4634–4640 Nucleic Acids Research, 1994, Vol. 22, No. 22

Evidence for intramolecularly folded i-DNA structures in biologically relevant CCC-repeat sequences

G.Manzini*, N.Yathindra[1] and L.E.Xodo
Department of Biochemistry, Biophysics and Macromolecular Chemistry, University of Trieste, Via Giorgieri 1, I-34127 Trieste, Italy and [1]Department of Crystallography and Biophysics, University of Madras, India

Received August 1, 1994; Revised and Accepted October 13, 1994

ABSTRACT

The structural behaviour of repetitive cytosine DNA is examined in the oligodeoxynucleotide sequences of $(CCCTAA)_3CCCT$ (HTC4), $GC(TCCC)_3TCCT(TCCC)3$ (KRC6) and the methylated $(\underline{C}C\underline{C}T)_3T\underline{C}CT(\underline{C}C\underline{C}T)_3C$ (KRM6) by circular dichroism (CD), gel electrophoresis (PAGE), and ultra violet (UV) absorbance studies. All the three sequences exhibit a pH-induced cooperative structural transition as monitored by CD. An intense positive CD band around 285 nm develops on lowering the pH from 8 to slightly acidic condition, indicative of the formation of base pairs between protonated cytosines. The oligomers are found to melt in a fully reversible and cooperative fashion, with a melting temperature (T_m) of around 50°C at pH 5.5. The melting temperatures are independent from DNA concentration, indicative of an intramolecular process involved in the structural formation. PAGE experiments performed with ^{32}P-labeled samples as well as with normal staining procedures show a predominantly single band migration for all the three oligomers suggestive of a unimolecular structure. From pH titrations the number of protons required for generating the structures formed by HTC4, KRC6 and KRM6 results to be around six. These findings strongly suggest that all the three sequences adopt an intramolecular i-motif structure. The demonstration of i-motif structure for KRC6, a critical functional stretch of the c-ki-*ras* promoter proto-oncogene, besides the human telomeric sequence HTC4, may be suggestive of larger significance in the functioning of DNA.

INTRODUCTION

Tandemly repeated sequences of cytosines are normally found at the 5'-ending strands of telomeres of eukaryotic chromosomes (1,2). They are also found to occur rather frequently in other non-coding regions of eukaryotic DNA, including promoter sites, satellite DNAs and introns. A few examples that can be provided are: intergenic sites of human $\alpha 2$ and $\alpha 1$ globins (3), chicken $\alpha 2$(I) collagen gene, (4) human c-*myc* gene (5) and murine c-ki-*ras* -proto-oncogene promoter (6). Also, *Drosophila virilis* DNA is known to contain a $(CT)_{8-18}(C)_{4-5}$ repeat as long as 400 base pairs (7). Even RNAs such as in the genomes of encephalomyocarditis virus and several other cardioviruses contain many stretches with over 75% cytosines (8). A search in the EMBL-NEW 1 and GENBANK 80 sequence data banks (involving about 180 million residues belonging to over 163,000 sequences), for the polypyrimidine stretch (TCCCTCCCTCCCTCCTTCCC-TCCCTCCC) and its complementary, that are found in the c-ki-*ras* promoter, reveals the abundance of $(CCCT)_n$ repeats, with n higher than six, in different locations that include repetitive DNA regions, coding genes and also regions of unknown functions. In fact this sequence repeated as many times as 48 is found in the promoter region of the oncogene-inducible T64 gene (9). Therefore, it is sufficiently important to examine the functional significance of these repetitive elements by studying their structural behaviour. While the complementary G-rich repeat of the telomeric strand is known to form a number of quadruplex structures (10–12) based on Hoogsteen hydrogen-bonded guanine quartets, much less is known about the conformational abilities of the C-containing sequences. Quadruplexes of the type seen for G-rich repeats are not expected for the C-repeats, in view of the rather limited hydrogen bonding versatility of cytosines. With the aim of elucidating the structural polymorphism of DNA oligonucleotides (ODN's) composed by C-block sequences, we have investigated the following three oligodeoxy-nucleotides containing C-repeats:

(i)	$(CCCTAA)_3CCCT$	22mer	HTC4
(ii)	$GC(TCCC)_3TCCT(TCCC)_3$	30mer	KRC6
(iii)	$(\underline{C}C\underline{C}T)_3T\underline{C}CT(\underline{C}C\underline{C}T)_3C$	29mer	KRM6

The 22-mer (HTC4) represents the telomeric repeat found in the human/vertebrate chromosome ends, while the 30-mer (KRC6) is a part of the 160 b.p. murine c-ki-*ras* proto-oncogene promoter (6). The 29-mer (KRM6) is the inverted sequence of KRC6 with alternating cytosines being methylated ($\underline{C}$) and lacking the terminal guanine residue. The CCC block repeats itself four times in HTC4 and six times in the other two sequences. The

*To whom correspondence should be addressed

latter two are intervened by a stretch of only two cytosines in the middle. KRC6 has been shown to be a necessary component of the promoter of the c-ki-*ras* proto-oncogene and its absence supresses the promoter activity (13). Moreover, this region is also found to be a good substrate for S1 nuclease and a nuclear protein factor binding to it has been identified (13). These observations make the structural investigations of ODNs comprising CCC-blocks important to understand their biological implications not only at the telomeric ends but also at other regions of the genome as well.

