Biochemistry **1993,** *32,* 6019–6031

Acid Multimers of Oligodeoxycytidine Strands: Stoichiometry, Base-Pair Characterization, and Proton Exchange Properties[†]

Jean-Louis Leroy,* Kalle Gehring,[‡] Abdelali Kettani, and Maurice Guéron

Groupe de Biophysique de l'Ecole Polytechnique et de l'URA D1254 du CNRS, 91128 Palaiseau, France

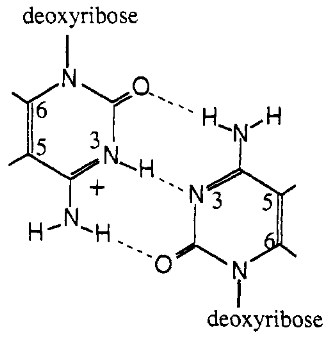
Received January 5, 1993; Revised Manuscript Received March 31, 1993

ABSTRACT: The structure recently proposed for the acid form of the oligonucleotide 5′-d(TC_5) is a four-strand "tetrad" in which two parallel-stranded, base-paired duplexes are intimately associated, with their hemiprotonated C·C$^+$ base pairs face-to-face and fully intercalated, in a so-called "i-motif" (Gehring et al., 1993). We use the amino and imino proton spectra to establish the structure and symmetry of the base pairs, properties which are a primary element in the resolution of the acid form described above. The amino proton spectrum gives the best lower limit (8×10^4 s^{-1}) on the rate of the imino proton jumping process which is responsible for the base-pair symmetry. The stoichiometry of the acid form of other deoxycytidine sequences is studied by gel filtration chromatography and in one case by an NMR equilibrium titration. In all cases, i.e., d(C_{12}), d($T_2C_8T_2$), d(C_4TC_4), d(TC_5), d(C_5), d(C_4), d(TC_4), d(TC_3T), and d(TC_3), the acid form elutes as a tetramer. A single-strand component is also present in some cases. But no dimer is observed, except for some samples prepared by quenching from high temperatures. The characteristic H1′–H1′ interresidue NOESY cross-peaks of the d(TC_5) structure (Gehring et al., 1993) are also found in all the tetramers where they have been searched for, i.e., those of d($T_2C_8T_2$), d(C_4TC_4), d(TC_3T), and d(TC_3) (not shown), suggesting that these tetramers also are built on the i-motif and that such structures may be formed generally by strands containing a stretch of as little as three deoxycytidines. From the NMR titration of d(TC_3), we derive a free energy of −7.6 kJ/mol per cytidine base pair for the formation of the tetramer from single strands. The free energy released by packing a *base pair* into the i-motif is comparable to that released in *forming* the base pair itself. Imino proton exchange is limited by base-pair opening, thanks to efficient intrinsic exchange catalysis: this explains the lack of effect of added catalysts. The base-pair lifetime is hundreds of times longer than in any DNA duplex, presumably due to the base-pair intercalation geometry. The variation of the lifetime along the sequence of the d(TC_5) tetramer provides support for the recently proposed structure. The internal amino proton exchanges from the open state of the C·C$^+$ pair, at a rate compatible with a p*K* of 9 appropriate for C$^+$. But the external proton exchanges from the closed state, as with a p*K* of 17! Comparable p*K* shifts occur upon transfer of organic acids from an aqueous to an organic solvent. We also show that exchange of the internal amino proton is opening-limited, even though it is eight times slower than imino proton exchange. The lifetime of the open state is at least 1 μs, 100 times longer than in duplexes.

Although the kinetics and dissociation constant of single base-pair opening in DNA duplexes are known in some detail (Guéron et al., 1989; Leroy, 1990), the structure of the open state from which imino protons exchange remains poorly understood, with little progress since the presence of a bridging water molecule was proposed for explaining the process of intrinsic catalysis (Guéron et al., 1987). For instance, we do not know if base-pair opening is a local fluctuation or if it occurs during a long-range distortion of the duplex such as unwinding or curvature, nor do we know if opening is toward the minor or the major groove.

The properties of structures different from double-stranded DNA may shed light on base-pair opening (and vice versa). Indeed the base-pair kinetics are altered in complexes of DNA with intercalating drugs (Leroy et al., manuscript in preparation), with bis-intercalating drugs (Leroy et al., 1992a), and with drugs which bind in the minor groove (Leroy et al., 1991), as well as in triple-stranded DNA (Kettani and Leroy, manuscript in preparation).

Scheme I



In this spirit, we undertook a study of proton exchange in triple helices formed of one G-rich and two C-rich strands. The preparation of these triple helices requires some care, since G-rich strands can autoassociate (Henderson et al., 1987; Smith & Feigon, 1992), as can C strands at acid pH (Akinrimisi et al., 1963). We therefore routinely inspect the oligonucleotides by proton NMR before mixing them. After dialysis against a low-salt, unbuffered medium, the C-strand solution is acid, and the NMR spectrum displays the imino proton resonances expected from the hemiprotonated C·C$^+$ base pairs represented in Scheme I (Akinrimisi et al., 1963). What is unexpected is that the imino proton spectrum persists

---

[†] This study was supported by grants to M.G. from the Ministère de la Recherche et de la Technologie and from the Association pour la Recherche contre le Cancer. Partial support to K.G. was provided by Ecole Polytechnique and by the Ministère des Affaires Etrangères.

* Correspondence should be addressed to this author at BIOP Polytechnique, 91128 Palaiseau, France. Phone, 33-1-6933-4412; fax, 33-1-6933-3004.

[‡] Present address: Laboratoire de RMN, Institut de Chimie des Substances Naturelles, CNRS, 91198 Gif, France.

0006-2960/93/0432-6019$04.00/0

at high temperatures and that long proton exchange times are found even in a DNA sequence which is short and designed to avoid longitudinal, out of step aggregation, e.g., d(TC_3).

We set out to explore the acid form of oligodeoxycytidine sequences. The spectra gave strong evidence for a duplex of identical parallel strands with a well-specified base-pair structure or (with the benefit of hindsight) *for any completely symmetric association of such duplexes*. But we did not seriously consider the possibility of associations, most of which, for instance two duplexes stacked one on the other (Genest et al., 1987), would not be very important.

We might have been alerted by at least two features of the 2D NOESY[1] spectra: short-time interresidue H1′–H1′ cross-peaks and interresidue H1′–H6 cross-peaks occurring in both directions (5′ to 3′ and 3′ to 5′). But the former are hard to see because they are close to the diagonal, and the latter are seen only in NOESY spectra with long mixing times (Gehring et al., 1993).

We did not even suspect a problem with the duplex model until we undertook to determine its handedness. In the course of that work, assignment of the phosphorus spectrum led to the realization that the order of base stacking differs from the covalent order, and to the model of the 5′-d(TC_5) "tetrad" structure. This structure is indeed based on a parallel-stranded, base-paired duplex, but it includes *two* of these, intimately associated with the $C{\cdot}C^+$ base pairs face-to-face and fully intercalated (Gehring et al., 1993). The feature of fully intercalated base pairs is designated the "i-motif". The topology of the structure is represented in Figure 1.

In the present work, we examine the stoichiometry of the acid multimers of strands containing different deoxycytidine sequences. Using in most cases gel filtration chromatography, and in one case an NMR equilibrium titration, we find that under our conditions, the equilibrium acid form of all sequences studied is tetrameric. The exchangeable proton spectrum is used to establish base-pair structure and the equivalence of bases within a pair, properties which are a primary element for the determination of the "tetrad" structure of the 5′-d(TC_5) tetramer (Gehring et al., 1993).

We also investigate the kinetics of proton exchange and of base-pair opening. This system is distinguished by the symmetry of the base pairs, by the low p*K*s of *both* the imino and amino protons, and by the structure in which the base pairs are incorporated.

The NMR stoichiometry measurement was performed on the 5′-d(TC_3) multimer. The most extensive exchange study was carried out on the acid form of d($T_2C_8T_2$), a sequence designed with two thymidines at each extremity in order to prevent overlap of staggered strands. The sequence dependence of base-pair lifetimes was examined on the d(TC_5) tetrad. We also used the tetramer formed by d(C_4TC_4) to study whether thymidines inside the run of Cs are base-paired.

MATERIALS AND METHODS

Nomenclature. In the present text, cytidine and $C{\cdot}C^+$ are used for both the riboside and for the deoxyriboside derivatives, unless stated otherwise. Sequences are listed starting from the 5′ end. "Tetramer" and "dimer" express strand stoichiometry, without structural connotation. "Duplex" is a two-strand, base-paired structure. "Tetrad" refers to the model of the d(TC_5) tetramer (Gehring et al., 1993) or to a similar model of another oligonucleotide. "i-motif" refers to an arrangement of intercalated base pairs, of which the first example is the one found in the tetrad.

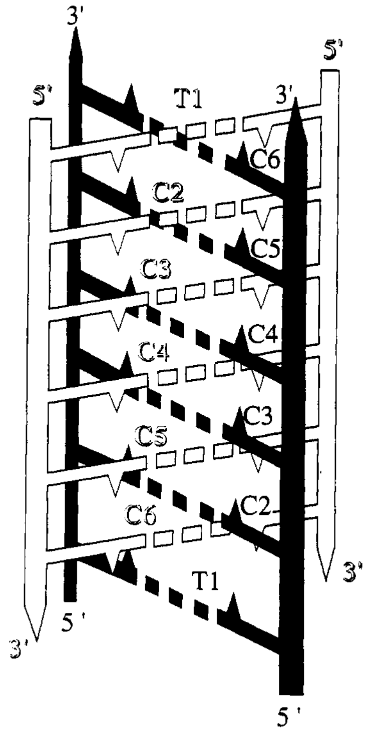


FIGURE 1: Sketch of the tetrad structure of d(TC_5) described in the article of Gehring et al. (1993). It includes two duplexes (ladders), antiparallel to one another. Each duplex is made of two strands, parallel and in register. The rung labeled C*n*, represents the $Cn{\cdot}Cn^+$ base pair, which is symmetric on the NMR time scale. The spikes indicate the base-pair orientation. The thymidines (T1) are represented as paired. The stacking order is 1, $\underline{6}$, 2, $\underline{5}$, 3, $\underline{4}$, 4, $\underline{3}$, 5, $\underline{2}$, 6, $\underline{1}$.

Deoxyoligonucleotide Synthesis and Preparation. The oligonucleotides d(C_{12}), d($T_2C_8T_2$), d(C_4TC_4), d(TC_5), and d(TC_3) were synthesized by the β-cyanoethylphosphoramidite method, on a 10-μmol scale, with a Pharmacia Gene Assembler DNA synthesizer. After detritylation, the resin was soaked in 2 mL of 25% ammonia for 30 min at room temperature so as to release the oligonucleotides. The supernatant was separated by centrifugation, maintained at 55 °C for 10 h to ensure complete deprotection of the oligonucleotides, and lyophilized. The oligonucleotides were purified by high-pressure chromatography as previously described (Kochoyan et al., 1990) and dialyzed against a 20 mM NaCl solution, using a 6–8-kDa cutoff Visking tube.

