1600–1606 Nucleic Acids Research, 1994, Vol. 22, No. 9

Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an *i-motif*

Jean-Louis Leroy, Maurice Guéron*, Jean-Louis Mergny[1] and Claude Hélène[1]
Groupe de Biophysique de l'Ecole Polytechnique et de l'URA D1254 du CNRS, 91128 Palaiseau and [1]Laboratoire de Biophysique, INSERM U201, CNRS UA481, Muséum National d'Histoire Naturelle, 43 rue Cuvier, Paris 5, France

Received February 10, 1994; Revised and Accepted April 11, 1994

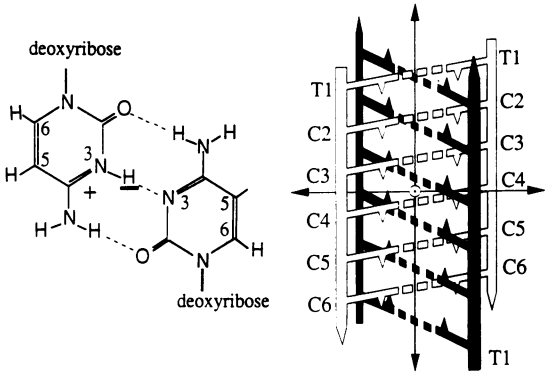
ABSTRACT

In the recently discovered *i-motif*, four stretches of cytosine form two parallel-stranded duplexes whose C·C$^+$ base pairs are fully intercalated. The *i-motif* may be recognized by characteristic Overhauser cross-peaks of the proton NMR spectrum, reflecting short H1′–H1′ distances across the minor groove, and short internucleotide amino-proton-H2′/H2″ across the major groove. We report the observation of such cross-peaks in the spectra of a fragment of the C-rich telomeric strand of vertebrates, d[CCCTAA]$_3$CCC. The spectra also demonstrate that the cytosines are base-paired and that proton exchange is very slow, as reported previously for the *i-motif*. From UV absorbance and gel chromatography measurements, we assign these properties to an *i-motif* which includes all or nearly all the cytosines, and which is formed by intramolecular folding at slightly acid or neutral pH. A fragment of telomeric DNA of *Tetrahymena*, d[CCCCAA]$_3$CCCC, has the same properties. Hence four consecutive C stretches of a C-rich telomeric strand can fold into an *i-motif*. Hypothetically, this could occur *in vivo*.

INTRODUCTION

A solution structure of the multimer formed by d(5′-TCCCCC) was recently obtained using NMR spectroscopy (1). It is a tetramer of equivalent strands, which presents the novel feature of intercalation of the C·C$^+$ pairs of two parallel-stranded duplexes (scheme I). Characteristic inter-residue NOESY patterns of this so-called *i-motif* include H1′–H1′ cross-peaks and reciprocal H1′–H6 cross-peaks, the latter originating from the face-to-face orientation of contiguous intercalated base pairs. The authors of that report suggested that a strand carrying four copies of the cytosine-rich telomeric repeat could fold into an intramolecular *i-motif*, an intriguing possibility given that the complementary guanine-rich sequence is well suited to fold into a so-called G-tetrad (2,3).

Such C-rich sequences, e.g. [CCCCAA]$_4$ and [CCCCAAAA]$_4$, had been reported to form compact, intramolecular structures involving C·C$^+$ base pairs at acid or even neutral pH, on the basis of UV absorbance melting curves, chemical modification and non-denaturing gel electrophoretic migration (4). These observations were not sufficient for defining the structure. In a recent article (5) from the same laboratory, which appeared after submission of the present work, a human (vertebrate) telomeric 4-repeat d[CCCTAA]$_4$ is studied by the same methods, together with a 2-repeat. The 4-repeat folds on itself, while the 2-repeat forms a dimer. An NMR measurement provides evidence that this dimer includes an *i-motif*, by the observation of H1′–H1′ NOESY cross-peaks. It is proposed by analogy that the 4-repeat monomer also forms an *i-motif*.



We now present compelling evidence that a fragment of telomeric DNA of vertebrates, namely d[CCCTAA]$_3$CCC, does indeed form an intramolecular *i-motif*. So does a fragment of telomeric DNA of *Tetrahymena*, d[CCCCAA]$_3$CCCC. The structures are stable at slightly acidic or neutral pH, a feature which allows for a hypothetical *i-motif* component of telomeres *in vivo*.

MATERIALS AND METHODS

Nomenclature

The terms stoiciometry, multimer, n-mer etc. refer to the association of molecules, and not to properties of a single molecules or of a structure resulting from intramolecular folding.

*To whom correspondence should be addressed

Oligonucleotide designates a single strand, as usual. To avoid confusion, repeats within a strand are written with square brackets, and curved brackets are used for multimers. Thus, $[CCCCAA]_4$ is a single strand, 24 nucleoside long, whereas $(CCCCAA)_4$ is the association of four strands, each 6 nucleoside long.

