VOLUME 117, NUMBER 35
SEPTEMBER 6, 1995

JOURNAL
OF THE
AMERICAN
CHEMICAL
SOCIETY

Intramolecular Folding of Pyrimidine Oligodeoxynucleotides into an i-DNA Motif

Jean-Louis Mergny,*,† Laurent Lacroix,† Xiaogang Han,‡ Jean-Louis Leroy,‡ and Claude Hélène†

Contribution from the Laboratoire de Biophysique, Muséum National d'Histoire Naturelle INSERM U201, CNRS UA481, 43, rue Cuvier 75005 Paris, France, and Groupe de Biophysique, Ecole Polytechnique, URA D1254 CNRS 91128 Palaiseau, France

Received April 3, 1995[⊗]

Abstract: At slightly acidic or even neutral pH, oligodeoxynucleotides which include a stretch of cytidines form a tetrameric structure in which two parallel-stranded duplexes have their hemiprotonated C.C+ base pairs face to face and fully intercalated, in a so-called i-motif, first observed for d-**TCCCCC**. Cytosine-rich pyrimidine oligodeoxynucleotides, 14−30 bases long, can form an intramolecular "i-motif", in a manner similar to that observed in strands of telomeric repeat sequences (Leroy et al. *Nucleic Acids Res.* **1994**, *22*, 1600−1606). This intramolecular i-motif can accommodate as few as eight cytosines (four intercalated C.C+ base pairs). Evidence for intramolecular pairing was provided by thermal denaturation, nondenaturing gel electrophoresis, and gel filtration experiments and was characterized by specific NMR cross-peaks. The stability of a $C.C^+$ base pair is very high at pH 5 and still detectable at neutral pH. Formation of the i-motif, whether intramolecular or intermolecular, should be taken into account as a competing structure in the design of triplex-forming oligodeoxynucleotides.

Introduction

It is well-known that cytosine-rich polymers may adopt a non-B structure at acidic pH. The formation of $C.C^+$ base pairs was originally reported 30 years ago in crystals of acetyl cytosine[1] and later in polydeoxy- or polyribonucleotides.[2−6] The proposed hemiprotonated base pair is displayed in Figure 1A. Solution studies of polydeoxycytidylic acid have pointed to the importance of hemiprotonation. Fiber diffraction studies resulted in a model of a double-helical structure with parallel strands and hemiprotonated $C.C^+$ base pairs in which all bases are in the anti conformation. Overall, the results showed that DNA sequences containing a significant stretch of oligo-C could adopt complex pH-dependent conformations. These structures were generally thought to be double-stranded. Recently, the structure and stoichiometry of the complex formed by the hexamer d-**TCCCCC** has been described.[7,8] This so-called *i-motif* is a tetramer of equivalent strands, which presents the novel feature of intercalated $C.C^+$ base pairs of two-parallel-stranded duplexes. Characteristic inter-residue NOESY patterns of this i-motif include H1′-H1′ cross-peaks and H2′/H2″ amino

† Laboratoire de Biophysique.
‡ Groupe de Biophysique.
* To whom correspondence should be addressed: fax (33-1) 40 79 37 05, e-mail: faucon@mnhn.fr.
§ Keywords: oligonucleotide; cytosine; cytidine; 5-methyl cytosine; protonated cytidine; i-motif; telomer; tetrad; hemiprotonated base pair, $C.C^+$ base pair; unusual DNA structure; parallel-stranded DNA.
⊗ Abstract published in *Advance ACS Abstracts*, August 15, 1995.

(1) Marsh, R. E.; Bierstedt, R.; Eichhorn, E. L. *Acta Crystallogr.* **1962**, *15*, 310−316.
(2) Akinrimisi, E. O.; Sander, C.; Tso, P. O. P. *Biochemistry* **1963**, *2*, 340−344.
(3) Langridge, R.; Rich, A. *Nature* **1963**, *198*, 725−728.
(4) Inman, R. B. *J. Mol. Biol.* **1964**, *9*, 624−637.
(5) Fasman, G. D.; Lindblow, C.; Grossman, L. *Biochemistry* **1964**, *3*, 1015−1021.
(6) Guschlbauer, W. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *57*, 1441−1448.
(7) Gehring, K.; Leroy, J. L.; Guéron, M. *Nature* **1993**, *363*, 561−565.
(8) Leroy, J. L.; Gehring, K.; Kettani, A.; Guéron, M. *Biochemistry* **1993**, *32*, 6019−6031.

0002-7863/95/1517-8887$09.00/0

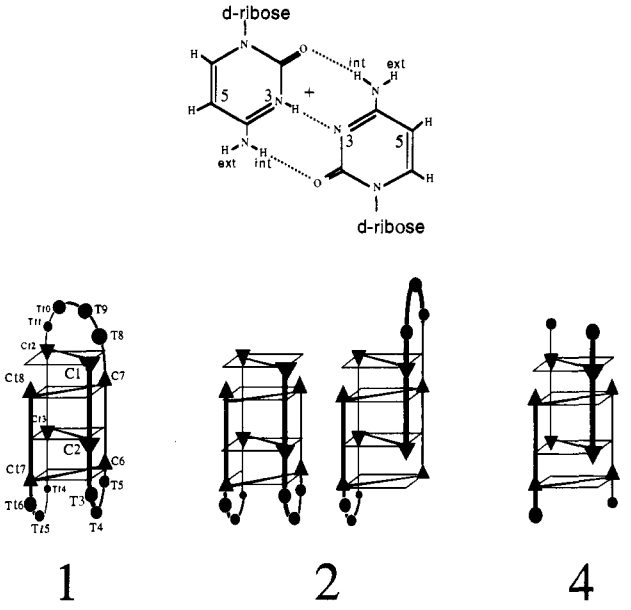

Figure 1. (A) The hemiprotonated C.C$^+$ base-pair. The external and internal amino protons are identified by the "ext" and "int" labels, respectively. (B) Schematic drawing of the i-motif: (1) intramolecular folding of d-**CCTTTCCTTTTCCTTTCC**; (2) two possible dimeric isomers obtained with d-**CCTTTCC** (4) tetramer formation by d-**TCC**. Cytosines and thymines are symbolyzed by triangles and circles, respectively. Four C.C+ base pairs are formed in each case. The i-motif involves the formation of two duplexes, antiparallel to each other. Each duplex is made of two strands in register with parallel orientation.

protons. Individual parallel-stranded duplexes are right-handed and underwound, and the two duplexes are "zipped together" in an antiparallel fashion (Figure 1B). Crystallographic data on the d-**CCCC** and d-**CCCT** oligodeoxynucleotides confirmed this structure.[9,10] The molecule is flat with two very wide grooves and two very narrow grooves. The molecules twist slowly in a right handed manner, and the distance between the nearest base pairs is only 3.1 Å. Such distance, which is lower than B-DNA (3.4 Å), was already reported for poly d(C) 30 years ago.[11] A strand carrying four copies of the cytosine-rich telomeric repeat may form an intramolecular i-motif.[12,13]

Oligodeoxynucleotides have been widely used in the antisense or triple-helix strategy.[14] Intermolecular triplex formation occurs upon binding of the third strand to the major groove of double-stranded DNA.[15,16] Many of the triple-helices which have been studied to date contain oligopyrimidine third strands, which require protonation of the cytosine residues at the N3 position in order to form the C.G*C$^+$ base triplet. The third strand binds in a parallel orientation with respect to the purine strand of the duplex. Sequence specificity is achieved by Hoogsteen hydrogen bonding interactions of thymine with A·T, and protonated cytosine with G·C Watson-Crick base pairs.[17-21] Thus, the C.C$^+$ base pair of the i-motif and the C·G*C$^+$ base triplet both require protonation of one cytosine at the N3 position. The i-motif could interfere with triplex formation, by trapping the third strand in a stable folded conformation. For these reasons we wanted to investigate the factors that could affect the stability of the i-motif and its competition with triplex formation. In this study, we have determined the stability of the folded form(s) of different single-stranded oligodeoxynucleotides, 11–29 bases in length. Most of these oligodeoxynucleotides, presented in Tables 1–3, were solely composed of pyrimidines, to mimick the strands used in the triplex strategy.

Material and Methods

Nomenclature. The terms stoichiometry, multimer, *n*-mer, etc. refer to the association of several molecules and not to properties of a structure resulting form intramolecular folding. The term i-motif refers to an arrangement of intercalated C.C+ base pairs.[7,8]

The sequences of most of the pyrimidine oligodeoxynucleotides investigated in the following experiments are shown in Tables 1–3. We adopted the following convention for all the oligodeoxynucleotides: the first number indicates the length of the oligodeoxynucleotide; the last number indicates the number of cytosines on each strand. The central letter allows to differentiate different oligodeoxynucleotides with the same length and cytosine content. Thus, 17b8 is a 17-mer containing eight cytosines. Oligodeoxynucleotides of the C_2, C_3, C_4, and C_5 families contain four clusters of respectively 2, 3, 4, or 5 cytosines.

Oligodeoxynucleotides. Unmodified oligodeoxynucleotides were synthesized by Eurogentec (Belgium) on the 0.2 μmol scale or by Oligo express (France) on the 50 nmol scale. They were suspended in 200 μL of bidistilled water, precipitated with 5 vol ethanol in the presence of 0.2 M sodium acetate, and washed with ethanol. The pellet was lyophilized, resuspended in bidistilled water, and then gel-filtrated on a Sephadex Quick spin G25 column. Oligodeoxynucleotides for NMR applications were synthesized on a larger scale (15 μmol) with a Pharmacia gene assembler DNA synthesizer, using the β-cyanoethyl phophoramidite method. Oligodeoxynucleotides were then reprecipitated and resuspended in bidistilled water. Concentration of all oligodeoxynucleotides were estimated by UV absorption at 30 °C, in a pH 8.0 buffer, using the sequence-dependent absorption coefficients given by Cantor & Warshaw.[22] All concentrations were expressed in strand molarity.