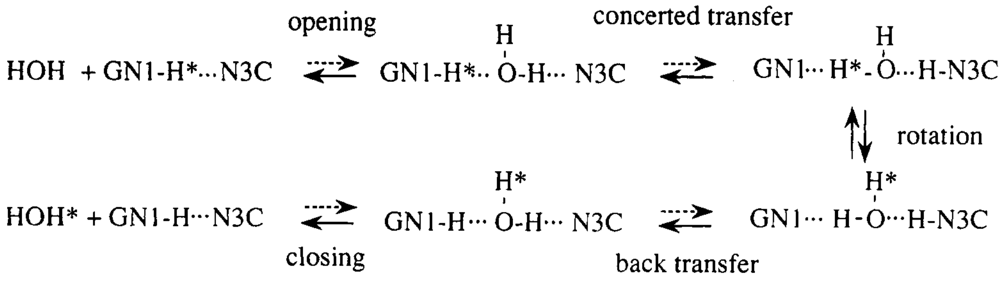
Proton exchange studies were performed in 0.1 M NaCl solutions with strand concentrations ranging from 1 to 6 mM. The pH was measured with a microelectrode and adjusted with HCl or NaOH, 0.1–0.5 M. The strand concentration was determined from the absorbance measured at pH 8.5. We used $A^{260} = 7960\ \mathrm{M^{-1}{\cdot}cm^{-1}}$ per base for the extinction coefficient of single-stranded d(C_{12}) (Adler et al., 1966) and $A^{260} = 7100\ \mathrm{M^{-1}{\cdot}cm^{-1}}$ for d(TC_5). The latter value was obtained by measuring the optical density of a d(TC_5) solution whose concentration was measured by NMR relative to a thymidine solution, itself calibrated using $A^{267} = 9700\ \mathrm{M^{-1}{\cdot}cm^{-1}}$. The concentrations of multimers are expressed in terms of *strand* concentration.

For the NMR measurements, the multimers were usually prepared by lowering the pH of a solution at the final strand

---

[1] Abbreviations: 1D, one dimensional; 2D, two dimensional; DSS, 2,2-dimethylsilapentane-5-sulfonate; FID, free induction decay; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser enhancement spectroscopy.

Scheme II



concentration. Small aliquots of HCl were added, and the acid form was monitored by the emergence and buildup of the imino proton NMR spectrum and by the evolution of the methyl spectrum. In general, the spectrum indicated a single species. When minority species (up to 20% of the total) were detected (e.g., the 11.0 ppm peak in Figure 5), they could be reduced or eliminated by annealing overnight at room temperature or at 4 °C.

Acidification of the $d(C_4TC_4)$ oligonucleotide in this manner produced a gel. The acid form was therefore prepared from a dilute solution (strand concentration <0.1 mM) at pH 8. The solution was dialyzed at room temperature against water maintained at pH 5, lyophilized, and redissolved at 4 °C to the concentration of the NMR experiment. Irreversible line broadening indicative of aggregation occurred at temperatures above 15 °C.

Gel Filtration Chromatography. The size, and hence stoichiometry, of the multimer was determined by high-pressure chromatography performed at room temperature with Beckman equipment. The columns were either a Lichrospher 100 diol column (250 mm × 4 mm i.d.) from Merck or a Synchropack GPC 6O column (250 mm × 4.6 mm i.d.) from Interchim. They were calibrated with the mononucleosides cytidine and thymidine; with single-stranded deoxyoligonucleotides [d(TGG), d(CGCTTGCG)], with deoxyoligonucleotides which were shown by NMR to form hairpins (unpublished) [$d(GGA_{10}T_{10}CC)$, $d(CCA_{10}T_{10}GG)$], with duplexes of the self-complementary deoxyoligonucleotides [$d(A_4GCGCT_4)$, d(GCGCATCGCG), $d(A_2GCGCT_2)$, d(AGCATGCT), $d(AT)_7$, $d(GGA_6T_6CC)$], with $tRNA^{Phe}$, and with 5s RNA. The elution medium was buffered by 20 mM sodium acetate at pH 4.2 or by 20 mM potassium phosphate at pH 7.2, and it contained salt (NaCl, 0.3 M) to reduce DNA–matrix interactions. The flow rate for optimal resolution was 0.4 mL/min. The chromatography samples for calibration or study consisted typically of 20 μL of a 1 OD/mL oligonucleotide solution. The elution profiles were measured by the absorbance at 280 nm.

In Figure 2, the elution volume is plotted versus the size in nucleotides of the different nucleic acid species in the nondenaturing elution medium. The plots show that the elution volume is primarily determined by the total number of residues. All the calibration samples were run both at pH 4.2 and 7.2, and the elution volumes were the same.

NMR Spectroscopy. The NMR measurements were performed on a home-built 8.45-T spectrometer controlled by a 386 PC-type computer networked with similar computers for data processing. Exchangeable proton measurements were carried out in 90% H_2O/10% D_2O solution with the "jump and return" (JR) sequence for solvent signal suppression (Plateau & Guéron, 1982; Guéron et al., 1991, 1992). The chemical shift reference was 2,2-dimethylsilapentane-5-sulfonate.

Imino and amino proton exchange rates were measured as previously described (Kochoyan et al., 1987; Leroy et al., 1988*a*). Exchange times longer than 1 min were determined in real time: a concentrated protonated sample was diluted 12-fold into D_2O, and the NMR spectrum was recorded at intervals. Exchange times shorter that 2 s were derived from the rate of magnetization transfer from water (Forsen & Hoffmann, 1963; Leroy et al., 1988*a*). The contribution of added catalyst to the exchange rate was obtained from the change in line width or longitudinal relaxation upon addition of catalyst.

The FID was processed for combined sensitivity and resolution, via multiplication by an exponential and a phase-shifted sine (Guéron, 1978), i.e.,

$$\exp(-\pi\delta t)\,\sin[2\pi(0.7)\delta t + 0.25]$$

with $\delta = 3$ Hz. Integral intensities were measured on spectra obtained with an inter-FID delay of at least 3 times the longest relaxation time and corrected for the frequency response to JR excitation (Guéron et al., 1991). The intensities of the thymidine imino, methyl, or H6 resonances were used as the reference for integration.

Theory of Exchange of Cytidine Protons. We summarize the theory of proton exchange (Eigen, 1964; Guéron & Leroy, 1992) in view of its application to the imino and amino protons of cytidine. Proton transfer from a donor (DH) to an acceptor (A) occurs via formation of a hydrogen-bonded complex (DH···A). The rate of proton transfer to A is

$$k_{tr} = k_{col}[A](1 + 10^{pK_D - pK_A})^{-1} \qquad (1)$$

where [A] is the acceptor concentration and the last factor represents the yield of AH^+ upon dissociation of DH···A. The rate constant for formation of the collision complex DH···A, k_{col}, is typically in the range of $10^9\ s^{-1}\ M^{-1}$ when complex formation is diffusion-controlled.

(a) For an imino proton in a Watson–Crick pair, for instance G·C, exchange requires the disruption of the base pair, followed by chemical exchange from the open pair. The latter reaction is catalyzed by proton acceptors such as NH_3 (Leroy et al., 1985), but it occurs even at neutral pH, in the absence of added catalyst (AAC), due to an intrinsic catalyst, most probably the cyclic nitrogen of the other nucleoside of the pair, acting as the proton acceptor. Transfer occurs in the open state, via a bridging water molecule (Guéron et al., 1987), as indicated in Scheme II, where the full arrows indicate the faster reactions.

The yield for concerted transfer from GN1 to CN3 across the water molecule is equal to $(1 + 10^{\delta pK})^{-1}$, where δpK is equal to donor pK ($pK_{GN1} = 9.3$) minus acceptor pK ($pK_{CN3} = 4$). The rate of concerted transfer may be expressed by

$$k_{tr}^{int} = R/(1 + 10^{\delta pK}) \qquad (2)$$

where the constant *R* is a pseudocollision frequency which replaces the product $k_{col}[A]$ of eq 1. When base-pair opening is not limiting (many openings being required for exchange), the exchange contribution of this process is

$$k_{\mathrm{AAC}} = k_{\mathrm{tr}}^{\mathrm{int}}K_{\mathrm{d}}/(1 + K_{\mathrm{d}}) \approx k_{\mathrm{tr}}^{\mathrm{int}}K_{\mathrm{d}} \qquad (3)$$

K_d being the base-pair dissociation constant ($K_d \ll 1$). From the values measured for the G·C imino proton in B-DNA at 15 °C ($k_{\mathrm{AAC}} \approx 1\ \mathrm{s}^{-1}$; $K_d \approx 10^{-6}$), we find $k_{\mathrm{tr}}^{\mathrm{int}} \approx 10^6\ \mathrm{s}^{-1}$, whereas the lifetime of the open state is typically 10^{-8} s. Hence, the probability for concerted transfer is only about 1% per opening event. This explains why exchange with water can be enhanced by proton acceptors such as ammonia or Tris (Leroy et al., 1985). In terms of the base-pair lifetime τ_0 and of the apparent base-pair dissociation constant, the complete expression for imino proton exchange is

$$\tau_{\mathrm{ex}} = \tau_0 + 1/[(k_{\mathrm{tr}}K_{\mathrm{d}}) + (k_{\mathrm{tr}}^{\mathrm{int}}K_{\mathrm{d}})] \qquad (4)$$

where k_{tr} is proportional to the concentration of added catalyst (eq 1). The base-pair lifetime is obtained as the limit of the exchange time at infinite catalyst concentration, i.e., under conditions where exchange occurs at each opening event. The base-pair dissociation constant is obtained by comparison with the exchange rate, k_{tr}, of the isolated nucleoside (Guéron et al., 1989).

(b) We now apply these considerations to the *imino proton of a C·C⁺ pair*. We assume that the water bridging model of intrinsic catalysis is valid. The p*K* difference between the donor (C^+) and the acceptor (C) is zero, rather than ca. 5 in the Watson–Crick pairs. Hence, the rate of intrinsic catalysis in the open state is increased by five orders of magnitude to $k_{\mathrm{tr}}^{\mathrm{int}} \approx 10^{11}\ \mathrm{s}^{-1}$, 3 orders of magnitude *larger* than the closing rate of Watson–Crick base pairs. If the closing rate of $C{\cdot}C^+$ pairs is comparable (or even up to 10^3 times larger), the imino proton exchanges at every base-pair opening, and added catalysts have no effect.

