Raman Signature of the Four-Stranded Intercalated Cytosine Motif in Crystal and Solution Structures of DNA Deoxycytidylates d(CCCT) and d(C_8)†

James M. Benevides,‡ ChulHee Kang,§ and George J. Thomas, Jr.*,‡

Division of Cell Biology and Biophysics, School of Biological Sciences, University of Missouri−Kansas City, Kansas City, Missouri 64110-2499, and Department of Biochemistry and Biophysics, Washington State University, Pullman, Washington 99164-4660

Received December 13, 1995; Revised Manuscript Received March 6, 1996⊗

ABSTRACT: The Raman spectral signature of the four-stranded cytosine structure formed by intercalation of two hemiprotonated and parallel-stranded oligodeoxycytidylate duplexes (so-called *i* motif) has been obtained from the crystal structure of d(CCCT) [Kang, C. H., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R., & Rich, A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91,* 11636−11640]. Identification of Raman markers diagnostic of the cytosine quadruplex is complemented by results obtained in a pH titration of 2′-deoxycytosine-5′-monophosphate (5′-dCMP) to show that the spectral fingerprint associated with N3 protonation of cytosine is distinct from that of quadruplex formation. The Raman spectrum thus provides a definitive basis for evaluating quantitatively both the extent of cytosine quadruplex formation and the degree of cytosine N3 protonation in DNA. Application to aqueous d(CCCT) and d(C_8) demonstrates that the four-stranded intercalated structure is formed by both of these oligodeoxycytidylates in aqueous solution. Whereas both 5′-dCMP and the d(CCCT) quadruplex exhibit a midpoint of titration (apparent pK_C) of 4.5 ± 0.2 at 10 °C, cytosine protonation in d(C_8) is shifted significantly toward the physiological range, with pK_C = 5.8 ± 0.2. The difference in pK_C between the two quadruplexes is equivalent to a free energy difference of 1.7 kcal/mol at 10 °C. The present findings extend the library of Raman conformation markers to deoxycytidylate residues in the novel *i* quadruplex. The significance of these results for probing solution conformations of telomeric DNA sequences is also considered.

Nucleic acid polymorphism, which is dependent upon base sequence and local molecular environment, is an essential structural property of the genetic material. Recent investigations have focused on the capability of gene regulatory proteins (Steitz, 1990) and specific counterions (Cheng & Pettitt, 1992) to alter the global structure of DNA. The precise interplay between ionic environment and nucleic acid conformation is of particular interest and importance, as has been illustrated in the novel quadruplex structures of telomeric DNA sequences and in the variety of triplex structures accessible to purine/pyrimidine sequences. Recent surveys have been given by Williamson (1994), Leroy and Guéron (1995), and Frank-Kamenetskii and Mirkin (1995). An intriguing feature of many multistranded nucleic acids is the capacity to undergo large conformational changes with relatively small changes in local concentrations of protons or metal cations. For example, in the case of guanine quadruplexes of telomeric DNA (Sen & Gilbert, 1990), switching between alternative quadruplexes in solution can be induced by relatively small changes in concentration of Na^+ and K^+ ions (Miura *et al.*, 1995). Similarly, nucleic acid polymorphism can depend sensitively upon solution pH.

Conformational switching of DNA under the control of solution pH is the result of protonation (or deprotonation) of specific base sites, which can be inferred from the apparent base p*K* values. Among the four bases, the p*K* value closest to neutrality and most likely to be of biological significance is that of cytosine. The p*K* governing cytosine N3 protonation (pK_C) is nominally about 2.5 units below neutral pH but can be elevated to the physiological range when the base is incorporated within a nucleic acid secondary structure (Inman, 1964). While cytosine N3 protonation eliminates Watson−Crick pairing and therefore destabilizes the *B* DNA duplex, it can enable alternative pairing schemes to compensate in part for the loss of Watson−Crick pairs. An example is pairing between protonated and unprotonated cytosines to yield the hemiprotonated base pair, $(C{\cdot}C)^+$, depicted in Figure 1A. Stabilization of secondary structure by $(C{\cdot}C)^+$ pairs may be especially important for nucleic acid sequences rich in cytosine residues.