MATERIALS AND METHODS

The oligonucleotides used in this work were synthesized according to standard solid-phase phosphoramidite chemistry:

$(CCCTAA)_3CCCT$	22mer	HTC4
$GC(TCCC)_3TCCT(TCCC)_3$	30mer	KRC6
$(\underline{CC}\underline{C}T)_3T\underline{C}CT(\underline{CC}\underline{C}T)_3C$	29mer	KRM6
$(AGGGTT)_3AGGG$	22mer	HTG4
$(GGGA)_3AGGA(GGGA)_3GC$	30mer	KRG6

Purification was carried out by gel-permeation chromatography using a G50 Sephadex resin and 10 mM ammonium bicarbonate as eluent.

Ultraviolet absorption (UV) measurements have been performed on a Cary 2200 (Varian) spectrophotometer, equipped with thermostated cell holder. Denaturations and renaturations were carried out by incrementing the temperature at a rate of 0.3°C/min using a Haake PG 20 temperature programmer, connected to a Haake water circulating bath equipped with a refrigerator. Oligomer concentrations were determined by absorption at 260 nm, at T = 80°C, assuming a molar extinction coefficients of 8000 $M^{-1}cm^{-1}$ per pyrimidine base and 15000 $M^{-1}cm^{-1}$ per adenine. Thermodynamic data have been obtained by analysing reversible melting curves with an *all-or-none* model .

Circular dichroism (CD) spectra were obtained with a Jasco J-500 A spectropolarimeter, equipped with a thermostated cuvette holder that allows measurements at controlled temperature. The dichrograph was connected to a Jasco DP500 N data processor. Spectra are presented as $\Delta\epsilon = (\epsilon_L - \epsilon_R)$ in units of $M^{-1}cm^{-1}$. The spectra were obtained from DNA solutions at mM strand concentrations in buffer as specified in the figure captions. For pH titrations the CD spectrum of each sample was measured initially at pH around 8 and pH was gradually lowered and the spectrum was obtained each time. The pH titrations were carried out by both acidification and basification. Starting from Tris buffer at pH near 8, successive aliquots of acetic acid were added to gradually decrease the pH to a value near 5. After each addition the CD spectrum was recorded. A similar procedure was adopted in going from acidic to basic conditions. Essentially superimposable results were obtained for all the three samples in the case of both increasing and decreasing pH. Thermal denaturations as a function of pH were also determined.

Polyacrylamide gel electrophoresis (PAGE) was carried out on gels obtained from buffer solutions containing 20%(w/v) acrylamide (3.3% bis-acrylamide), 0.07% ammonium persulphate. The buffers used were sodium acetate (pH = 5.5) and Tris−HCl (pH 8). In general 2−4 μg of DNA were loaded in each lane. Electrophoresis was performed at controlled temperature (15°C) by means of thermostated apparatus. Gels were run at constant voltage (6V/cm) and bromophenol blue was used as marker. Electrophoretic bands were stained with 'stains-all' dye in water/ formamide (1:1) and the gels photographed with a Polaroid camera. The oligonucleotides were labeled with [γ^{32}P]-ATP (DuPont) using T4 polynucleotide kinase (Pharmacia). After electrophoresis the gels were dried (90 min at 85°C) and autoradiographed on a Kodak film.

RESULTS

It is well known that oligo- and poly-nucleotides built up by regular C- containing repeats assume secondary structures at slightly acidic pH and this process is accompanied by marked CD changes (14−20). The CD spectra for the three sequences (HTC4, KRC6, KRM6) have been recorded at room temperature at different pH values. CD spectra have been obtained in Tris-acetate buffer (50mM) in the presence of 20mM KCl. CD spectra at various pH values are shown in Fig.1. The magnitudes, band shapes and crossovers for the three sequences are very similar. The red shift followed by an intense increase in the magnitude of the positive band in the long wavelength region near 285 nm, concomitant with the appearance of the negative band near 260 nm, are seen as the pH is gradually lowered. A small negative band of rather low intensity at around 230 nm is detected for the sequences KRC6 and KRM6 and not for HTC4. This could be attributed to absence of adenosine residues in them, as the presence of even a single adenosine residue in the repeating moiety tends to eliminate this signal. This can be inferred by inspection and comparison of the CD spectra of poly (dC) (14), $(CCTT)_5$ (19), $T_6C_6T_6$ (19) on one hand, and $(AACC)_5$ (19), $(CCATT)_5$ (20) and $A_6C_6A_6$ (19) on the other. The CD changes occur in a very narrow range of pH, as seen from Fig. 2, and are indicative of co-operative transitions. The half transitions occur at similar values of pH for the three sequences considered, the longer ones at slightly higher pH values of around 6.6, the shorter at pH = 6.2. The CD spectral features of the present

Table I. Thermodynamic and pH titration data for the CCC-block repeat oligomers

	pH[a]	n	T_m[b] (K)	ΔH° [c] (kcal/mol)	ΔS° [c] (cal/mol·K)
HTC4	6.2	5.8 ± 0.4	326	65 ± 8[d]	200 ± 25
KRC6	6.55	5.6 ± 0.4	333	90 ± 9	270 ± 30
KRM6	6.6	4.6 ± 1	337	77 ± 9	230 ± 30

[a] at semitransition
[b] at pH = 5.1
[c] at pH = 5.1, modeling the process simply as a two state transition between structured and unstructured forms of the oligomers
[d] to be compared with the data of 60 kcal/mol for the almost identical sequence d(CCCTAA)3CCC (ref. 25)