(c) Amino proton exchange in the neutral nucleoside is controlled by OH^- catalysis above pH 8. The rate is about 100 times slower than expected for a diffusion-controlled process (McConnell & Politowski, 1984). This indicates (eq 1) that the amino proton p*K* is 2 units larger than that of OH^-, hence a p*K* of 17.7. In N^3-methylated cytidine, which provides a model for N^3-protonated cytidine, and in N^3-protonated cytidine itself, the p*K* of the amino group is lowered to about 9 (McConnell, 1978; McConnell & Politowski, 1984). Exchange of the amino protons of N^3-methylated cytidine is catalyzed by OH^- at a rate $k_{\mathrm{ex}} = 5 \times 10^9[\mathrm{OH}^-]\ \mathrm{s}^{-1}$ at 0 °C (Leroy & Kettani, manuscript in preparation).

RESULTS

Multimer Stoichiometry

(a) Figure 2A shows the chromatographic determination of the multimer stoichiometry. The label "N" indicates the elution volume of a sample of $d(T_2C_8T_2)$, prepared by diluting a 1-μL aliquot of the sample whose imino proton spectrum is shown in Figure 7. It corresponds to 48 ± 5 residues, indicating four-strand stoichiometry.

The same stoichiometry was found for samples (1 OD/mL) of $d(TC_5)$ ("N" in Figure 2B) and $d(C_4TC_4)$ (not shown), similarly related to the NMR spectra of Figures 5 and 8, respectively. Samples of $d(C_5)$ and $d(C_4)$ (1 OD/mL) eluted as a mixture of tetramer and single strand, the latter corresponding to a proportion of 10% for $d(C_5)$, 50% for $d(C_4)$. At 3 OD/mL, $d(TC_4)$ eluted as a tetramer. At 6 OD/mL, $d(TC_3T)$ and $d(TC_3)$ both eluted as 20% tetramer and 80% single strand. The observation of distinct elution peaks for the multimer and single strand shows that there was no interconversion during the run. Elution of a dimer was not observed in any of the samples.

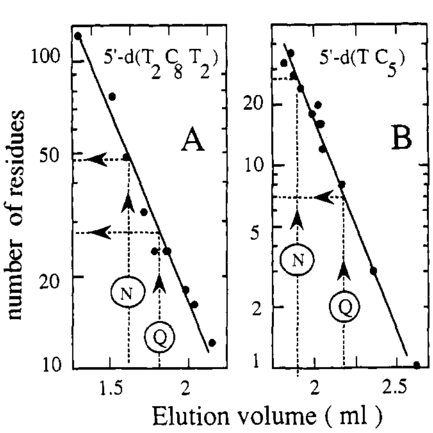


FIGURE 2: Calibration and use of the gel filtration columns. For calibration, the log of the known number of residues of several nucleic acid species under nondenaturing conditions (see the experimental section) is plotted vs the elution volume. The straight line is a fit by eye to the measurements (●). (A) Calibration of the Lichrospher 100 diol column and application to the native (N) and quenched (Q) forms of $d(T_2C_8T_2)$. The elution volumes correspond, respectively, to 48 residues (a tetramer) and to 24 residues (a dimer). The sample (1 OD/mL) is derived from the NMR sample whose spectrum is displayed in Figure 8. (B) Calibration of the Synchropack GPC 60 column and application to the native (N) and quenched (Q) forms of $d(TC_5)$. These elute respectively as 24 and 7 residues, corresponding to a tetramer and to a presumed single strand, within the uncertainties of the calibration procedure. Experimental details are given in the text.

(b) We also studied samples of $d(T_2C_8T_2)$ (label "Q" in Figure 2A) and of $d(C_4TC_4)$ which had been held 3 min at 100 °C and acid pH, had then been quickly cooled (quenched) in liquid nitrogen, and were then warmed and kept at 18 °C. In both cases, the product eluted at first as a dimer. It subsequently converted in a matter of weeks to other species and finally to the same tetramer as above. The same procedure was applied to $d(TC_5)$ (6 OD/mL). At first, the elution volume (label "Q" in Figure 2B) indicated seven residues, presumably a single strand. It then changed to a tetrameric form with a time constant of 3 days.

(c) At pH 7.2, all the oligonucleotides eluted as single strands.

(d) An NMR titration of the multimer–single strand equilibrium of $d(TC_3)$ established the multimer stoichiometry independently of any empirical calibration. The conversion of this product as a function of temperature or pH change was always faster than the time resolution of the experiments (minutes) and reversible. The NMR spectrum of $d(TC_3)$ in D_2O (100 mM NaCl, 20 mM sodium succinate, pH 4.2, T = 6 °C) is shown in Figure 3. It consists of two components in a concentration-dependent ratio. The spectral component which dominates at low concentrations is assigned to the single strand, consistent with its lack of imino proton resonances and in agreement with the chromatography results. The two spectral components are in slow exchange, and their intensities change reversibly as a function of the total concentration which is determined with reference to thymidine, as explained under Materials and Methods. Hence the system is at equilibrium. The proportions of free strand and multimer were determined by the intensity of the well-resolved lines of cytidine H6, thymidine H6, and methyl (Figure 3).

The log–log plot of multimer concentration (counted in single strands) vs single-strand concentration (Figure 4) shows that the multimer concentration increases as the fourth power

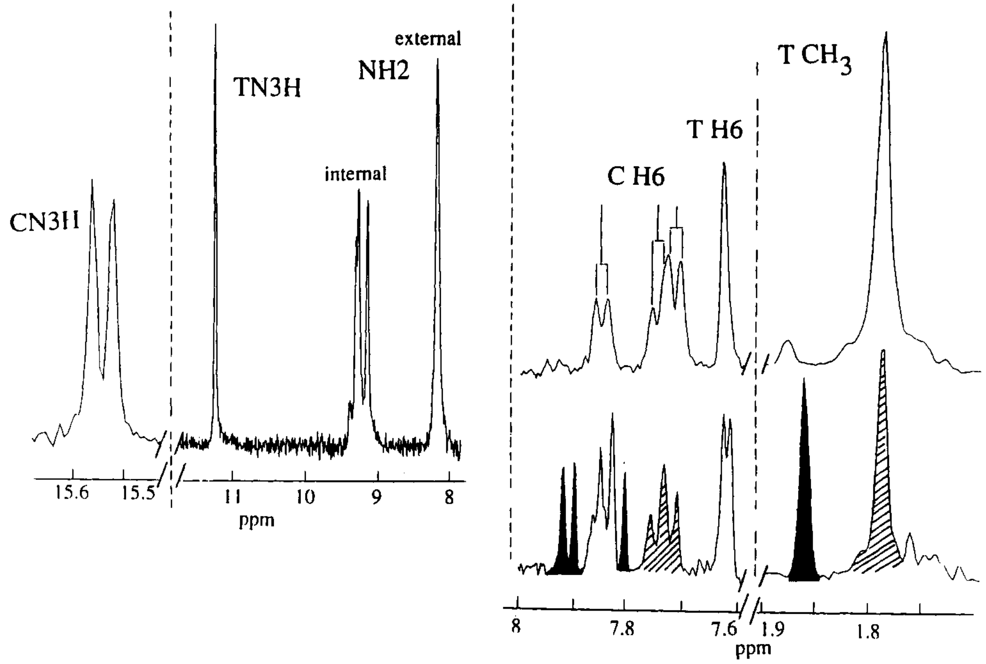


FIGURE 3: NMR spectra of d(TC_3). (Left side) Spectrum of exchangeable protons, in H_2O. Two cytidine imino proton resonances are visible. The three internal amino protons are nearly resolved. By reference to thymidine (one imino proton per strand), the imino proton lines correspond to 0.5 proton each, the internal amino proton lines to one proton each, and the external amino proton peak to three protons. The chemical shifts of the amino protons fall midway between those for neutral and protonated cytidine, in a condition of fast chemical exchange (see text). (Right side) H6 and methyl proton regions, measured in D_2O. At a total strand concentration of 4 mM (upper trace), d(TC_3) forms a multimer. At 1 mM (lower trace), the multimer and the single strand are both present in measurable concentrations. The relative concentrations were obtained from the areas of multimer peaks (hatched) and of single-strand peaks (black). The absolute concentration was determined as explained in the text. The multimer spectrum gives no indication of any difference between strands. Solution conditions: pH 4.2; sodium succinate, 20 mM; NaCl, 120 mM; T = 6 °C.

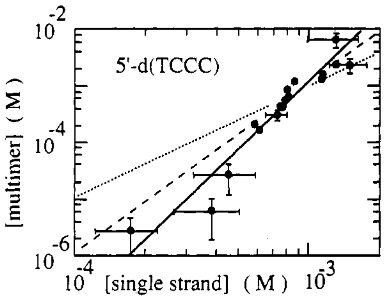


FIGURE 4: NMR titration of the multimer formed by d(TC_3), using the spectral peaks indicated in Figure 3. The concentration of the multimer (in strands) is plotted vs the concentration of single strand in a log–log plot. Lines of slope 2 (···), 3 (- - -), and 4 (—) are drawn through the data, which strongly favor a stoichiometry of 4. Error bars are given on six selected data points.

of the single-strand concentration, indicating the formation of a tetramer. The equilibrium constant is (1400 μM)^3 at 6 °C.

Formation and Spectral Characterization of the Acid Form

d(TC_5). Figure 5 shows the exchangeable proton region of an NMR spectrum taken under the conditions in which the tetrad structure was solved (Gehring et al., 1993). The signal at 11.36 ppm is assigned to the imino proton of thymidine by its NOE connectivity to the methyl protons and to the H6 proton of thymidine. In view of their chemical shift, the signals between 15 and 16 ppm are assigned to imino protons of protonated cytidines (Becker et al., 1965; Dos Santos et al., 1989). At −5 °C, four peaks are resolved, with relative intensities 1, 2, 1, and 1. When the thymidine imino proton peak at 11.3 ppm is normalized to 1, their total intensity is 2.3 ± 0.3, i.e., 0.5 per cytidine.

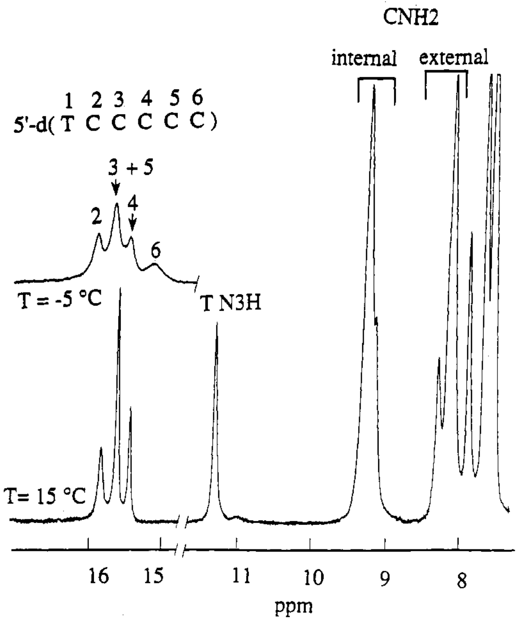


FIGURE 5: Proton NMR spectrum of d(TC_5) in H_2O (90%), pH 4.9. At −5 °C, the spectrum displays four cytidine imino proton peaks with intensities of 0.5 (1, 2, 1, and 1) per strand, showing that all C·C$^+$ base pairs are formed. The assignments were obtained as indicated in Table I. At 15 °C, the imino proton peak from base pair C6·C6$^+$ is broad beyond detection, due to exchange with solvent. The internal and the external amino proton clusters each integrate to 5 ± 1 proton per strand. The NaCl concentration is 100 mM.