Oligodeoxynucleotide synthesis

The oligodeoxynucleotides d(TCCC), $[CCCTAA]_3CCC$, $[CCCTAA]_7CCC$, CCCCAA and $[CCCCAA]_3CCCC$ were synthesized on a 10-μmol scale by the β-cyanoethyl phosphoramidite method with a Pharmacia Gene Assembler DNA synthesizer, and purified as described (6). Sequences CCCTAACCC and $[CCCTAA]_2CCC$ were from Eurogentec (Belgium). They were purified by alcohol precipitation and gel filtration chromatography (Amersham). Concentrations were determined from the absorbance at neutral pH, which was computed with a nearest-neigbour model (7), leading to absorbances (A_{260}) of 30600, 185900, 404685, 46800, 18860O, 76500 and 131190 $M^{-1}cm^{-1}$ respectively. All concentrations are expressed in strand molarity.

UV absorbance

The variation of ultra-violet absorbance with temperature was measured with a KONTRON-UVIKON 940 spectrophotometer. The solutions were vortexed, introduced in 1 cm path-length quartz cells and overlaid with a thin layer of paraffin oil to prevent evaporation. The temperature of the 6-cell holder was regulated by circulation of a glycerol−water coolant (20%−80%), using a HAAKE cryo-thermostat, and monitored by a thermistor immersed in a cell containing only buffer. Heating and cooling rates were set by a HAAKE PG20 temperature programmer at 9°C/hour. Temperature and absorbance were measured every 4 minutes. Water condensation was avoided by gently blowing dry air in the cell compartment.

Gel filtration chromatography

The size, and hence stoichiometry, of the n-mers ($n \geq 1$) was determined by high-pressure gel filtration chromatography performed at room temperature with Beckman equipment. The column was a Synchropack GPC 100 (250 mm×4.6 mm I.D.) from Interchim, calibrated with oligonucleotides as described (6). The elution medium was buffered by 20 mM sodium acetate at pH 4.2, and it contained salt (NaCl, 0.3 M) to reduce DNA−matrix interactions. The flow rate for optimal resolution was 0.4 mL/minute. The chromatography samples for calibration or study consisted typically of 20 μL of a 1 OD/mL oligonucleotide solution. The elution profiles were measured by the absorbance at 280 nm.

Nuclear magnetic resonance

The NMR samples were prepared by lowering the pH of a solution at the final concentration. Small aliquots of HCl were added and the acid form was monitored by the build-up of the imino proton NMR spectrum. Heat treatments are described below. The concentration of the NMR samples was limited to 2.1 mM, in view of aggregation.

The NMR measurements were performed on a home-built 8.45 T spectrometer, as described (6). The 'jump and return' (JR) sequence was used for solvent signal suppression. The 1D spectra were processed with automatic subtraction of Lorentzians at the solvent frequency and amplitude correction for the JR response (8−10). For measuring slow imino and amino proton exchange rates, a concentrated sample in H_2O was diluted 12-fold into D_2O and the NMR spectrum was recorded at intervals. Integral intensities were measured on 1D spectra obtained with an inter-FID delay of at least 3 times the longest relaxation time. Proton NOESY data were obtained in the hypercomplex mode (11) on the same spectrometer, and transferred to an IRIS workstation (Silicon Graphics, Inc.) for processing with FELIX 1.1 (courtesy of Hare Research).

The NOESY pulse sequence for samples in D_2O consisted in 90° pulses for preparation, mixing and readout. The spectral width was 3.5 kHz, and the acquisition time was 0.3 s. The t_1 delay was incremented from 0 to 60 ms (200 values of t_1). The inter-FID delay was 1.5 s.

For NOESY measurements in H_2O, a 10 ms homo-spoil gradient was applied at the start of the mixing period, and a JR sequence with maximum sensitivity at 13.3 ppm was used for readout. The spectral width was 10 kHz, and the acquisition time was 0.1 s. The t_1 delay was incremented from 0 to 20 ms. The inter-FID delay was 1 s.

Data processing in the t_2 dimension included exponential broadening (1 Hz in D_2O, 3 Hz in H_2O) and sine-bell apodization (12), with phase-shifts of 45° for D_2O solutions and 60° for H_2O, and a FELIX skew factor of 0.8. A digital shift accumulation correction (13) was applied to reduce the residual water signal, and the base line was corrected with a polynomial function. Data processing in the t_1 direction consisted in sine-bell apodization with the same phase and skew as in the t_2 direction.

