine tracts, the nucleotide sequences between them, and the experimental conditions such as pH, temperature, strand and salt concentrations (Gehring et al., 1993; Chen et al., 1994; Leroy et al., 1994; Leroy & Guéron, 1995; Berger et al., 1995; Kang et al., 1994, 1995; Mergny et al., 1995; Nonin & Leroy, 1996; Nonin et al., 1997; Gallego et al., 1997; Castati et al., 1997; Kanaori et al., 1998; Cai et al., 1998; Han et al., 1998; Weil et al., 1999).

Monomeric and dimeric i-motif structures may be biologically relevant as elements of DNA self-organization or of *cis* or *trans* interactions between C-rich tracts. NMR studies have shown that d(5mCCTTTCCTTTACCTTTCC) and d(5mCCTTTACC), which contain repetitions of tracts of two cytidine bases as found in centromeric human satellite III, form monomeric and dimeric i-motif structures, respectively (Han et al., 1998; Nonin et al., 1997). Similarly, a fragment of the C-rich strand of the human insulin minisatellite, d(CCCCTGTCCCC), forms a dimeric i-motif structure up to pH 7 (Castati et al., 1997). The intramolecular structure of a human telomeric fragment, d(CCCTAACCCTAACCCTAACCCT), has been solved recently (Phan et al., 2000).

The centromere is a heterochromatic region defined by cytologists as the primary constriction of the chromosomes. The centromeric DNA of most species contains repetitive blocks of sequences. In humans, the centromere encompasses about 10 % of the genome and includes mainly seven satellites, of which satellites α and III contain cytidine-rich tracts (Frommer et al., 1982; Willard, 1990), which are potential candidates for i-motif formation. Satellite α (about 5 % of the genome) is present in all chromosomes close to the primary constriction. It is composed of 171 base-pair long repeats. Within the repeated sequence, the 17 base-pair CENP-B box, d(TCCCGTTTCCAACGAAG) on the C-rich strand, is the binding site of the CENP-B protein (Masumoto et al., 1989). An 11-base fragment of this box, d(TCCCGTTTCCA), forms an i-motif dimer *in vitro* (Gallego et al., 1997, 1999).

Like satellite α, satellite III could be part of the functional centromere (Grady et al., 1992; Therkelsen et al., 1997). It includes the tandem repeats $(CCATT)_n$ $(AATGG)_n$. The G-rich strand can form several stem-loop structures *in vitro* (Jaishree & Wang, 1994; Castati et al., 1994; Gupta et al., 1994; Chou et al; 1996; Zhu et al., 1996).

We have investigated the structure of the C-rich strand, for which no structure has been yet reported. In satellite III, the ATT linker is some-times replaced by TTT (Prosser et al., 1986). Placing the TTT linker in the first or the third position of a 3.5 repeat (*i.e.* CCTTTCCATTCCATTCC or CCATTCCATTCCTTTCC) breaks the quasi-symmetry of the sequence, resulting in better spectral dispersion. The spectral properties of these two sequences are similar, and we chose to focus on the sequence d(5mCCATTCCATTCCTTTCC) henceforth designated as Mcent-TTT:L3 (a centromeric sequence modified (M) by methylation on the first cytidine base, whose third spacer is TTT. cf. Table 1). Cytidine methylation is commonly used to increase the spectral dispersion in the amino proton region and to provide a landmark for assignments thanks to the strong amino-methyl cross-peaks (Leroy & Guéron, 1995). The spectral dispersion is further increased in Mcent-U, a version of Mcent-TTT:L3 modified by T to U substitutions at positions 9 and 13 (Table 1). The melting temperatures of Mcent-TTT:L3 and Mcent-U are identical. Their properties, in particular their pH-dependent behavior and the inter and intra-residue NOESY cross-peaks, are very similar showing that both oligonucleotides adopt the same structures. Together with the higher spectral resolution, this makes Mcent-U an excellent substitute for the structural study of d(CCATTCCATT CCTTTCC).

We show that Mcent-U can fold into i-motif structures. Depending on pH, it forms two intra-molecular structures with different intercalation topologies, *i.e.* with external C·C$^+$ pairs formed either by cytidine bases at the 3' end of the C-stretches (3'E intercalation; Phan et al., 2000), or at the 5' end (5'E intercalation). This structural modulation is absent in d(5mCCATTCCT-TACCTTTCC), which differs from d(5mCCATTC-CATTCCTTTCC) by a T/A inversion in the second spacer.

Results

pH-dependent structure of Mcent-U

The 1D spectrum of Mcent-U presents imino and amino proton resonances (in the ranges of 16-15 ppm, and 10.5-8 ppm, respectively) characteristic of hemiprotonated C·C$^+$ base-pairs (Figure 1). Two structures coexist, in a pH-dependent ratio. According to gel-filtration chromatograms and to the invariance of the proton spectrum upon sample dilution from 3.3 to 0.1 mM (not shown), they are monomeric at the concentration of 1 mM used in the NMR experiments. Both species are in comparable proportions at pH 4.6. They are henceforth designated η and λ for the principal structures above and below pH 4.6, respectively (η as high pH and λ as low pH). The η-form is stable up to pH 7 at -4°C and pH 6.5 at 15°C.

Table 1. Oligonucleotide sequences

Name	Sequence
Mcent-TTT:L3	5mCCATTCCATTCCTTTCC
Mcent-U	5mCCATTCCAUTCCUTTCC
Mcent-L2	5mCCATTCCTTACCTTTCC

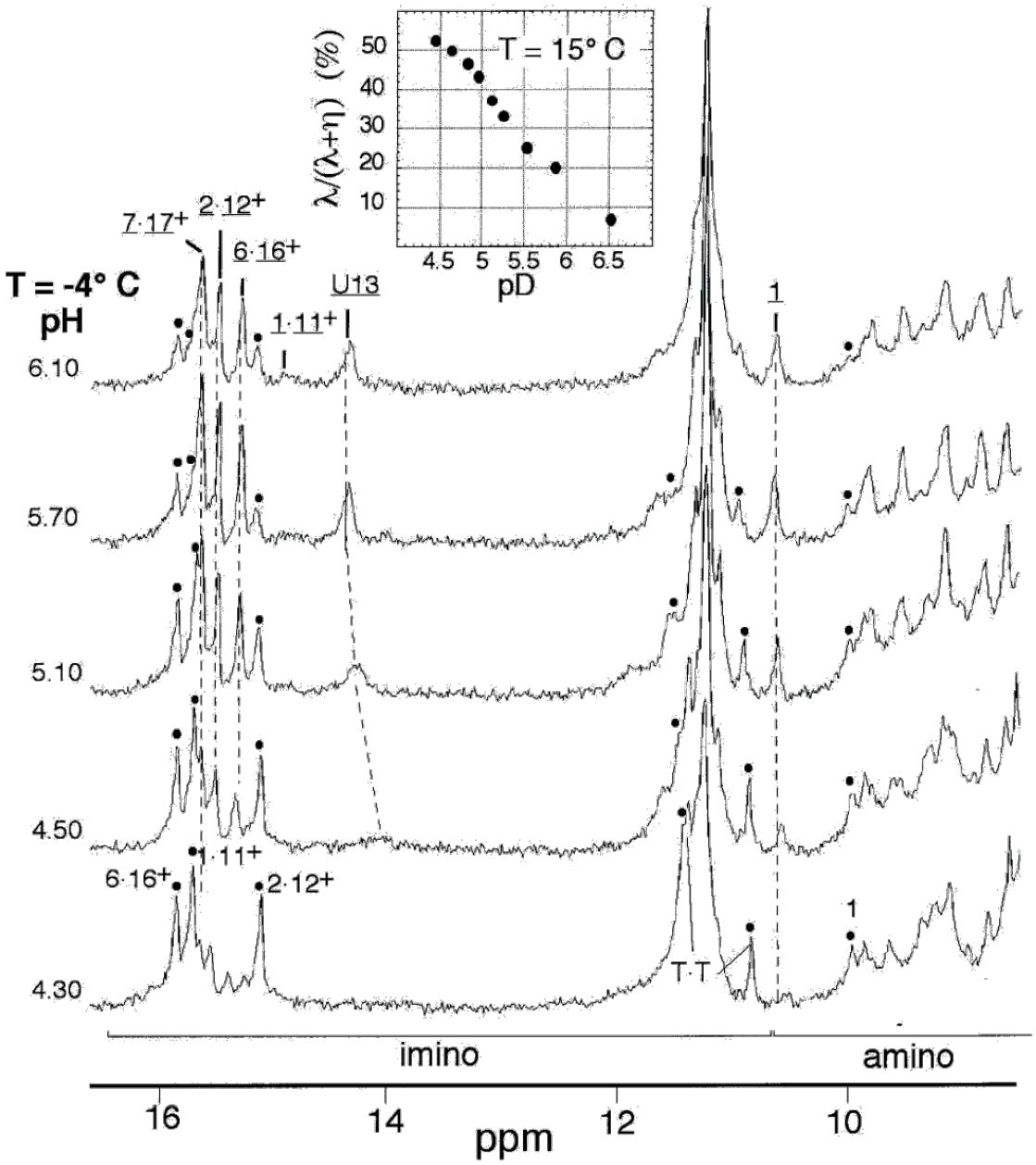

Figure 1. The 1D exchangeable proton spectrum of the Mcent-U oligomer recorded versus pH at -4°C, in 95 % H_2O, 5 % ^2H_2O, 1 mM EDTA solution. The spectra disclose two i-motif species in slow equilibrium on the NMR timescale, whose proportions are pH-dependent. The resonances of form λ, the major i-motif species observed at pH 4.3, are labeled by black circles. The four C·C$^+$ imino protons of form η the major species above pH 5, are resolved. The *cis* 5mC1 amino proton (1) at about 10.6 ppm is a good marker of structure η, about 0.7 ppm downfield from its position in form λ. Form η is associated with the imino resonance at 14.3 ppm of an A U base-pair, and form λ with two imino protons of a T·T pair. Inset: Proportion of form λ versus pD at 15°C derived from volume measurements of cytidine H5-H6 TOCSY cross-peaks (30 ms mixing time, ^2H_2O solution). Forms λ and η are in equal proportions at pD 4.6.