The amino proton resonances are distinguished from the aromatic protons by a real-time exchange measurement [not shown; see Figure 8 for the corresponding characterization in the d($T_2C_8T_2$) tetramer]. They appear as two mostly unresolved clusters at 9.2 and 8.2 ppm, which are assigned

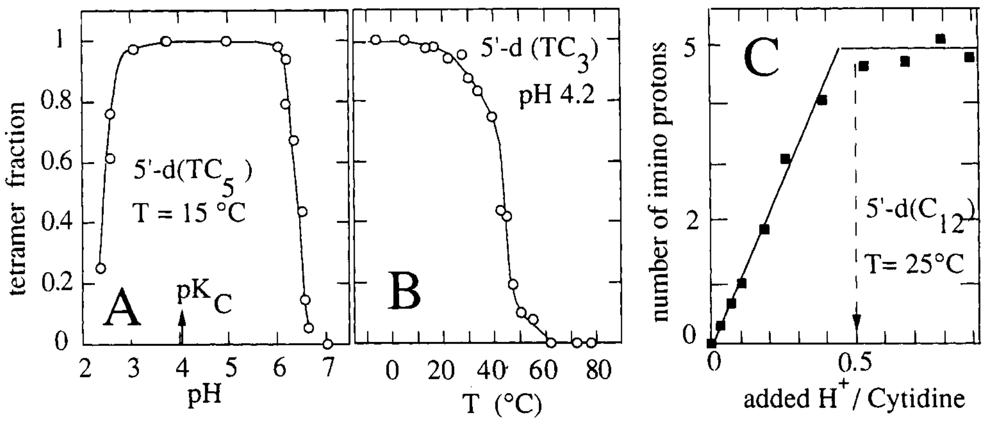


FIGURE 6: (A) pH dependence of the d(TC$_5$) tetramer/single-strand equilibrium. The tetramer fraction, counted in strands, was obtained from the intensity of its T methyl NMR line, which is in slow exchange with that of the of the single strand. (B) Melting of the d(TC$_3$) tetramer. (C) Titration of the imino protons of the d(C$_{12}$) tetramer. The area of the imino proton spectrum levels off at 0.5 H$^+$ per cytidine, i.e., one proton per C·C$^+$ pair. The titration is started at pH 7.5. The end point is at pH 6.6.

respectively to the internal and external position by their NOE connectivities with the imino and the H5 protons (not shown). Each cluster integrates to 5 ± 1 protons per strand.

Figure 6A shows a study of the equilibrium between tetramer and single strand vs pH, using the well-resolved peaks of the T methyl proton (at 1.68 ppm in the single strand and at 1.61 ppm in the tetramer, data not shown) or those of cytidine H6 or of amino protons. At a concentration of 10 mM (single strand), the tetramer is the main species between pH 6.5 and 3. At lower pH, it dissociates and the NMR spectrum shows signals from the tetramer and from single strands with protonated cytidines. The concentration of any other species is a small fraction of the total.

At pH 4.9, no single-strand spectrum was detectable below 40 °C. The aromatic and methyl proton resonances are in slow exchange between the tetramer and single-strand forms up to and even above the melting temperature (T_m = 74 °C). The line broadening is less than 2 Hz at this temperature, corresponding to a tetramer lifetime longer than 0.16 s.

Among the cytidine imino proton peaks, that of C6 broadens out below 15 °C (Figure 5), and this is ascribed to exchange with water. Up to the melting temperature, the total area of the other resonances varies with pH and temperature as does the tetramer fraction, as determined from the intensities of the tetramer and single strand methyl peaks. At 0 °C, the longitudinal relaxation times, the line widths, and the chemical shifts are independent of pH.

Other Tetramers. The other cytidine strands studied here behave similarly to d(TC$_5$) as regards the pH and temperature dependence of the single-strand to tetramer transition. The melting temperature is generally lower for shorter strands. Except for unresolved peaks and for exchange broadening, the imino proton spectrum contains one peak for each cytidine, with an intensity of 0.5 proton per base. Tetramer formation was studied in particular detail in d(TC$_3$) and in d(T$_2$C$_8$T$_2$).

The NMR equilibrium dilution measurement of d(TC$_3$) was presented above (Figures 3 and 4). Figure 6B shows the melting of the tetramer (strand concentration 5 mM, in a solution of NaCl, 30 mM, at pH 4.2). The melting temperature is 42 °C, and the activation enthalpy is 380 kJ/mol. The melting temperature falls if the salt concentration is increased (T_m = 32 °C in 1.5 M NaCl) or if the pH is raised (T_m = 0 °C at pH 5.7).

At a strand concentration of 4 mM, the tetramer of d(T$_2$C$_8$T$_2$) forms cooperatively around pH 6.5, and this value is nearly temperature-independent between 0 and 35 °C. At −5 °C, the cytidine imino proton region, normalized to the thymidine imino proton peak, shows a cluster of seven nearly resolved NMR lines integrating to 3.5 ± 0.3 protons and a line at 14.88 ppm of intensity 0.5 ± 0.1 per single strand (Figure 7). This line, and no others, has an NOE connectivity with the imino proton peaks of thymidine (not shown). Each of the two amino proton clusters integrates to 8 ± 1.5 protons, or a total of 1 per cytidine, as expected (Figure 9).

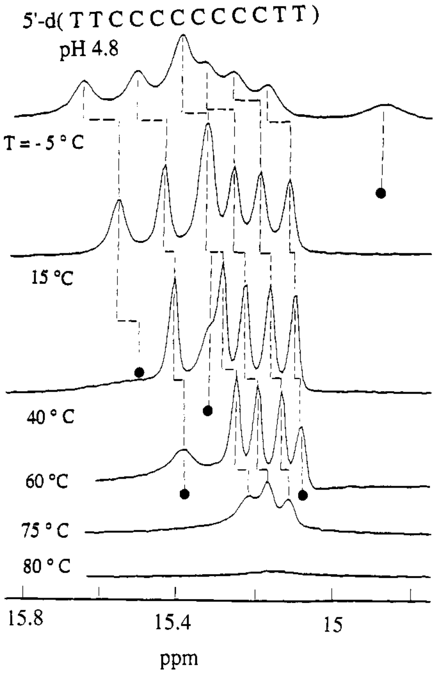


FIGURE 7: Variation of the imino proton spectrum of the d(T$_2$C$_8$T$_2$) tetramer with increasing temperature. Termination of a dotted connecting line (●) indicates disappearance of the corresponding peak, due to broadening. Three peaks are still visible at 75 °C, with nearly equal widths. The fraction of the material in the single-strand form, as determined from the intensity of the single-strand spectrum (cytidine CH6 and T methyl lines; not shown) is less than 10% at 60 °C. It reaches 50% at 78 °C. Experimental conditions: pH 4.8; NaCl, 100 mM; concentration in strands, 5 mM. The chemical shift reference is DSS.

The pH dependence for the formation of the d(C$_4$TC$_4$) tetramer is similar. At 0 °C and pH 4.9 the integrated area of the C imino proton lines (17–15 ppm) is 3.5 protons when the T imino proton line (at 11 ppm) is normalized to 1 (Figure 8). The area of the amino proton clusters (7 ± 1 internal amino protons ca. 9.2 ppm and 6 ± 2 external amino protons at 8.2 ppm) indicates that most of the cytidines contribute.

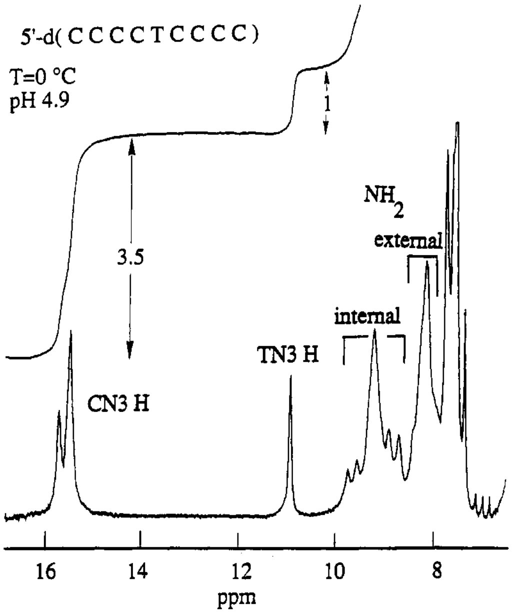


FIGURE 8: Proton NMR spectrum of the d(C_4TC_4) tetramer. Integration of the C imino proton NMR lines (17–14 ppm) corresponds to seven C·C$^+$ pairs. Solution conditions: pH 4.9, 100 mM NaCl.

Table I: Spectral Assignments and Exchange Times of the Imino Protons in the d(TC_5) Tetramer

imino proton[a,b]	ppm	τ_{ex} (s) T = 0 °C	T = 15 °C
T1[c]	11.36	0.5	0.15
C.C6	15.13	0.006	≪0.001
C.C2	15.86	0.8	0.18
C.C5 and C.C3	15.64	1200; 9000	500; 700
C.C4	15.43	2280	300

[a] The assignments are based on long mixing-time NOESY cross-peaks of the imino protons with base and sugar protons and on imino–imino cross-peaks (manuscript in preparation). The latter agree with the stacking order of base pairs in the tetrad structure, 1, $\underline{6}$, 2, $\underline{5}$, 3, $\underline{4}$ (Gehring et al., 1993). [b] The pairs are listed in the order of stacking in the tetrad. The exchange times of the cytidine amino protons, each equal to the base-pair lifetime, increase sharply as one enters the tetrad. The T·T base pair is not formed or, if formed, has a lifetime smaller than 1 ms. [c] In the case of thymidine, catalysis is the limiting step for exchange.

In the d(C_{12}) tetramer, the imino protons form a broad unresolved line at 15.6 ppm. When calibrated with the cluster of H6 resonances, the intensity corresponds to 5 ± 2 protons per strand at 25 °C. The formation of the C·C$^+$ pairs titrates at 0.5 proton per cytidine (Figure 6C). The midpoint is ca. pH 6.8, and the area of the imino proton NMR line reaches a pH-independent plateau at pH 6. The resonances of nonexchangeable protons are anomalously broad, suggesting aggregation by base pairing between staggered strands.