RESULTS

UV absorbance

The formation of structure in a C-rich oligonucleotide results in hyperchromism at 260 nm (4). In order to monitor the structural transition without interference from the change of absorption due to protonation of cytidine (14), we recorded the absorbance at

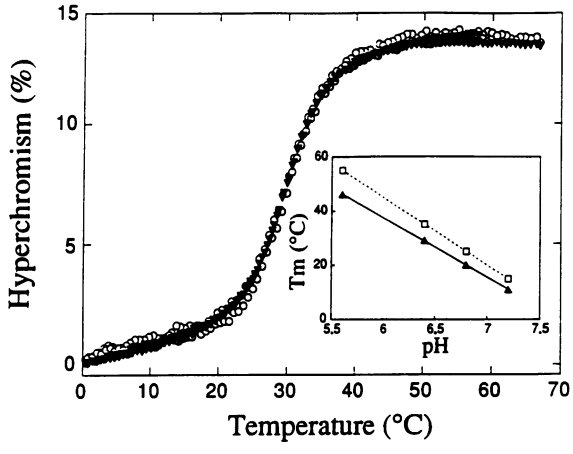


Figure 1. Denaturation profiles for the vertebrate telomeric sequence at pH 6.4. The profiles were recorded at 265 nm for two different concentrations (triangles: 10 μM; circles 1.2 μM). The transitions are reversible. The invariance of the denaturation profile vs. concentration indicates that the melting process involves no change in multimericity. Insert: The effect of pH on T_m, for the vertebrate (triangles) and *Tetrahymena* (open squares) sequences. The profiles, measured at a rate of 6°C/hour, were reversible except at pH 7.2, where hysteresis was observed.

the isobestic wavelength for C protonation (265 nm). At pH 6.4, the transition of the vertebrate telomeric sequence was monophasic and reversible, with a half-transition temperature T_m of 29°C (Fig. 1). At pH 7.2, hysteresis was observed, a consequence of slow kinetics (15).

The structural transition was unaffected by a ten-fold increase in concentration (Fig. 1). Therefore it involves no change in stoichiometry, i.e. it is an internal rearrangement of the observed species. One supposes that the species is a monomer, and this is indeed confirmed by the gel filtration measurements.

Plotting the log of the equilibrium constant vs (temperature)$^{-1}$ yields a straight line, confirming that the melting curve reflects a single reaction. The enthalpy and entropy for the formation of the low-temperature form at pH 6.4 are −251 kJ/mol and −0.79 kJ/mol/K.

As expected for the formation of $C \cdot C^+$ base pairs, the melting profile is strongly pH-dependent, with a reduction of T_m by 22°C per pH unit for the vertebrate sequence, and 25°C for the *Tetrahymena* sequence (Fig. 1, insert).

We also studied shorter and longer fragments of the human telomeric sequence, with respectively 2, 3 and 8 cytosine stretches. The melting transition of the shorter fragments was concentration-dependent (T_m=8°C and 4°C at pH 6.4, 5 μM concentration), indicating intermolecular association. The behavior of the 8-repeat fragment $[CCCTAA]_7CCC$ was very similar to that of the 4-repeat, indicating that its folded form resembles two contiguous folded structures of the 4-repeat. The T_m was 30°C at pH 6.4. These results are those expected if the folded structure includes an *i-motif* formed by four CCC stretches.

We investigated the effect of various cations on the stability of the vertebrate sequence $[CCCTAA]_3CCC$. At pH 6.4, the T_m was highest at low ionic strength (10 mM cacodylate buffer with no other salt). It was 10°C less in 0.1 or in 0.3 NaCl. The effect of Li^+ and K^+ was similar, with a reduction of T_m by 6° for a concentration of 30 mM. The effect of divalent cations (Mg^{2+}, Ca^{2+}) was larger, −15° for 30 mM.

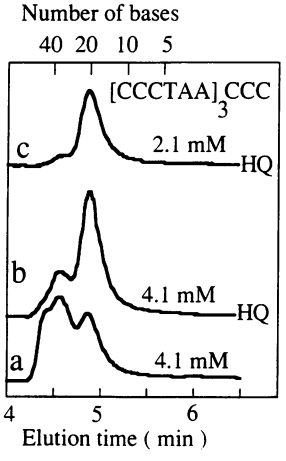


Figure 2. Gel filtration profile of the vertebrate telomeric C strand repeat. (**a**) Sample prepared by lowering the pH of a 4.1 mM solution from 7.5 to 4.7. It contains a mixture of multimer, dimer and monomer. (**b**) After heating up to 100° and quenching at 0° (HQ label), only monomer and dimer remain (80% and 20% respectively). (**c**) In a sample at the 2.1 mM concentration of the NOESY experiments, the same treatment reduces the dimer concentration to 8%. Profiles were stable for days. Experimental conditions: T= 18°; NaCl, 0.3 M, acetate 20 mM, pH 4.2 .

Gel exclusion chromatography

At the concentrations of the UV absorption experiments (1.2 and 10 μM), the chromatography profile of the vertebrate repeat sequence $[CCCTAA]_3CCC$ displayed only monomers (not shown). In contrast, acidification of a 4.1 mM solution produced a mixture of monomers, dimers and multimers (Fig. 2, trace a), which was stable for days at room temperature. The solution was then heated to 100°C and quenched by immersion in an ice-water bath. Only monomers (21 bases) and dimers (42 bases) remained, in proportions of 80% and 20% respectively (trace b).