A parallel-stranded duplex stabilized by $(C{\cdot}C)^+$ pairs was first proposed by Langridge and Rich (1963) on the basis of fiber X-ray diffraction studies of poly(rC). The salient features of this structure have been confirmed by both infrared (Hartman & Rich, 1965) and Raman spectroscopy (Chou & Thomas, 1977). Other secondary structure models incorporating $(C{\cdot}C)^+$ pairs have been described recently (Robinson *et al.*, 1992; Robinson & Wang, 1993; Jaishree & Wang, 1993; Liu *et al.*, 1993; Wang & Patel, 1994). Among these, the four-stranded (quadruplex) intercalated structure formed by oligodeoxcytidylate sequences and termed the *i* motif (Gehring *et al.*, 1993; Leroy *et al.*, 1993) is of particular interest. In the *i* form, two hemiprotonated

† This is paper LX in the series Raman Spectral Studies of Nucleic Acids. Support by NIH Grant GM54378 is gratefully acknowledged (G.J.T.).

* Author to whom correspondence may be addressed.
‡ University of Missouri−Kansas City.
§ Washington State University.

⊗ Abstract published in *Advance ACS Abstracts,* April 15, 1996.

S0006-2960(95)02942-4 CCC: $12.00

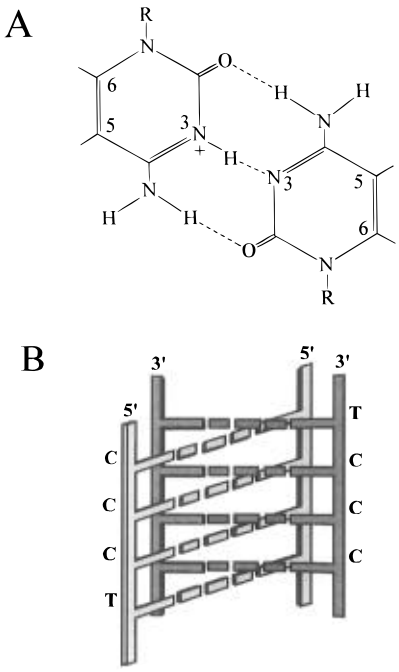

FIGURE 1: (A) Hemiprotonated cytosine base pair. (B) d(CCCT) in the four-stranded intercalated structure incorporating hemiprotonated cytosine base pairs and antiparallel arrangement of parallel duplexes (*i* motif).

and parallel-stranded cytosine duplexes are intercalated with one another in an antiparallel orientation, as depicted in Figure 1B. Details of this DNA structure have been proposed on the basis of NMR spectroscopy and X-ray crystallography (Leroy *et al.*, 1994; Manzini *et al.*, 1994; Berger *et al.*, 1995; Kang *et al.*, 1994,1995). The novel cytosine quadruplex is also of biological interest because a cytosine-rich sequence occurs naturally as a complement to the guanine-rich strand of telomeric DNA.

Raman spectroscopy has been demonstrated as an effective probe of nucleic acid secondary structure for both solution and crystalline samples [reviewed recently by Thomas and Tsuboi (1993)]. Raman spectra have been particularly useful in characterizing the solution conformations (Miura & Thomas, 1994), dynamics (Miura & Thomas, 1995), and phase transitions (Miura *et al.*, 1995) of the guanine-rich telomeric repeat (Oxy-4) of the ciliate *Oxytricha nova*. In the Oxy-4 sequence, which consists of four tandom repeats of T_4G_4, two quadruplex secondary structures have been identified on the basis of Raman conformation markers which are different from one another and distinct as well from those of canonical *B* DNA. The complementary Oxy-4 repeat, A_4C_4, has also been investigated (Miura & Thomas, 1994) and is of interest as a potential *i*-forming sequence.

In the present work we establish Raman markers diagnostic of the four-stranded intercalated cytosine motif by examination of the Raman spectrum of the authenticated *i* motif in the single crystal of d(CCCT). This structure has been solved at 1.4 Å resolution by X-ray crystallography (Kang *et al.*, 1994). The Raman markers identified for the *i* form are complemented by results obtained in a Raman pH titration of 2′-deoxycytosine-5′-monophosphate (5′-dCMP) to demonstrate that the spectral fingerprint associated with N3 protonation of cytosine is distinct from that of quadruplex formation. The results are used to establish the solution secondary structures and pK_C values in both d(CCCT) and d(C_8). The feasibility of Raman spectroscopy as a probe of both cytosine and guanine quadruplexes in telomeric DNA is demonstrated.