UV Absorption Spectrophotometry. Absorbance versus temperature heating and cooling curves were obtained using a KONTRON-UVIKON 940 spectrophotometer. The solutions were vortexed, introduced in quartz optical cells, and overlaid with a thin layer of paraffin oil to prevent evaporation. The optical pathlength was 1 cm. The temperature of the 6-cells holder was regulated by a circulating liquid (80% water–20% glycerol) using a HAAKE cryothermostat and monitored by a thermoresistance immersed in an accompanying cell containing only buffer. Constant heating or cooling rates (dT/dt) were obtained using a HAAKE PG20 temperature programmer. The rate of temperature changes was set between 3 and 15 °C/h. Absorbance and temperature were recorded every 5 or 6 min. Water condensation on the cell walls at low temperatures was

(9) Chen, L.; Cai, L.; Zhang, X. H.; Rich, A. *Biochemistry* **1994**, *33*, 13540–13546.
(10) Kang, C. H.; Berger, I.; Lockshin, C.; Radliff, R.; Moyzis, R.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11636–11640.
(11) Hartman, K. A.; Rich, A. *J. Am. Chem. Soc.* **1965**, *87*, 2033.
(12) Leroy, J. L.; Gueron, M.; Mergny, J. L.; Hélène, C. *Nucleic Acids Res.* **1994**, *22*, 1600–1606.
(13) Ahmed, S.; Kintanar, A.; Henderson, E. *Nature Struct. Biol.* **1994**, *1*, 83–88.
(14) Hélène, C.; Toulmé, J. J. *Biochem. Biophys. Acta* **1990**, *1049*, 99–125.
(15) Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J.-L.; Thuong, N. T.; Lhomme, J.; Hélène, C. *Nucleic Acids Res.* **1987**, *15*, 7749–7760.
(16) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645–650.
(17) de los Santos, C.; Rosen, M.; Patel, D. *Biochemistry* **1989**, *28*, 7282–7289.
(18) Rajagopal, P.; Feigon, J. *Biochemistry* **1989**, *28*, 7859–7870.
(19) Rajagopal, P.; Feigon, J. *Nature* **1989**, *339*, 637–640.
(20) Radhakrishnan, I.; Gao, X.; de los Santos, C.; live, D.; Patel, D. J. *Biochemistry* **1991**, *30*, 9022–9030.
(21) Radhakrishnan, I.; Patel, D. J. *Structure* **1994**, *2*, 17–32.
(22) Cantor, C. R.; Warshaw, M. M. *Biopolymers* **1970**, *9*, 1059–1077.

Table 1. Effect of pH on the Thermodynamic Parameters of the 16b8 d-TCCTCCTTTTCCTCCT and 16m8 Oligodeoxynucleotide (Same Sequence as 16b8 with 5-Methylcytosines Instead of Cytosines)[a]

	16b8				16m8			
pH	$\Delta H°$ (kcal·mol^{-1})	$\Delta S°$ (cal·mol^{-1}·K^{-1})	T_m (°C)	$\Delta G°_{20°C}$ (kcal·mol^{-1})	$\Delta H°$ (kcal·mol^{-1})	$\Delta S°$ (cal·mol^{-1}·K^{-1})	T_m (°C)	$\Delta G°_{20°C}$ (kcal·mol^{-1})
4.4	−38	−120	38	−2.8	−29	−95	38	−1.2
5.6	−49	−155	38	−3.5	−47	−150	40	−3.0
6.4	−50	−169	24.5	−0.8	−43	−144	25.5	−0.8
6.8	−50	−174	16	+0.7	−46	−160	16.5	+0.6

[a] T_m values are given at ±1 °C; $\Delta G°$ values at ± 0.1 kcal/mol; $\Delta H°$ values at ± 3 kcal/mol. Average of three independent experiments. These values were calculated from the denaturation profiles at 265 nm (see Material and Methods).

Table 2. Thermodynamic Parameters of Different C_2 Oligodeoxynucleotides at pH 6.4[a]

name	sequence (5′→3′)	folding	T_m (°C) (pH 6.4)	dT_m/dpH (pH 6.4)	$\Delta G°_{20°C}$ (kcal·mol^{-1})	$\Delta H°$ (kcal·mol^{-1})	$\Delta S°$ (cal·mol^{-1}·K^{-1})
11a8	CCTCCTCCTCC	bi-					
11b8	TCCCCTCCCCT	bi-					
14a8	CCTCCTTTTCCTCC	intra-	22.0	−20	−0.29	−51	−174
14b8	CCTTCCTTCCTTCC	intra-	9.5 ± 2	−23 ± 2	+0.89	−25 ± 5	−89
15a8	CCTCCTTTTCCTCTC	intra-	3.5 ± 3	−29 ± 4	+1.17	−20 ± 7	−70
16a8	CCTTCCTTTTCCTTCC	intra-	21.5	−21	−0.29	−50	−169
16b8	TCCTCCTTTTCCTCCT	intra-	24.5	−20	−0.84	−50	−169
16m8	T*C*CT*C*CTTTT*C*CT*C*CT	intra-	25.0	−20	−0.81	−43	−144
17a8	CCTTTTCCTCCTTTTCC	intra-	17.5	−19	+0.55	−54	−185
17b8	CCTTTCCTTTCCTTTCC	intra-	20.5	−18	−0.07	−51	−175
18a8	CCTTTCCTTTTCCTTTCC	intra-	23.5	−18	−0.74	−61	−206
10a4	CTCTTTTCTC	none					
12c6	CCTCTTTTCCTC	none					

[a] Except otherwise noted, T_m values are given at ±1 °C; $\Delta G°$ values at ± 0.1 kcal/mol; $\Delta H°$ values at ± 3 kcal/mol. (Average of three measurements). Folding was shown to be intramolecular from concentration-independent denaturation profiles (in the range 0.5−15 μM), except for the first two oligodeoxynucleotides, 11a8 and 11b8, which showed a concentration-dependent hysteresis. T_m values for these two oligodeoxynucleotides could not be determined at low ionic strength, because of very slow association kinetics. Thermal denaturation profiles for bimolecular processes could only be analyzed in different conditions, i.e., at higher ionic strength (0.3 M NaCl, as in Figure 9). The last two oligodeoxynucleotides, 12c6 and 10a4 showed no transition, and thus, no parameters could be determined. All values were calculated from the denaturation profiles at 295 nm. Identical results were obtained from the denaturation curves recorded at 265 nm.

Table 3. Thermodynamic Parameters of C_n ($2 \leq n \leq 5$) Oligodeoxynucleotides at pH 6.0[a]

name	sequence (5′→3′)	folding	T_m (pH 6) (°C)	$\Delta G°_{37°C}$ (kcal·mol^{-1})	$\Delta H°$ (kcal·mol^{-1})	$\Delta S°$ (cal·mol^{-1}·K^{-1})	hypochromism (%)
17b8	$(CCTTT)_3CC$	intra-	27	+1.65	−51	−170	9
21a12	$(CCCTTT)_3CCC$	intra-	45	−1.56	−72	−226	12
21n12	$(CCCTAA)_3CCC$	intra-	39	−0.45	−67	−215	13
25p16	$(CCCCTTT)_3CCCC$	intra-	54.5	−4.92	−93	−283	11
29a20	$(CCCCCTTT)_3CCCCC$	intra-	59.5	−7.57	−113	−340	13

[a] (Average of two to three independent experiments; see Table 2 for conditions). Folding was shown to be intramolecular from concentration-independent denaturation profiles (in the range 0.5−15 μM). All values were calculated from the denaturation profiles at 295 nm. Identical results were obtained from the denaturation curves recorded at 265 nm. Hypochromism was calculated at 265 nm.

prevented by gently blowing a stream of dry air in the cell compartment.

Analysis of the Reversible Denaturation Curves. From the classical relation $\Delta G° = -RT \ln(K) = \Delta H° - T\Delta S°$ one can deduce that $\ln(K) = -(\Delta H°/R)*(1/T) + (\Delta S°/R)$. Thus, $\ln(K)$ can be expressed as a linear function of $1/T$. The equilibrium constant K can be written as $K = \theta/(1 - \theta)$ for an *intra*molecular equilibrium (θ is the fraction of folded oligodeoxynucleotide) or $K = \theta/(2c_o(1 - \theta)^2)$ for a *bi*molecular equilibrium involving two identical strands, respectively (c_o is the total strand concentration). θ (in the range 0.04−0.96), and thus K, can be easily determined at each temperature from the denaturation profiles. $\ln(K)$ is then plotted as a function of $1/T$ (in Kelvin). If the all or none model is in agreement with the experimental data, one should obtain a straight line (r routinely ≥0.998), whose slope is $-\Delta H°/R$ and y-intercept $\Delta S°/R$ (see Figure 5B for an example).

At neutral pH (6.8−7.2) the thermal dissociation curves (heating) of some oligodeoxynucleotides, which belonged to the C_3, C_4, and C_5 families, were shifted toward higher temperatures with respect to the association (cooling) curves, and an hysteresis phenomenon was obtained. Such a behavior, which is the result of slow association and dissociation kinetics, has already been described for triple helix formation.[23] This hysteresis was concentration independent, indicating a slow intramolecular process. Quantitative analysis of the kinetic parameters of i-motif formation is beyond the scope of this paper, and will be presented elsewhere.

Gel Filtration. 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 oligodeoxynucleotides as previously described.[8] Oligodeoxynucleotides were eluted with a 20 mM sodium acetate, 300 mM NaCl buffer, at pH 4.7. Lower salt concentration increased DNA−matrix interactions and decreased the resolution. The optimal flow rate was determined to be 0.4 mL/min, giving an elution time of 4 to 6 min for a typical experiment.

Nondenaturing Gel Electrophoresis. Oligodeoxynucle-

otides were labeled with T4-polynucleotide kinase (New England Biolabs) and ^{32}P ATP (ICN). The absence of smaller, contaminating species was first checked by 20% polyacrylamide denaturating gel electrophoresis. Radiolabeled oligodeoxynucleotides (10 nM strand concentration) were incubated at 4 °C in presence or in absence of a large excess of an unlabeled oligodeoxynucleotide (2−200 μM strand concentration) in a pH 5.6; 50 mM MES buffer or in a pH 8.3; 50 mM Tris Borate EDTA (TBE) buffer. After incubation for 0.5−100 h, samples were loaded on a nondenaturing 12% polyacrylamide gel. Migration at 4 °C lasted 2−4 h (3W) at 4 °C. The running buffer was recirculated during electrophoresis. Gels were dried and analyzed on a phosphorimager instrument (Molecular Dynamics). The migration of C-rich oligodeoxynucleotides was compared with control oligodeoxynucleotides (oligothymidylates: T_{22}, T_{16}, T_{14}, or T_9).

Nuclear Magnetic Resonance. The d-**CCTTTCCTTTTC-CTTTCC** NMR sample was prepared by lowering the pH of a 6.9 mM solution from neutrality to pH 4.8 with aliquots of HCl. Although no salt was added to the NMR sample, the Na^+ concentration due to the counterions was about 0.13 M.

The NMR experiments were performed on a 360 MHz home-built spectrometer. The JR sequence was used for solvent signal suppression.[24] The NOESY data were obtained on the hypercomplex mode. The data were transferred to an IRIS workstation (Silicon graphics Inc.) and processed with Felix 1.1 (courtesy of Hare Research). Acquisition and processing procedures used in 2D experiments have been described in ref 12. The chemical shifts are referenced to DSS.

To measure the exchange rate of amino protons, a concentrated H_2O solution of the hemi-protonated structure (about 12 mM) formed by d-**CCTTTCCTTTTCCTTTCC** was 10-fold diluted in D_2O at 0 °C and immediately introduced in the spectrometer. The exchange rate was obtained from spectra recorded at different time intervals.[25]

Mung Bean Nuclease Cleavage. A P^{32} labeled oligodeoxynucleotide (T16, or 16b8) was incubated overnight at 15 °C in a 50 mM sodium acetate, 30 mM sodium chloride, 1 mM zinc sulfate pH 5.5 buffer. 10 units of mung bean nuclease were then added. 7 μL alliquots were taken up at various amounts of time, and the reaction was stopped by addition of 2 mM EDTA and incubation at 95 °C for 5 min. Products of cleavage were analyzed by denaturing PAGE.