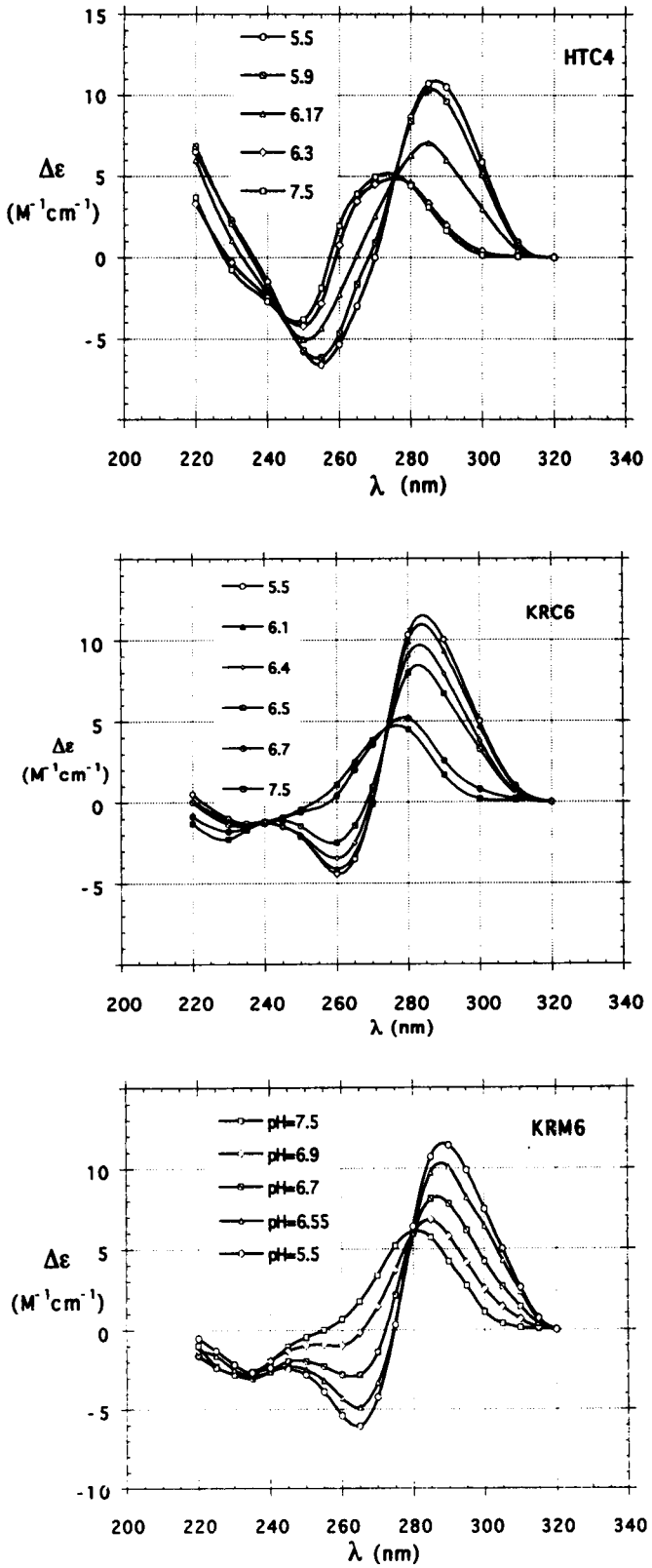

Figure 1. Circular dichroism spectra of HTC4, KRC6 and KRM6 as a function of pH and at room temperature. The spectra were taken in the presence of 20 mM KCl and 50 mM Tris acetate buffer with sample concentration around 0.1 mM in nucleotides. Ellipticity is expressed in units of $M^{-1}\,cm^{-1}$, where M is given as (mol nucleotide)$\cdot L^{-1}$.