MATERIALS AND METHODS

1. Samples for Raman Spectroscopy. The mononucleotide 5′-dCMP (Sigma Chemical, St. Louis, MO) was dissolved to 0.5 M in 0.15 M NaCl solution. An initial sample volume of 200 μL at pH 8.0 $\pm$ 0.1 was titrated stepwise, in pH increments of 0.5, to a final pH of 1.0 by addition of 0.1 N HCl. An aliquot of 5 μL was removed at each step and transferred to a glass capillary (Kimax no. 34507), which was sealed for subsequent Raman analysis.

The cytosine oligomers, d(CCCT) and d(C_8), were synthesized on an Applied Biosystems model 381A synthesizer. Purifications were performed on an ISCO model 2350 HPLC system. The initial trityl-on separation was performed on a PRP-1 reverse phase column (Hamilton, Reno, NV). After recovery, samples were detritylated and again separated. The second separation resolved the desired detritylated sequence from contaminating byproducts of the detritylation step. Further details of oligomer purification protocols have been described (Benevides *et al.*, 1994). Each purified oligomer was dissolved to 35 μg/μL in 0.15 M NaCl (total volume $\approx$20 μL). As noted above, the pH was lowered incrementally to 2.5 by addition of 0.1 N HCl and aliquots ($\approx$2 μL) were removed and sealed in glass capillaries for subsequent Raman analysis. The sodium salt of d(CCCT) was dissolved to 2 mM in a solution containing 20 mM $MgCl_2$, 0.12 M spermine, 80 mM $SrCl_2$, 40 mM strontium cacodylate buffer (pH 6.0−7.0), and 10% 2-methyl-2,4-pentanediol (MPD). Crystals were grown by the hanging drop method over 30% MPD. The rectangular brick-shaped crystals appeared after 3 weeks and eventually grew to approximate dimensions of 0.5 $\times$ 0.5 $\times$ 1.0 mm. After X-ray authentication of the d(CCCT) crystal structure (Kang *et al.*, 1994), the crystal was transferred with mother liquor to a capillary which was sealed for analysis by Raman microspectroscopy.

2. Raman Spectroscopy of Solutions. Raman spectra of solutions of 5′-dCMP, d(CCCT), and d(C_8) were excited at 514.5 nm using an argon laser (Innova-70, Coherent, Inc., Santa Clara, CA) and were collected on a triple spectrograph (model 1877, Spex Ind., Edison, NJ) equipped with a liquid nitrogen-cooled CCD detector (model LN-CCD-1152UV, Princeton Instruments, Princeton, NJ). Usually, five exposures of 120 s duration were accumulated and five accumulations were averaged to produce the traces shown below. Spectra were collected with band resolution of 5 cm^{-1}. For each sample, the buffer spectrum was also collected and used to compensate the solution spectrum for contributions of the solvent (Benevides *et al.*, 1984).

Raman frequencies of well-resolved bands are accurate to within $\pm$1.5 cm^{-1}. For d(CCCT) and d(C_8), spectral intensities were normalized to the phosphodioxy stretching band at 1092 cm^{-1}, which is essentially invariant to pH change in the range of present interest. For 5′-dCMP, however, the corresponding phosphotrioxy stretching band (980 cm^{-1}) is sensitive to secondary ionization of the phosphate group and is not a reliable intensity standard over the entire pH range of interest. Accordingly, $KClO_4$ was added to 5′-dCMP solutions to provide an independent Raman intensity standard (928 cm^{-1} band of ClO_4^-).