The same procedure was used for the NMR samples. Trace c shows the profile of an aliquot of an NMR sample (concentration 2.5 mM), taken before the NMR measurement. The composition is 92% monomers and 8% dimers. An aliquot taken after the NMR measurement had the same profile. Hence the NMR spectrum is mostly that of monomer species.

The behavior of the *Tetrahymena* repeat $[CCCCAA]_3CCCC$ was quite similar. After heating and quenching as above, a 1.8 mM oligomer solution contained only monomers and dimers, in proportions of 90 % and 10 % respectively (Fig. 3, left panel, trace a).

For comparison, elution profiles of the CCCCAA fragment are also shown (traces b and c). They display a mixture of monomer and tetramer (6 and 24 bases respectively) in proportions dependent on the concentration. The tetrameric stoichiometry is confirmed by the equilibrium titration (right panel of Fig. 3): the multimer concentration increases as the fourth power of the monomer concentration. The dissociation constant is $(1.2\ 10^{-5})^3\ M^3$. We expect that the tetramer of CCCCAA has the *i-motif* structure, but this was not investigated.

NMR spectroscopy

The NMR spectrum of the vertebrate telomeric repeat, the NOESY spectrum and the NMR exchange kinetics are quite

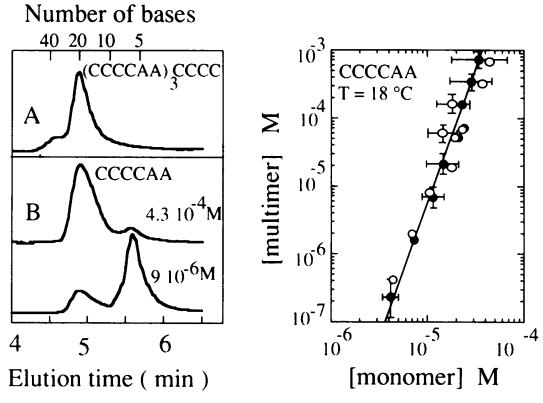


Figure 3. Left panel: Comparison of (**a**) the elution profile of the *Tetrahymena* telomeric repeat $[CCCCAA]_3CCC$, with (**b,c**) profiles of a CCCCAA fragment. The telomeric repeat sample was prepared by dilution of an aliquot of the NMR sample (concentration 1.8mM). The proportions of dimer (42 bases) and monomer (21 bases), respectively 10% and 90%, are constant for days. The CCCCAA fragment elutes as a tetramer or a monomer, in proportions which depend on the concentration. Right panel: Titration of the multimer-monomer equilibrium of CCCCAA. The concentrations of multimer and monomer were measured from elution profiles like those in the left panel. The titrations were performed after 2-day incubation at room temperature of samples which were initially mostly tetrameric (filled symbol) or monomeric (empty symbol). A line of slope 4, corresponding to a tetramer−monomer equilibrium with a dissociation constant of $(1.2\times10^{-5})^3\ M^3$, is drawn through the experimental points. Same experimental conditions as in Fig.2.

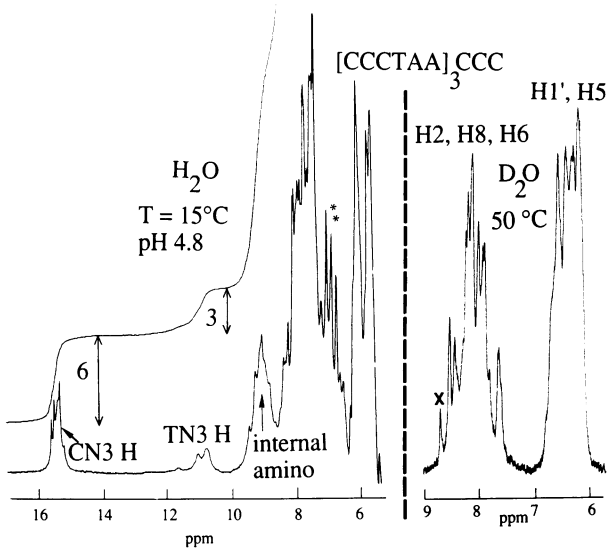


Figure 4. NMR spectra of the vertebrate telomeric oligomer. Left panel: By comparison with the spectrum of the three thymine imino protons (10 to 12 ppm), the cytosine imino proton spectrum around 15.5 ppm integrates to 6 protons, corresponding to 6 $C \cdot C^+$ pairs . The NMR lines around 7 ppm labeled by stars are from contaminant NH_4^+. Right panel: By comparison with the integrals of the (H2,H8,H6) and (H1',H5) clusters, which correspond to 27 and 33 protons respectively, the peak at 8.6 ppm corresponds to 0.5 protons. This indicates the existence of at least two intra-molecular configurations.