The NOESY spectra provide evidence for two i-motif species. A striking difference is that form η has an A·T/U base-pair, and that form λ has a T·T pair, as will be shown below (Figures 2 and 3).

Titration of the cytidine and uridine H5-H6 TOCSY cross-peaks (Figure 1, inset) shows that the proportion of the η-species is higher than 90 % above pD 6.2

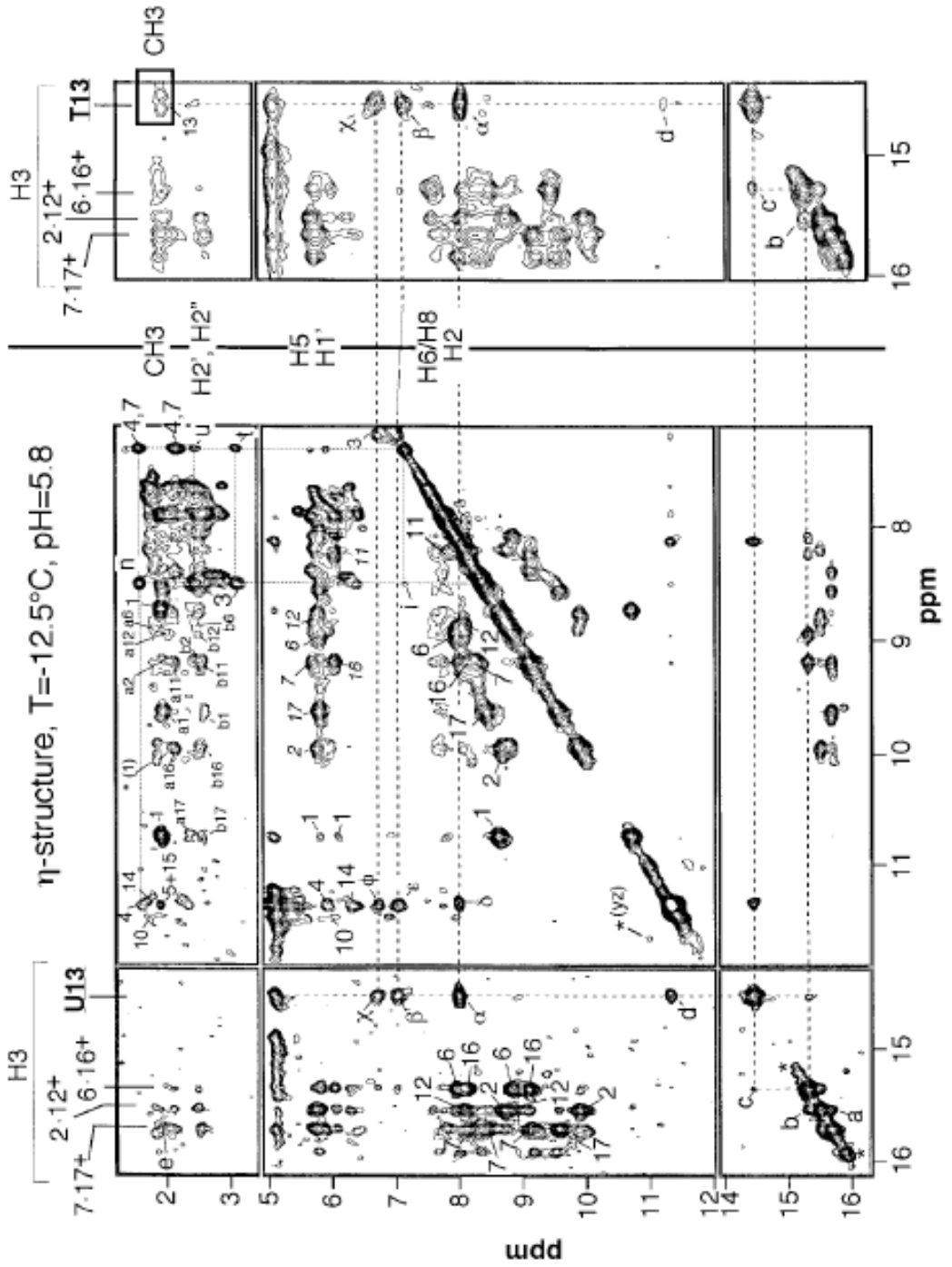

Figure 2. The 250 ms NOESY spectra recorded at -12.5°C of the η-forms of Mcent-U (left panel) and Mcent-TTT:L3 (right panel). The intra-residue cross-peaks are labeled by the residue number (in italics for the H5-H4*cis* cross-peaks), the inter-residue cross-peaks by letters. Left: The cross-peaks labeled by a star arise from form λ. The strong 5mC1(CH_3)-(H4*cis*,H4*trans*) and the cytidine H5-H4*cis* cross-peaks have been used as starting points for the assignment. The inter-residue amino H2' and amino H2'' are labeled a and b, respectively, followed by the residue number of the H2' and H2'' protons. Pair C6·C16$^+$ connects to U13(H3) at about 14.5 ppm (cross-peak c), which connects in turn to the imino proton of T4 (cross-peak d). The U13(H3) chemical shift and its cross-peaks with A3(H2) (α), and A3(H6*cis*, H6*trans*) (β and χ, respectively) show that U13 and A3 form a pair. Cross-peaks n, u, and t between A3(H8)-T4(CH_3), T4(H6)-A3(H2') and T4(H3)-A3(H2''), respectively, on one hand, and cross-peaks δ and (ε, ϕ) between T4(H3)-A3(H2) and T4(H3)-A3(H6*cis*,H6*trans*), set the relative orientation of A3 and T4. Right: Detailed region of the imino proton resonances in the McentTTT:L3 spectrum, labeled as in left panel. Solution conditions: pH 5.8 in 95 % H_2O-5 % ^2H_2O, 1 mM EDTA, 9 % (v/v) $C^2H_3O^2H$.

Below pH 4.3, the apparition of new TOCSY H5-H6 cross-peaks, most probably from the fully C-protonated unstructured strand, is an obstacle to the structural NMR study of form λ. Higher salt concentrations broaden the spectrum and do not favor one species. Adding an A or an AT dinucleotide at the 3' end of the sequence also broadens the spectra (data not shown).

We have determined the solution structure of the η-form of Mcent-U and established a model of its λ-form.

Mcent-U in the η-form

Spectral assignment

The procedure was caried out as before (Phan et al., 2000) using the following experiments: TOCSY, NOESY, and natural abundance HMBC experiments. The 17 expected spin systems were identified and connected together sequentially by a 50 ms hetero-TOCSY experiment. All cytidine bases display strong H6-H3' cross-peaks, indicating anti glycosidic angle and N-type sugar pucker. A striking feature of the spectrum is the up-field chemical shift of almost all the protons of C7 compared to those of the other cytidine bases (Figure 2).

The cytidine imino protons are assigned as usual by the NOESY intra-residue cross-peaks to the amino protons, themselves assigned by connection with H5. These cross-peaks show that residues 7-17, 2-12 and 6-16 are base-paired. The amino protons of 5mC1, and the imino protons T4 and T14 are NOESY-connected to their respective methyl group (Figure 2). The 1D spectra display a broad peak around 14 ppm (Figure 1). This peak was assigned to U13(H3) by comparison with the spectrum of Mcent-TTT:L3, where the corresponding imino proton resonance is connected with a methyl group (Figure 2, right panel, cross-peak 13). The chemical shift of U13(H3) and its strong NOE-connection with A3(H2) (cross-peak a, Figure 2) indicate H-bonding to A3(N1). Two exchangeable protons, which are connected together (cross-peak 3, Figure 2), to U13(H3) (cross-peaks b and w), and to T4(H3) (cross-peaks e and f) are assigned to the amino protons of A3.

i-Motif intercalation

According to Lavery's definition (Lavery et al., 1992), the ``black'' and the ``white'' faces of an *anti* nucleotide are oriented in the 5' and 3' directions, respectively. The i-motif core involves two alternating steps: steps with adjacent cytidine bases facing each other by their white faces and by their black faces, respectively (stressed by a bold black line in Figure 4(c)). The Mcent-U η-form displays the characteristic inter-residue NOESY cross-peaks of the i-motif (schematized in Figure 4(c)), re¯ecting the short amino- H2'/2'' (Leroy & Guéron, 1995) and H1'-H2'/2'' (Castati et al., 1997) distances across the wide and narrow grooves, respectively, at white steps, and the short H1'-H1' distances (Gehring et al., 1993) across the narrow grooves at both white and black steps. These two categories of inter-residue connectivities provide a redundant description of the intercalation topology, which is complemented by imino-imino proton cross-peaks. The patterns of amino-H2' cross-peaks (between residues 1 and 17, 11 and 7, 2 and 16, and 12 and 6; Figure 2), of imino-imino cross-peaks between pairs C7·C17$^+$ and C2·C12$^+$, and pairs C2·C12$^+$ and C6·C16$^+$ (Figure 2), of H1'-H2'' cross-peaks (between residues 11 and 17, 1 and 7, 6 and 2, and 16 and 12; schematized in red in Figure 4(c)), and of H1'-H1' cross-peaks (between residues 1 and 7, 7 and 2, 2 and 6, 11 and 17, 17 and 12, and 12 and 16; in red, Figure 4(c)), define the following intercalation: (5mC1·C11$^+$)/(C7·C17$^+$)/ (C2·C12$^+$)/(C6·C16$^+$) (Figure 4(a)). The connectivities used for structure determination are schematized in Figure 4(c).