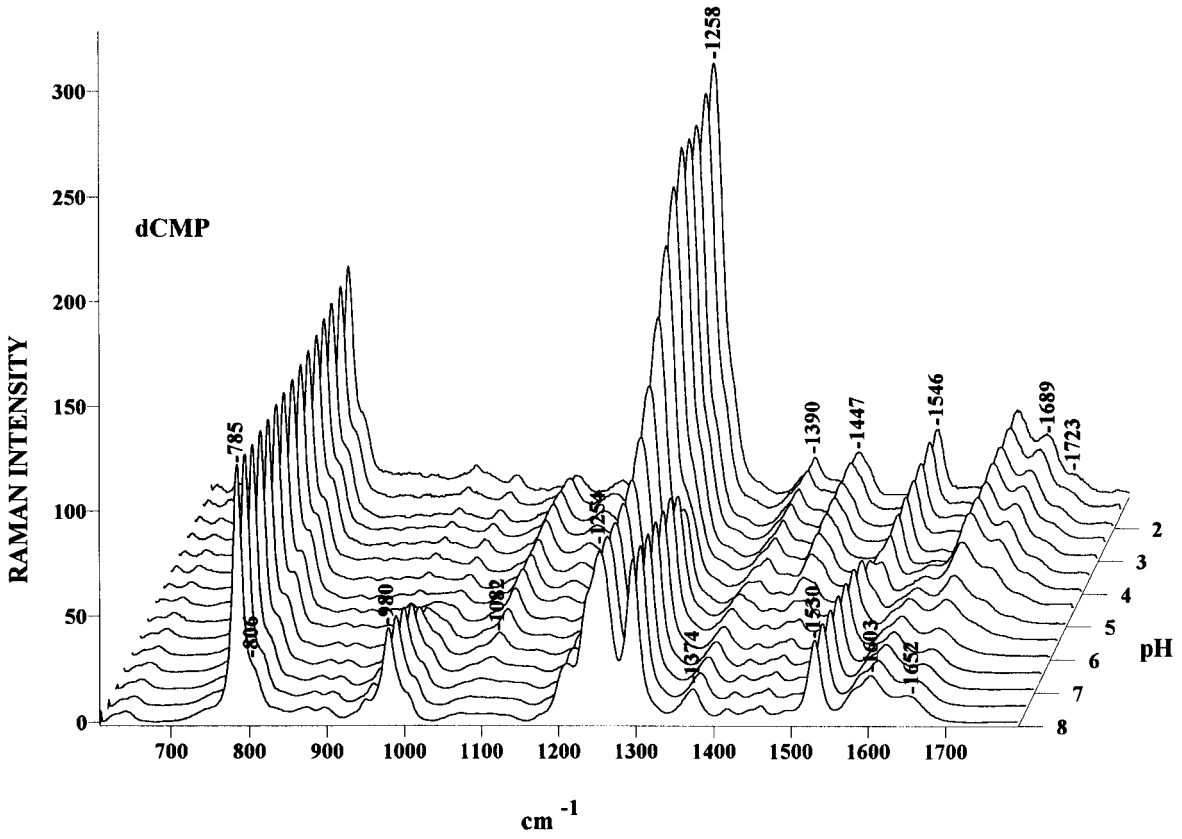

FIGURE 2: Raman spectra of aqueous 5′-dCMP as a function of pH.

3.Raman Microspectroscopy of d(CCCT) Crystals. Raman spectra of single crystals of d(CCCT) were obtained with 514.5 nm excitation (Innova-2, Coherent, Inc.) using the previously described Jobin−Yvon triple spectrograph incorporating a CCD detector and Olympus microscope (Thomas *et al.*, 1995). Laser power at the sample was always maintained below 15 mW, and the crystal was thermostated at 4 °C on a Physitemp model TR-4 thermoelectrically cooled cold-plate anchored to the microscope stage. Spectra shown below, collected at 5 cm^{-1} resolution, represent unsmoothed averages of 20 exposures (120 s/exposure). Methods of processing the spectral data have been described (Benevides *et al.*, 1993).