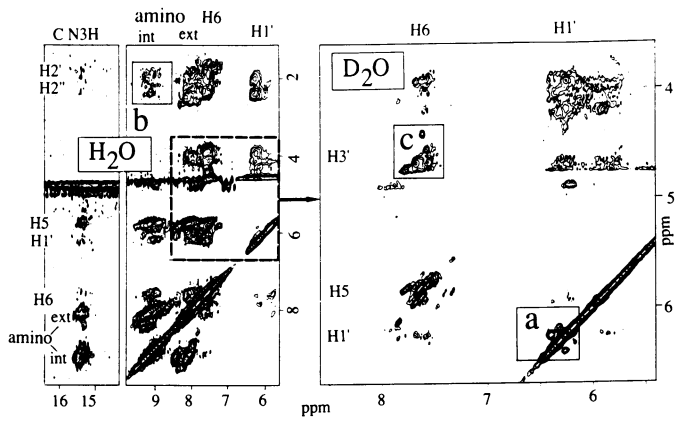


Figure 5. NOESY spectra of the vertebrate telomeric strand 5'-d([CCC-TAA]$_3$CCC) in H_2O (left) and D_2O (right). Mixing time 200 ms, T= 15 °C, pH 4.7. The H1'−H1' cross-peaks (a box) and the cross-peaks between amino and H2'/H2" protons (b box) are characteristic of the *i-motif*. Strong H6−H3' (c box) and weak H1'−H6 cross-peaks indicate an anti conformation and 3'-endo sugar puckering.

similar to those of the tetramer of TCCCCC (1,6) and therefore provide strong evidence for an *i-motif* structure.

The 1D spectrum includes peaks at 15 to 16 ppm (Fig. 4, left) which are assigned to imino protons of protonated cytosines on the basis of their chemical shift (16, 17). The chemical shifts of the amino protons, mid-way between those for neutral and protonated cytosines, show that the protonated cytosines participate in symmetrical hemi-protonated $C \cdot C^+$ base pairs (6). By comparison with the thymine imino proton spectrum ca. 11 ppm, which is assumed to correspond to three protons, the integrated intensity of the cytosine imino proton spectrum corresponds to six protons per strand, hence all 12 cytosines are base-paired. The spectrum is very similar to that of the tetramer of TCCC (6), which we find from NMR to have an *i-motif* structure (not shown). This has also been observed by K.Gehring (private communication).

The spectrum on the right of Fig. 4 shows the aromatic region of the spectrum in D_2O. The (H2, H8, H6) and (H1', H5) clusters correspond to 27 and 33 protons respectively, and their integrals are indeed in this ratio (not shown). By comparison, the resolved peak at 8.6 ppm corresponds to 0.5 proton. This indicates that the telomeric strand is present in at least two conformations, which, according to the chromatography measurements, are intra-molecular.

Due to the presence of multiple forms and to the lack of symmetry, the NMR spectrum includes more peaks and is therefore more crowded than that of *i-motif* structures formed by tetramers such as $(TCCCCC)_4$ whose strands are all equivalent (1). Furthermore, the chemical shift dispersion is actually smaller. Nevertheless, the NOESY spectrum displays the characteristic H1'−H1' cross-peaks of the *i-motif* (Fig. 5, box a).

The face-to-face geometry of base pairs in the *i-motif* gives rise to indirect H1'(i)−H6(j) and H1'(j)−H6(i) cross-peaks of comparable intensities between protons of nucleotides i and j facing each other across the minor groove (1). In the present case the cross-peaks were unobservably weak (the reason for this is given below), and hence they could not be compared.

We have recently found a third, even more striking NOESY pattern of the *i-motif*, consisting in strong cross-peaks between amino protons and H2" and H2' protons. They could be expected from the short distances occuring in the *i-motif* between amino protons of one strand and the H2'/H2" protons belonging to the anti-parallel strand located across the major groove (1). The cross-peaks are located in a region of the spectrum which is empty in the case of B-DNA oligomers, even with very long mixing times. But they are easily detected in NOESY spectra of for instance TCC or T(5-methyl-cytosine)CC tetramers, even with very short mixing times (in preparation).

These characteristic cross-peaks appear clearly in the 200 ms NOESY spectrum of the vertebrate telomeric sequence (Fig. 5, box b). The assignment to amino protons and to H2' and H2" protons is unambiguous, based on exchange of the amino protons, on chemical shifts, and on the NOESY connectivities. The cross-peaks are also observed with a mixing time of 100 ms, small enough to restrict the influence of spin diffusion in a molecule of 21 nucleotides. They can therefore be ascribed to a direct Overhauser effect, indicating a short inter-proton distance.

The D_2O NOESY spectrum also shows strong intra-nucleotide H3'−H6 cross-peaks, reflecting an *anti* orientation of the glycosidic linkage and a 3'-endo sugar pucker (Fig. 5, box c). Thus the nucleoside conformation is similar to that found in the *i-motif* of the TCCCCC tetramer (1).