Results

NMR Evidence for the Intramolecular i-Motif. The exchangeable proton region of the NMR spectrum of 18a8 (d-**CCTTTCCTTTTCCTTTCC**) displayed in Figure 2 exhibits cytidine imino protons (16−15 ppm) and amino protons (10−8 ppm) characteristic of hemi-protonated C.C$^+$ base pairs.[8,26] The NOESY spectrum in H_2O solution (Figure 3) shows eight partially resolved intraresidue cross-peaks between internal and external amino protons (box A).

For four potential C.C$^+$ pairs (Figure 1), we detect three cytidine imino protons. The missing imino proton probably exchanges too rapidly at −7 °C to be detected. The cytidine imino proton at 15.2 ppm is exchange-broadened and disappears above 0 °C.

The pattern of NOE connectivities observed between amino and imino protons in d-**CCTTTCCTTTTCCTTTCC** differs from that previously observed in the symmetrical structures formed by intercalation of four identical oligo-C strands. In symmetrical C.C$^+$ pairs, the connectivities between the imino

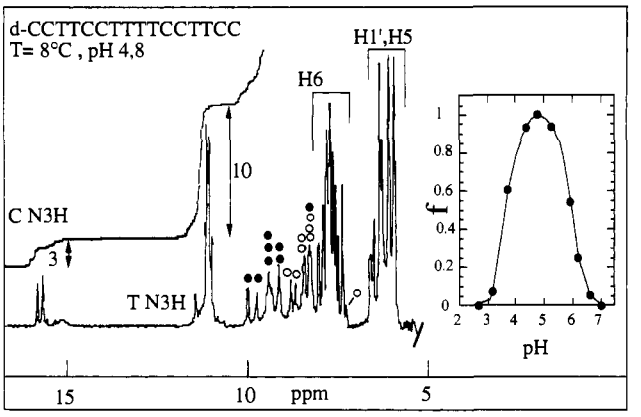

Figure 2. 1D NMR spectrum of the exchangeable proton region of d-**CCTTTCCTTTTCCTTTCC**. By reference to the thymidine imino protons cluster, the area of the cytidine imino region shows that at least 3 C.C$^+$ pairs are formed. The internal and the external amino protons, identified by the experiment displayed in Figure 3 are labelled (black circles) and (open circles) respectively. The FID was multiplied by a shifted sine bell function[68] and the spectrum was corrected for the frequency response to JR excitation.[69] The inter-FID delay was 5 s. The fraction, f, of the hemi-protonated structure which is displayed as a function of pH in the insert was obtained by measuring the area of the 15.8 and 15.6 ppm imino proton peaks vs pH. Experimental conditions: H_2O 90%, pH 4.8, T = 7 °C. Strand concentration 6.9 mM.

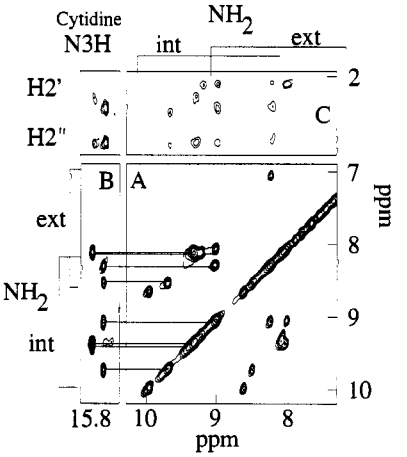

Figure 3. NOESY spectrum of d-**CCTTTCCTTTTCCTTTCC** in 90% H_2O (mixing time 250 ms). Amino-amino and amino-imino connectivities characterize non symmetrical hemi-protonated C.C$^+$ pairs. Note the upfield shift of the amino protons of one cytidine. Unusual H2′/H2″-amino proton cross peaks are characteristic of the i-motif.[26] Experimental conditions: pH 4.8, T = 7 °C, strand concentration: 6.9 mM.

proton and two identical amino groups give rise to two cross peaks.[26] In the structure formed by d-**CCTTTCCTTTTC-CTTTCC**, the four cross-peaks which connect the 15.6 ppm imino proton to amino protons indicates a nonsymmetrical base-pair. Due to overlaps in the amino proton region, the 15.8 ppm imino proton shows only two cross-peaks with intensities twice larger: one with two unresolved external amino protons and one with two unresolved internal amino protons.

The amino protons of one cytidine (8.3 and 7.1 ppm, respectively) exhibit an unusual upfield shift. Nevertheless, the observation that the position of these protons is independent of pH in a range (pH 6 to 3), where protonation of the N3 position

(23) Rougée, M.; Faucon, B.; Mergny, J. L.; Barcelo, F.; Giovannangéli, C.; Garestier, T.; Hélène, C. *Biochemistry* **1992**, *31*, 9269−9278.
(24) Plateau, P.; Guéron, M. *J. Am. Chem. Soc.* **1982**, *104*, 7310−7311.
(25) Guéron, M.; Leroy, J. L. *Methods Enzymol.* **1995**, *in press*,
(26) Leroy, J. L.; Guéron, M. *Structure* **1995**, *3*, 101−120.

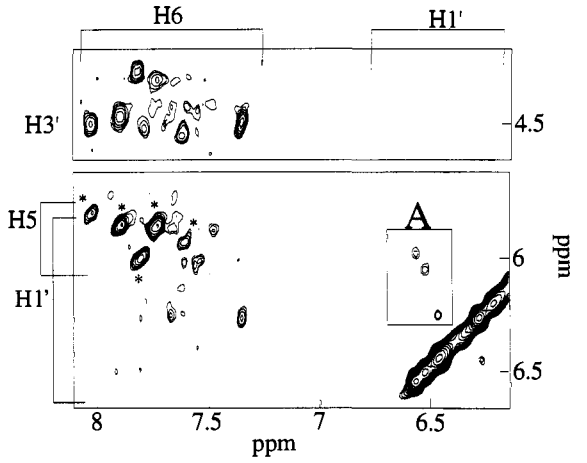

Figure 4. NOESY spectrum in D_2O (mixing time 50 ms) at 7 °C. H5-H6 cross peaks, assigned from a TOCSY experiment are indicated by stars. The strong H1′-H1′ cross-peaks (box "A") are typical of the structure formed by intercalation of two hemi-protonated oligo-C duplexes. Strong H6-H3′ (upper panel) cross-peaks show that several nucleosides have a sugar puckering in the 3′ endo conformational range. Weak H6-H1′ cross-peaks indicates an anti-conformation. Experimental conditions: pH 4.8, $T = 7$ °C, Strand concentration: 6.9 mM.

would shift the amino protons by about 1.5 ppm,[27] suggests that the N3 position is protected and that this cytidine also belongs to a C.C+ pair.

Melting occurs at 42 °C in the conditions of the NMR experiment. The intensity of the cytidine imino proton peaks at 15.8 and 15.6 ppm decreases during the melting transition as the fraction of the hemi-protonated structure. The NMR titration displayed in the insert of Figure 2 shows that the stability of the hemi-protonated structure is maximal at pH 4.8. Proton line widths and chemical shifts in the hemi-protonated structure are independent on pH. The random coil species, which dominates below pH 3.5 and above pH 6 in the condition of the NMR experiment, is readily characterized on the NMR spectra by its amino protons which exhibit the same chemical shift vs pH than those of the amino protons in the cytidine monomer.[27]

The TOCSY spectrum (data not shown) and the NOESY spectrum in D_20 (Figure 4) show eight unresolved H5/H6 cross-peaks (two peaks with intensities corresponding to a single H5/H6 couple and three peaks with intensities twice larger).

The i-motif stucture may be recognized by characteristic cross-peaks which are absent in B DNA spectra. H1′-H1′ cross peaks reflect the short interstrand distance accross the minor groove[7,8] and cross-peaks between H2′/H2″ and amino protons connect nucleosides belonging to anti-parallel strands across the major groove.[7] The observation of these cross-peaks on the NOESY spectra of d-**CCTTTCCTTTTCCTTTCC** establishes that this oligomer adopts the i-motif. H1′-H1′ cross-peaks appear in the 50 ms NOESY spectrum (box "A" of Figure 4). Most of the amino protons are connected to H2′/H2″ protons in the NOESY spectrum displayed in Figure 3. The strong H3′-H6 cross peaks (upper region of Figure 4) indicate that several nucleosides adopt a N sugar puckering. The weak H1′-H6 cross-peaks show that the nucleosides are *anti*.

Proton Exchange. Using a series of NMR spectra recorded after dilution of a concentrated protonated d-**CCTTTCCTTTTC-CTTTCC** sample into D_2O, we measure that 5 ± 1 internal (i.e., hydrogen-bonded) amino protons exchange in 30 ± 10 min at 0 °C (data not shown). The internal amino proton at 8.3 ppm, the external amino protons and the imino protons are already exchanged in the first spectrum recorded one minute after the solvent change. Although the internal amino protons of d-**CCTTTCCTTTTCCTTTCC** exchange more rapidly than those of the intercalated motif formed with longer oligo-C tracts ($\tau_{ex} = 3$ h for the internal amino protons of the [d-**TCCCCC**]$_4$ tetramer[8]) the difference observed between the exchange rate of internal and external amino protons is typical of i-motif structure[8,12] and contrast with the case of cytidine in B[28] and Z DNA[29] whose internal and external amino protons exchange at comparable rates.

Typical Denaturation Profile of a C-Rich Oligodeoxynucleotide. Dissociation of the i-motif leads to an hyperchromism at 265 nm.[12,13,30] The 18a8 oligodeoxynucleotide had a T_m of 23.5 °C at pH 6.4. No signal was observed in a 7.8 cacodylate buffer, showing that this phenomenon was pH-dependent. Similar results were observed with other pyrimidine oligodeoxynucleotides, 14−18 bases long (Table 2). An example of a denaturation profile, obtained for the 16b8 pyrimidine oligodeoxynucleotide d-**TCCTCCTTTTCCTCCT** is presented on Figure 5A. This denaturation profile is characterized by a sharp increase of absorbance at 265 nm (open circles), whereas an inverted transition is observed at 295 nm (black triangles). Thus, dissociation of the structure ($T_m = 16$ °C) led to an increase of absorbance at 265 nm. Both curves could well be analyzed as an all-or-none intramolecular phenomenon, as shown in Figure 5B. Thermodynamic parameters determined form the profiles at 265 and 295 nm were identical within experimental error.