RESULTS AND DISCUSSION

1. Raman Spectroscopy and pH Titration (pK_C) of 5′-dCMP. In order to determine the pH sensitivity of Raman bands of cytosine, the Raman spectrum of 5′-dCMP was monitored in the range 8.0 > pH > 1.0. The results are shown in Figure 2 at pH increments of 0.5 units. The data indicate two discrete protonation events. The first, characterized by a shift of the Raman band at 980 cm^{-1} (PO_3^{2-} marker) to 1082 cm^{-1} [$P(OH)O_2^-$ marker], represents protonation of the 5′ phosphomonoester group. This spectral change provides a basis for Raman determination of the pK governing dissociation of the protonated phosphomonoester group (i.e., pK_P, for the equilibrium $P(OH)O_2^- = PO_3^{2-} + H^+$). The second protonation event, characterized by numerous Raman band frequency and intensity changes, represents protonation of the pyrimidine N3 site and provides a basis for Raman determination of the cytosine ring pK (i.e., pK_C, for the equilibrium $CH^+ = C + H^+$, where CH^+ refers to the N3-protonated mononucleotide). Table 1 catalogues the most important spectral changes accompanying the cytosine N3 protonation. In effect, Raman bands diagnostic of the neutral base at 1254, 1530, and 1603 cm^{-1} are replaced by bands diagnostic of the protonated base at 1258, 1390, 1546, 1689, and 1723 cm^{-1}.

On the basis of the Raman band intensity changes of Figure 2, we have determined the pK_P and pK_C values for 5′-dCMP. Typical Raman titration curves are shown in Figure 3; titration curves (not shown) for other pH sensitive bands of 5′-dCMP yield consistent results. The present values for pK_C (= 4.5 ± 0.2, based upon intensity decay of the 1254 cm^{-1} band) and pK_P (= 6.1 ± 0.2, based upon intensity decay of the 980 cm^{-1} band) are in good agreement with previous determinations (Hurst *et al.*, 1953; Cohn, 1955; Jordan, 1960; Borah & Wood, 1976; O'Connor *et al.*, 1976). We note that the band at 1254 cm^{-1} should be well suited to pK_C determination in oligonucleotides and nucleic acids owing to its high intrinsic intensity and great sensitivity to pH. The band at 1530 cm^{-1} is also favored for Raman spectrophotometric determination of pK_C by virtue of its separation from other bands in spectra of nucleic acids.

The Raman spectra of Figure 2 also reveal a pH dependent intensity change near 806 cm^{-1}. Because the 800−850 cm^{-1} interval in spectra of nucleic acids is critical to evaluating phosphodiester backbone conformation (Thomas & Tsuboi, 1993), it is important to assess whether the 806 cm^{-1} band of 5′-dCMP reflects a protonation event occurring in the base or in the phosphate group. The Raman titration curve, shown in Figure 3, reveals clearly that the protonation event corresponds to pK_P and not pK_C. The 806 cm^{-1} band of 5′-dCMP is therefore unrelated to N3 protonation and presumably reflects formation of the 5′-phosphomonoester monoanion [$COP(OH)O_2^-$]. A similar band with identical

Table 1: Raman Frequencies, Intensities and Assignments for dCMP, d(C_8) and d(CCCT) at High and Low pH[a]