Connectivities involving loop residues

Three loops enclose the i-motif core. The first (ATT) and third (UTT) loops are designated loops L1 and L3, and the central AUT loop L2. The connectivities between the intercalated cytidines show that stretches 5mC1-C2 and C6-C7 on one hand and stretches C11-C12 and C16-C17 on the other, are aligned in anti-parallel orientations along the sides of the narrow grooves. Hence loops L1 and L3 cross the narrow grooves, and loop L2 a wide groove. The connectivities between the loop residues corresponding to distances shorter than 4.7 Å

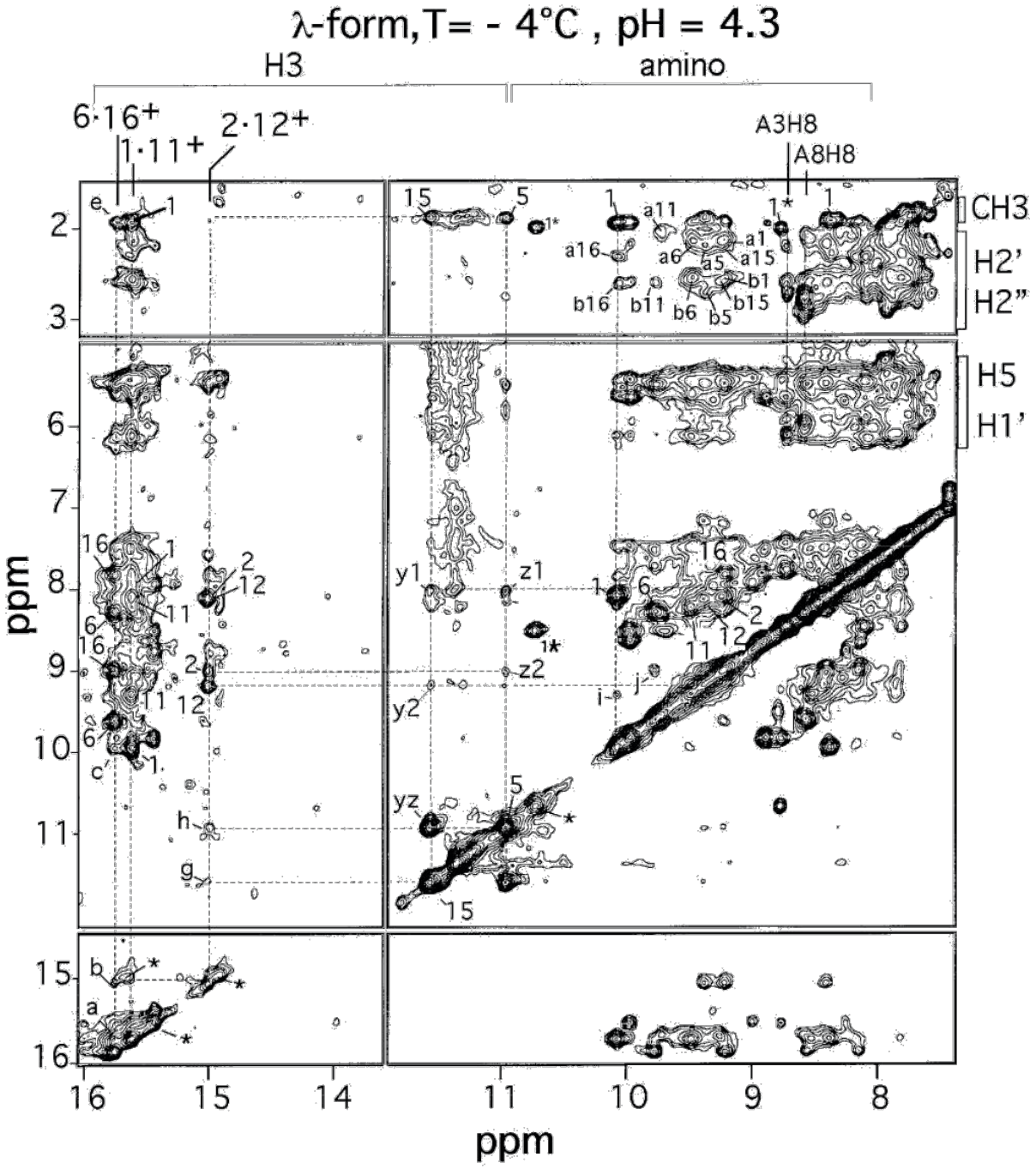

Figure 3. The 250 ms 2D NOESY spectrum of the λ-form of Mcent-U. Labeling is the same as for Figure 2. The strong intra-residue 5mC1(CH_3)-(H4cis,H4trans) cross-peaks have been used as starting points for the assignment of the exchangeable protons of pair 5mC1·C11$^+$. The imino protons of 5mC1·C11$^+$ and C6·C16$^+$ are connected by cross-peak a, and those of C6·C16$^+$ and C2·C12$^+$ by cross-peak b. The imino protons of pair T5·T15 are connected to each other by the strong yz cross-peak, and to C2·C12$^+$ by cross-peaks g, h, y1, z1, y2 and z2 for T15(H3)-C2 C12 (H3), T5(H3)-C2 C12 (H3), T15(H3)-C2(H4*trans*), T5(H3)-C2(H4*trans*), T15(H3)-C2(H4*cis*) and T5(H3)-C2(H4*cis*), respectively. Conditions: pH 4.3 and -4°C in 95 % H_2O-5 % ^2H_2O.

are displayed in Figure 4(c). Typical i-motif NOESY cross-peaks connect the i-motif core to the residues of loops L1 and L3, showing that the A3·U13 pair extends the i-motif core at its lower end, under C6·C16$^+$. The A3(H8) resonance is broader than that of A8. Its line width decreases from -12.5°C to 25°C, while A3(H2) remains narrow in this range. The overall NOESY cross-peak pattern in ^2H_2O is identical at -9°C and 15°C (data not shown). The orientation of T4 is given by the cross-peaks T4(H3)-A3(H6*cis*)/ (H6*trans*) (cross-peaks ε and ϕ, Figure 2) and T4(H3)-U13(H3) (cross-peak d, Figure 2). In loop L2, the H8-H1' pathway of connectivities can be

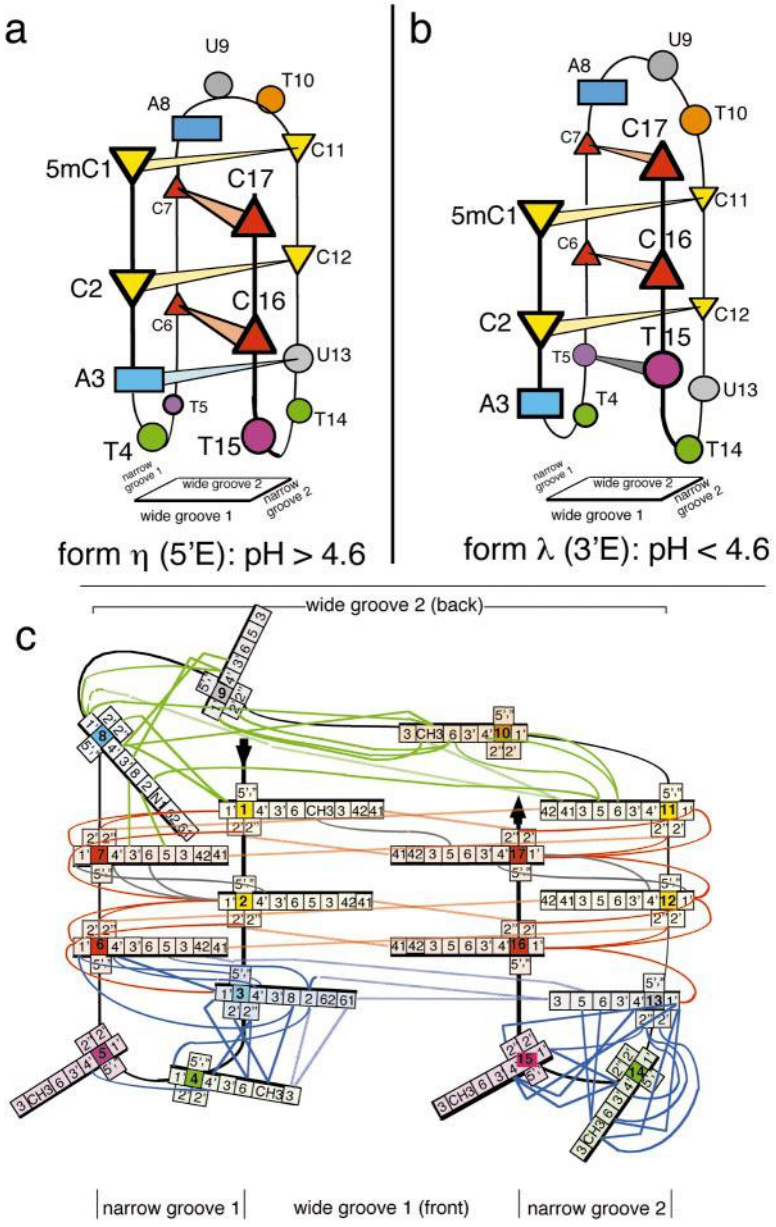

Figure 4. Mcent-U schematic i-motif structures. The C·C$^+$ base-pairs of the i-motif core are in red and yellow, with the cytidine bases symbolized by triangles, adenine by rectangles, and thymidine and uridine by circles. (a) The η-form has the 3'E topology. The inter-loop A3·U13 base-pair extends the i-motif core. (b) The λ-form has the 5'E topology. The inter-loop T5·T15 base-pair (in pink) extends the i-motif core. (c) A representation of the inter-residue distance restraints used for computation of the η-form of Mcent-U. The residues are labeled by their sequential number in bold characters on color-filled squares, with the same color code as above. The hydrogen numbers are those of the chemical formula. Distances involving exchangeable and non-exchangeable protons are symbolized by dotted and continuous lines, respectively: red for characteristic short distances of the i-motif; green and blue for distances involving loop residues; gray for the others.

followed from A8 to T10 (not shown). A8(H2) displays cross-peaks with C11(H5) (very strong), and with C7(H5) (strong) (cf. Figure 4(c)). The absence of i-motif-type connectivities between pair 5mC1 C11 and A8, U9 or T10, shows that the i-motif core is not extended on top in loop L2. The intra-residue U9(H6)-(H1') cross-peak appears stronger in the ^2H_2O NOESY spectrum than the other intra-residue H6-H1' peaks because it over-laps with a cytidine H5-H6 from the λ-form (not shown).

pH-dependent chemical shifts

The U13 imino proton shifts up-field and exchanges faster as pH decreases from 5.2 to 4.3 (Figures 1 and 5): 3 ms at pH 4.7 versus 18 ms at pH 6 and 0°C. In order to assess a potential adenine protonation in the η-form, we have examined the chemical shift variations ϖερσυσ pH of the protons of A3 and A8, and of two neighboring residues, T4 and C7. The pH-induced chemical shift variations of these protons are among the largest observed for the η-form. The aromatic protons A8(H2), A3(H2), C7(H6) and T4(H6) shift down-field by about 0.1 ppm between pH 5.8 and 4.3 (Figure 6). For comparison, the H8 and H2 protons of the dA monomer shift down-field by about 0.1 ppm between pH 6 and 3, with a mid-titration at pH pK_a(N1), *i.e.* 3.8 (not shown).