Absorption Spectrum of 16b8 Oligodeoxynucleotide as a Function of pH. Absorption spectra of the 16b8 oligodeoxynucleotide (16b8; d-**TCCTCCTTTTCCTCCT**) were recorded at different pHs and at high temperature where all structures were unfolded (Figure 6A). At a fixed temperature of 58 °C, an isosbestic point was obtained at 263 nm: at this wavelength, the absorption coefficient is independent of the protonation state of the cytosines. Depending on the oligodeoxynucleotide primary sequence, this isosbestic point varied between 263 and 266 nm (not shown). The temperature of 58 °C was high enough to insure that the oligonucleotide was unfolded at all pH values. A d-T_{16} control oligodeoxynucleotide showed not variation of its absorption spectrum over a pH range of 3.6−7.2. From absorbance spectra, the fraction of protonated cytosines could be calculated at each pH. These data were analyzed assuming independent protonation of the eight cytosines and gave a pK_a of 4.8 ± 0.2 for each N3 group (not shown). A similar pK_a value was obtained for the 22d10 oligodeoxynucleotide d-**TTCCCTCCTTTTTTCCCTCCTT** with an isosbestic point at 265 nm. This pK_a value is in good agreement with the pK_a determined for a shorter pyrimidine oligodeoxynucleotide d-**CTTCCTCCTCT**.[31] At higher ionic strength (100 mM NaCl, 10 mM acetate or cacodylate) these curves were shifted to lower pH values giving a pK_a of 4.5 ± 0.2 for the 16b8 oligodeoxynucleotide and 4.2 ± 0.3 for the 22d10 oligodeoxynucleotide. We did not investigate the effect of temperature on the pK_a of cytosine. It was previously reported that the pK_a of cytosine decreased by 0.1 unit when temperature increased by 10 °C.[31] At low temperature (1 °C), all absorbance spectra were nearly superimposed (Figure 6B),

(27) Raszka, M. *Biochemistry* **1974**, *13*, 4616−4622.
(28) Kettani, A. Thesis, Paris VII, 1993.
(29) Kochoyan, M.; Leroy, J. L.; Guéron, M. *Biochemistry* **1990**, *29*, 4799−4805.
(30) Manzini, G.; Yathindra, N.; Xodo, L. E. *Nucleic Acids Res.* **1994**, *22*, 4634−4640.
(31) Manzini, G.; Xodo, L. E.; Gasparotto, D.; Quadrifoglio, F.; van der Marel, G. A.; van Boom, J. H. *J. Mol. Biol.* **1990**, *213*, 833−843.

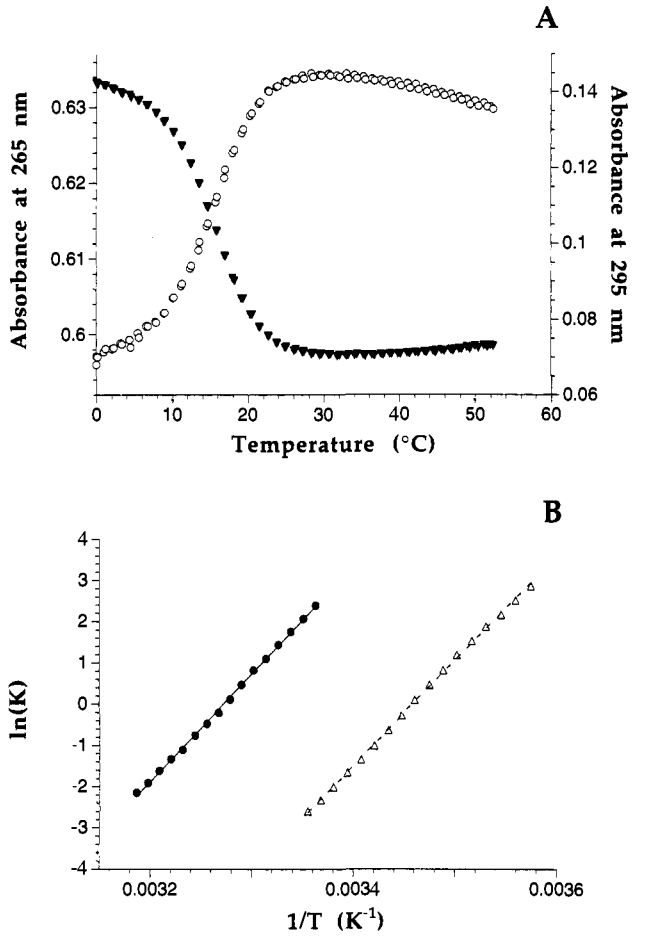

Figure 5. (A) Denaturation profiles obtained at pH 6.8 for the 16b8 d-TCCTCCTTTTCCTCCT oligodeoxynucleotide at two different wavelengths: 265 nm (open circles) and 295 nm (black triangles). Left vertical axis: absorbance values at 265 nm; right vertical axis: absorbance values at 295 nm. Experimental conditions: 10 mM cacodylate buffer. The profiles were reversible at this pH, as indicated by the surimposition of one denaturation and one renaturation profiles for each wavelength. (B) Thermodynamic analysis of the denaturation profile: plot of $\ln(K)$ versus $1/T$ at two different pH (triangles: pH 6.8; circles: pH 6).

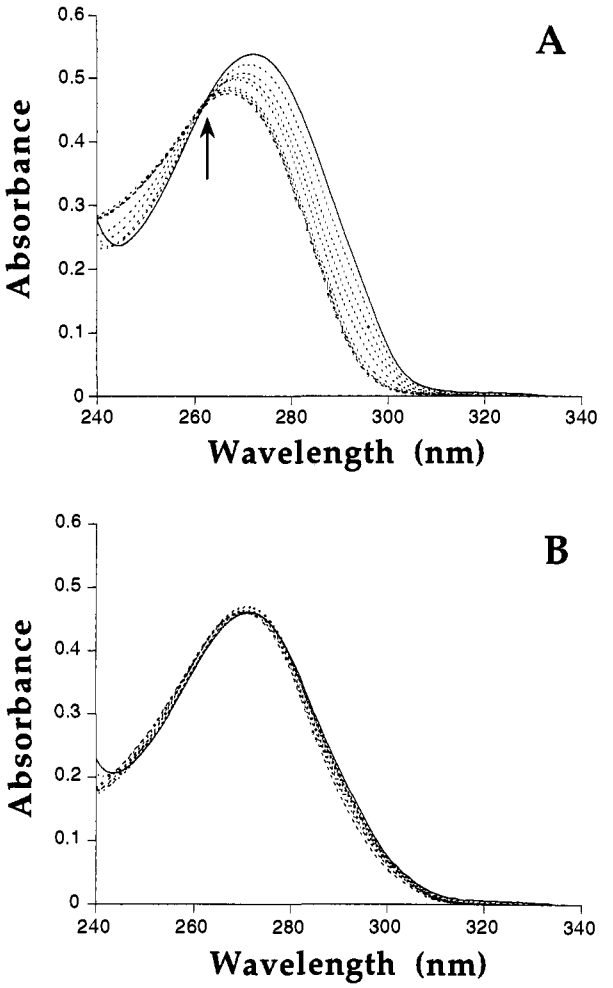

Figure 6. Absorption spectrum of d-TCCTCCTTTTCCTCCT (16b8) as a function of pH. (A) (upper panel) T = 58 °C; (B) (lower panel) T = 1 °C. All spectra were recorded in a 10 mM cacodylate (pH 5.6–6.8) or acetate (pH 3.6–5.6) buffer. The isosbestic point at 265 nm (for T = 58 °C) is indicated by an arrow. Solid line: absorbance profile at pH 3.6. Dotted lines: absorbance profiles at regular pH intervals (0.4 units) between 4 and 6.8.

suggesting that the folded conformation involved the same number of protonated cytosines, in the pH range 3.6–6.8.

These data explain the inverted denaturation curves obtained at 295 nm (Figure 5A). The shorter wavelength (265 nm, close to the isosbestic point of the pH titration) gives an information which is independent of the protonation state of the cytosines. The longer wavelength (295 nm) should tell us whether the formation of a structure requires protonation/deprotonation of some cytosines: at this wavelength, the difference between protonated and nonprotonated cytosine is maximum, as infered from absorbance spectra at different pH values (Figure 6A).

Effect of pH. As shown in Figure 7, the thermal stability of the structure was pH-dependent. The melting temperature was maximum at a pH value close to the pK_a of cytosine (4.8 in a low-salt buffer), in good agreement with the NMR pH titration presented in Figure 2 (inset). In the 5.2 to 7.2 pH range, the T_m of the structure was a linear function of pH, as previously observed for a telomer oligodeoxynucleotide.[12] An increase of pH of 1 unit led to a decrease of the melting temperature of 20 °C or more. Such a pH dependence was previously observed for cytosine-rich polyribonucleotides[2] and polydeoxyribonucleotides[4] as well as oligodeoxynucleotides.[12]

The effect of pH was not limited to a modification of the melting temperature. The shape of the transition, especially at

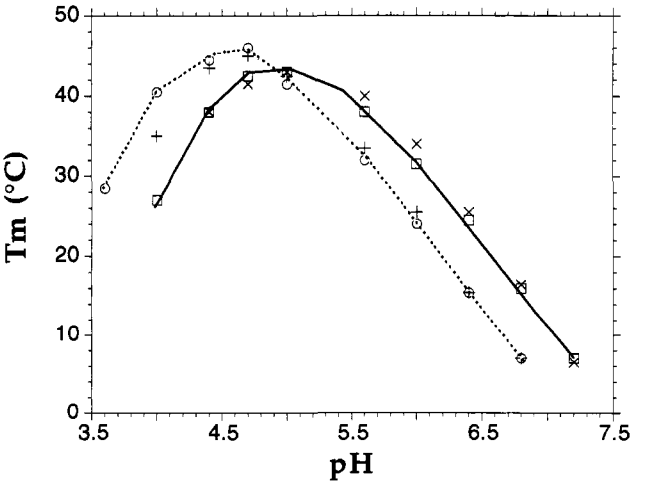

Figure 7. Effect of pH and ionic strength on T_m value for the 16b8 and 16m8 oligodeoxynucleotides in a 10 mM acetate or cacodylate buffer. squares: 16b8; crosses: 16m8; circles: 16b8 + 0.1 M NaCl; vertical crosses: 16m8 + 0.1 M NaCl.

295 nm, was also pH dependent. At pH 5.2 or more, the absorbance at 295 nm decreased upon dissociation of the complex (see for example the profile at pH 6.8, on Figure 5A). This melting profile was a mirror-image of the melting profile

recorded at 265 nm. At pH 4.8, the absorbance at 295 nm was almost independent of temperature. At pH 4.4 or lower, the absorbance at this wavelength increased when the complex was unfolded (not shown). These results are in agreement with the requirement for protonation of half of the cytosines. At pH = pK_a, no net protonation is required, whereas at pH < pK_a, deprotonation of some of the cytosines is required (see Discussion).

The variation of absorbance at 265 nm was always positive upon dissociation of the complexes at all pH values. However, the slope of the transition, and thus the $\Delta H°$ of the reaction, was pH-dependent (see Table 1 for a recapitulation of the results of melting experiments, and Figure 5B for an example of thermodynamic analysis of the melting profiles). The formation of this structure was strongly enthalpy driven at near neutral pH: the slope was maximum at pH 6−7, and dropped below pH 5.5. It is instructing to compare the thermodynamic parameters of the 16b8 oligodeoxynucleotide (Table 1) at pH 4.4 and 5.6. The T_m were the same (38 °C), but the calculated $\Delta H°$ and $\Delta S°$ were very different. At pH 4.4, the enthalpy of folding was less negative, and thus less favorable (−29 kcal/mol instead of −49 kcal/mol at pH 5.6). This was in part compensated by a less negative, and thus more favorable, entropy of formation (−155 instead of −174 cal/mol/K; see discussion). Raising the pH above 5.6 did not induce significant changes in $\Delta H°$, whereas the $\Delta S°$ was more and more negative (Table 1, and Figure 5B: the slope of the ln(K) curves at pH 6.0 and 6.8 were identical within experimental error).