On the other hand, the intra-nucleotide H1'−H6 cross-peaks are much weaker than in $(TCCCCC)_4$. This shows that the conformations are more *anti* in the present case. It also explains the weakness of the *inter*-nucleotide H1'−H6 cross-peaks: these are indirect, and the diffusion pathway is expected to include an intra-nucleotide H1'−H6 connection.

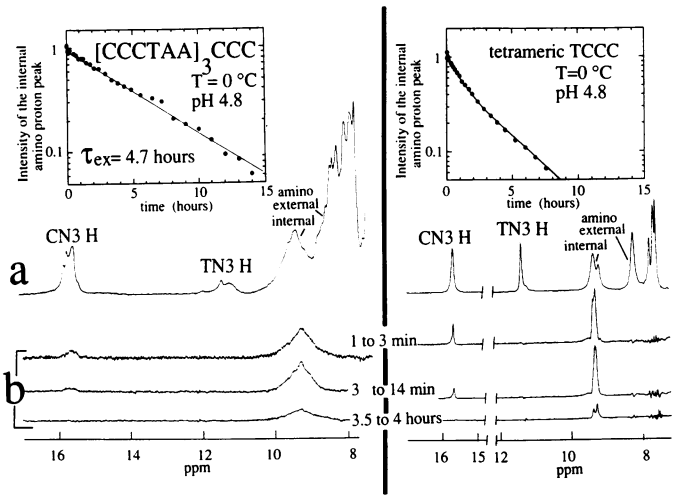


Figure 6. Real-time proton exchange experiments in the vertebrate telomeric strand and in the tetramer of TCCC. The similarity of spectra and exchange rates suggests similar structures. Left panel: (**a**) Reference spectrum in 90% H_2O at 0°C. The exchange experiment started when a concentrated protonated solution of the oligomer was diluted into D_2O. (**b**) Difference spectra obtained by subtracting the spectrum recorded 19 hours after changing solvent from those recorded during the time intervals indicated on the figure. The number of slowly exchanging internal amino protons is 10.5 ± 1.5, by reference to the non-exchangeable resonances. The insert shows that these protons have the same exchange time of 4.7 hours (a relative vertical scale is used). The external amino protons exchange in less than one minute. Right panel: The same experiment was done on the TCCC tetramer which has an *i-motif* structure. One internal amino proton exchanges in less than 3 minutes. Exchange of the two others is slow, with exchange times of hours. Their linewidth diminishes as the external amino proton exchanges for deuterium.

Proton exchange

Figure 6 (left) shows a real-time exchange measurement of the vertebrate sequence. The external (i.e. non-hydrogen-bonded; scheme I) amino protons exchange in less than 1 minute. But most of the internal (hydrogen-bonded) amino protons (10.5 ± 1.5 out of 12) exchange extremely slowly and at the same rate (t_{ex} = 4.7 hours). The imino protons which are observable in the subtracted spectra (b) have exchange times in the range of 10 to 60 s. As is apparent by comparison with Fig. 6b, the exchange behaviors of amino protons and of imino protons are remarkably similar to those of the tetramer of TCCC and of other C·C$^+$ oligomers which were extensively studied previously (6). In the earlier study, it was observed and explained that in the case of C·C$^+$ base pairs, and in contrast to the case of Watson–Crick pairs, the imino proton exchange time is equal to the base-pair lifetime, due to efficient internal catalysis. The same is expected also in the present case.

Tetrahymena repeat

All the experiments carried out on the vertebrate repeat were duplicated on the *Tetrahymena* repeat [CCCCAA]$_3$CCCC, with only slight differences in results. The melting temperatures were higher by 5 to 10°C. NMR spectra and NOESY cross-peaks were comparable. The internal amino protons exchanged slowly, 8 ± 2 in 0.2 hrs, and 8 ± 2 in 2.2 hrs. Approximately 2 of the 8 cytosine imino protons exchanged in 20 minutes.

DISCUSSION

Structure of C-rich strands

The results reported above constitute strong evidence that a fragment d[CCCTAA]$_3$CCC of the cytosine-rich strand of the vertebrate telomere folds into an intramolecular *i-motif* at slightly acidic pH. This is also the case for the *Tetrahymena* telomeric sequence [CCCCAA]$_3$CCCC.