Imino proton exchange and i-motif lifetime

We have measured the imino proton exchange times by magnetization transfer from water versus temperature, pH (Figure 5) and catalyst concen-

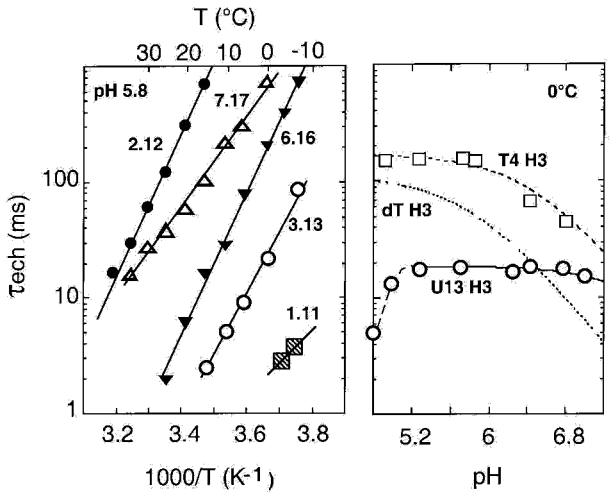

Figure 5. Imino proton exchange times of the Mcent-U η-form. Left panel: C·C$^+$ base-pair lifetimes and U13(H3) exchange time ωερσυσ 1000/T. Right panel: The imino proton exchange times of U13, T4, and dT versus pH. U13(H3) displays the pH-independent plateau due to intrinsic catalysis, and its exchange time decreases below pH 5.2. T4(H3) is slightly protected from hydroxide catalysis in comparison with the mononucleoside.

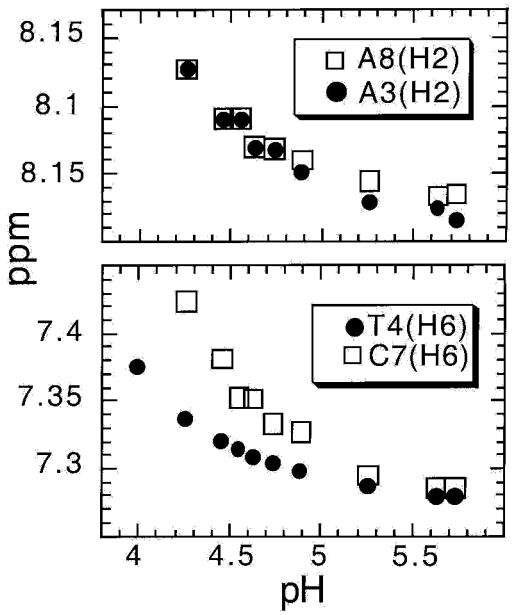

Figure 6. Chemical shift titration in the η-form ωερσυσ pH at 15 C from 1D spectra recorded in ^2H_2O. Open squares and filled circles are for upper and lower loop residues, respectively. Upper panel: A8(H2) and A3(H2). Lower panel: C7(H6) and T4(H6).

tration (not shown). Due to intrinsic catalysis, the imino proton exchange time in a C C pair is equal to the base-pair lifetime (Leroy et al., 1993). The times are 720, 11, 16.5 and less than 1 ms at 15°C for pairs C2·C12$^+$, C7·C17$^+$, C6·C16$^+$ and 5mC1·C11$^+$ respectively (Figure 5, left panel), with activation energies of 118, 74, 113 and 57 kJ/ mol.

Between pH 5.2 and 6.8, the A3·U13(H3) exchange displays the pH-independent plateau of intrinsic catalysis characteristic of base-paired imino protons (Figure 5, right panel). For pH <6.5, the U13(H3) exchange time is shorter than that of the uridine monomer (not shown), which is typical of a poorly stable base-pair (Nonin et al., 1995). Its line width at pH 4.5 (Figure 1) shows that the lifetime of A3·U13 is shorter than 1 ms. The dis-sociation constant of the base-pair, 9.5 $10^{ÿ3}$ at 0°C, was determined by comparison of the catalyzed exchange at pH 6.9 of dU and U13 by $HPO^{2}{}_{4}^{-}$. The hydroxide catalysis of T4(H3) is slowed by a factor of 6 by comparison with that of dT(H3) (Figure 5, right).

The lifetime of the η-form was derived from the build-up of exchange cross-peaks between A8(H8), A8(H2), C7(H5) and C17(H6) in the folded structure and in the melted strand (Macura et al., 1994), measured for mixing times ranging from 150 ms to 1 s. It is about 11 s at 25°C, pH 5.8.

Computed η-form

The structure was computed with the simulated annealing protocol described in Materials and Methods. T5, which exhibits only one inter-residue NOE, was positioned mainly by seven repulsive restraints. The i-motif has two wide and two narrow grooves. The cytidine bases are *anti* and their sugar pucker is N-type. The i-motif core is extended below the C6·C16$^+$ end by the inter-loop reverse Watson-Crick pair A3·U13 (Figure 7). A3 and T4 are stacked within loop L1. The strand turns in loop L1 between T4 and T5, and the ring of T5 is oriented outwards. The turn in loop L3 is more progressive. The N3H vectors of T14, T15, and T4 point inward in contrast to T5. On top, the three residues of loop L2 are unstacked. The base of A8 is turned towards wide groove 2 between pairs 5mC1·C11$^+$ and C7·C17$^+$, with H2 pointing towards C11(H4*trans*). The U9 base points out-wards, while T10 is partly stacked on A8. The glycosidic angle of U9 is not well defined, reflecting the fact that the intramolecular U9(H6)-(H1') cross-peak is not resolved. The violations and deviations from ideal geometry of the ten best computed structures are given in Table 2.

Mcent-U in the λ-form

Intercalation topology

The NOESY and TOCSY spectra of the Mcent-U λ-form are hard to evaluate because of the contributions of the η-form and of the unstructured protonated single strand that appears at low pH. The 250 ms NOESY spectrum in H_2O, pH 4.3 (Figure 3) displays characteristic features of an i-motif structure

Table 2. Violations and deviations from ideal geometry in structural computation of the Mcent-U η-structure

Number of NOE violations larger than 0.2 Å	3
Largest NOE violation (Å)	0.24
RMSD of input distance restraints (Å)	0.05
Number of violations on dihedral angles ε larger than 1°	2
Largest dihedral angle ε violation (deg.)	7
RMSD from ideal ε (deg.)	1.6
Number of angle violations greater than 6°	2
Largest angle violation (deg.)	6.1
RMSD from ideal bond angle (deg.)	1.3
Number of improper violations larger than 5°	0
RMSD from ideal improper (deg.)	0.34
Bond violation larger than 0.05 Å	0
RMSD from ideal bond (Å)	0.005

(C·C$^+$ imino proton resonances between 15 and 16 ppm, and amino-H2'/H2'' cross-peaks).

Three C·C$^+$ imino protons resonances are observed at -4°C (Figure 1). At -10°C, a broad fourth resonance is detected at 14.9 ppm. It is ascribed to the fourth C·C imino proton. The 5mC1·C11$^+$ imino proton is assigned as in the η-form by the CH_3-amino, amino-imino and CH_3-imino cross-peaks. It is down-field shifted by about 1 ppm as compared to its position in the η-form, and its exchange time (about 100 ms at 15 C) is much longer than that usually found for external C·C$^+$ pairs (*cf*. Figure 5, left panel for comparison with the η-form). These observations suggest that pair 5mC1 C11 is internal to the i-motif core, and therefore, that the intercalation topology is 3'E and the stacking order as follows: (C7·C17)/(5mC1·C11)/(C6·C16)/(C2·C12). The C6·C16$^+$ and C2·C12$^+$ imino protons are then identified by imino-imino connectivities (cross-peaks a and b, Figure 3).

Two resolved thymidine imino proton resonances connected by cross-peak yz, disclose a T·T pair. They are also connected to the C2·C12$^+$ imino proton at 15 ppm (cross-peaks g, h, y2, z2, z1 and z2, Figure 3) indicating that the T·T pair closes the i-motif core.

Comparison with the Mcent-L2 structure

To help for the structural determination of the λ-form, we studied d(5mCCATTCCTTACCTTTCC) (named Mcent-L2), which differs from Mcent-TTT:L3 by substitution of TTA for ATT in loop L2. The NOESY spectra of these sequences, which are almost superimposable, indicate closely related structures. Mcent-L2 forms a single pH-independent monomeric i-motif with narrow and well-resolved NMR lines. Its spectrum, which greatly resembles that of d(5mCCTTTCCTTTACCTTTCC) (Han et al., 1998) was easily assigned by the procedures described above for the η-form of Mcent-U. Its assignments were carried over to the very similar spectrum of λ -form of Mcent-U.

The Mcent-L2 intercalation topology is 3'E. The central i-motif is sandwiched by two pairs; T8·A10 in the central loop atop C7·C17$^+$ and T5·T15 below C2·C12$^+$. The TTA spacer crosses the wide groove, and the ATT and the TTT spacers cross the two narrow grooves. Each cross-peak in the Mcent-L2 spectrum has its counterpart in the spectrum of the λ-form of Mcent-U, confirming the intercalation topology (C7·C17$^+$)/(5mC1·C11$^+$)/ (C6·C16$^+$)/(C2·C12$^+$)/(T5·T15) for the λ-form. The loop topologies across the grooves are identical in the two structures.