The pH value also had an impact on the kinetics of intramolecular folding. At neutral pH (6.8−7.2), pyrimidine oligodeoxynucleotides, containing 12 cytosines or more, did not melt in a reversible fashion. Such a phenomenon was previously reported for the folding of the C-rich strand of human telomeres[12] and were indicative of slow kinetics of folding at neutral pH.

Effect of Ionic Strength. As discussed previously, the pK_a of cytosine was shifted to higher values in a low-salt buffer. Thus, at pH ≥ 4.8, increasing NaCl concentration from 0 to 100 mM led to a destabilization of the structure (Figure 7). Again, no difference between the 16b8 and the 16m8 oligonucleotide was evidenced. Further increase in ionic strength (from 0.1 to 0.3 NaCl) had no effect on the melting temperature. A similar behavior was reported for polycytidylic acid.[2]

We also investigated the effect of various monovalent or divalent cations on the stability of this structure. The addition of 5 mM $MgCl_2$, $CaCl_2$, $ZnCl_2$ or 10 mM LiCl or KCl had no effect on the T_m of the 18a12 oligodeoxynucleotide in a 10 mM cacodylate, 100 mM NaCl, pH 6.4 buffer.

Intramolecular versus Intermolecular Pairing. The electrophoretic mobility of C-rich oligodeoxynucleotides was compared with oligodeoxythymidilates on a nondenaturing gel. At basic pH (8.3), no anomaly of migration was observed (Figure 8B). All C-rich oligodeoxynucleotides nearly co-migrated to a position close to that of a control T_n oligodeoxynucleotide of identical length. This is in clear contrast with the abnormally fast migration observed at pH 5.6 (Figure 8A). One of the oligodeoxynucleotides (see lane 2 on Figure 8A) showed a complex migration profile, with two close bands, suggesting the presence of several intramolecular isomers.

All these phenomena were concentration-independent for most of the long oligodeoxynucleotides (14 mer and above). Addition of increasing amounts of "cold" oligodeoxynucleotide (2 to 200 μM, strand concentration) had no effect on the migration (not shown). Shorter oligodeoxynucleotides (11 mers)

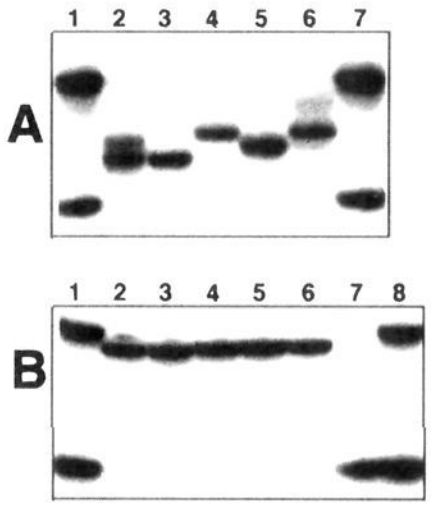

Figure 8. (A) Nondenaturing gel electrophoresis in a pH 5.6 MES buffer, 4 °C. Lanes 1 and 7: (Size markers) mixture of T_{22} and T_9 oligodeoxynucleotides. Lane 2 to 6: (from right to left) 22a12 d-**CCCTCTCCTTTTTTCCCTCTCC**; 22b12 d-**CCCTCCCTTTTTT-TTCCCTCCC**; 22c10 d-TTTT**CCCCCTTTTCCCCC**TTTT; 22d10 d-TT**CCCTCCTTTTTTCCCTCC**TT; and 22e10 d-TT**CCCCCTTTT-TTTTCCCCC**TT oligodeoxynucleotides, 20 nM strand concentration. All samples were diluted to the final desired concentration, heated at 95 °C for 5 min, incubated at the gel temperature for 2 h, and then loaded onto the gel. (B) Non denaturing gel electrophoresis in a pH 8.3 TBE buffer, 37 °C. Lanes 1 and 8: (size markers) mixture of T_{22} and T_9 oligodeoxynucleotides. Lane 7: T_9 oligodeoxynucleotide alone. Lane 2 to 6: (from left to right) 22a12; 22b12; 22c10; 22d10; 22e10 oligodeoxynucleotides, 20 nM strand concentration.

displayed a more complex behavior. The 11a8 oligodeoxynucleotide exhibited two bands on a non-denaturing gel. The fastest species was only present at very low concentration (10 nM). In the 2−200 μM strand concentration range, a slower-migrating species was observed, with an intermediate mobility as compared with the markers.

The nature of the folded form was confirmed by T_m measurements at various strand concentrations. Typical denaturation profiles are presented in Figure 9. Most of the oligodeoxynucleotides showed a concentration-independent profile, shown in Figure 9B for the 17b8 oligodeoxynucleotide. Such behavior was indicative of intramolecular pairing. The 11a8 oligodeoxynucleotide, which is composed of three short loops (each composed of a single thymine) displayed a concentration-dependent profile (Figure 9A). A 5-fold increase in concentration led to an increase of T_m as well as a decrease of the hysteresis (difference between heating and cooling curves). Such concentration dependence is typical of bi- or multi-molecular association. It should be noted that the hysteresis observed in Figure 9A for an intermolecular process is concentration dependent, as compared with the hysteresis phenomenon obtained for the intramolecular folding of long pyrimidine oligodeoxynucleotides at neutral pH. Other oligodeoxynucleotides, such as 16a8 (Figure 9C) showed a slightly concentration-dependent denaturation profile, especially at high ionic strength. At the same concentrations, the profiles were perfectly superimposed in a 10 mM cacodylate buffer (without NaCl). Thus, a decrease in ionic strength had a destabilizing effect on intermolecular base-pairing, as compared to intramolecular folding.

Gel filtration experiments also confirmed the stoechiometry of the complexes. The chromatography performed with an aliquot of d-**CCTTTCCTTTTCCTTTCC** directly collected from the NMR tube showed that 90% of the sample migrates as a monomer (Figure 10A). On the other hand, the elution profile obtained with an aliquot of d-**CCTTCCTTTTCCTTCC** prepared in the same way showed that 60% of this oligomer forms a dimer (Figure 10B). This might explain the poor resolution of the NMR spectrum of this oligomer (data not

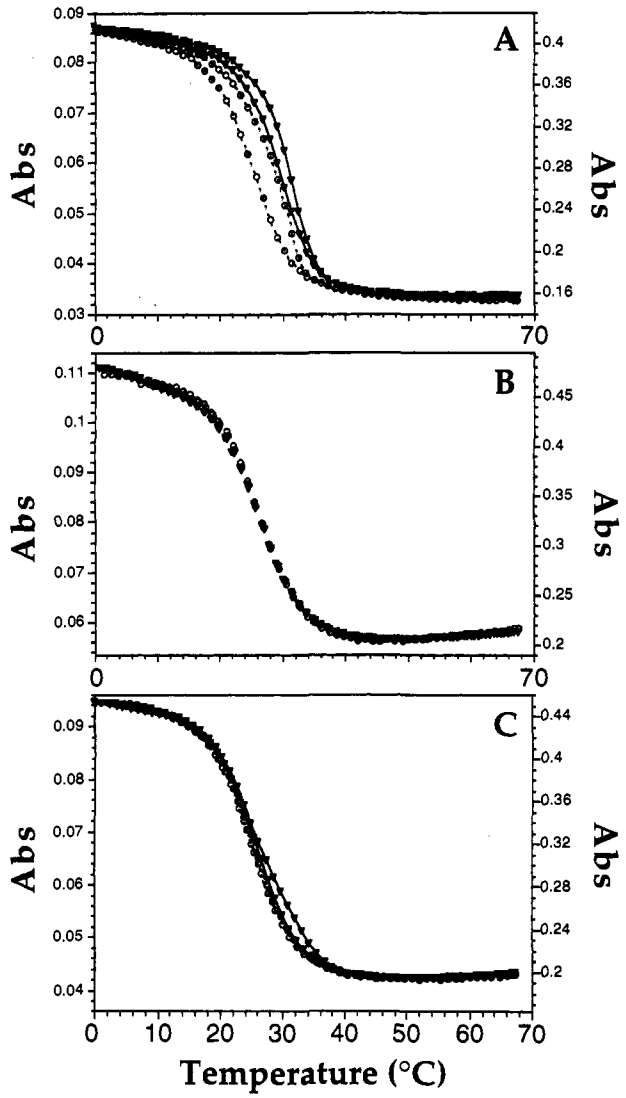

Figure 9. Evidence for intramolecular or intermolecular base pairing. Concentration dependence of the denaturation/renaturation profiles for three different oligodeoxynucleotides at pH 5.6 in a 10 mM cacodylate; 300 mM NaCl buffer. Upper panel 11a8, d-**CCTCCTCCTCC**; Middle panel: 17b8, d-**CCTTTCCTTTCCTTTCC**; Lower panel 16a8, d-**CCTTCCTTTTCCTTCC**. Left vertical axis: absorbance values at 2 μM strand concentration (open circles); Right vertical axis: absorbance values at 10 μM strand concentration (triangles). In each case, the denaturation and renaturation profiles were recorded at 265 nm.

shown). The profile recorded with an aliquot derived from the NMR sample (6 mM) immediately after heating at 100 °C and quenching in ice-water showed only a monomer (Figure 10C). The NMR spectrum of the quenched sample was then nearly identical to that of d-**CCTTTCCTTTTCCTTTCC** (not shown). Further gel filtration experiments performed with aliquots collected from the NMR sample maintained at room temperature after quenching showed that the dimer/monomer proportion reverts in a few hours to the initial composition. In dilute solutions (5 μM) the folded monomer of d-**CCTTCCTTTTC-CTTCC** was stable for days. This is in agreement with thermal denaturation experiments which showed that intramolecular folding was fast at pH < 7, and intermolecular folding slow, as shown by the hysteresis phenomenon presented in Figure 9A. A quick drop in temperature favored the fastest folding. Slow conversion to a more stable structure was then observed when the sample was kept at low temperature.