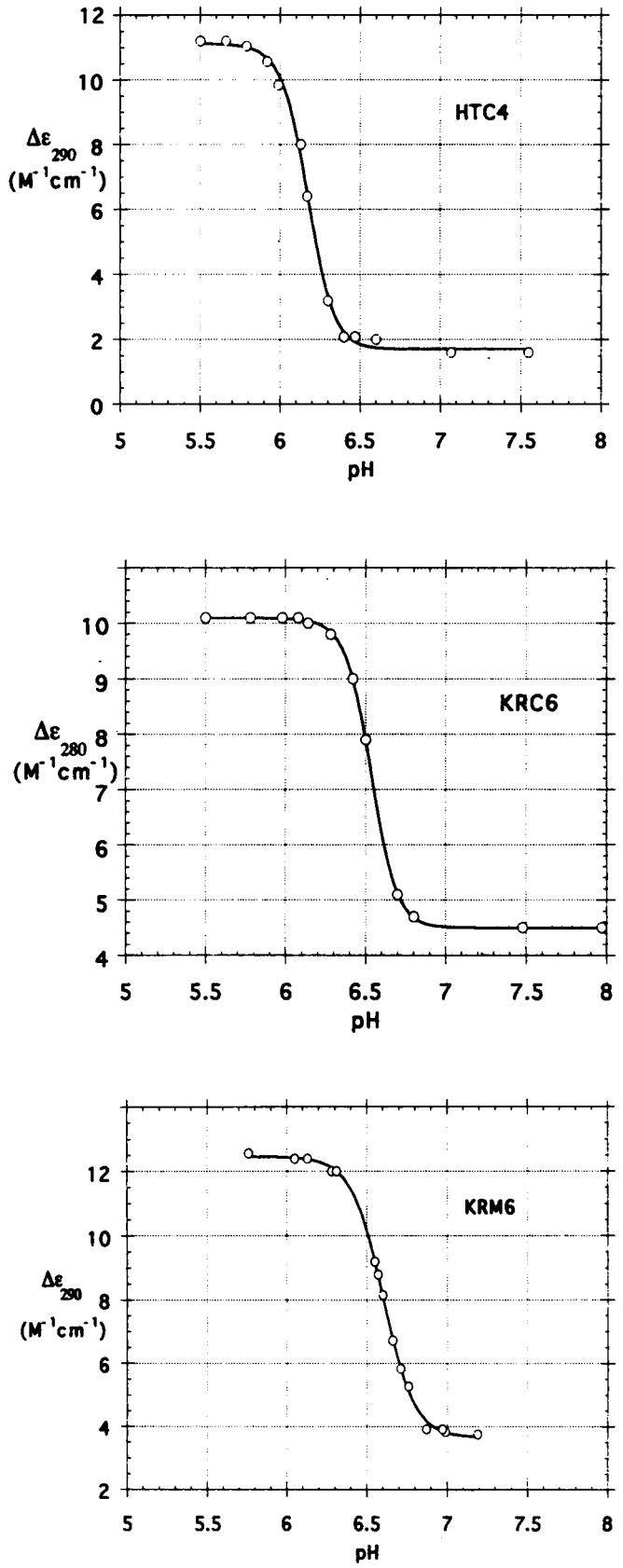

Figure 2. Variation of C.D. signal at constant wavelength as a function of pH for HTC4, KRC6 and KRM6, at room temperature. Full lines represent the best fits with eq. 3.

oligonucleotides match nicely with those found in several cytosine-rich oligo- and poly-nucleotide systems (14–20), and have been considered earlier as demonstrating the formation of hydrogen bonded pairs between protonated cytosines.

The temperature dependence of the CD spectra has also been examined for all the three ODNs and a similar behaviour is noticed in all of them. Hence, representative CD spectra obtained for HTC4 alone, as a function of temperature, as well as a melting profile are shown in Fig. 3. One clearly defined co-operative transition is found for all the three sequences. UV melting profiles have been recorded in parallel, with analogous results. The processes were reversible at the heating and cooling rate of the experiments (0.3°C/min). Melting profiles, at a number of different oligomer concentrations, have also been recorded: the semi-transition temperatures were found to be nearly independent of sample concentration, clearly indicative of a predominantly intramolecular process. The reciprocal of the melting temperatures determined at different pH values for the three sequences are shown in Fig. 4. Increase in T_m values on lowering pH is expected in relation to the formation of the $C:C^+$ base pairs.

Since useful information about the molecularity and structure of ODN's in nondenaturing conditions can also be inferred through electrophoresis, a number of experiments have been performed. Fig. 5 shows the autoradiograph of a PAGE experiment conducted at pH 5.5. Each of the three samples, HTC4, KRC6 and KRM6 migrated as a single species. HTC4 (lane d) migrated slower than its complementary strand HTG4 (lane c), which is known to form an intramolecularly folded G-quartet structure (11). HTC4 exhibited a mobility comparable to that of the reference 14-mer duplex d(CGCACGCGCGTGCG) (lane a, the slower band corresponding to the full length duplex, the faster band corresponding to the hairpin form of this palindromic reference). This points to the presence of a monomolecular species for HTC4. It may be noted that the reference 14-mer duplex (28 nucleotides) is larger than the monomolecular HTC4; it has however 28 negative charges compared to the 16 net negative charges (−22+6) of the 22-mer HTC4, if 6 out of 12 of its cytosines are protonated. This fact is relevant for comparing the expected relative mobilities of a normal duplex and a hemiprotonated structure. In lane b, an equimolar mixture of labeled HTC4 and unlabeled complementary HTG4 was run: the absence of a band corresponding to the WC duplex suggests that at this pH, at least, one of the structures formed by the two oligomers is sufficiently stable to prevent fast duplex formation. KRC6 and KRM6 (lanes e and f) migrated also as single bands, with a mobility lower than that of HTC4, as expected for larger species of the same molecularity. Experiments were repeated at higher sample concentrations for pH values of 5.5 (Fig. 6a) and 8 (Fig. 6b), using unlabeled oligomers and the gels were stained with 'stains-all' dye. In this case, however, also minor amounts of multimeric species were detected at acidic pH (Fig. 6a, lanes a and c), whose relative abundance is difficult to assess through dye staining. At pH 8, only one band is detected in each case (Fig. 6a, lane a for KRC6 and lane d for KRM6), as expected for unstructured single strands. In these experiments the sequence KRG6, the complementary of KRC6, has been run in lane b, as well as the equimolar mixture of the two (KRC6 and KRG6) in lane c. However, the mixture has been annealed at neutral pH and the subsequent acidification in the experiment of Fig. 6a evidently did not disrupt the duplex in favour of the intramolecular structures of the components. Incidentally, it can be observed that KRG6 did not fold intramolecularly into a G-quartet structure, since its mobility almost coincides with that of the corresponding WC duplex. Whether this suggests simply