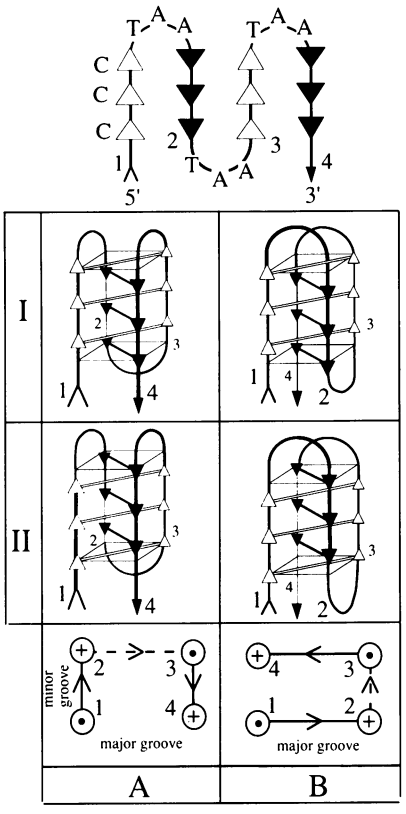
The evidence includes: the melting curves and gel chromatography migration profiles showing the formation of a pH-sensitive intramolecular structure; the imino and amino proton NMR spectra demonstrating the formation of C·C$^+$ base pairs; the very slow imino and amino proton exchange rates, indicative of an *i-motif* structure involving most, if not all, of the cytosine moieties. the 3′-*endo*/*anti* nucleoside conformation previously observed in *i-motif* structures; and the NOESY cross-peaks indicative of short H1′–H1′ distances and short amino proton-H2′/H2″ inter-residue distances: the *i-motif* is the only known structure in which such distances are found.

This evidence consists primarily in qualitative features of the NOESY spectra, and our conclusions are therefore compelling even in the absence of detailed assignments. The poor resolution of the spectrum of the intramolecular, non-symmetrical structure of telomeric 4-repeats suggests that such assignments will not come easily. A 600 MHz spectrum did not improve matters considerably (not shown). It seems more important to look for sequences which fold mostly or entirely in one form.

The *i-motif*, which involves the intercalation of two double, parallel-stranded helixes, can be formed by the association of four strands carrying a cytosine stretch (e.g. TCCCCC), or presumably of two strands carrying two such stretches (6), or, as shown here, by the intramolecular folding of one strand carrying four such stretches.

In the case of one strand, there are four different configurations in which all cytosines are base-paired and all base pairs are intercalated, as indicated by the NMR measurements. We start with the cytosine stretch 1 in a given vertical position, and place stretch 2 anti-parallel to it (scheme II). For intercalation with stretch 1, stretch 2 must be offset vertically, either down (row I) or up (row II), giving two possibilities. Furthermore, stretch 2 may be positioned (and its nucleosides oriented) in two ways: either accross the minor groove of the *i-motif* with respect to stretch 1 (column A), or accross the major groove (column B). This gives a total of 2×2 = 4 possibilities. Once stretches 1 and 2 are positioned, the positions of stretch 3 (paired to 1) and 4 (paired to 2) follow. Hence there are four configurations, defined by the four dispositions of stretch 2.

There is no obvious energetic difference among the four configurations: two of them have two loops across the major grooves and one across a minor groove, and this arrangement may or may not be less stable than the other one (two loops across minor groove, one across a major groove). Thus we do not know how many and which configurations are present in our samples. The melting measurements are consistent with one species, but do not exclude more. In the NMR conditions, there is evidence for more than one species, from the 0.5 proton intensity of the resolved line at 8.7 ppm (Fig. 4, right side). The *Tetrahymena* sequence [CCCCAA]$_4$ also forms more than one intramolecular structure, according to the electrophoretic migration pattern (4). One could attempt to design a strand favoring a single configuration by adjusting the length and composition of the different loops.



Complementary G-rich and C-rich strands

The G tetrads include stacks of planar structures of four hydrogen-bonded guanines. They form as tetramers of a strand which includes one stretch of three or four guanines, or as dimers of two strands carrying two such stretches, or again by the folding on itself of a strand carrying four such stretches. The Watson–Crick complements of these elements have C-stretches which are expected to form an *i-motif*. In particular, the sequence d(AGGG[TTAGGG]$_3$) folds on itself as a G tetrad whose solution structure has been solved by NMR (3). This sequence is the complement of the human telomeric sequence studied here, except for the initial adenine.

The ends of eucaryotic chromosomes consist in ten- to thousand-fold repeats of sequences in which a C-rich sequence is paired to its Watson–Crick G-rich complement, terminated typically by an overhanging, non base-paired, dual repeat of the G-rich sequence which could play a role in telomere–telomere associations (reviewed in references 18 and 19). Sundquist and Klug (20) studied oligonucleotide dimers consisting of a strand carrying a series of the *Tetrahymena* TTGGGG repeat, associated with a shorter series of the complementary strand, resulting in a double-helical section followed by a single-stranded 3′-terminal overhang of two G-rich repeats. They showed that two overhangs associate to form G-tetrads, while the rest of the molecules retain the Watson–Crick structure.