Thus, in the 0.01−100 μM concentration range and at low ionic strength, the most abundant species resulted from intra-

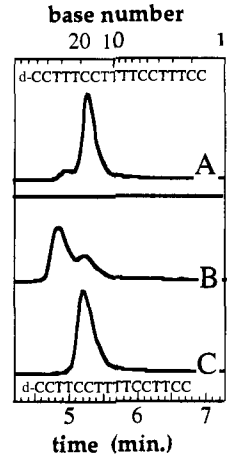

Figure 10. Characterization of the oligomer stoichiometry by gel filtration chromatography. The sample were prepared by dilution of an aliquot of the concentrated solutions used in NMR experiments: [d-**CCTTTCCTTTTCCTTTCC**] = 6.9 mM and [d-**CCTTCCTTTTC-CTTCC**] = 6 mM. (A) The elution profile of d-**CCTTTCCTTTTC-CTTTCC** shows that 90% of this oligomer migrates as a monomer and that 10% forms a dimer. (B) For d-**CCTTCCTTTTCCTTCC**, the proportion of dimer amounts to 60%. (C) The gel filtration profile of an aliquot of the NMR sample of d-**CCTTCCTTTTCCTTCC** obtained immediately after heating at 100 °C and quenching at 0 °C shows only the monomer species. The quenched NMR sample (6 mM) reverts to the same proportion of dimer as in B in a few hours. Experimental conditions: NaCl = 0.3 M, pH 4.7, acetate 20 mM, T = 20 °C.

strand folding except for a few oligodeoxynucleotides (e.g., 11a8). However, one should note that at higher strand concentrations and/or ionic strength, (in the millimolar range for strand concentration compatible with NMR experiments) dimeric structures were observed for some oligodeoxynucleotides (e.g., 16a8). The quantitative extent of dimeric *vs* monomeric folding depends on the primary sequence of the oligodeoxynucleotide (compare Figures 10A and 10B).

To provide us with a better understanding of the parameters that could affect intramolecular folding, the comparison between cytosine-rich oligonucleotides presented in Tables 1−3 was performed at low ionic strength. A summary of the stoechiometry of the complexes is given in Tables 2 and 3. This stoichiometry is valid for strand concentrations below 100 μM.

Substitution of Cytosines by 5-Methylcytosines. The pK_a of the 16m8 oligodeoxynucleotide (d-T***C***CTC***C***TTTT***C***CTC***C***T; ***C*** = 5-methyl cytosine) was determined using the same method as the one presented for the 16b8 oligodeoxynucleotide. The isosbestic point between the protonated and deprotonated forms of 5-methylcytosines was shifted to longer wavelengths (270 nm). No significant difference of pK_a with the non-methylated oligonucleotide was observed (pK_a = 4.8 at low ionic strength, pK_a = 4.55 in a 100 mM NaCl), in good agreement with previous observations on nucleosides[32] and oligodeoxynucleotides.[33] Denaturation profiles of the 16m8 oligodeoxynucleotide were compared with those of 16b8 (an oligodeoxynucleotide having the same sequence, with cytosines instead of 5 methyl cytosines). No significant difference of T_m (±2 °C) was observed between these two oligodeoxynucleotides at all pH values between 4.4 and 7.2 (Figure 7). These results were in good agreement with the pK_a values: the optimal stability was obtained at pH = pK_a. It has recently been shown that the i-motif can accommodate some methylated cytosines.[26,30] The

(32) Hall, R. H. In *The modified nucleosides in nucleic acids*; Colombia University Press: New York, 1971; pp 192−194.
(33) Xodo, L. E.; Manzini, G.; Quadrifoglio, F.; van der Marel, G. A.; van Boom, J. H. *Nucleic Acids Res.* **1991**, *19*, 5625−5631.

present study demonstrates that, at least for the 16b8/16m8 sequence, this substitution has little, if any, effect on the stability of the i-motif.

Sequence Requirements. We have compared several oligodeoxynucleotides of the C_2 family (Table 2). The presence of four repeats of (at least) two cytosines is an absolute requirement for i-motif formation. The 12c6 and 10a4 oligonucleotides did not form any structure at any pH between 5.6 and 7.2.

An interruption of one of the cytosine tracts by a single thymine had a strong destabilizing effect on i-motif stability (Table 2). The 15a8 oligodeoxynucleotide d-**CCTCCTTTTC-CTCTC** had a much lower T_m (3.5 °C) than the corresponding 14a8 oligodeoxynucleotide d-**CCTCCTTTTCCTCC** (T_m = 22 °C under the same experimental conditions). The shape of the denaturation profile was also very different (less "cooperative") leading to a $\Delta H°$ value of only −20 kcal/mol for 15a8.

The length of the loops also played an important role on the stability of the i-motif. The 11a8 oligonucleotide d-**CCTC-CTCCTCC** was unable to form the i-motif in an intramolecular fashion, demonstrating that three short loops (each composed of one thymine) were not compatible with intramolecular folding. The 14b8 oligonucleotide d-**CCTTCCTTCCTTCC** could form an intramolecular structure, however less stable than the 17b8 oligonucleotide d-**CCTTTCCTTTCCTTTCC** (Table 2).

The 14a8 d-**CCTCCTTTTCCTCC**, 16a8 d-**CCTTCCTTTTC-CTTCC** and 18a8 d-**CCTTTCCTTTTCCTTTCC** oligodeoxynucleotides formed intramolecular structures with similar stabilities (T_m= 22, 21.5, and 23.5 °C, respectively). This shows that, as long as the central loop contains four thymines, the two other loops may accommodate a single base with no marked energetic penalty.

The different length requirements of the three loops could be explained by the geometry of the i-motif. One base is sufficient to join two strands across the minor grooves, but is not sufficient to join two strands across the major grooves.

Length of the Cytosine Tract. We compared the melting profiles of four oligodeoxynucleotides containing four clusters of a variable number of cytosines (between 2 and 5). As shown in Table 3, increasing the length of the cytosine repeats had a dramatic effect on the T_m of the i-motif. The length of the cytosine tract also had a strong impact on the thermodynamic parameters (Table 3). Four of the five oligonucleotides had identical loops, composed of three thymines, between the cytosine repeats, and thus the difference in stability could not arise from the length or the nature of the loops.

From the thermodynamic parameters presented in Table 3, one can calculate that each $C.C^+$ base pair decreases the $\Delta H°$ by 10.5 ± 0.5 kcal·mol^{-1} and $\Delta S°$ by 57 ± 1 cal·mol^{-1}·K^{-1}. At 37 °C, addition of one cytosine in each repeat led to a decrease in $\Delta G°$ of about 3 kcal·mol^{-1}. This value corresponds to an increase of the equilibrium constant for the unfolded-folded equilibrium of more than a hundred fold.

Nature of the Loops. Telomers are specialized chromatin domains at the ends of chromosomes that contain simple repetitive elements.[34-36] The repetitive motif is **CCCTAA** for most vertebrates and **CCCCAA** for some ciliates. In a previous study, we investigated the folding of fragments of the C-rich strand of the human telomer sequence.[12] These oligodeoxynucleotide fragments were (**CCCTAA**)$_3$**CCC** and **CCCTAAC-CC**. The first oligodeoxynucleotide was able to form the i-motif

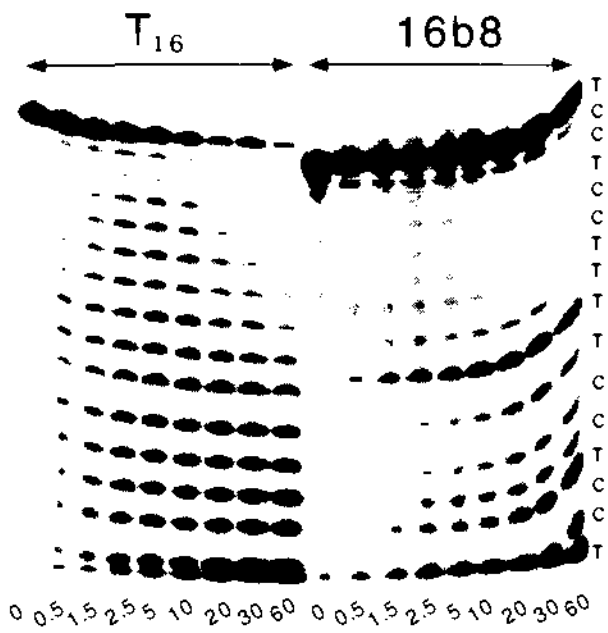

Figure 11. Resistance of the i-motif against mung bean nuclease degradation. A ^{32}P labeled 16 mer (16b8) d-**TCCTCCTTTTCCTCCT**, shown on the right part of gel, and a control t6 mer d-TTTT-TTTTTTTTTTTT, on the left part of the gel were incubated at pH 5.6 in a 50 mM MES, t mM $ZnCl_2$ buffer, in presence of 10 units of mung bean nuclease for various amounts of time (in minutes, indicated below each lane).

in an intramolecular fashion with each loop composed of one thymine and two adenines. We have synthesized the corresponding oligodeoxynucleotides (**CCCTTT**)$_3$**CCC** (21a12) and **CCCTTTCCC**, where all adenines have been replaced by thymines. Thermal denaturation studies showed that 21a12 was able to form an intramolecular structure more stable (ΔT_m = +3 °C) than the telomeric oligodeoxynucleotide (Table 3). Changing the TAA repeats into TTT led to stabilization of the structure by 1.1 kcal/mol at 37 °C. Such difference, which is presented in Table 3 at pH 6, was preserved at pH 7.2 (not shown).

A similar conclusion was reached with the **CCCTTTCCC** oligodeoxynucleotide, which formed a more stable bi-molecular structure than the **CCCTAACCC** oligodeoxynucleotide (ΔT_m = 6 °C at 10 μM; not shown). A closely related oligodeoxynucleotide (**CCCTTTTCCC**) also formed a bimolecular structure.[37] The (**CCCCTT**)$_3$**CCCC** oligodeoxynucleotide also formed a more stable intramolecular i-motif, as compared to the (**CCCCAA**)$_3$**CCCC** oligodeoxynucleotide (not shown).

Mung Bean Nuclease Resistance of the i-Motif. The mung bean nuclease cleaves single-stranded DNA at acidic pH. We wanted to investigate whether i-motif formation could interfere with a single-strand specific nuclease. As shown in Figure 11, the i-motif increases the stability of a pyrimidine oligodeoxynucleotide towards a pH sensitive nuclease. The control oligonucleotide, dT_{16}, which is unable to form any structure, is quickly digested by the nuclease, with a regular degradation pattern. This is in clear contrast with the degradation pattern of the 16b8 oligonucleotide, where only a few bases are susceptible to mung bean nuclease attack. These exposed bases are in two of the three putative loops of an intramolecular i-motif. The half life of the 16b8 oligonucleotide was determined to be 40 min at 15 °C, as compared to less than 4 min for the T16 control. Thus, intramolecular folding of the 16b8 oligodeoxynucleotide leads to a partial protection against nuclease cleavage at acidic pH.

Discussion

i-motif formation has been demonstrated so far by NMR and crystallography for the following sequences: d-**TCC**;[26] d-

(34) Blackburn, E. H. *Nature* **1991**, *350*, 569−573.
(35) Broccoli, D.; Cooke, H. *Am. J. Hum. Genet.* **1993**, *52*, 657−660.
(36) Bebichov, D. V. *Genetika* **1993**, *29*, 373−387.
(37) Pilch, D. S.; Shafer, R. H. *J. Am. Chem. Soc.* **1993**, *115*, 2565−2571.