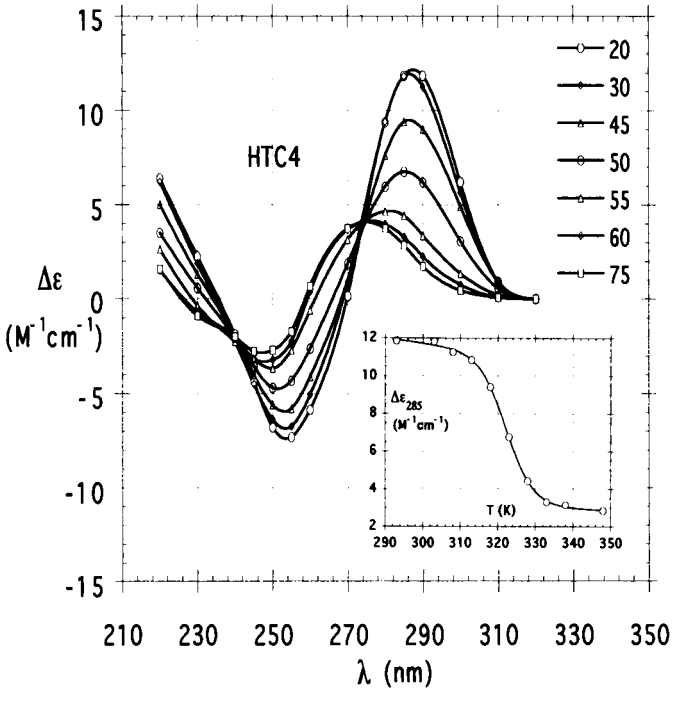

Figure 3. C.D. spectra obtained as a function of temperature for a representative CCC repeating ODN (HTC4) under the same experimental conditions of Fig. 1, but at pH 5.5. Insert: melting profile of HTC4, obtained by plotting the ellipticity at 285 nm (pH 5.5) as a function of temperature.

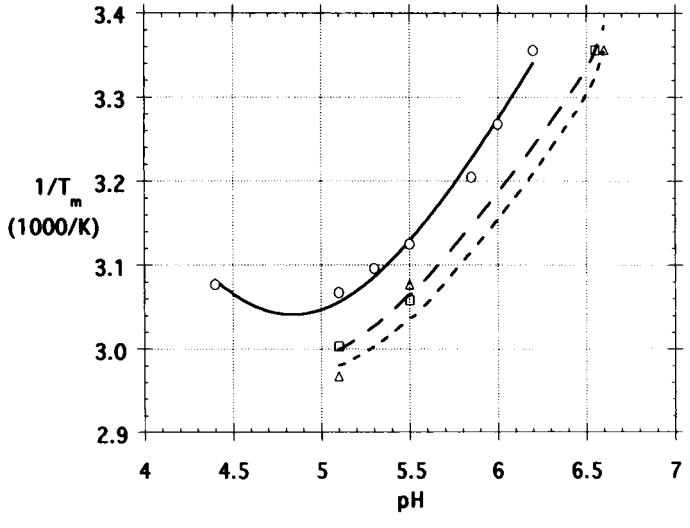

Figure 4. Dependence of $1/T_m$ on pH (○) HTC4 (|) KRC6 and (Δ) KRM6. The experimental conditions were 20 mM KCl and 50 mM Tris buffer. The curves are the best fits of the data using eq. 3, where $K_f = \exp[-(\Delta H°/RT) + (\Delta S°/R)] = \exp[(1-(T_m/T))\Delta S°/R]$ and $pK_a = 4.8$, assumed independent of temperature. The adjustable parameters were n, T_m and $\Delta S°$.

an unfeatured single strand or implies a different structure remains to be established; however, the presence of minor amounts of low mobility species, irrespective of pH, is suggestive of intermolecular G-quartet forms.

The cooperativity and the reversibility of the structural transitions observed at 25°C as a function of pH prompted us to evaluate the number of protons involved in the process from the data presented in Fig. 2. Assuming that the process can be described, in the first approximation, as an equilibrium between unstructured (unprotonated) (U) and structured (protonated) (SH_n) forms we can write the following equation:

$$U + nH^+ = SH_n \text{ with } K_s = [SH]_n/[U][H^+]^n \qquad \text{(eq.1)}$$

where n is the number of protons involved in the formation of the structure and K_s is the equilibrium constant. Plotting of the data in terms of a standard Hill plot, ln {[SH_n]/[U]} *versus* pH, a straight line is obtained with a slope equal to 2.3 n, where n is the number of protons involved. Values of n around 6 for HTC4 (with four CCC blocks) and slightly less for the other two sequences, KRC6 and KRM6 (with six CCC blocks) were obtained. This evidence supports an intramolecular transition involving hemi-protonation of cytosines for HTC4, and about 30% cytosine protonation for the KRC6 and KRM6. The number of protons involved in such intramolecular processes could be inferred also from the dependence of T_m on pH, since, at T_m,

$$\ln \{[SH_n]/[U]\} = 0$$

It follows then:

$$\Delta G° = \Delta H° - T_m \Delta S° = -RT_m\, 2.3\ n \cdot pH$$

and

$$(\Delta H°/T_m) = \Delta S° - R \cdot 2.3 \cdot n \cdot pH$$

After evaluating $\Delta H°$ from the absorption-temperature profiles through a standard method for intramolecular transitions, a value of $n = 6 \pm 0.5$ was obtained from the plot of $1/T_m$ versus pH, for pH values higher than 5.5 (Fig. 5), i.e. at pH values sufficiently higher than pK_a of cytosines, which is about 4.3 for monomeric cytosine and is expected to be slightly higher in oligomers. In fact, the formation of hemiprotonated $C:C^+$ pairs implies that this simple model is adequate only in such range of pH. Maximum thermal stability can be expected at pH around pK_a, i.e. at hemiprotonation of the unfolded sequence, as both full protonation and deprotonation should destroy the structure. A relatively simple way to model this feature is to describe the system with a partition function, that includes all possible protonated forms for the unstructured oligomer, and assume that only the partially protonated chains fold into a lower free energy structure. For HTC4, with four CCC repeats, the partition function can be written as