In the cell, telomeric DNA is usually associated with proteins, and its structure is not known. It could vary according to physiological conditions, e.g. the phase in the cell cycle. Association of chromosomes via the G overhang has not been demonstrated *in vivo*. There are indications that the structure of the double-stranded part of telomeric DNA could differ from a B-DNA duplex. For instance, the transcriptional regulator repressor activator protein (RAP1) binds to telomeric repeats of *S.cerevisiae*, and induces a distortion of the double helix structure (21). Also, the radiosensitivity of intra-chromosomal telomeric sites of a Chinese hamster ovary (CHO) cell line is greater than that of the rest of the genome, a feature which might be related to structure (22).

No folding of a piece of the G-rich strand into a G-tetrad, or of its complementary sequence into an *i-motif* has been detected *in vivo* at this time. Nevertheless, one can but notice that a stretch of four repeats of the two complementary sequences, each of equal length, can each fold onto itself in structures involving their G and their C bases respectively, thus providing for some or all of the telomeric repeats an alternative structure to the Watson–Crick duplex: each *i-motif* structure (scheme II) would face a G-tetrad, to which it could associate by Watson–Crick pairing of the bases in the loops. This double-knobbed structure would be compact and maybe stable. It would present specific structures for molecular recognition which could play a role in such functions as chromosome stabilization or as the anchoring of chromosomes to the nuclear envelope (19). Analogous structures could also be formed by connecting the C-rich strand of one chromosome with the G-rich strand of another, again via the loops.

The marginal stability of the *i-motif* at neutral pH is not a strong argument against such speculations. First, telomere-binding proteins—which are known to exist—could shift the pK of formation of the *i-motif* to a more basic value than that of isolated strands, ca. 6 to 7 (Fig. 1). Inversely, the marginal stability of the *i-motif* structure of the C-strand could be useful for the modulation of the telomeric structure by protein–nucleic acid interactions.

ACKNOWLEDGEMENTS

We thank Dr Ali Kettani and Monique Leblond for the synthesis of the DNA sequences, Isabelle Corcelle for assistance in the equilibrium titrations and Dr Michel Rougée for helpful discussions. The 2D acquisition software was written by K.Gehring.

REFERENCES

1. Gehring,K., Leroy,J.L., and Guéron,M. (1993) Nature, 363, 561–565.
2. Henderson,E., Hardin,C.C., Walk,S.K., Tinoco,I.Jr. and Blackburn,E.H. (1987) Cell 51, 899–908.
3. Wang,Y. and Patel,D.J. (1993) Structure 1, 263–282.
4. Ahmed,S. and Henderson,E. (1992) Nucleic Acids Research 20, 507–511.
5. Ahmed,S., Kintanar,A. and Henderson,E. (1994) Nature Structural Biology 1, 83–88.
6. Leroy,J.L., Gehring,K., Kettani,A., and Guéron,M. (1993) Biochemistry, 32, 6019–6031.
7. Cantor,C.R. and Warshaw,M.M. (1970) Bioplymers 9, 1059–1077.
8. Plateau,P. and Guéron,M. (1981) J. Amer. Chem. Soc. 104, 7310–7311.
9. Guéron,M., Plateau,P. and Decorps,M. (1991) Progress in NMR Spectroscopy 23, 135–209.
10. Guéron,M., Plateau,P., Kettani,A. and Decorps,M. (1992) J. Magn. Reson. 96, 541–550.
11. States,D., Haberkorn,R.A. and Ruben,D.J. (1982) J. Magn. Reson. 48, 286–292.
12. DeMarco,A. and Wüthrich,K. (1976) J. Magn. Reson. 11, 201–204.
13. Roth,K., Kimber,B.J. and Feeney,J. (1980) J. Magn. Res. 41, 302–309.
14. Manzini,G., Xodo,L.E., Gasparotto,D., Quadrifoglio,F., van der Marel,G.A., van Boom,J.H. (1990) J. Mol. Biol. 213, 833–843.
15. Rougée,M., Faucon,B., Mergny,J.L., Barcelo,F., Giovannangeli,C., Montenay-Garastier,T. and Hélène,C. (1993) Biochemistry , 31, 9269–9278.

16. Becker,E.D., Miles,H.T. and Bradley,R.B. (1965) J. Amer. Chem. Soc. 87, 5575–5582.
17. Dos Santos,C., Rosen,M. and Patel,D.J. (1989) Biochemistry 28, 7282–7288.
18. Blackburn,E.H. (1992) Annu.Rev. Biochem., 61, 113–129.
19. Gilson,E., Laroche,T. and Gasser,S.M. (1993) Trends in Cell Biology, 3, 128–134.
20. Sundquist,W.I. and Klug,A. (1989) Nature, 342, 825–829.
21. Gilson,E., Roberge,M., Giraldo,R., Rhodes,D. and Gasser,S.M. (1993) J. Mol. Biol., 231, 293–310.
22. Alvarez,L., Evans,J.W., Wilks,R., Lucas,J.N., Brown,J.M. and Giacca A.J. (1993) Genes, Chromosomes and Cancer, 8, 8–14.