^{m}CCT;[26] d-CCCT;[10] d-TCCC;[8] d-CCCTAA;[38] d-CCCAAT,[39] d-CCCC;[9] d-TCCCCC;[7] d-TTCCCCCCCCTT;[8] and d-(CCCTAA)$_3$CCC.[9,10] Except for the last sequence, the i-structure were formed by association of four different strands. Each of the two parallel duplexes constitutive of the i-motif may accommodate between 2 and (at least) 8 C.C+ base pairs. The i-motif also seems to have a specific circular dichroism signature.[30] In this study, we have shown that many cytosine-rich oligodeoxynucleotides could adopt a folded conformation at slightly acidic or neutral pH, as shown by thermal denaturation and non denaturing gel electrophoresis experiments. As expected, pH plays a crucial role on the stability of such a structure. The melting temperature of the complexes, whether intra- or intermolecular, was extremely pH-dependent: lowering the pH by one unit would typically lead to a T_m increase of 20 to 25 °C. The NMR study of the 18a8 oligomer demonstrates the formation at low pH of a structure containing hemiprotonated C.C$^+$ pairs whose NOE connectivities (H1′-H1′ and H2′/H2″-amino proton) are characteristic of the four-stranded structure formed with two intercalated oligo-C duplexes.[7,8] Several nucleosides have the N sugar puckering required for base intercalation.[26] The amino proton exchange kinetics (slow exchange for internal amino protons and faster exchange for external amino protons) indicates a restricted rotation for the amino group, a feature characteristic of the i-motif.[8] These arguments, together with those derived from the gel filtration study demonstrate that 18a8 adopts a back folded i-motif. There is no indication on the NMR spectra for the existence of more than one intercalated configuration.

The i-motif can be formed in an intramolecular fashion as soon as eight cytosines are present on the same oligodeoxynucleotide, but the length and base sequence of the loops will also have a significant impact on the stability of the folded structure. As a result of folding, two "intramolecular duplexes" each containing two C.C$^+$ base pairs are formed. The connecting loops may be very short: a single thymidine is sufficient for two of these loops, whereas the central loop must be longer (four bases for optimal stability).

Protonation of the Cytosines. There are some similarities between the i-motif, Hoogsteen duplexes and triple helices for cytosine-containing oligodeoxynucleotides: all these structures require protonated cytosines at N3 position. As the pK_a of this group is low (between 4.2 and 4.8, depending on temperature and ionic strength) formation of these structures at neutral pH is accompanied by protonation of the cytosines, and the stability of these complexes is extremely pH-dependent.[40]

Protonation of the C.C$^+$ base pair makes the observed thermodynamics parameters dependent on the buffer species as well as the pH. A similar conclusion was obtained for triplex formation involving C.G*C$^+$ triplets.[41] The $\Delta H°_{buffer}$ (corresponding to the protonation of the buffer species) influences the apparent thermodynamic parameters. To perform experiments in the 3.6−7.2 pH range, we have chosen acetate and cacodylate buffers; their $\Delta H°_{buffer}$ is close to zero. As a consequence, the pH of these buffers is not temperature dependent, making it appropriate for melting experiments. One can distinguish three limit cases (Figure 12).

(I) At neutral pH (pH ≫ pK_a) the observed reaction for the formation of an intramolecular i-motif can be represented as

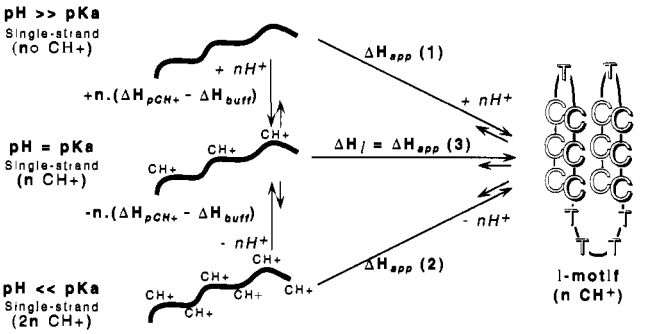

Figure 12. Thermodynamic scheme of i-motif formation at different pH.

follow: S + n•BH$^+$ ↔ iDNA(nH$^+$) + n•B where n is the number of protonated cytosines involved in the i-motif (one per C.C$^+$ base pair); B and BH$^+$ are the deprotonated and protonated buffer species; S is the single strand (with no protonated cytosine at pH ≫ pK_a); and iDNA(nH$^+$) is the intramolecular i-motif, containing nC.C$^+$ base pairs.

On a thermodynamical point of view, the reaction can be decomposed as follows:

(1) nBH$^+$ ↔ n•B + n•H$^+$ $\quad -n•\Delta H°_{buffer}$

(2) S + n•H$^+$ ↔ S(nH$^+$) $\quad + n•\Delta H°_{pCH^+}$

(3) S(nH$^+$) ↔ iDNA(nH$^+$) $\quad \Delta H°_{i\text{-}motif}$

From the melting curves, one can determine an apparent $\Delta H°_{app}$ which is $\Delta H°_{app} = n•(\Delta H°_{pCH^+} - \Delta H°_{buffer}) + \Delta H°_{i\text{-}motif}$.

(II) At (very) acidic pH (pH ≪ pK_a), all cytosines on the free strand are protonated, and the observed reaction for the formation of an intramolecular i-motif can be represented as follow: S(2nH$^+$) + n•B ↔ iDNA(nH$^+$) + n•BH$^+$ where S(2nH$^+$) is the single strand (with all protonated cytosine at pH ≪ pK_a) and iDNA(nH$^+$) is the intramolecular i-motif, containing nC.C$^+$ base pairs.

From the melting curves, one can then determine an apparent ΔH_{app} which is $\Delta H°_{app} = -n(\Delta H°_{pCH^+} - \Delta H°_{buffer}) + \Delta H°_{i\text{-}motif}$.

(III) At pH = pK_a, half of the cytosines on the single-strand are protonated. This is an optimal situation for the i-motif, as no net protonation or deprotonation is required.[42] The reaction will be S(nH$^+$) ↔ iDNA(nH$^+$). In that case, $\Delta H°_{app} = \Delta H°_{i\text{-}motif}$.

It is thus interesting to compare the melting profiles at low and high pH. Table 1 summarizes the thermodynamic parameters of 16b8 at different pH values. From a comparison of the thermodynamic parameters at near neutral pH (case I) and pH = pK_a (case III), it is possible to determine n ($\Delta H°_{pCH^+} - \Delta H°_{buffer}$). In a cacodylate buffer, the second term $\Delta H°_{buffer}$ is negligible. One can then determine the relative contributions of $\Delta H°_{i\text{-}motif}$ and $\Delta H°_{pCH^+}$. We obtained a value of −5.0 kcal/mol for $\Delta H°_{pCH^+}$, assuming $n = 4$. This value is in excellent agreement with the values documented for free cytosine,[43] and only slightly more negative than the one calculated by Manzini *et al.*[31] $\Delta H°_{pCH^+}$ also plays a role in triplex formation, as each C.G*C$^+$ triplet involves the protonation of a cytosine in the third strand.[41] A value of −6.0 kcal/mol for $\Delta H°_{i\text{-}motif}$ was calculated. Table 3 shows that the $\Delta H°_{app}$ for the formation of C.C+ base pairs are roughly additive, with a $\Delta H°_{app}$ of −10−

(38) Kang, C. H.; Berger, I.; Lockshin, C.; Ratliff, R.; Moyzis, R.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3874−3878.

(39) Berger, I.; Kang, C.; Fredian, A.; Ratliff, R.; Moyzis, R.; Rich, A. *Nature Struct. Biol.* **1995**, *2*, 416−425.

(40) Hüsler, P. L.; Klump, H. H. *Arch. Biochem. Biophys.* **1995**, *317*, 46−56.

(41) Wilson, W. D.; Hopkins, H. P.; Mizan, S.; Hamilton, D. D.; Zon, G. *J. Am. Chem. Soc.* **1994**, *116*, 3607−3608.

(42) *Even though, the distribution of the protonated cytosines plays a role: the distribution of protons in the i-motif is only a subset of the proton distribution in single strands.*

(43) Izatt, R. M.; Christensen, J. J.; Rytting, J. H. *Chem. Rev.* **1971**, *71*, 439−481.

12 kcal/mol base pair, (at pH above 5.5) in good agreement with the values documented for the human telomeric sequence.[12,30]

Type of Structures Adopted by C-Rich Oligodeoxynucleotides. Isomers of the i-Motif. For most of the oligonucleotides presented in this study, denaturation curves could well be analyzed as a one step equilibrium, suggesting, but not demonstrating, that a single folded species was obtained. Polyacrylamide gels demonstrated that some oligodeoxynucleotides (see lane 2 on Figure 8A) migrated as two bands, suggesting the presence of intramolecular isomers. Such a possibility was mentioned for the folding of a telomeric C-rich strand.[12] It should be noted that the i-motif has considerable conformational flexibility, in the orientation of its phosphate groups but also in the helical arrangement of the successive cytidine residues. The phosphate-phospate distance across the two wide grooves can vary between 12 and 17 Å.[9] The P−P distance across the two narrow grooves varies from 6 to 9 Å. Although, in theory, the molecule could assume several axes of symmetry, analysis of the crystal structure of d-**CCCT** revealed considerable asymmetry. For example, the two narrow grooves are not identical. This could explain why in Figure 11, only one of the two short loops is exposed to nuclease digestion. The thymines in each loop are probably in a different conformation. Some of the possible conformations could be imposed by the nature and length of the three loops. One nucleotide is probably sufficient to bridge the two narrow grooves (see Table 2), but not to bridge the two wide grooves. The 17a8 olgodeoxynucleotide has two loops of 4 thymines at each end, and a central loop of only one thymine, imposing a conformation where two loops bridge the wide grooves. Other oigodeoxynucleotides, such as 18a8, have three long loops, and have more flexibility.