Discussion

We have studied the solution behavior of Mcent-U, d(5mCCATTCCAUTCCUTTCC), a modified version of the cytidine-rich sequence d(CCATTC-CATTCCTTTCC) found in the human centromeric satellite III, whose substitutions are designed so as to improve the spectral resolution and do not affect the structure. We have found two structures depending on pH, the η-form at ``high'' pH and the λ-form at ``low'' pH. The i-motif core is extended below by an interloop A·U pair in the η-form, and a T·T pair in the λ-form. To our knowledge, such intercalation changes with pH have not been reported elsewhere.

The i-motif core in the Mcent-U η-form

The Mcent-U η-form has the 5'E intercalation topology like the telomeric fragment (Phan et al., 2000Th e geometry of the i-motif core of form η is comparable to that of other i-motif structures studied so far. Like the intercalated cytidine bases in the i-motif core, C6-A3 and C16-U13 are connected by H1'-H1' cross-peaks across narrow grooves. The A3 U13 pair may be thus considered as part of the i-motif core. It has a buckle (-19(±6)°) slightly higher than that of the C·C$^+$ pairs, but comparable propeller twist (-149(10)°) and stacking interval (3.3(± 0.4) Å). The negative helical twist between C2 and A3 shows that the helix is distorted locally

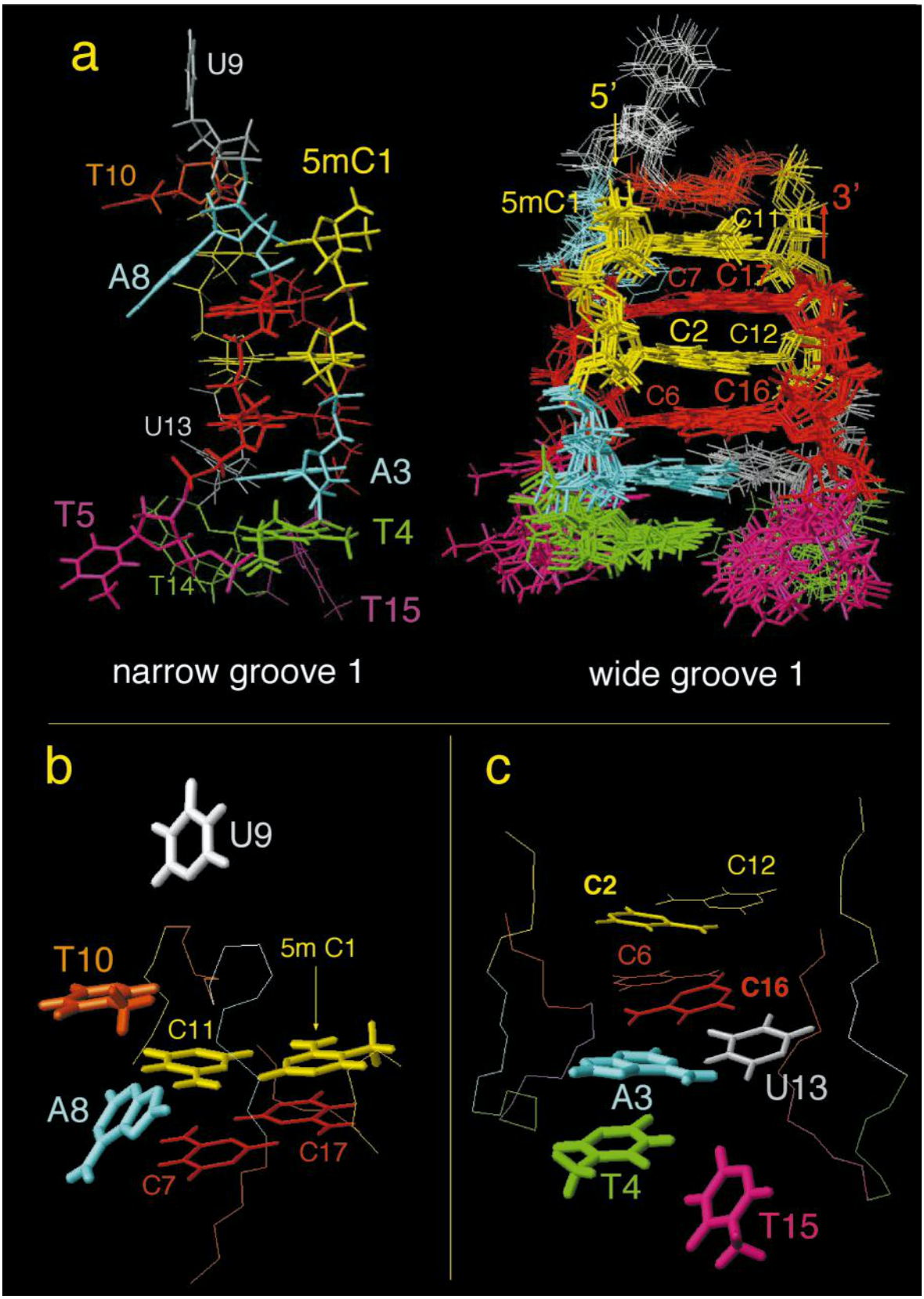

Figure 7. Structure of the Mcent-U η -form with the same color coding as for Figure 5. (a) Right: Superposition of the eight conformers of lower energy. View from wide groove 1. The central i-motif core (in red and yellow) is extended on the C6·C16$^+$ side by the A3·U13 base-pair. Left: View from narrow groove 1 of the best η-structure.
(b) Close up of loop L2 and neighboring base-pairs. A hydrogen bond may exist between A8(N3) and the non-bonded C11 amino proton. The imino protons of U9 and T10 point outwards. (c) Partial close-up view of loops L1 and L3, and neighboring base-pairs. T14 and T5 have been omitted for clarity. The A3·U13 pairing is reverse Watson-Crick. The lower loops can accommodate a T4·T15 base-pair.

to accommodate the A3·U13 pair below the i-motif core. The average P-P distances across the narrow and the wide grooves are in the range of those usually observed. The interstrand A3-U13 C1'-C1' distance is about 2 Å larger than the corresponding distances within the i-motif core, as required for accommodation of the purine.

The lifetimes of the terminal C·C$^+$ pairs fully exposed to the solvent are generally shorter than 1 ms at 0 C. The C6·C16$^+$ lifetime, 16.5 ms at 15°C, thus reflects protection by pair A3·U13. Similar effects have been observed in the folded i-motif of d(5mCCTTTCCTTTACCTTTCC), whose external C·C pairs, protected by an interloop T·T pair at one end and by an intra-loop T·A at the other, have lifetimes in the range of 3 ms at 15°C (Han et al., 1998). In the Mcent-U η-form, the stabilization brought by pair A3·U13 spreads to C2·C12$^+$, whose lifetime (720 ms at 15 C) is much longer than that of C7·C17$^+$ (111 ms).

The fact that the η-form lifetime (11 seconds at 25°C) is much longer than the C·C$^+$ pair lifetimes, shows that the base-pair opening motions are not limited by the dissociation of the structure.

Loop foldings in the Mcent-U η -form

The central AUT loop spans wide groove 2 and does not disclose any intra-loop base-pair as predicted by its sequence directionality (Hilbers et al., 1994). A8 and T10 residues are *anti*; A8, U9 and T10 sugar puckers are C4'-*exo*. As shown in Figure 7(b), the N3H vectors of U9 and T10 point outwards. This leaves one face of pair C1·C11$^+$ fully exposed to the solvent. A8 is turned towards major groove 2, between C7 and C11. The up-field chemical shifts of the C7 protons are consistent with the ring currents generated by A8 according to Giessner-Prettre's shielding tables (Giessner-Prettre et al., 1976). The relative orientation of A8 and C11 bases is supported by the up-field chemical shift of the C11 cis amino proton, which senses the adenine ring currents. The averaged A8(N3)-C11(H4*trans*) distance, 2.0 Å, and (A8(N3)-C11(H4*trans*);C11(H4*trans*)-C11(N4)) angle, 8.8°, allow H-bonding of C11(H4*trans*) to A8(N3). The expected cross-peak between A8(H8) and C11(H4*trans*) is not observed because these resonances are not resolved. Loops L1 and L3 extend the i-motif core by the inter-loop reverse Watson-Crick A3·U13 pair. T4 also contributes to stabilization of the structure by stacking interactions with A3. The fact that A3(H8) is broad at -14°C and becomes narrower as temperature is increased, suggests a motion despite the fact that A3 is base-paired. Although they are not coplanar, T4 and T15 rings are in proper positions to form a base-pair, which could account for the fact that T4(H3) exchange is slowed by a factor of about 6 by com-parison with the thymidine monomer (Figure 5). The similarity of chemical shifts between T4 and T15 imino protons prevents observation of a potential cross-peak between them. A simulated computation enforcing T4·T15 pairing yielded a set of structures with energies associated with NOE violations comparable to those of the best structures computed without this restraint.

Effect of pH on the η form

Below pH 5.2, U13(H3) is up-field shifted and its exchange time decreases sharply (Figures 1 and 5). Similar effects are observed for terminal A·T pairs in B-DNA duplexes (unpublished results). The imino proton, which appears as a single resonance, exchanges rapidly between its positions in the close pair (about 14 ppm) and in the open state (about 11 ppm). Thus, the effects observed below pH 5.2 reflect the increasing proportion of the open pair.

The pH-induced increase of the A3·U13 dissociation constant involves protonation of the structure. The protonation sites with the nearest pK are the N1 of the two adenosine bases (pK_a(N1) 3.8). The pH-dependence of A3(H2), which is comparable to that of dA(H2), suggests protonation of A3(N1). This protonation destabilizes the base-pair by preventing its closing. It also affects the neighboring T4, which is stacked to A3 above pH 5.2 (Figure 7). The T4(H6) down-field shift below pH 5.2 (Figure 6) can be explained either by the proximity of the protonated A3, or by partial destacking from A3$^+$. C7, which is involved in a hemi-protonated pair cannot protonate around pH 4. The pH-dependent chemical shift of its H6 proton observed below pH 5.2 (Figure 6) probably reflects structural changes induced by partial protonation of A8(N1).