dCMP		d(C_8)		d(CCCT)		
pH 8.0	pH 3.0	pH 8.3	pH 5.0	pH 7.5	pH 4.2	assignment[b]
628 (0.3s)		624 (0.5)		624 (0.3s)		C
	617 (0.2)		619 (0.5)		618 (0.3)	C^+, $(C{\cdot}C)^+$
643 (0.4)	643 (0.3)	643 (0.5b)	649 (0.9)	638 (0.3b)	643 (0.8)	C, C^+, $(C{\cdot}C)^+$
				668 (0.2)	668 (0.5)	T
746 (0.5s)	739 (0.6s)	743 (0.3s)		751 (1.0s)	741 (0.8)	d, T
785 (10)	787 (10)	785 (10)	788 (10)	782 (10)	788 (10)	C, C^+, $(C{\cdot}C)^+$, bk
806 (2.0s)	805 (3.6s)					bk [C-O-P]
			804 (2.7)		806 (3.4)	bk [O-P-O]
833 (0.5)						d
		848 (0.1)	849 (1.0)		849 (0.9)	bk [O-P-O]
873 (0.6)	875 (0.6)					d
		885 (0.3)	886 (1.0)		886 (1.0)	bk [d]
897 (0.6)	897 (0.6)					d
					906 (0.5s)	bk [d]
	924 (0.7s)	924 (0.1)		913 (0.1)		d
951 (0.9s)	951 (0.8)	950 (0.4s)		946 (0.3)		d
	966 (0.6s)	965 (0.8)	963 (1.3)	963 (0.4)	965 (1.4)	bk
980 (3.6)						$[PO_3^{-2}]$
		983 (0.8)		989 (0.7)		bk
1005 (1.1s)	1005 (0.7s)	1008 (0.7s)	1006 (0.8)		1003 (0.8s)	d
				1020 (0.7s)	1012 (1.2)	d
1074 (0.4b)	1059 (0.5s)	1058 (0.8s)	1046 (0.5b)	1058 (1.1s)	1052 (0.9s)	d [C−O]
	1082 (1.6)					$P(OH)O_2^-$
		1092 (2.1)	1092 (2.0)	1092 (1.8)	1092 (2.1)	bk [PO_2^-]
1136 (0.2s)	1140 (0.6)	1134 (0.3s)	1134 (0.4)	1137 (0.4s)	1140 (0.6)	d, T
	1199 (0.7s)	1189 (0.8s)	1175 (0.9)	1187 (1.1s)	1174 (1.2)	C, C^+, $(C{\cdot}C)^+$, T
1211 (2.3)		1212 (2.4)	1217 (1.8s)	1208 (2.4)	1207 (2.0)	C, C^+, $(C{\cdot}C)^+$, T
1238 (5.1s)		1240 (5.7s)		1238 (6.3)		C, T
1254 (6.6)	1258 (18)	1250 (6.2)	1254 (7.8)	1248 (5.5)	1257 (8.5)	C, C^+, $(C{\cdot}C)^+$
		1265 (5.1s)	1266 (9.2)	1265 (3.8)	1267 (8.6)	C, C^+, $(C{\cdot}C)^+$
1296 (6.3)	1279 (5.7s)	1292 (5.9)	1291 (5.2)	1292 (4.9)	1292 (5.0)	C, C^+, $(C{\cdot}C)^+$
	1315 (0.5)		1318 (0.9s)		1320 (0.4s)	C^+, $(C{\cdot}C)^+$
1374 (1.3)	1372 (0.9s)	1374 (1.3)	1372 (1.4)	1374 (3.0)	1376 (4.6)	T, C, C^+, $(C{\cdot}C)^+$
	1390 (1.6)		1388 (2.5)		1387 (3.3)	C^+, $(C{\cdot}C)^+$
1417 (0.5)	1415 (0.7s)	1414 (0.4)	1424 (1.0b)	1417 (0.8b)	1424 (1.0)	d
1447 (0.5s)	1447 (1.9)	1438 (0.4s)		1440 (0.6s)		C, C^+, $(C{\cdot}C)^+$
1461 (0.7)	1458 (1.3s)	1458 (0.6)	1458 (1.2)	1460 (0.8)	1457 (1.5)	d
1495 (0.4)					1477 (0.6s)	C, T
1530 (3.2)		1530 (3.0)		1530 (2.8)		C
	1546 (2.6)		1543 (2.0)		1543 (1.7)	C^+, $(C{\cdot}C)^+$
1581 (1.3s)		1583 (0.8s)		1583 (1.1s)		C
	1592 (0.3)					C^+, $(C{\cdot}C)^+$
1603 (1.8)		1603 (2.0)		1606 (2.2)		C
1652 (1.1)	1652 (3.4)	1659 (3.2)	1659 (5.8)	1660 (4.1)	1657 (5.1)	C, C^+, $(C{\cdot}C)^+$, T [C═O]
	1689 (2.4)					C^+, $(C{\cdot}C)^+$ [C═O]
	1723 (0.8)					C^+, $(C{\cdot}C)^+$ [C═O]

[a] Low pH forms correspond to N3-protonated cytosine in dCMP and N3-hemiprotonated base pairs in d(CCCT) and d(C_8). Frequencies are in cm^{-1} units. Numbers in parentheses are relative intensities with an arbitrary value of 10 assigned to the band near 785 cm^{-1} in each spectrum. [b] Abbreviations: b, broad; s, shoulder; C, cytosine residue; T, thymine residue; d, deoxyribose ring; bk, deoxyribose−phosphate backbone. Brackets indicate probable vibrational assignments to specific atoms or subgroups of the backbone or base residues.