$$Z = \{1 + ([H^+]/K_a)\}^{12} + K_f \cdot ([H^+]/K_a)^n \qquad \text{(eq. 2)}$$

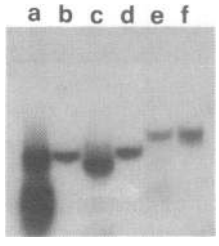

Figure 5. Non-denaturing 20% polyacrylamide gel electrophoresis in 100 mM sodium acetate buffer (pH 5.5) of: ^{32}P-labeled 14-mer palindromic reference sequence (lane a); HTC4 plus non-labeled HTG4 in excess (lane b); HTG4 (lane c); HTC4 (lane d); KRM6 (lane e); and KRC6 (lane f). The gel was run at 15°C at a constant voltage of 10V/cm.

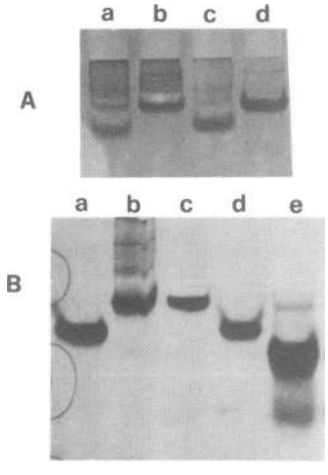

Figure 6. (**A**) Non-denaturing 20% polyacrylamide gel electrophoresis in 50 mM sodium acetate (pH 5.5) of KRC6 (lane a); KRG6 (lane b); KRM6 (lane c); and equimolar KRC6 plus KRG6 (lane d). The gel was run at 15°C at a constant voltage of 10V/cm; (**B**) Non-denaturing 20% gel in 50 mM Tris−HCl (pH 8), 50 mM NaCl of KRC6 (lane a); KRG6 (lane b); equimolar mixture of KRG6 and KRC6 (lane c); KRM6 (lane d); and 14-mer palindromic reference sequence (lane e). Both gels were stained with stains-all dye. The gel was run at 15°C at a constant voltage of 10V/cm.

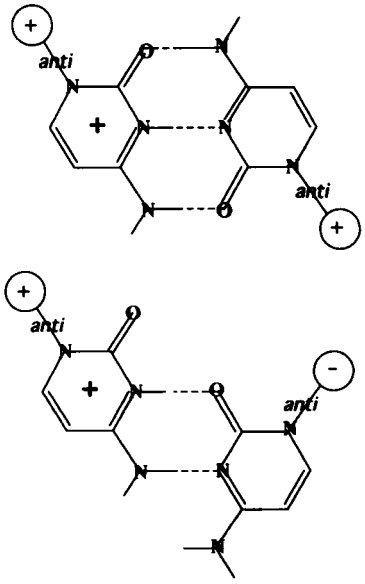

Figure 7. Possible pairing schemes between cytosine and protonated cytosine: (**a**) *trans* pair with hemiprotonated bases (**b**) *cis* pair with one of the cytosines protonated.

n being the number of protons necessary to form the folded structure and K_f the constant of folding the (already) n-protonated single strand. The equilibrium constant of the simpler model, K_s, differs from K_f in that it includes, besides folding, also hemiprotonation.

The fraction of structured oligomers is then

$$\theta = K_f \cdot ([H^+]/K_a)^n/Z \qquad \text{(eq. 3)}$$

The data of Fig. 2a corresponding to HTC4 have been analysed with this model assuming for pK_a the value of 4.8, that has been chosen on the basis of the results presented in Fig. 5 (see below). The fitting of the data provided the following results: n = 5.8 ± 0.4 and log K_f = 8.2 ± 0.6. The value of n is consistent with that reported above from the data of Hill plot. Confident of the validity of the model, optimised log K_f and pK_a values of 7.4 ± 1.1 and 5.0 ± 0.2, with n fixed at the expected hemiprotonation value of 6, were obtained for HTC4.

The data of Fig. 2b relative to KRC6 and KRM6 have been similarly processed. In these cases, values of n = 5.6 ± 0.4 and log K_f = 9.8 ± 0.6 have been obtained for the first sequence and n = 4.6 ± 1 and log K_f = 8 ± 1 for the second. Table I summarises the data. A fairly good agreement is found between our results for HTC4 and those of Leroy *et al.* (25) for the almost identical sequence $d(CCCTAA)_3CCC$.

DISCUSSION

It is evident from the data presented above that all the three sequences, containing four and six blocks of CCC repeats, assume intramolecularly folded structures stabilised by hydrogen bonds between hemiprotonated cytosines. In the case of HTC4 the number of protons that has been found indicates that just hemiprotonation is needed for the formation of the structure. Since each of the other two sequences contains six blocks of CCC (differently from the four present in HTC4), hemiprotonation should require 9 protons (n = 9). Having deduced a significantly lower value clearly rules out a hairpin structure based on $C:C^+$ pairing for both the sequences.