Mcent-U at low pH (λ-form)

The λ-form has the same 3'E intercalation topology and loop folding as the i-motif structures of Mcent-L2 and of the d(5mCCTTTCCTTT ACCTTTCC) sequence (Han et al., 1998). Lower loops L1 and L3 bridge the minor grooves, and upper loop L2 the major groove. Furthermore, the i-motif core is extended in the three cases by an inter-loop T·T pair. But, in contrast with the two other structures, the central adenine base of Mcent-U in loop L2 is not base-paired and the C C pair at the top of the stack is exposed. This can account for the low melting temperature: Mcent-U and Mcent-TTT:L3 λ–structures melt around 28°C at pH 4.5. For comparison, the T_m of d(5mCCTTTCCTTTACCTTTCC) is around 45°C at pH 4.2 (Han et al., 1998).

The pH-dependent intercalation topology

The pH-dependence of the equilibrium between the η and the λ-forms indicates a difference in protonation. The mid-titration point of the η / λ equilibrium is about pH 4.6. No intermediate structure (*i.e.* with the 5'E intercalation topology and a T·T pair, or with the 3'E intercalation topology and an

A·T/U pair) is observed. The lack of exchange cross-peaks shows that the interconversion between the two forms is slow. It probably requires unfolding of the structures.

All cytidine bases are hemi-protonated in the two Mcent-U forms. Based on the pH-dependent changes observed on U13(H3) below pH 5.2, we propose that protonation of A3(N1) disrupts pairing with U13. At low pH, the formation of a T·T pair in the lower loops, as observed in form λ, would leave a gap of intercalation between that pair and the bottom C6·C16$^+$ pair if the 5'E topology of the η-structure were maintained. At high pH, the formation of an A·U pair in the 3'E topology is probably not favored for the same reason. Such arguments may explain the shift from the η-form at pH 6 to the λ-form when A3 is protonated. The preference for the 5'E topology at high pH values suggests that the stabilization by an inter-loop A·U pair is higher than by an inter-loop T·T.

The 3'E intercalation topology of Mcent-L2 allows the simultaneous formation of the T8·A10 pair across loop L2 and of the T5·T15 pair between loops L1 and L3. It is pH-independent despite the possibility of the A3 and A10 protonations as proposed for Mcent-U. This is probably because the i-motif core is stabilized from the other end by a pH-independent T·T pair. The pH-independent behavior of d(5mCCTTTCCTTTACCTTTCC) (Han et al., 1998) may have the same origin.

Intercalation topology of the i-motif core: comparison with other i-motif structures

The Mcent-U and McentTTT:L3 η-forms have the same topology (5'E) as the monomeric telomeric i-motif (Phan et al., 2000); however, the loop folding is different because in the later structure, loops 1 and 3 span the wide grooves and loop 2 one narrow groove. The Mcent-U and McentTTT:L3 λ-forms have the same intercalation and folding topologies as the i-motif formed by d(5mCCTTTCCTTTACCTTTCC) (Han et al., 1998).

In solution or in crystals, some short C-rich oligonucleotides form a mixture of tetrameric structures, the various intercalation topologies of which sometimes do not maximize the number of intercalated pairs (Leroy & Guéron, 1995; Cai et al., 1998; Kanaori et al., 1998). The case of the monomeric i-motifs studied so far is different. They adopt either the 3'E or the 5'E, but in each case, the intercalation of the C·C$^+$ pairs is maximal. This is therefore the extra intercalated base-pair, and thus the sequence of the linkers, which determine the intercalation.

Biological Implications

The centromere is a prominent heterochomatic region, defined by cytologists as the primary constriction of the chromosomes and by geneticists as the point from which the recombination distances are measured. It is the site of attachment of the microtubules during the cell division. It mediates several mitotic and meiotic functions (spindle attachment, kinetochore nucleation and sister chromatid cohesion). The centromeric region is associated with many proteins (for a recent review, see Dobbie et al., 1999). The centromere function is essential, and its loss results in chromosomic instability, wrong chromosome partitioning during cell division, and chromosome loss in subsequent cell divisions.

By contrast to the telomeres (the ends of the chromosomes), which display remarkable conservation between species, mammalian centromeric DNAs share very little homology with one another. This raises the question of what comprises the centromere. The sequence requirements are not clear. In human, the centromere encompasses about 10 % of the genome and contains several tandemly repeated sequence families called satellites I to IV, α, β and γ. Which of these satellites are essential to the centromeric function is not known. Flowing from the idea that the functional centromere must be present in all 23 pairs of chromosomes, the most important candidate is satellite α, which is present on all chromosomes close to the primary constriction. It is composed of 171 base-long repeats including the 17 base-pair CENP-B box (CTTCGTTGGAAACGGGA) whose C-rich counterpart forms an i-motif structure in vitro (Gallego et al., 1997). The X-ray structure of the CENP-B/CENP-B box complex has been published recently (Iwahara et al., 1998). It shows that the CENP-B protein binds to the CENP-B box in a region that is not rich in cytidine. However, this does not argue against a biological role for the i-motif, which may act as a regulator.

Several cases of neocentromeres evolving from DNA sequences sharing no homology at all with a DNA have been reported, supporting the idea that the primary sequence is not the major determinant for the function. Tertiary structure might thus be important for the function (Mitchell, 1996; Sunkel & Coello, 1995), and in this context, tandemly repeated sequences or satellite DNAs, which are major components of the eukaryotic genome and widespread in the centromeric region, are interesting candidates.

Satellite III, whose consensus sequence is $(ATTCC)_nATTCGGGTTG$ (n 1-13) (Prosser et al., 1986) has been proposed to be part of the functional centromere (Grady et al., 1992). It has been recently identified in all human chromosomes by the primed in situ labeling (PRINS) technique, except in chromosomes 6, 8, 11, 12, 18, 19 and X (Therkelsen et al., 1997), but the authors have raised the possibility that these chromosomes could either share a lower degree of homology with the consensus sequence or, as assumed before by Grady, too low a number of copies for detection by PRINS. The fact that satellite III-like pentameric repeats are found in the proterminal region of human chromosomes (Vocero-Akbani et al., 1996),

at the kappa locus of human immunogobulin light chains (Pargent et al., 1991), and at the end of the 3' LTR in the proviral sequence of HTLV1 (Seiki et al., 1982), in regions with high genetic recombination potential or with functional roles such as pro-motion of transcription and viral DNA integration is stimulating for further studies.

The variability of i-motif structures observed in vitro for short C-rich fragments of the human satellite III, and of the associated exposed residues in the loops, is not inconsistent with a functional role. In vivo, the l and the η-forms of the satellite III repeats might be preferred depending on the intranuclear medium. In the η-form, unpaired loop L2 residues may provide an easy access for ligand recognition. The variability in intercalation topology might be important if the satellite III sequence adopts different conformations in response to differing biological conditions. These structures may be stabilized differently *in vivo* than *in vitro* by intranuclear ligands, inter i-motif inter-actions (for instance, loop-loop interactions) and/ or by the effective concentration of the high number of d(CCATT) repeats. The fact that the comp-lementary G-rich strand, once separated from its C-rich counterpart, also adopts different kinds of structures (hairpin, stem-looped, duplex) in sol-ution (Jaishree & Wang, 1994; Castati et al., 1994; Chou et al., 1996) and that on this strand also a single nucleotide substitution has dramatic effect on the structure (Zhu et al., 1996), is consistent with this idea.

Materials and Methods

Sample preparation

The oligonucleotides were synthesized on a 2.5 mm scale using solid-phase b-phosphoramidite chemistry and purified as described (Leroy & Guéron, 1995). The collected fractions were adjusted to neutral pH and dialyzed several times against 10 mM NaCl solution and finally against H_2O. After lyophilization, the samples were dissolved in 95 % H_2O, 5 % ^2H_2O or 99.98 % ^2H_2O solutions containing 1 mM EDTA and 0.1 mM dimethyl silapentane sulfonate. The sample pH was measured at room temperature for most of the experiments, or at 0 C for samples used in exchange time measurements. It was adjusted using 0.1 to 1 M NaOH and HCl stock solutions. $C^2H_3O^2H$ was added (9 % (v/v)) for the experiments recorded at temperature lower than ÿ7 C so as to avoid sample freezing. Before all 2D experiments, the samples were heated up to 100 C and quenched in an ice-water bath to favor the formation of monomeric species. The strand concentration was computed from the absorbance at 260 nm, using a nearest-neighbor model (Cantor & Warshaw, 1970).

Stoichiometry determination

Stoichiometry was measured by HPLC using a Synchropack GPC 100 size-exclusion chromatography column (Synchrom, Lafayette, IN, USA) and calibrated using a set of reference samples as described (Leroy et al., 1994; Han et al., 1998). It was deduced also from the comparison of 1D spectra recorded versus the oligonucleotide concentration.

NMR experiments and data processing

All the NMR experiments were performed on a 500 MHz Varian Unity Inova spectrometer equipped with a 5 mm Penta probe. The spectra in H_2O were recorded using for detection the ``jump and return`` (JR) sequence (Plateau & Guéron, 1982) adjusted with two weak pulses for 90 pulse phase and length corrections. The maxi-mum frequency response was set to 13 ppm. In ^2H_2O solvent, the residual H^2HO resonance was pre-saturated during the recycle delay by low-power irradiation, which was switched off 0.3 seconds before excitation to avoid saturation of the near-water fast relaxing H3' resonances (Phan et al., 2000).