Proton Exchange in the Tetramers

We studied the exchange process of imino and amino protons in the d($T_2C_8T_2$) and in the d(TC_5) tetramers. Both tetramers give similar results.

Imino Protons. The exchange times of the d(TC_5) imino protons at 0 and 15 °C are shown in Table I. Those of cytidines C3 to C5 are very long (many minutes), whereas those of C6 and C2 are shorter than 1 s. In the proposed tetrad structure (Gehring et al., 1993), the order of stacked base pairs is 1, $\underline{6}$, 2, $\underline{5}$, 3, $\underline{4}$, 4, $\underline{3}$, 5, $\underline{2}$, 6, $\underline{1}$ (Figure 1). One immediately notes that the C6 pair, which has the shortest exchange time, is the most outward cytidine pair, next to the thymidines (numbered 1), and that the C2 pair follows.

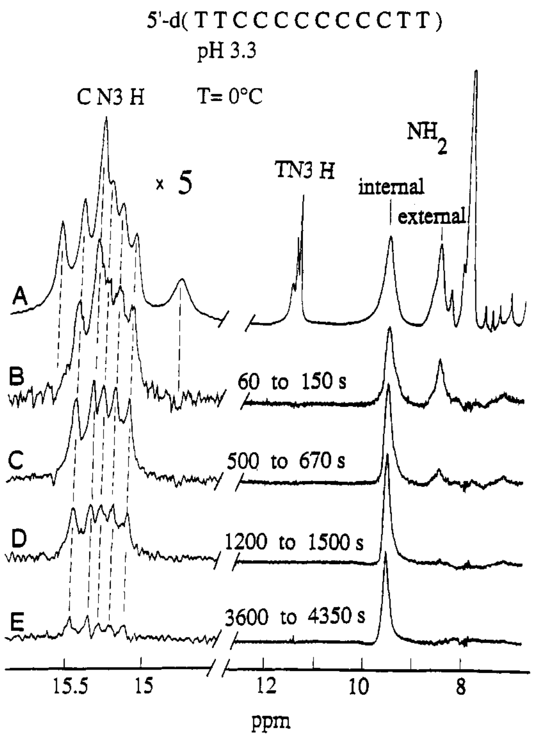


FIGURE 9: Real-time solvent exchange experiment in the d($T_2C_8T_2$) tetramer. (A) Reference spectrum in 90% H_2O. The protonated tetramer was diluted into D_2O at t = 0. Each of the spectra B, C, D, and E is the difference between a spectrum recorded in the time interval indicated and the spectrum recorded 24 h after dilution (not shown). Two imino protons exchange in less than 60 s and one in about 2 min. The remaining five have the same exchange time, 25 min. The internal and external amino protons exchange in 360 and 7 min, respectively. Note the progressive narrowing of the internal amino proton cluster, due at least in part to the loss of proton–proton couplings upon deuteration at the imino and at the external amino positions. Experimental conditions: 100 mM NaCl; pH 3.3; T = 0 °C. The measurement was carried out at low pH in order to slow down OH$^-$-catalyzed exchange of the external amino proton (see Figure 10).

Figure 9 shows a real-time exchange measurement in the d($T_2C_8T_2$) tetramer, whose protons have not been assigned. In the cluster of slowly exchanging protons, all the exchange times are comparable, ca. 25 min, except for one line at 15.3 ppm whose exchange time is about 2 min. The two peaks at the left and right of the cluster, at 14.88 and 15.6 ppm, have short exchange times, 15 ms and ca. 1 s, respectively. The positions in the imino proton spectrum and the exchange times are analogous to those of the C6 and C2 imino protons of d(TC_5), suggesting that, correspondingly, these peaks of d($T_2C_8T_2$) belong to C10 and C3, and that the structure is similar to that of the d(TC_5) tetrad, starting with thymidines and followed by cytidines in the order $\underline{10}$, 3, ... (or $\underline{3}$, 10, ...). The one and only NOE connectivity of the 14.88 ppm peak to the thymidine imino protons supports this suggestion.

The exchange times of the cytidine imino protons are not affected by the proton acceptors acetate (1 M $CH_3CO_2^-$) and phosphate (0.1 M PO_4H^{2-}), even though their pKs, larger than that of the cytidine mononucleoside, make them efficient catalysts. The exchange times are nearly independent of pD (uncorrected) even at pD 3 and 6.9, values for which the tetramer is partly dissociated (Figure 10).

Exchange of the thymidine imino protons is catalyzed by OH$^-$. At 0, 15, and 25 °C, the efficiency of the catalyst is 5 times lower for the d(TC_5) and d($T_2C_8T_2$) tetramers, and 2

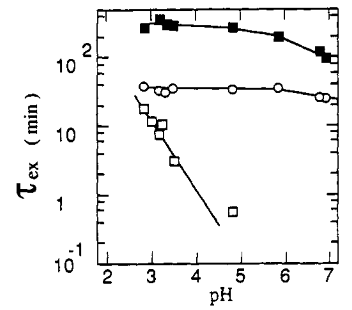


FIGURE 10: Effect of pH on proton exchange in the $d(T_2C_8T_2)$ tetramer at 0 °C. The five imino protons with the slow and common exchange rate (see Figure 9) (○); internal amino protons (7 ± 1 protons per strand) (■); external amino protons (7 ± 1 protons per strand) (□).

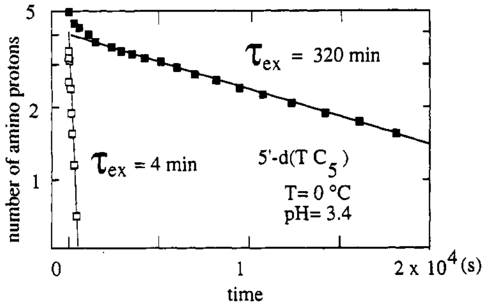


FIGURE 11: Intensity of the amino proton clusters as a function of the time after injection of a concentrated H_2O solution of the (TC_5) tetramer in D_2O. The initial intensities (40 s after transfer) are determined by reference to the intensity of the H6 proton lines. Four or five of the five external protons exchange in 4 min. Four of the five internal amino protons have the same exchange time (320 min), and one exchanges more rapidly. Experimental conditions: pH 3.2; 100 mM NaCl; T = 0 °C.

times lower for the $d(C_4TC_4)$ tetramer, than in the mononucleoside (Guéron et al., 1988; Guéron & Leroy, 1992).

Amino Protons. The exchange behavior is similar in all cases. In the $d(T_2C_8T_2)$ tetramer, exchange of the internal amino protons is slow, unaffected by added catalysts, and weakly pH-dependent (Figure 10). At 60 °C, pH 5.8, there is no effect of sodium acetate (1 M $CH_3CO_2^-$), phosphate (0.1 M PO_4H^{2-}), or Tris (5 mM Tris base). At 0 °C, pH 5.8, exchange is actually slowed 3-fold in 1 M solutions of potassium phosphate or sodium acetate, and also of sodium chloride, indicating that the effect is nonspecific and related to ionic strength.

The results of a real-time exchange experiment on the $d(TC_5)$ tetramer at 0 °C are plotted in Figure 11. The cluster assigned to the internal amino protons has an intensity of five protons (per strand). The decay corresponds to an exchange time of 20 min for one proton and of 320 min for the four others. Under the same conditions, 7 ± 1 of the eight internal amino protons of $(T_2C_8T_2)$ exchange in 300 min.

In contrast, exchange of the external amino protons is much faster and is sensitive to added catalysts. In the $d(TC_5)$ tetramer, four out of the five protons exchange in 4 min at 0 °C and pH 3.4 (Figure 11). The fifth may exchange faster. Exchange is catalyzed by OH^- according to $\tau_{ex} = 1/(1.7 \times 10^8[OH^-])$ s (Figure 10), a rate moderately smaller (20 times) than when OH^- acts on the amino protons of N^3-methyl cytidine (Kettani and Leroy, manuscript in preparation). In contrast, the rates of catalysis by acetate, phosphate, Tris, and citrate at 15 and 35 °C, and by phosphate at 0 °C, are reduced by factors of 10^3 to 10^4 as compared to the mononucleoside (McConnell & Politowski, 1984)!

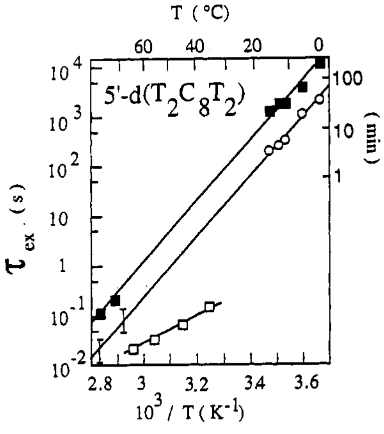


FIGURE 12: Proton exchange times in the $d(T_2C_8T_2)$ tetramer versus the inverse of temperature at pH 5.7: internal amino protons (■); external amino protons (□); the most slowly exchanging imino protons (○) and vertical segments. The latter span the distribution of individual exchange times. The low-temperature data are from real-time measurements; the high temperature data are from T_1 measurements.

Temperature Dependence of Proton Exchange. Imino proton spectra of the $d(T_2C_8T_2)$ tetramer at different temperatures are shown in Figure 7. The frequency shifts (ca. 0.003 ppm/°C for C and 0.006 ppm/°C for T) are comparable to those observed in Watson–Crick duplexes. The line widths of C^+ imino protons increase with temperature due to exchange. Strand dissociation is less than 10% at 60 °C; it is 30% at 75 °C and 70% at 80 °C in this experiment. Three imino peaks remain clearly visible at 75 °C.

The activation energies derived from the Arrhenius plot of exchange times between 0 and 70 °C (Figure 12) are the same for the imino protons and for the internal amino protons: 110 kJ/mol. The activation energy for exchange of the external amino protons, measured at pH 5.5 between 40 and 70 °C, is different: 65 kJ/mol.

DISCUSSION

The formation of C·C^+ base pairs was originally reported 30 years ago by Akinrimisi et al. (1963) and by Langridge and Rich (1963), who proposed the hemiprotonated base pair displayed in Scheme I, and by Hartman and Rich (1965), who established that poly(rC) exhibits a structural transition upon binding one proton for two cytidines. The C·C^+ pair had been observed in crystals of cytosine-5-acetic acid (Marsh et al., 1962) and was later found in crystals of 1-methylcytosine hemiiodide hemihydrate (Kistenmacher et al., 1978), of an intercalation complex of the parallel $r(CA)_2$ duplex and proflavine (Westhof & Sundaralingam, 1980; Westhof et al., 1980), and of ammonium deoxycytidylyldeoxyguanosine (Cruse et al., 1983).