Other Type of Structures. It has been shown that alternating (**C**-T)$_n$ sequences may adopt at least three different conformations: an anti-parallel-stranded duplex with C:T base pairs, an antiparallel-stranded duplex with CH^+:T base pairs, or a parallel stranded duplex with $C.C^+$ and T:T base pairs, depending on the pH.[44] Alternating d(G-A:T-C) sequences are structurally polymorphic, even at neutral pH.[45] Thus, i-motif formation, although frequent for C-rich oligodeoxynucleotides, might not be the only possible folded conformation, especially when the cytosines are dispersed on the primary sequence. The i-motif might also co-exist with other structures, such as a DNA hairpin loop on the same oligodeoxynucleotide.[46] Most of the oligodeoxynucleotides presented in this study contained clusters of cytosines. CT-alternating oligodeoxynucleotides probably adopt a different conformation.[44,47,48]

We could not investigate by NMR all the oligodeoxynucleotides presented in Tables 1, 2, and 3. For these reasons, we cannot formally exclude in all cases that parallel duplexes are formed instead of the i-motif. However, such a structure is less likely. First, it is difficult to imagine *intra*molecular parallel duplexes for short oligodeoxynucleotides: a simple hairpin would bring the two strands in an antiparallel conformation, and we have shown by several techniques (gel filtration, thermal denaturation and gel retardation) that most oligodeoxynucleotides do form intramolecular structures. Second, most of the oligodeoxynucleotides had similar, if not identical, behaviors towards the techniques we used in this study. Third, the i-motif offers several energetic advantages over a parallel duplex: (i) the tetrad allows more stacking interactions, (ii) extensive sugar van der Waals contacts are absent in a duplex, and (iii) in the i-motif, base stacking interval averages to 3.1 Å between adjacent cytosine residues along the axis, as intercalation produces a stacking of the bases in which only the exocyclic atoms are involved in the stacking.[9,10]

Biological Significance of the i-Motif. pH-Dependence. Considering the low pK_a of cytosine, it has been suggested that any type of base-pairing involving cytosine protonation would not be stable at neutral pH. $C.C^+$ base pair stability is maximal at pH = pK_a, when half of the cytosines on an unstructured oligodeoxynucleotide are protonated. However, $C.C^+$ base pair formation is not excluded at neutral pH: free H^+ ions are then taken up upon formation of the i-motif. At pH 7, where a negligible fraction of the cytosines are protonated in the single stranded structure, one proton is taken up for the formation of each hemiprotonated base pair. This base pair is, of course, less stable at pH 7 than at pH 5. But this reduced stability at neutral pH did not prevent i-motif crystals to grow in a cacodylate buffer at pH 7,[10] and many of the oligodeoxynucleotides presented in this study, showed a thermal transition at near neutral pH (see Figure 5A or 7 for example). The fact that the i-motif appears to be stable at neutral pH enhances the possibility of its being found under physiological conditions, especially if some proteins specifically bind to this structure, and stabilize it in vivo.

Implications for Telomers. All telomeric repeats have a peculiar distribution of guanines and cytosines resulting in one strand being G-rich in comparison with the complementary C-rich strand.[34,49] Several studies have already revealed that the G-rich strand is capable of forming a tetra-stranded structure in which four guanines are hydrogen bonded in one plane.[50−52] Single-stranded oligodeoxynucleotides that mimick the C-rich strand of vertebrate telomers are known to self associate in vitro.[53] This structure was shown to be the i-motif[12,13] and the crystal structure of a related oligodeoxynucleotide has been published recently.[38] Intramolecular folding of the G-rich and/or the C-rich strand could play a role in telomere maintenance and function, a possibility that remains to be tested.

Apart from telomeric regions, eukaryotic genomes contain stretches of cytosine-rich sequences that are located within regions of functional importance. Some of the triplet repeat extensions observed in some pathological situations contain 66% of cytosine in one strand (e.g., the **CCG** repeat of human fragile X syndrome). d-**CCCT** repeats are overrepresented in the human genome.[30] Most of these sequences are compatible with i-motif formation.

Interference of i-DNA Formation with Oligonucleotide-Based Therapeutic Strategies. Cellular Uptake and Stability. Such a pH-sensitive structure might also play a role in the chemical stability or cellular uptake of oligodeoxynucleotides used *in vivo*. Endocytosis seems to be the commom mechanism of oligodeoxynucleotide entry into a living cell. Oligodeoxynucleotides, when added to a cell culture medium, are efficiently taken up by cells, and accumulate in intracellular vesicles (endosomes). The pH of such vesicles is low (pH ≈ 6), even before fusion with lysosomes. We have shown that the i-motif

(44) Jaishree, T. N.; Wang, A. H. J. *Nucleic Acids Res.* **1993**, *21*, 3839−3844.
(45) Casasnovas, J. M.; Huertas, D.; Ortizlombardia, M.; Kypr, J.; Azorin, F. *J. Mol. Biol.* **1993**, *233*, 671−681.
(46) Rohozinski, J.; Hancock, J. M.; Keniry, M. A. *Nucleic Acids Res.* **1994**, *22*, 4653−4659.
(47) Jaishree, T. N.; Wang, A. H. J. *FEBS Lett* **1994**, *337*, 139−144.
(48) Brown, D. M.; Gray, D. M.; Patrick, M. H. *Biochemistry* **1985**, *24*, 1676−1683.
(49) Blackburn, E. H. *Cell* **1994**, *77*, 621−623.
(50) Kang, C. H.; Zhang, X.; Ratliff, R.; Moyzis, R.; Rich, A. *Nature* **1992**, *356*, 126−131.
(51) Smith, F. W.; Feigon, J. *Nature* **1992**, *356*, 164−168.
(52) Wang, Y.; Patel, D. J. *J. Mol. Biol.* **1993**, *234*, 1171−1183.
(53) Ahmed, S.; Henderson, E. *Nucleic Acids Res.* **1992**, *20*, 507−511.

increases the stability of a pyrimidine oligodeoxynucleotide towards a pH sensitive nuclease (Figure 11). Thus, the i-motif could play a role in determining the lifetime of an oligodeoxynucleotide in some cellular compartments. It is also important to note that the strand concentration of oligonucleotides in these vesicles can be extremely high, and would also favor inter-strand pairing. Cell uptake could also be affected by the folded conformation. In fact, it has been recently shown that a cytosine-rich oligonucleotide had a different fate in HL60 cells as compared with a control oligonucleotide of similar length[54] and was trapped in acidic compartments where i-motif formation is the most likely.

Interference with Triplex Formation or Antisense Activity. The folding of the third strand might also interfere with triplex formation. Several parameters have a differential impact on triplex and i-DNA stability. To design a triplex-forming oligodeoxynucleotide, the choice of the target sequence is of course the first constraint. If possible, one should avoid a repetitive motif containing several cytosines in a row. Triplex formation is mostly limited to oligopurine-oligopyrimidine stretches. Although fairly abundant in the human genome,[55] these sequences are often repetitive like $(\mathbf{CCT})_n$, $(\mathbf{CCCT})_n$ or $(\mathbf{CCCCT})_n$. Any sequence of the type $\mathbf{N}_a\mathbf{C}_w\mathbf{N}_n\mathbf{C}_x\mathbf{N}_p\mathbf{C}_y\mathbf{N}_m\mathbf{C}_z\mathbf{N}_b$ (with $a, b \geq 0$; $n, m \geq 1$; $w, x, y, z \geq 2$ and $p \geq 3$) is potentially able to form an intramolecular i-motif. In other words, most of the oligodeoxynucleotides containing four repeats of at least two cytosines are able to form an intramolecular structure. The sequence requirements for a bi-molecular i-motif (see Figure 1B for an example) are even less stringent, as only two stretches of two cytosines are required. Many of the potential binding sites for triplex forming oligodeoxynucleotides belong to this family: the promoter regions of important genes contain repeating sequence motifs (e.g., **ACCCTCCCC** for the human c-myc promoter;[56] $(\mathbf{CCCT})_n$ for the murine c-Ki-Ras promoter[30]). Thus, the i-motif, and other self-associated structures, act as a potential limitation to the sequence repertoire for the triplex strategy.

Formation of the i-motif should not only be taken into account for triplex formation, but also in antisense applications, as soon as the oligodeoxynucleotide contain several cytosines in a row. A quick overview of the literature on antisense revealed that many oligodeoxynucleotides studied are potentially able to from the i-motif. For example, several antisense oligonucleotides directed against HIV, such as ISIS 3466 which is complementary to gp120 messenger RNA,[57] have a sequence compatible with i-motif formation. The effect on the i-motif stability of chemical modifications that stabilize oligodeoxynucleotides against nuclease degradation remains to be tested. Preliminary results suggest that cytosine-rich phosphorothioate oligodeoxynucleotides could form a very stable intramolecular structure.

All the oligonucleotides presented in this study were unmodified, single-stranded pieces of DNA. Many research groups are designing oligonucleotides analogs to improve the efficiency of such molecules as antisense or antigene agents. Such chemical modifications could also have an impact on i-motif stability, or more generally on any type of competing self-structure. In this study, we only investigated the substitution of cytosines by 5-methylcytosines, which did not affect the stability of the i-motif. This replacement has been shown to favor triplex formation[33,58-60] as a result of an increase in base stacking.[61] This modification could then favor triplex over self-association of the third strand. On the other hand, RNA oligoribonucleotides do not form a stable i-motif (Lacroix et al., in preparation) but do form triple-helices on a DNA target[62-65] and are promising agents to form triple-helices on C.G rich DNA targets. Thus, the choice of an oligodeoxynucleotide for triplex formation must take into account not only its ability to form a triplex, but also the ability of the third strand to form an undesired, competing structure such as G-quartets,[66] GA-duplexes,[67] or i-DNA.

Acknowledgment. We thank Pr. T. Garestier and M. Rougée for helpful discussions. The T_m experiment software was writen by J. S. Sun and B. Sun.

JA9510626

(54) Tonkinson, J.; Stein, C. A. *Nucleic Acids Res.* **1994**, *22*, 4268–4275.
(55) Behe, M. J. *Nucleic Acids Res.* **1995**, *23*, 689–695.
(56) Firulli, A. B.; Maibenco, D. C.; Kinniburgh, A. J. *Arch. Biochem. Biophys.* **1994**, *310*, 236–242.
(57) Perlaky, L.; Saijo, Y.; Busch, R. K.; Bennett, C. F.; Mirabelli, C. K.; Crooke, S. T.; Busch, H. *Anti-Cancer Drug Des.* **1993**, *8*, 3–14.
(58) Lee, J. S.; Woodsworth, M. L.; Latimer, L. J. P.; Morgan, A. R. *Nucleic Acids Res.* **1984**, *12*, 6603–6614.
(59) Povsic, T. J.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 3059–3060.
(60) Sun, J. S.; François, J. C.; Montenay-Garestier, T.; Saison-Behmoaras, T.; Roig, V.; Chassignol, M.; Thuong, N. T.; Hélène, C. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9198–9202.
(61) Hausheer, F. H.; Singh, U. C.; Saxe, J. D.; Flory, J. P.; Tufto, K. B. *J. Am. Chem. Soc.* **1992**, *114*, 5356–5362.
(62) Roberts, R. W.; Crothers, D. M. *Science* **1992**, *258*, 1463–1466.
(63) Escudé, C.; Sun, J. S.; Rougée, M.; Garestier, T.; Hélène, C. *C. R. Acad. Sci. Paris, Serie III* **1992**, *315*, 521–525.
(64) Escudé, C.; François, J. C.; Sun, J. S.; Ott, G.; Sprinzl, M.; Garestier, T.; Hélène, C. *Nucleic Acids Res.* **1993**, *21*, 5547–5553.
(65) Han, H.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 3806–3810.
(66) Olivas, W. M.; Maher, L. J. *Biochemistry* **1995**, *34*, 278–284.
(67) Noonberg, S. B.; François, J. C.; Garestier, T.; Hélène, C. *Nucleic Acids Res.* **1995**, *25*, 1956–1963.
(68) Guéron, M. *J. Magn. Res.* **1978**, *30*, 515–520.
(69) Guéron, M.; Plateau, P.; Decorps, M. *Prog. NMR Spectrosc.* **1991**, *23*, 135–209.