The 2D data were acquired in the hypercomplex mode in 256 steps (States et al., 1982). The NOESY, COSY and TOCSY in ^2H_2O were recorded with a spectral width of 4.1 kHz, an acquisition time of 249.8 ms and repetition delays varying from 1.2 seconds at ÿ12.5 C to 2.2 seconds at 25 C. NOESY spectra in H_2O were recorded with a spectral width of 12 kHz, an acquisition time of 85 ms, a repetition delay of 1.3 seconds and mixing times of 30, 50, 80, 150 and 250 ms. A small Z-gradient pulse (6.5 G/cm, 2 ms) was applied before JR detection to cancel out spurious transversal magnetization. For short mixing times, the first 90 pulse was phase-shifted by 45 so as to avoid the ÿ180 orientation during phase cycling (Smallcombe, 1993). The TOCSY and NOESY experiments in ^2H_2O were recorded with mixing times of respectively 30 and 70 ms, and 35, 55, 75, 90, 150 and 300 ms. The TOCSY used MLEV-17 repetitions without trim pulses (Bax & Davis, 1985) during the mixing period. The ^1H-^{31}P hetero-TOCSY experiment used for sequential assignment was recorded with a ^{31}P spectral width of 1.5 kHz, an acquisition time of 683 ms, a repetition delay of 2.2 seconds, and a 50 ms DIPSY-2 mixing pulse sequence (Kellogg, 1992). The ^1H-^{31}P hetero-COSY (Sklenar & Bax, 1987) used for J-coupling measurements was acquired with a ^{31}P spectral width of 1.2 kHz, an acquisition time of 853 ms and a repetition delay of 1.75 seconds. The adenine aromatic protons were correlated via carbon C4 by a natural abundance HMBC experiment (van Dongen et al., 1996a,b) lasting 64 hours at 15 C for a strand concentration of 1.8 mM, with t set to 40 ms.

The imino proton exchange times were measured by magnetization transfer (Guéron & Leroy, 1995). The i-motif lifetime was determined at 25 C from the measurement of the volumes of the diagonal peaks and of the exchange cross-peaks of A8(H8), A8(H2), C7(H5) and C17(H6) with the unstructured strand in ^2H_2O NOESY experiments recorded with mixing times (tm) of 0.15, 0.4, 0.5, 0.6, 0.8 and 1 second assuming a null volume for tm=0 (Macura et al., 1994).

For 1D experimental data, the T1 and the saturation transfer from water measurements were processed as described (Nonin et al., 1997), using a home-made pro-gram running on PC computers. The spectra recorded in H_2O were corrected for the JR frequency response. The remaining water signal was suppressed by post-acquisition processing in the time domain (Guéron et al., 1991). The 2D experiments were processed using the Felix 97.2 package software (Molecular Simulations, Inc) running on an Indigo 2 workstation (Silicon Graphics).

The residual water signal in H_2O NOESY spectra was reduced by a digital shift procedure (Roth et al., 1980).

Distance restraints

The distances were derived from the build-up of NOE cross-peaks measured at -4°C in H_2O and 15°C in ^2H_2O with mixing times of 30, 50 and 80 ms, and of 35, 55, 75, and 90 ms, respectively. The cross-peaks were manually defined, and the volumes integrated in Felix 97.2. Cross-peaks in H_2O were corrected for the digital shift procedure and the JR excitation response. The cross-peak volumes were scaled by reference to the H5-H6 or H5-CH_3 volumes (corresponding to distances of 2.45 and 2.9 Å, respectively). The inter-residue distances were sorted into three categories, with lower and upper bounds of 1.8-2.7 Å, 2.3-3.7 Å and 3.2-4.7 Å. The intra-residue distances were restrained to the measured distance 15 to 30 % depending on the peak resolution. Upper bound restraints involving methyl or C imino protons were increased by 0.5 Å to account for the CH_3 rotation or the jump of the C imino proton between the two N3 of the pair. The four C.C+ base-pairs were enforced by restraining the inter-base *cis* amino proton-O2 and N3-N3 distances to 1.74(±0.1) Å and 2.76(±0.1) Å, respectively. No base-pairing restraint was imposed between A3 and U13.

We have used 139 attractive restraints (including 79 inter-residue restraints) derived from NOESY measurements, and 12 ambiguous restraints involving H5'/H5'' protons. A list of the interproton distances shorter than 5 Å was generated from the ten best structures using the MOLMOL 2.4 routines (Koradi et al., 1996), and confronted to the NMR data. Short inter-proton distances inconsistent with the NOESY spectra were excluded by 29 repulsive distance restraints longer than 4.2 or 4.7 Å depending on the peak resolution. Because of the strong spin diffusion that occurs between geminal H2'/H2'' or between amino protons, the shortest measured distance involving one of these protons was restrained, and the longest distance was used as a lower bound. We have included 38 such restraints in the structure computation.

Dihedral angle restraints

Heteronuclear $J_{H3'-P}$ coupling constants were measured manually for all residues except for 5mC1, A3, C16 and C17, from the splitting of the outer components of the cross-peak multiplets in the ^{31}P-^1H hetero-COSY experiment. The scalar couplings fall within the range of 3 to 15 Hz. According to a modified Karplus relation (Lankhorst *et al*., 1984), this excludes three ranges of values of ε backbone angle, $20° < \varepsilon < 100°$, $-60° < \varepsilon < -20°$ and $140° < \varepsilon < 180°$. These exclusions were applied during the structure calculation.

Structure computation and visualization

The structures were computed by X-PLOR 3.851 (Brünger, 1990) running on an INDIGO 2 workstation (Silicon Graphics). Calculations started from an extended, single-stranded DNA structure generated using the DNA-builder of Quanta 95 (Molecular Simulation Inc.), and further randomized in X-PLOR. The covalent geometry was enforced by the standard harmonic potential. Potential energy terms related to electro-statics and empirical dihedral were omitted. The randomized starting structure was first energy-minimized by five Powell cycles. In a second step, molecular dynamics was run for 2000 steps of 2 ps with an initial velocity corresponding to 2400 K, and the van der Waals force constant was increased after step 1500. The third step consisted in cooling the system to 300 K in steps of 25 K each, followed by 2 ps dynamics. Base-pair restraints, experimental interproton distance and dihedral restraints on angles were introduced during the cooling step with force constants of 500, 50, and 50 kcal mol^{-1} $Å^{-2}$, respectively. Base-pair planarity was not enforced. In final step, the structures thus generated entered an energy minimization procedure of 1000 Powell cycles. We then selected and gathered the ten out of 100 conformers with the lowest NOE-related energies and aligned them by minimization of the RMSD between corresponding cytidine bases C1' and N1 positions. The structures were visualized using MOLMOL 2.4, whose nucleotide library was adapted to recognize 5-methyl and protonated cytidine bases and deoxyuridine bases. The residues and helix parameters were computed with X-PLOR 3.851 and CURVE 5.1 (Lavery & Sklenar, 1989). The RMSD values were determined for the base and the sugar-phosphate moieties by pairwise comparison of the aligned conformers.

Coordinates deposition

The coordinates for the lowest energy Mcent-U Z-structure, and the restraints used in the computation, have been submitted to the RCSB Protein Data Bank (PDB and RCSB ID codes are 1G22 and RCSB012136, respectively).

Acknowledgments

We thank Dr Maurice Guéron for helpful discussions. This work was supported by Centre National de la Recherche Scientifique (CNRS) of France and, by grant 9272 (19 December 1997) from the Association pour la Recherche contre le Cancer.

References

Ashley, C. T. & Warren, S. T. (1995). Trinucleotide repeat expansion and human disease. Annu. Rev. Genet. 29, 703-728.

Bax, A. & Davis, D. G. (1985). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. J. Magn. Reson. 65, 355-360.

Berger, I., Kand, 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). Nature Struct. Biol. 2, 416-425.

Brünger, A. T. (1990). X-PLOR Version 3.1, A System for X-ray Crystallography and NMR, Yale University Press, New Haven and London.

Cai, L., Chen, X., Raghavan, S., Ratliff, R., Moyzis, R. & Rich, A. (1998). Intercalated cytosine motif and novel adenine cluster in the crystal structure of the tetrahymena telomere. Nucl. Acids Res. 29, 4696-4705.

Cantor, C. R. & Warshaw, M. M. (1970). Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleotides. Biopolymers, 9, 1059-1077.

Castati, P., Gupta, G., Garcia, A. E., Ratliff, R., Hong, L., Yau, P., Moyzis, R. K. & Bradbury, E. M. (1994). Unusual structures of tandem repetitive DNA sequences located at human centromeres. Biochemis-try, 33, 3819-3830.

Castati, P., Chen, X., Deaven, L. L., Moyzis, R. K., Bradbury, E. M. & Gupta, G. (1997). Cytosine-rich strands of the insulin minisatellite adopt hairpins with intercalated cytosine cytosine pairs. J. Mol. Biol. 272, 369-382.

Chen, L., Cai, L., Zhang, X. & Rich, A. (1994). Crystal structure of a four-stranded intercalated DNA: d(C4). Biochemistry, 33, 13540-13546.

Chou, S.-H., Zhu, L. & Reid, B. (1996). On the relative ability of GNA triplets to form hairpins versus self-paired duplexes. J. Mol. Biol. 259, 445-457.

Dobbie, K. W., Hari, K. L., Maggert, K. A. & Karpen, G. H. (1999). Centromere proteins and chromosome inheritance: a complex affair. Curr. Opin. Genet. Dev. 9, 206-217.

Frommer, M., Prosser, J., Tkachuck, D., Reisner, A. H. & Vincent, P. C. (1982). Simple repeated sequences in human satellite DNA. Nucl. Acids Res. 10, 547-563.

Gallego, J., Chou, S.-H. & Reid, B. R. (1997). Centro-meric pyrimidine strands fold into an intercalated motif by forming a double hairpin with novel T:G:G:T tetrad: solution structure of the d(TCCCGTTTCCA) dimer. J. Mol. Biol. 273, 840-856.

Gallego, J., Golden, E. B., Stanley, D. E. & Reid, B. R. (1999). The folding of centromeric DNA strands into intercalated structures: a physicochemical and computational study. J. Mol. Biol. 285, 1039-1052.