Investigations by optical rotary dispersion (Fasman et al., 1964; Guschlbauer, 1967; Brahms et al., 1967; Gray et al., 1980; Brown et al., 1985; Edwards et al., 1990), small-angle X-ray scattering (Gulick et al., 1970), and UV and IR spectroscopy (Hartman & Rich, 1965) showed that poly(rC), poly(dC), and a number of C-rich oligonucleotides form a helical structure when the pH is lowered to 6.5–6, i.e., more than 2 pH units above the cytidine pK. NMR spectroscopy has been applied to the alternating oligonucleotide $d(CT)_3$ (Sarma et al., 1986) and to $d(A_5C_5)$ (Luo et al., 1992). Preliminary reports on similar oligonucleotides, including their exchange properties, have also been issued by this laboratory (Leroy et al., 1990, 1992*b*). All the polynucleotides and oligonucleotides were either found or assumed to be double-stranded.

In contrast, it has now been shown that the acid form of $d(TC_5)$ is tetrameric rather than dimeric (Gehring et al., 1993). The resolution of the tetrameric structure rests mainly on NOESY and ^{31}P–^{1}H hetero-TOCSY spectra, together with spectral features of imino and amino protons that are reported in the present work. The structure, which has been denoted "tetrad", includes two parallel-stranded, base-paired duplexes, intimately associated antiparallel to one another, with the $C{\cdot}C^+$ base pairs face-to-face and fully intercalated (Gehring et al., 1993). The new structural feature of fully intercalated base pairs is designated the "i-motif". The structure is sketched in Figure 1.

Stoichiometry and Stability of the Acid Multimers

The equilibrium dilution experiment with electrophoresis monitoring (Gehring et al., 1993) demonstrates directly that $d(TC_5)$ forms a tetramer at acid pH. The similar experiment with NMR monitoring which is presented here leads to the same conclusion for $d(TC_3)$. The gel filtration experiments generalize the existence of tetramers to every sequence studied. In situations of presumed equilibrium, dimers were absent or, if present, were a minor fraction at most. The $d(T_2C_8T_2)$ dimer, produced by quenching, transformed spontaneously into the tetramer.

Thus, tetramer formation by strands containing stretches of deoxycytidine may be a general phenomenon. The stretch may be as short as three units, as in $d(TC_3)$. The formation of tetramers in other sequences is an open question. As suggested elsewhere (Gehring et al., 1993), the i-motif could also form within one strand, for instance in a repeated sequence such as $d(C_4A_2)_4$, which is found in telomeres. One should consider whether tetramers could be formed generally by association of parallel-stranded duplexes such as those recently reported (Rippe et al., 1992).

The equilibrium constant $K_{1,4}$ of $(1400\ \mu M)^3$, measured above for the dissociation of the $d(TC_3)$ tetramer into single strands, corresponds to a free energy of –45.5 kJ/mol, or –7.6 kJ/mol per $C{\cdot}C^+$ base pair. The corresponding numbers for $d(TC_5)$ are $(3.8\ \mu M)^3$ at 10 °C and –8.48 kJ/mol per $C{\cdot}C^+$ base pair (Gehring et al., 1993). These free energies are therefore comparable to that for the formation of the B-DNA duplex d(CGCGATCGCG), –6 kJ/mol per base pair (Leroy et al., 1988).

The corresponding constant $K_{2,4}$ for the dissociation of the tetramer into two dimers is also of interest. An upper limit is provided by the observation that no dimer of $d(TC_3)$ is observed in the experiments of Figure 3, and we take this to mean that the dimer concentration, in strands, is less than 10% of the total concentration. This gives an upper limit of 30 μM for $K_{2,4}$, which corresponds to at least –24.1 kJ/mol or –4.0 kJ/mol per $C{\cdot}C^+$ base pair of the tetramer. By difference, the free energy of formation of a dimer from single strands should be at most –3.6 kJ/mol per $C{\cdot}C^+$ pair. Thus, the free energy released by packing base pairs into the i-motif of a tetramer is at least as large as that released in forming the base pair itself.

Assuming similar stacking, the linear charge density is the same in a duplex and in the tetramer formed by base-pair intercalation of two identical duplexes (Figure 1), hence the duplex to tetramer transformation is electrostatically indifferent. But the formation of a charged multimer from single strands which are neutral (as is the case when the cytidines are protonated, i.e., when the pH is well below the p*K* of 4.0 for cytidine) is electrostatically unfavorable, hence the multimers should be stabilized by salt. Conversely, the same multimer, when formed from single strands with a larger total charge than the multimer (above pH 4.0), should be destabilized by salt. Preliminary observations (not shown) are consistent with this expectation, in line with similar observations regarding acid poly(rC) (Hartman & Rich, 1965; Guschlbauer, 1967).

Base Pairing and Strand Direction in Tetramers

The structure of the $d(TC_5)$ tetrad (Gehring et al., 1993) is built on symmetrical hemiprotonated base pairs (Scheme I) formed by corresponding residues of parallel strands. Although this base-pairing mode is an obvious possibility for a duplex, tetramers allow more possibilities. It is therefore important to establish Scheme I as firmly and precisely as possible. The amino proton spectrum (Figure 5) provides the crucial evidence for the structure and symmetry of the base pairs. It also gives the best lower limit on the rate of the imino proton jumping process which is responsible for this symmetry.

The key properties of the amino proton spectrum are (1) the arrangement in only two clusters, one for the internal and one for the external amino protons, (2) the values of the chemical shifts, and (3) the small line widths.

The NMR spectrum is analyzed by comparison with the amino proton spectra of *neutral cytidine* in the C·G Watson–Crick pair of B-DNA or Z-DNA and with the spectrum of *protonated cytidine* in the Hoogsten $C^+{\cdot}G$ pair found in DNA triplexes. In B-DNA, the chemical shifts are around 8.5 and 6.8 ppm for the internal and external amino protons, respectively (Patel, 1976), and they are 8.4 and 6.0 ppm in Z-DNA (Kochoyan et al., 1990). In DNA triplexes, the C^+ resonances are found at 10.0 and 9.0 ppm (Rajagopal & Feigon, 1989). The downfield shifts of 2 ppm upon protonation are the same as in the mononucleotide (Raska, 1974; Kettani and Leroy, manuscript in preparation). In all cases the internal amino proton resonance is the most downfield, as in the $C{\cdot}C^+$ tetramers.

If the $d(TC_5)$ tetramer contained C and C^+ bases in slow exchange, there would be *four* amino proton clusters at chemical shifts close to the above. In contrast, the amino proton spectrum displays only *two* clusters at ca. 9.2 and 8.2 ppm. Each of these values is about midway between those given above. Therefore, the spectrum is readily interpreted as that of amino protons whose chemical shift switches quickly between the values for C and C^+, as expected if paired cytidines share one imino proton equally and quickly on the relevant NMR time scale. In agreement with this interpretation, the imino proton spectrum integrates to half a proton per cytidine. The spectra of other oligonucleotides present the same properties (Figures 3, 8, and 9).

A lower limit on the jumping rate of the imino proton is derived from the width of the amino proton spectrum. The spectrum of the $d(TC_3)$ tetramer exhibits a fully resolved internal amino proton line at 9.1 ppm (Figure 3). The 15 Hz full width at half-height is 9 Hz broader than the H6 proton peak of thymidine (6 Hz) at 6.3 ppm. Thus it is safe to assume that the broadening δ due to switching between C and C^+ is less than 10 Hz, and it could be much less. (In the other tetramers also, there is no visible exchange broadening of the amino proton peaks, but the lines are broader so that the lower limit obtained is less stringent.) In the fast exchange regime, the exchange broadening δ is equal to $(\pi/2)\Delta^2\tau$, where

Δ is the amino proton shift due to N3 protonation, and τ is the lifetime of each of the C and C^+ states (Abragam, 1961). With a Δ value ca. 2 ppm, hence 720 Hz, and assuming that $\delta < 10$ Hz, we obtain a lower limit of 8×10^4 s^{-1} for τ^{-1}.

The equal intensities of the amino proton clusters, the different rates of exchange with water of the two amino protons, and the observability of the imino proton spectrum (which requires protection against exchange with water), are all consistent with only one (Scheme I) among the possible hydrogen-bonding configurations for $C{\cdot}C^+$ pairs (Arnott et al., 1974). In particular, they are incompatible with a configuration proposed earlier for the base pairs formed by $d(C_4A_4T_4C_4)$ (Gray & Cui, 1984), since this configuration has one internal amino proton (N4H···N3) and three external amino protons per base pair.

But structures other than base pairs must also be considered. For instance, in a manner similar to that observed for guanosine in sequences such as $d(G_4T_2)_4$ (Aboul-ela et al., 1992; Kang et al., 1992; Smith & Feigon, 1992), four cytidines could form a planar, heme-like ring, each base donating one hydrogen bond by its amino group to one neighbor and accepting one by its O2. The four N3 nitrogens located inside the ring would share two imino protons. This configuration is incompatible with the observed imino proton spectrum because the ring is large and accessible to water molecules with which the imino protons could exchange rapidly.

Altogether, we feel that Scheme I is very strongly supported by the exchangeable proton spectrum and that it is therefore a sound foundation for the resolution of the tetramer structure via 2D NMR (Gehring et al., 1993).

Scheme I may be built either on antiparallel strands with one cytidine in the *anti* conformation and one in the *syn* conformation or on parallel strands with both cytidines *anti*. The latter is proved by the H6–H3′ NOESY cross-peaks (Gehring et al., 1993); hence the backbones of the paired bases must be parallel.

Which bases are paired together? This question is solved by consideration of the imino proton spectrum. We have just seen (from the amino proton spectrum) that each base switches between the protonated and the unprotonated states faster than 8×10^4 s^{-1}. This conversion is due to transfer of the imino proton between the two bases. The rate is fast on the scale of frequency splittings in the imino proton spectrum, and this spectrum should consist of one peak per base pair.

In the case of $d(T_2C_8T_2)$ at −5 °C, the imino proton spectrum consists of six peaks with an intensity of 0.5 proton per base, and one with an intensity twice as large which splits in two around 15 °C (Figure 7). This corresponds nicely to one peak per cytidine residue of the single strand, showing that each base is paired to the corresponding one of another strand. If we were dealing with a duplex, we would conclude that each base of one strand is paired to the corresponding one on the other (parallel) strand. In particular, an antiparallel duplex is excluded even without reference to the H6–H3′ NOESY cross-peaks, because it would give rise, by symmetry, to an imino proton spectrum of only four peaks.