It is important to remember that CD spectra similar to those exhibited by these sequences have been observed earlier in a number of ODNs, containing different amounts of cytosines (18−20), and structures that include parallel (18−20) and antiparallel duplexes (18), as well as parallel duplexes with looped out bases (17), stabilised by $C:C^+$ pairs, have been proposed. Analogous models including hairpin structures have also been put forward on the basis of different evidences (21). While the triple hydrogen bonded $C:C^+$ *trans* pair, concomitant with *anti* N-glycosidic conformations for all the cytosines, leads to parallel alignment of the sugar-phosphate backbone chains, a different $C:C^+$ pair with only two hydrogen bonds (Fig. 7), could provide an antiparallel duplex (18) as well as a hairpin structure. However, nmr studies (22,23) on a number of cytosine rich ODN's, TC_6, TC_5, TC_4, TC_3, TC_3T, $C_4TC_4T_2C_8T_2C_{12}C_5$, and C_4 have revealed a novel and unusual four stranded structure arising from antiparallel intercalation of two $C:C^+$ paired (triple hydrogen bond) parallel stranded duplexes, called i-DNA. Moreover, recently, evidence has been provided that sequences containing four blocks of CCC of the human telomeric repeat adopt an intramolecular i-motif based structure (24,25). They are essentially equivalent to HTC4, from which they differ only for the presence of two extra adenosines at the end (24) or for the absence of the terminal thymidine (25). In the light of these findings, and also of the evidences provided here by the estimated number of protons involved in the folding process, gel mobility and CD spectra, one can argue that intramolecularly folded structures of the same type are assumed not only by the four CCC containing HTC4, but also by the other two oligomers containing six CCC blocks.

In this context it should be made clear that the six CCC-block containing sequences, KRC6 and KRM6, have a feature that is absent in the four CCC-block containing HTC4. This pertains to the possibility of generating a number of different intramolecular foldings based on i-motif by using four out of the six available CCC blocks. In fact, besides the obvious folding based on CCC blocks 1, 3, 4 and 6, which implies the role of loops for the blocks 2 and 5, along with the central TCCT moiety,

GCTCCCTCCCTCCCTCCTTCCCTCCCTCCC
1 2 3 4 5 6

several other folding schemes are possible once it is recognised the feasibility of one-nucleotide loops across the minor grooves. A model building study infact reveals the stereochemical feasibility of one nucleotide loop across the minor groove but not across the major groove. The study further shows that the loop bases can participate in both base pairing and stacking interactions (unpublished results).

Another evidence deserving a comment is the higher stability of the six CCC-block repeats (KRC6 and KRM6) than that of the four CCC-block HTC4. This could be attributed to the presence in the former of either additional stabilising interactions between the cytosines of the two neighbouring loops or, alternatively, to the occurrence of one nucleotide loops across minor grooves, since it is known that, at least for hairpins with WC helical stems, shorter loops are associated with higher stability (26,27), provided that minimal steric requirements are fulfilled.

The evidence for the intramolecularly folded structure is also consistent with the single band migration in PAGE experiments. The minor amounts of slow moving bands that are present in the cases of KRC6 and KRM6, but not for HTC4, may be indicative of a number of multimeric forms based on i-motif caused by the presence of six CCC blocks as argued above.

In view of the evidences supporting the intramolecularly folded i-motif for the telomeric type repeats (24,25), it is imperative to assign to this structure the CD spectrum (Fig.1) exhibited by HTC4 at slightly acidic pH values. The CD spectrum is characterised by a strong positive band near 285 nm and a smaller negative one around 260 nm with a cross over near 270 nm. Since essentially the same type of spectrum is found also for the other two fragments KRC6 and KRM6, it appears reasonable, in line with other evidences, to assign the same type of structure to them too. It is noteworthy that this spectrum is very similar to those reported for a number of cytosine containing oligo- and polynucleotides at acidic pH: poly d(C) (14−16), poly $d(CT)_n$ (17), $d(C_8)$ (14), $d(CCTT)_5$ and $d(AACC)_5$ (19), and $d(CCATT)_n$ and $d(AATCC)_n$ (20). A number of structures have been proposed for these molecules, from the canonical parallel stranded double helix of poly(dC):poly(dC^+) to the $C:C^+$ base paired duplexes with extra thymines looped out (17). In the light of the present results and also of those recently reported (22−25) it is tempting to reconsider at least partially the interpretation of those CD data, since many of the above sequences could

conceivably adopt inter- or intra-molecular i-motif structures, in particular those with at least two cytosine repeats. However, it appears more demanding to propose a similar type of structure for $(CT)_n$ alternating sequences.

As noted earlier the sequence KRC6 having six CCC blocks forms a part of the 166 b.p. promoter fragment of the c-ki-*ras* proto-oncogene. It has been shown that the 30 b.p. stretch represented by KRC6 is necessary for optimal transcription, and deletion constructs of this repress promoter activity (6). The adjudication of the sequence KRC6 in terms of intramolecular i-motif structure points to the possible role of this structure in gene regulation. Although it has been shown (28) that an insert of the entire promoter in plasmid under superhelical and acidic conditions takes the intramolecular H-DNA form, the latter is also a necessary requirement for the formation the i-motif structure. It would be interesting to look for conditions that would facilitate the formation of intramolecular i-motifs in plasmids, as they might afford another structure-based means to influence gene regulation. The unpaired C-strand in intramolecular triplexes, especially those with predominant G*G:C triads (29) might offer a good system to seek for i-motifs. Infact the possible coexistence of i-motif and the G-quadruplex structure can be inferred from the existing data in the literature (30,31).