Gehring, K., Leroy, J. L. & GueÂron, M. (1993). A tetra-meric DNA structure with protonated cytosine-cytosine 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: a graphical approach. Biopolymers, 15, 2277-2286.

Grady, D. L., Ratliff, R. L., Robinson, D. L., Mc Canlies, E. C., Meyne, J. & Moyzis, R. K. (1992). Highly con-served repetitive DNA sequences are present at human centromeres. Proc. Natl Acad. Sci. USA, 89, 1695-1699.

Géron, M. & Leroy, J. L. (1995). Studies of base-pair kinetics by NMR measurement of proton exchange. Methods Enzymol. 261, 383-413.

Guéron, M., Plateau, P. & Decorps, M. (1991). Solvent signal suppression in NMR. In Progress in NMR Spectroscopy (Emslye, J. W., Feeney, J. & Sutcliff, J. H., eds), vol. 23, pp. 135-209, Pergamon Press, Oxford, UK.

Gupta, G., Garcia, A. E., Castati, P., Ratliff, R., Bradbury, E. M. & Moyzis, R. K. (1994). Stem-loop structures of repetitives DNA sequences located at human centromeres. In Structural Biology: The State of the Art. Proceedings of the Eighth Conversation, State University of New York, Albany, NY 1993 (Sarma, R. H. & Sarma, M. H., eds), pp. 137-154, Adenine Press, New York.

Han, X., Leroy, J. L. & GueÂron, M. (1998). An intramole-cular i-motif: the solution structure and base-pair opening kinetics of d(5mCCT$_3$CCT$_3$ACCT$_3$CC). J. Mol. Biol. 278, 949-965.

Hilbers, C. W., Heus, H. A., van Dongen, M. J. P. & Wijmenga, S. S. (1994). The hairpin elements of nucleic acid structure. Nucl. Acids Mol. Biol. 8, 56-104.

Iwahara, J., Kigawa, T., Kitagawa, K., Masumoto, H., Okasaki, T. & Yokoyama, Shigeyuki (1998). A helix-turn-helix structure in human centromere protein B (CENP-B). EMBO J. 17, 827-837.

Jaishree, T. N. & Wang, A. H.-J. (1994). Human chromo-somal centromere (AATGG)$_n$ sequence forms stable structures with unusal base-pairs. FEBS Letters, 347, 99-103.

Kanaori, K., Maeda, A., Kanehara, H., Tajima, K. & Makino, K. (1998). ^{1}H nuclear magnetic resonance study on equilibrium between two four-stranded solution conformations of short d(CnT). Biochemis-try, 37, 12979-12986.

Kang, C. H., Berger, I., Locksin, C., Ratliff, R., Moyzis, R. & Rich, A. (1994). Crystal structure of intercalated four-stranded d(C$_3$T) at 1.4 Å resolution. Proc. Natl Acad. Sci. USA, 91, 11636-11640.

Kang, C. H., Berger, I., Locksin, C., Ratliff, R., Moyzis, R. & Rich, A. (1995). Stable loop in the crystal struc-ture of the intercalated four-stranded cytosine-rich metazoan telomere. Proc. Natl Acad. Sci. USA, 92, 3874-3878.

Kellogg, G. W. (1992). Proton-detected hetero-TOCSY experiments with application to nucleic acids. J. Magn. Reson. 98, 176-182.

Koradi, R., Billeter, M. & WuÈtrich, K. (1996). MOLMOL: a program for display and analysis of macromol-ecular structures. J. Mol. Graph. 14, 51-55.

Lankhorst, P. P., Hasnoot, C. A. G., Erkelens, C. & Altona, C. J. (1984). Carbon-13 NMR in confor-mational analysis of nucleic acid fragments. 2. A reparametrization of the Karplus equation for viv-inal NMR coupling constant in CCOP and HCOP fragments. J. Biomol. Struct. Dynam. 1, 1387-1405.

Lavery, R. & Sklenar, V. J. (1989). De®ning the structure of irregular nucleic acids: convention and principles. J. Biomol. Struct. Dynam. 6, 655-667.

Lavery, R., Zakrzewska, K., Sun, J. S. & Harvey, C. (1992). A comprehensive classi®cation of nucleic acid structural families based on strand direction and base-pairing. Nucl. Acids Res. 20, 5011-5016.

Leroy, J. L. & GueÂron, M. (1995). Solution structures of the i-motif tetramers of d(TCC), d(5methylCCT) and d(T5methylCC): novel NOE connections between amino protons and sugar protons. Struc-ture, 3, 101-120.

Leroy, J. L., Gehring, K., Kettani, A. & GueÂron, M. (1993). Acid multimer of oligo-cytidine strands: stoi-chiometry, base-pair characterization and proton exchange properties. Biochemistry, 3, 6019-6031.

Leroy, J. L., GueÂron, M., Mergny, J. L. & HeÂleÁne, C. (1994). Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif. Nucl. Acids Res. 22, 1600-1606.

Macura, S., Westler, W. M. & Markley, J. L. (1994). Two-dimensional exchange spectroscopy of proteins. Methods Enzymol. 239, 106-144.

Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. & Okazaki, T. (1989). A human centromere antigen (CENP-B) interacts with a short speci®c sequence in alphoid DNA, a human centromeric satellite. Methods Enzymol. 261, 383-413.

Mergny, J. L., Lacroix, L., Han, X., Leroy, J. L. & HeÂleÁne, C. (1995). Intramolecular folding of pyrimidine oligonucleotides into an i-motif. J. Am. Chem. Soc. 117, 8887-8898.

Mitchell, A. R. (1996). The mammalian centromere: its molecular architecture. Mutat. Res. 372, 153-162.

Nonin, S. & Leroy, J. L. (1996). Structure and conversion kinetics of a bi-stable DNA i-motif: broken sym-metry in the d(5mCCTCC) tetramer. J. Mol. Biol. 261, 399-414.

Nonin, S., Leroy, J. L. & GueÂron, M. (1995). Terminal base-pairs of oligonucleotides: imino proton exchange and fraying. Biochemistry, 34, 10652-10659.

Nonin, S., Phan, A. T. & Leroy, J. L. (1997). Solution structure and base-pair opening kinetics of the i-motif dimer of d(5mCCTTTACC): a non canonical structure with possible roles in chromosome stab-ility. Structure, 5, 1231-1246.

Pargent, W., Meindl, A., Thiebe, R., Mitzel, S. & Zachau, H. G. (1991). The human immunoglobulin kappa locus. Characterization of the duplicated O regions. Eur. J. Immunol. 21, 1821-1827.

Phan, A. T., Géron, M. & Leroy, J. L. (2000). The sol-ution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere. J. Mol. Biol. 299, 123-144.

Plateau, P. & Guéron, M. (1982). Exchangeable proton NMR without base-line distorsion, using strong pulse sequences. J. Am. Chem. Soc. 104, 7310-7311.

Prosser, J., Frommer, M., Paul, C. & Vincent, P. C. (1986). Sequence relationships of three human satellites DNAs. J. Mol. Biol. 187, 145-155.

Roth, K., Kimber, B. J. & Feeney, J. (1980). Data shift accumulation and alternate delay accumulation techniques for overcoming the dynamic range problem. J. Magn. Reson. 41, 302-309.

Seiki, M., Hattori, S. & Yoshida, M. (1982). Human adult T-cell leukemia virus: molecular cloning of the provirus DNA and the unique terminal structure. Proc. Natl Acad. Sci. USA, 79, 6899-6902.

Sklenar, V. & Bax, A. (1987). Measurement of ^{1}H-^{31}P NMR coupling constant in double-stranded DNA fragments. J. Am. Chem. Soc. 109, 7525-7526.

Smallcombe, S. H. (1993). Solvent suppression with sym-metrically-shifted pulses. J. Am. Chem. Soc. 115, 4776-4785.

States, D. J., Haberkorn, R. A. & Ruben, D. J. (1982). A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. J. Magn. Reson. 48, 286-292.

Sunkel, C. E. & Coello, P. A. (1995). The elusive centro-mere: sequence divergence and functional conserva-tion. Curr. Opin Genet. Dev. 5, 756-767.

Therkelsen, A. J., Nielsen, A. & Kolvraa, S. (1997). Local-isation of the classical DNA satellites on human chromosomes as determined by primed in situ labeling (PRINS). Hum. Genet. 100, 322-326.

van Dongen, M. J. P., Wijmenga, S. S., Eritja, R., Azorin, F. & Hilbers, C. W. (1996a). Through-bond corre-lation of adenine H2 and H8 protons in unlabeled DNA fragments by HMBC spectroscopy. J. Biol. NMR, 8, 207-212.

van Dongen, M. J. P., Wijmenga, S. S., van der Marel, G. A., van Boom, J. H. & Hilbers, C. W. (1996b). The transition from a neutral-pH double helix to a low-pH triple helix induces a conformational switch in the CCCG tetraloop closing a Watson-Crick stem. J. Mol. Biol. 263, 715-729.

Vocero-Akbani, A., Helms, C., Wang, J.-C., Sanjurjo, F. J., Korte-Sarfaty, J., Veile, R. A., Liu, L., Jauch, A., Burgess, A. K., Hing, A. V., Holt, M. S., Ramachandra, S., Whelan, A. J., Anker, R. & Ahrent, L. et al. (1996). Mapping human telomere regions with YAC and P1 clones: chromosome-specific markers for 27 telomeres including 149 STSs and 24 polymorphisms for 14 proterminal regions. Genomics, 36, 492-506.

Weil, J., Min, T., Yang, C., Wang, S., Sutherland, C., Sinha, N. & Kang, C. (1999). Stabilization of the i-motif by intra-molecular adenine-adenide-thymine base triple in the structure of d(ACCCT). Acta Crys-tallog. sect. D, 55, 422-429.

Willard, H. F. (1990). Centromeres of mammalian chromosomes. Trends Genet. 6, 410-416.

Zhu, L., Chou, S.-H. & Reid, B. R. (1996). A single G-to-C change causes human centromeres TGGAA repeats to fold back into hairpins. Proc. Natl Acad. Sci. USA, 93, 12159-12164.