But again, in a tetramer, the eight-peak imino proton spectrum corresponds to 16 base pairs, and it could arise in more ways than one. Restricting our consideration to parallel-stranded base pairs, one can imagine, besides symmetric associations of two equivalent parallel-stranded duplexes, a circle built of misregistered strands and also a cruciform. The determination of the correct structure cannot rest on symmetry considerations alone. It requires distance information such as that provided by NOESY spectra. We have searched for the characteristic H1′–H1′ interresidue NOESY cross-peaks that are observed for the $d(TC_5)$ tetrad (Gehring et al., 1992). In all cases examined, i.e., $d(T_2C_8T_2)$, $d(C_4TC_4)$, $d(TC_3T)$, and $d(TC_3)$, the cross-peaks have been found (not shown), indicating that these tetramers are built, like the $d(TC_5)$ tetrad, on the i-motif of intercalated, face-to-face base pairs.

The same symmetry arguments apply of course to the $d(TC_5)$ tetramer itself (Figure 5). The cytidine imino proton spectrum may be interpreted as five peaks, each of intensity 0.5, with two peaks overlapping, whereas only three peaks, with intensities 1, 1, and 0.5, are expected for an antiparallel duplex with dangling T.

Lastly, we consider the information provided by the thymidine spectrum. There is only one T methyl line in $d(C_4TC_4)$, $d(TC_5)$, and $d(TC_3)$. This is compatible with structures in which all strands are identical. The very modest reduction of the exchange rate of the imino proton, as compared to isolated thymidine (see *d(TC₅)* under Results), indicates either that T·T base pairs (Koole et al., 1987, van Genderen et al., 1990) do not form in $d(TC_5)$, $d(T_2C_8T_2)$, and (C_4TC_4) tetramers or, if they do form, that their dissociation constant is larger than 0.2.

Exchange of the Labile Cytidine Protons with Water

Imino Protons. The exchange process of $C{\cdot}C^+$ imino protons is remarkable by its low rate and by the lack of effect of catalysts. This may be discussed with reference to two factors which distinguish $C{\cdot}C^+$ pairs from Watson–Crick pairs. First, the low p*K* of $CN3H^+$ promotes the efficiency of proton acceptors, *including water itself.* With a collision rate of 10^{10} s^{-1}, eq 1 gives a rate of 10^6 s^{-1} for direct proton transfer to water (p*K* = −1.7). Furthermore, the effect of 1 mM $CH_3CO_2^-$ or PO_4H^{2-}, computed with the collision rate of ca. 10^9 s^{-1}, derived from the rate of catalysis of the imino proton of uridine (Guéron et al., 1989), should be equivalent to that of water.

Second, the equivalence of the proton donor ($CN3H^+$) and of the intrinsic proton acceptor (CN3) makes *intrinsic catalysis* 10^5 times faster than in Watson–Crick base pairs, i.e., a rate of 10^{10} s^{-1}, 10^4 times faster than catalysis by water or by 1 mM acetate or phosphate and 10 times faster than by these anions at 1 M concentration. This explains the lack of effect of catalysts, as we now discuss.

There are three cases. In the first case, the lifetime of the open state is less than ca. 10^{-10} s, 100 times less than a typical lifetime for the open state of a Watson–Crick pair (Guéron & Leroy, 1992). During this short time, it is plausible that a tight water-bridged structure, such as that in Scheme II, will not be disrupted. If so, the intrinsic catalyst will dominate over added catalysts, as described above. In the second case, the open-state lifetime is longer than 10^{-10} s, but opening is still a single-pair event. One would expect the water-bridging structure to be maintained for the duration of the open state, as proposed for B-DNA. Exchange would then occur at each opening event. In the third case, opening causes the two bases to become widely separated, as could happen if opening involves many base pairs, for instance in a fraying mode. But the separation is not instantaneous. If a water-bridged structure is maintained for as little as 10^{-10} s, exchange by intrinsic catalysis would again occur at every opening.

Thus, for any opening event lasting more than 10^{-10} s, exchange will plausibly be open-limited. Since this time is very short in comparison to the duration of the open state in B-DNA, we conclude that *the imino proton exchange time is probably the same as the base-pair lifetime.* The latter is then remarkably long for many of the base pairs of the tetramers.

Amino Protons. The investigation of amino proton exchange in Watson–Crick base pairs is much more difficult than that of imino proton exchange. The main culprits are amino group rotation, high p*K*, and unfavorable chemical shift, resulting in complex and/or multiple exchange mechanisms and poorly resolved spectra. The $C{\cdot}C^+$ base pairs present us with a better case, since there is only one type of amino group, whose resonances fall further from water and in a less crowded spectral region. And since each base is protonated half of the time, the p*K* of the amino protons should be close (but see below) to that of the protonated mononucleoside, $pK \approx 9$ (McConnell & Politowski, 1984).

Little is known of the exchange pathway of C amino protons in DNA duplexes. On the basis of tritium/proton exchange measurements, Teitelbaum et al. (1975a,b) proposed that it involves protonation at the N3 position and hence requires opening of the base pair. The internal and external amino protons of cytidine exchange at the same rate in B-DNA (Teitelbaum et al., 1975b; Hartmann et al., 1982) and in Z-DNA oligonucleotides (Kochoyan et al., 1990) but at different rates in $C{\cdot}C^+$ tetramers, as shown in Figures 9, 10, 11, and 12. For instance, the exchange times are 320 and 4 min at 0 °C and pH 3.4 for the internal and external protons of the $d(TC_5)$ tetramer. This immediately constrains the rotation time of the amino group to 320 min or more vs 3 s for the mononucleoside when protonated and 0.1 s when neutral (Shoup et al., 1971; Kettani and Leroy, manuscript in preparation). Even more striking is the difference between the external amino proton in tetramers and the amino protons of N^3-methylcytidine, a model for protonated cytidine. In the model, the amino protons exchange as expected for a group with a p*K* of 9. In the tetramer, exchange catalysis by OH^- is 20 times less efficient, but catalysts such as acetate, citrate, phosphate, and Tris have an efficiency at least 1000 times smaller than in the model!

These data can be reconciled if the amino group p*K* of *protonated* cytidine is raised from 9 to 17 in the tetramers (eq 1). It will be argued elsewhere that such a p*K* shift may be expected between a nucleoside in water and one well stacked in a polynucleotide (Kettani and Leroy, manuscript in preparation). This is the situation which is relevant to exchange of the *external* amino proton, whose exchange rate is so large that exchange must occur from the *closed* base pair. We note that a similar increase of the p*K* of the amino group of *neutral* cytidine in DNA duplexes could explain the low efficiency (10^2 to 10^3 times lower than in the neutral mononucleoside) of OH^- as an exchange catalyst for the external amino proton of cytidine in Z-DNA (Kochoyan et al., 1990) and in B-DNA (Kettani and Leroy, manuscript in preparation). Acid catalysis of guanosine imino proton exchange (Guéron & Leroy, 1992) could be explained similarly via protonation on N7. Such p*K* shifts could also contribute to the dependence on structure of the sensitivity of nucleic acids to chemical attack. The exchange of the internal amino proton is discussed in the next section.

Internal Motions of Tetramers

Despite the differences with the Watson–Crick case, the imino proton exchange process in $C{\cdot}C^+$ base pairs can be understood within the same framework, as has been discussed above. On this basis, we assimilate the base pair lifetime with the imino proton exchange time.

In the $d(TC_5)$ (Table I) and $d(T_2C_8T_2)$ tetramers, some of the base-pair lifetimes are extraordinarily long, hundreds of times longer than those of Watson–Crick base pairs in B-DNA (Guéron & Leroy, 1992), B′-DNA (Leroy et al., 1988), or Z-DNA (Kochoyan et al., 1990). They are comparable to the longest lifetimes in tRNA (Leroy et al., 1985) and in complexes of bis-intercalating drugs (Leroy et al., 1992). Even in the very short $d(TC_3)$ tetramer, one base-pair lifetime is as long (a few seconds) at 0 °C as that of G·C pairs in Z-DNA.

The long lifetime is not a property of C^+, since it does not occur in the $C^+{\cdot}G$ component of $C^+{\cdot}G{\cdot}C$ triplets (Kettani and Leroy, manuscript in preparation). Nor is it a feature of parallel strands since it is not found in the reverse Watson–Crick A·T base pair of a parallel-stranded hairpin stem (Germann et al., 1989). Most probably, the very long lifetimes are related to the tetrad geometry and particularly to the feature of intercalated base pairs, the i-motif. Swinging out of a base may be inhibited because each groove of one duplex is occupied by one backbone of the other duplex. Alternatively, a pair may be clamped by those of the other duplex between which it is stacked, a situation similar to that produced by bis-intercalating drugs in B-DNA (Leroy et al., 1992*a*). [Long lifetimes of the base pairs which compose G-tetrads are suggested by the observation of long imino proton exchange times (Smith & Feigon, 1992), but no base-pair measurements have been reported.]

Does the opening mode concern a single base pair or is it connected to fraying? In the $d(TC_5)$ tetramer, the only one for which the imino protons are assigned, the lifetimes are many minutes for the C3, C4, and C5 pairs, but they are only 0.8 s for the C2 pair and 0.006 s for the C6 pair at 0 °C. The thymidine pair, if it is formed, lives less than 0.001 s. The observed lifetimes therefore increase nicely in the stacking order of the tetrad structure (Figure 1), if one ignores differences among the very long lifetimes.

The short lifetimes of the C2 and C6 pairs of $d(TC_5)$ may then be ascribed to fraying, as found for the Watson–Crick base pairs close to the extremities of duplexes. But we do not know whether the long-lived base pairs open individually or in a process related to fraying. The following observations argue against opening by fraying: in the $d(TC_5)$ tetramer, the long base-pair lifetimes are unequal; they are not greatly affected by ionic strength; the longest lifetime in the $d(TC_5)$ tetramer (9000 s, Table I) is longer than in the tetramer of the longer oligonucleotide $d(T_2C_8T_2)$ (1500 s, Figure 10); the exchange times in the $d(T_2C_8T_2)$ tetramer are similar when 70% of the oligonucleotide is in a single-stranded conformation (i.e., at pD 6.9, 0 °C) and when the proportion of single strand is negligible (i.e., between pH 3 and 6). The latter suggests that fraying by interaction with single strands, if it occurs, is slower than the opening mode which dominates imino proton exchange.

One may also consider the variation of the base-pair lifetimes with temperature. In the $d(T_2C_8T_2)$ tetramer, the lifetimes vary by a factor of 10 over 15 °C (Figure 12), and this corresponds to an activation enthalpy for opening of 110 kJ/mol. This value, which is also adequate for some of the long lifetimes in the $d(TC_5)$ tetramer (Table I), is moderately larger than the value of 79 kJ/mol for opening the A·T pair of the B-DNA duplex d(CGCGATCGCG) (Guéron & Leroy, 1992).

On the other hand, five of the eight base pairs of the duplexes of the $d(T_2C_8T_2)$ tetramer have the same lifetime (Figure 9). This suggests that these pairs open together, but it does not prove it. On the whole, the observations are in favor of single-pair opening at this time.