CONCLUSIONS

It is likely that i-DNA will be a recurring secondary structural moiety in C_n-block DNA repetitive sequences. Well formed G-quadruplex and i-motif structures as those assumed by the complementary strands of telomeric repeats, when mixed, form WC duplexes although with slow kinetics (unpublished results). While it may seem that the WC duplex is thermodynamically favoured, it is conceivable that with a little help from external factors, the individual strands can fold into their own characteristic structures. It is well established that the vast majority of eukaryotic telomeric sequences conform to a generally regularly repeating G_n (and C_n) block: in particular all vertebrates appear to share the (TTAGGG) (and CCCTAA) repeat. As clearly pointed out by Leroy *et al.* (25) in concluding their paper, it is suggestive to relate the fact that the telomeric repeat has been so strongly conserved through evolution and the observation that both the G-block repeat and its complementary C-block can assume unusual structures of their own.

ACKNOWLEDGEMENTS

This work has been financially supported by the Italian Ministry of University and Research and CNR. One of the authors (N.Y.) has carried out this work with the support of ICTP, Trieste (Italy).

REFERENCES

1. Blackburn, E.H., and Szostak, J.W. (1984) A. Rev. Biochem. 53,163–194.
2. Weiner, A.M. (1988) Cell 22, 155–157.
3. Shen, C.-K.J. (1983) Nucleic Acids Res.11, 7899–7910.
4. McKeon, C., Schmidt, A. and Crombrugghe, B.A. (1984) J. Biol. Chem. 255, 6636–6640.
5. Cooney, M., Czernuszewicz, G., Postel, E.H., Flint, S.J. and Hogan, M.E. (1988) Science 241, 456–459.
6. Hoffman, E.K., Trusko, S.P., Freeman, N., and George, D.L. (1990) Mol. Cell. Biol. 7, 2592–2596.
7. Tautz, D. and Renz, M. (1984) J. Mol. Biol. 172, 229–235.
8. Brown, F., Newman, J., Stott, J., Porter, A., Frisby, D., Newton, C., Carey, N. and Fellner, P. (1974) Nature 251, 342–344.
9. Michel, D., Gillet, G., Volovitch, M., Pessac, B., Calothy, G. and Brun, G. (1989) Oncogene Res. 4, 127–136.
10. Sen, D. and Gilbert, W. (1988) Nature 334, 364–366.
11. Williamson, J.R., Raghuraman, M.K. and Cech, T.R. (1989) Cell 59, 871–880.
12. Sundquist, W.I. and Klug, A. (1989) Nature 342, 825–829.
13. Hoffman, E.K., Trusko, S.P., Murphy, M. and George, D.L. (1990) Proc. Natl. Acad. Sci. 87, 2705–2709.
14. Gray, D.M. and Bollum, F.J. (1974) Biopolymers 13, 2087 -2102.
15. Marck, C., Thiele, D., Schneider, C. and Guschlbauer, W. (1978) Nucleic Acids Res. 5, 1979–1996.
16. Thiele, D., Marck. C., Schneider, C., and Guschlbauer,W. (1978) Nucleic Acids Res. 5, 3679–3695.
17. Gray, D.M., Vaughan, M., Ratliff, R.L. and Hayes, F.M. (1980) Nucleic Acids Res. 8, 3695–3707.
18. Gray, D.M., Cui, T., and Ratliff, R.L. (1984) Nucleic Acids Res. 12, 7565–7580.
19. Edwards, E.L., Ratliff, R.L. and Gray, D.M. (1988) Biochemistry 32, 6019–6031.
20. Edwards, E.L., Patrick, M.H., Ratliff, R.L., and Gray, D.M. (1990) Biochemistry 29, 828–836.
21. Ahmed, A. and Henderson, E. (1992) Nucleic Acids Res. 20, 507–511.
22. Gehring, K., Leroy, J.L. and Gueron, M.(1993) Nature 363, 561–565.
23. Leroy., J.L., Gehring, K., Kettani, A. and Gueron, M. (1993) Biochemistry 32, 6019–6031.
24. Ahmed, S., Kintanar, A. and Henderson, E. (1994) Structural Biology 1, 83–88.
25. Leroy., J.L., Mergny, J.L., Gueron, M. and Hélène, C. (1994) Nucleic Acids Res. 22, 1600–1606.
26. Orbons, L.P.M., van der Marel, G.A., van Boom J.H., and Altona, C. (1986) Nucleic Acids Res. 14, 4187–4195.
27. Xodo, L.E., Manzini, G., Quadrifoglio, F., van der Marel, G.A. and van Boom, J.H. (1988) Biochemistry 27, 6321–6326.
28. Pestov, D.G., Dayn, A., Siyanova, E.Y., George, D.L. and Mirkin, S.M. (1991) Nucleic Acids Res. 19, 6527–6532.
29. Dayn, A., Samadashwily, G.M. and Mirkin, S.M. (1992) Proc. Natl. Acad. Sci. 89, 11406–11410.
30. Lyamichev, V.I., Mirkin, S.M., Danilevskaya, O.N., Voloshin, O.N., Balatskaya, S.V., Dobrynin, V.N., Filippov, S.A. and Frank-Kamenetskii, M.D. (1989) Nature 339, 634–637.
31. Voloshin, O.N., Veselkov, A.G., Belotserkovskii, B.P., Danilevskaya, O.N., Pavlova, M.N., Dobrynin, V.N. and Frank-Kamenetskii, M.D. (1992) J. Biomol. Struct. Dyn. 9, 643–652.