Lastly, we discuss briefly how the exchange properties of the internal amino proton provide information on the lifetime of the open state. Figure 12 shows that the exchange time of the internal amino protons of d($T_2C_8T_2$) tetramer is in a constant ratio (of eight) to the imino proton exchange time, i.e., *to the base-pair lifetime*, when the latter varies by five orders of magnitude between 0 and 75 °C. This implies that exchange of the internal amino proton proceeds through base pair opening and hence from the open state (contrary to the case of the external amino proton). The constant ratio also implies that exchange is opening-limited, a proposition supported by the absence of effect of pH on the exchange rate (Figure 10). In the open state, the base is hydrated, and the p*K*s are probably close to those of cytidine in water. Hence the p*K* of the amino proton of C^+ is probably close to 9, just like the p*K* of the guanosine imino proton. Thus intrinsic catalysis as in Scheme II, but with a water bridge between the *amino* proton of C^+ and the N3 acceptor of the neutral base opposite, could occur as for the *imino* proton of G in a Watson–Crick G·C pair, at the same rate of 10^6 s^{-1} (as described in *Theory of Exchange of Cytidine Protons*). The constraint of opening-limited exchange then puts *a lower limit of* 10^{-6} *s on the lifetime of the open state*, 100 times longer than in the Watson–Crick base pairs.

We must still explain why exchange of the internal amino proton is eight times slower than the base pair opening event that allows imino proton exchange, rather than equal to it. We note in passing that the exchange is *slower*, not faster than that of the imino proton. This excludes any "barely-open" mode which would promote exchange of the internal amino proton separately from that of the more deeply buried imino proton. A look at Scheme I suggests that the opening event may involve a hinging motion that breaks open only one of the two amino hydrogen bonds. A given amino proton will then exchange only during one out of two such opening events. Also, exchange will occur efficiently only if the imino proton is on the same base as the free amino proton, so that the latter, being on C^+, has a p*K* of 9 rather than 17.7. If the imino proton ends up randomly on one side or the other, and assuming that it remains on the same side for most of the lifetime of the open state, it is only one out of four opening events which will produce exchange of a given amino proton. Moreover, the distribution may not be quite random. The hydrogen bond of the amino group on C^+ (that with the low p*K*) may be stronger than that of the amino group on the neutral base and may not be broken quite as frequently, thus giving rise to a ratio larger than four.

CONCLUSION

The formation of tetramers by stretches of deoxycytidine at acid pH is shown to be a rather general phenomenon. Similar imino proton spectra, NOEs, and exchange properties indicate that all the tetramers studied here have a structure similar to that of the d(TC_5) tetramer, with cytidine base pairs organized as an i-motif (Gehring et al., 1993).

The properties of exchangeable protons are precious for characterizing the base-pair structure and symmetry, as well as the strand directions. Imino proton exchange from C·C^+ pairs is predicted and found to be unaffected by catalysts, in agreement with the model of intrinsic catalysis originally developed for the interpretation of imino proton exchange in Watson–Crick base pairs. The base-pair lifetimes in the tetramers are extremely long, a property probably related to the structure of intercalated base pairs.

For the first time in a structural study, the amino protons provide crucial information, particularly on the dynamical symmetry of the C·C^+ pairs and on the lifetime of the open state, which is apparently at least 100 times longer than in Watson–Crick duplexes.

ACKNOWLEDGMENT

We thank A. Moraillon and M. Leblond for synthesis and purification of many of the oligonucleotides used in this study, Dr. J. M. Schmitter for advice on gel filtration chromatography, and Professor A. Wishnia for discussions in the early stages of this work.

REFERENCES

Aboul-ela, F., Murchie, A. I. H., & Lilley, D. M. J. (1992) *Nature (London) 360*, 280–282.

Abragam, A. (1961) *The Principles of Nuclear Magnetism*, Chapter 10, Oxford University Press, London.

Adler, A., Grossman, L., & Fasman, D. G. (1967) *Proc. Natl. Acad. Sci. U.S.A. 57*, 423–430.

Akinrimisi, E. O., Sander, C., & Tso, P. O. P. (1963) *Biochemistry 2*, 340–344.

Becker, E. D., Miles, H. T., & Bradley R. B. (1965) *J. Am. Chem. Soc. 87*, 5575–5582.

Brahms, J., Maurizot, J. C., & Michelson, A. M. (1967) *J. Mol. Biol. 25*, 465–480.

Brown, D. M., Gray, D. M., Patrick, M. H., & Ratliff, R. L. (1985) *Biochemistry 24*, 1676–1683.

Cruse, W. B. T., Egert, E., Kennard, O., Sala, G. B., Salisbury, S. A., & Viswamitra, M. A. (1983) *Biochemistry 12*, 1833–1839.

Dos Santos, C., Rosen, M., & Patel, D. J. (1989) *Biochemistry 28*, 7282–7288.

Edwards, E. L., Patrick, M. H., Ratliff, R. L., & Gray, D. M. (1990) *Biochemistry 29*, 828–836.

Fasman, G. D., Lindblow, C., & Grossmann L. (1964) *Biochemistry 8*, 1015–1021.

Forsèn, S., & Hoffman, R. A. (1963) *J. Chem. Phys. 39*, 2892–2901.

Gehring, K., Leroy, J. L., & Guéron, M. (1993) *Nature (London)* (in press).

Genest, D., Mazeau, K., & Ptak, M. (1987) *J. Biomol. Struct. Dyn. 5*, 67–78.

Germann, M. W., Vogel, H. J., Pon, R. T., & van de Sande, J. H. (1989) *Biochemistry 28*, 6220–6228.

Gray, D. M., & Cui, T. (1984) *Nucleic Acids Res. 12*, 7565–7580.

Guéron, M. (1978) *J. Magn. Reson. 30*, 515–520.

Guéron, M., & Leroy, J. L. (1992) in *Nucleic Acids and Molecular Biology* (Eckstein, F., & Lilley, D. M., Eds.) Vol. 6, Springer-Verlag, New York.

Guéron, M., Kochoyan, M., & Leroy, J. L. (1987) *Nature (London) 382*, 89–92.

Guéron, M., Charretier, E., Hagerhorst, J., Kochoyan, M., Leroy, J. L., & Moraillon, A. (1989) in *Biological Structure, Dynamics, Interactions & Expression* (Sarma, R. H., & Sarma, M. H., Eds.) Proceedings of the Sixth Conversation in Biomolecular Stereodynamics, Vol. 2, Adenine Press, New York.

Guéron, M., Plateau, P., & Decorps, M. (1991) *Prog. Nucl. Magn. Reson. Spectrosc. 23*, 135–209.

Guéron, M., Plateau, P., Kettani, A., & Decorps, M. (1992) *J. Magn. Reson. 96*, 541–550.

Gulik, A., Inoue, H., & Luzzati, V. (1970) *J. Mol. Biol. 53*, 221–238.

Gupta, M. H., & Sarma R. H. (1986) *FEBS Lett. 205*, 223–229.

Guschlbauer, W. (1967) *Proc. Natl. Acad. Sci. U.S.A. 57*, 1441–1448.

Hartman, K. A., Jr., & Rich, A. (1965) *J. Am. Chem. Soc. 87*, 2033–2039.

Hartmann, B., Pilet, J., Ptack, M., Ramstein, J., Malfroy, B., & Leng, M. (1982) *Nucleic Acids Res. 10*, 3261–3277.

Henderson, E., Hardin, C. C., Volk, S. K., Tinoco, I., Jr., & Blackburn, E. (1987) *Cell 51*, 899–908.

Kang, C., Zhang, X., Ratliff, R., Moyzis, R., & Rich, A. (1992) *Nature (London) 356*, 126–131.

Kellogg, G. W. (1992) *J. Magn. Reson. 98*, 176–182.

Kistenmacher, T. J., Rossi, M., & Marzilli, L. G. (1978) *Biopolymers 17*, 2581–2585.

Kochoyan, M., Leroy, J. L., & Guéron, M. (1987) *J. Mol. Biol. 196*, 599–608.

Kochoyan, M., Leroy, J. L., & Guéron, M. (1990) *Biochemistry 29*, 4799–4805.

Koole, L. H., van Genderen, M. H. P., & Buck, H. M. (1991) *J. Am. Chem Soc. 109*, 3916–3921.

Langridge, R., & Rich, A. (1963) *Nature (London) 198*, 725–728.

Leroy, J. L. (1990) *Regard Biochim. 5*, 57–65.

Leroy, J. L., Broseta, D., & Guéron, M. (1985) *J. Mol. Biol. 184*, 165–178.

Leroy, J. L., Kochoyan, M., Huynh-Dinh, T., & Guéron, M. (1988) *J. Mol. Biol. 20*, 223–238.

Leroy, J. L., Kettani, A, Kochoyan, M., & Guéron, M. (1990) Abstracts of the XIV International Conference on Magnetic Resonance in Biological Systems, Warwick, U.K.

Leroy, J. L., Gao, X., Guéron, M., & Patel, D. J. (1991) *Biochemistry 30*, 5653–566.

Leroy, J. L., Gao, X., Misca, V., Guéron, M., & Patel, D. J. (1992a) *Biochemistry 31*, 1407–1415.

Leroy, J. L., Gehring, K., & Guéron, M. (1992b) Abstracts of the XV International Conference on Magnetic Resonance in Biological Systems, Jerusalem.

Luo, J., Sarma, M. H., Yuan, R., & Sarma, R. H. (1992) FEBS *Lett. 306*, 223–228.

Marsh, R. E., Bierstedt, R., & Eichhorn, E. L. (1962) *Acta Crystallogr. 15*, 310–316.

McConnell, B. (1978) *Biochemistry 17*, 3168–3176.

McConnell, B., & Politowski, D. (1984) *Biophys. Chem. 20*, 135–148.

Patel, D. J. (1976) *Biopolymers 15*, 533–558.

Plateau, P., & Guéron, M. (1981) *J. Am. Chem. Soc. 104*, 7310–7311.

Rajagopal, P., & Feigon, J. (1989) *Biochemistry 28*, 7859–7870.

Raszka, M. (1974) *Biochemistry 13*, 4616–4622.

Rippe, K., Fritsch, V., Westhof, E., & Jovin, T. M. (1992) *EMBO J. 10*, 3777–3786.

Shoup, R. R., Becker E. D., & Miles, H. T. (1971) *Biochem. Biophys. Res. Commun. 43*, 1350–1353.

Smith, F. W., & Feigon, J. (1992) *Nature (London) 356*, 164–168.

Teitelbaum, H., & Englander, S. W. (1975) *J. Mol. Biol. 92*, 79–92.

van Genderen, M. H. P., Hilbers, M. P., Koole, L. H., & Buck, H. M. (1990) *Biochemistry 29*, 7838–7845.

Westhof, E., & Sundaralingam, M. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1852–1856.

Westhof, E., Roa, S. T., & Sundaralingam, M. (1980) *J. Mol. Biol. 142*, 331–361.