pH behavior occurs in Raman spectra of 5′-monoribonucleotides (Hartman *et al.*, 1973). Thus, the 806 cm^{-1} band of 5′-dCMP, although occurring fortuitously in the same spectral interval as nucleic acid conformation markers (Thomas & Tsuboi, 1993), is due to the *phosphomonoester* monoanion and should not be confused with either a base vibration or a *phosphodiester* vibration of DNA.

2. pH Dependence of Raman Spectra of d(C_8) and d(CCCT). Raman spectra of d(CCCT) in neutral (pH 7.5) and acidic (pH 4.2) solutions are compared in Figure 4. Corresponding spectra for d(C_8) are compared in Figure 5. Table 1 lists the spectral frequencies and intensities for comparison with data of 5′-dCMP. Near neutral pH, both d(CCCT) and d(C_8) generate similar Raman spectra with respect to bands of the cytosine residues and phosphodiester backbone (*c*f. Figures 4 and 5). For both oligonucleotides at this pH, the data demonstrate that cytosines are in the neutral form and that the phosphodiester backbones lack Raman markers diagnostic of a uniformly ordered structure of the *A* (807 cm^{-1}) or *B* (835 cm^{-1}) type (Erfurth *et al.*, 1972; Prescott *et al.*, 1984). We observe instead a very weak and broad profile of Raman intensity in the 800−850 cm^{-1} interval, which overlaps the high-frequency side of the prominent cytosine marker band at 785 cm^{-1}. This diffuse Raman scattering is considered indicative of a wide distribution of torsion angles in the backbones of both single-stranded oligonucleotides.

At acidic pH, d(CCCT) and d(C_8) again yield Raman spectra which are similar to one another (cf. Figures 4 and 5) yet strikingly different from their neutral pH counterparts. The difference spectrum computed between neutral pH and acidic pH for d(CCCT) (Figure 4, bottom trace), is qualitatively similar to that of d(C8) (Figure 5, bottom trace) indicating that the pH-induced structure change is the same

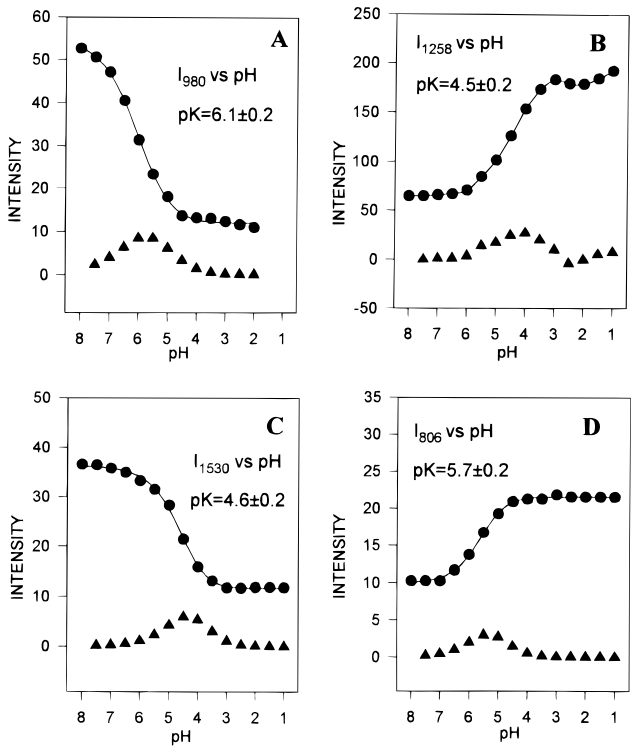

FIGURE 3: pH titration curves for selected Raman bands of 5′-dCMP. (A) 980 cm^{-1}; (B) 1258 cm^{-1}; (C) 1530 cm^{-1}; (D) 806 cm^{-1}. Filled circles indicate data points (normalized Raman intensities, arbitrary scale). First derivatives of the titration curves evaluated at the data points (filled triangles) are also shown.

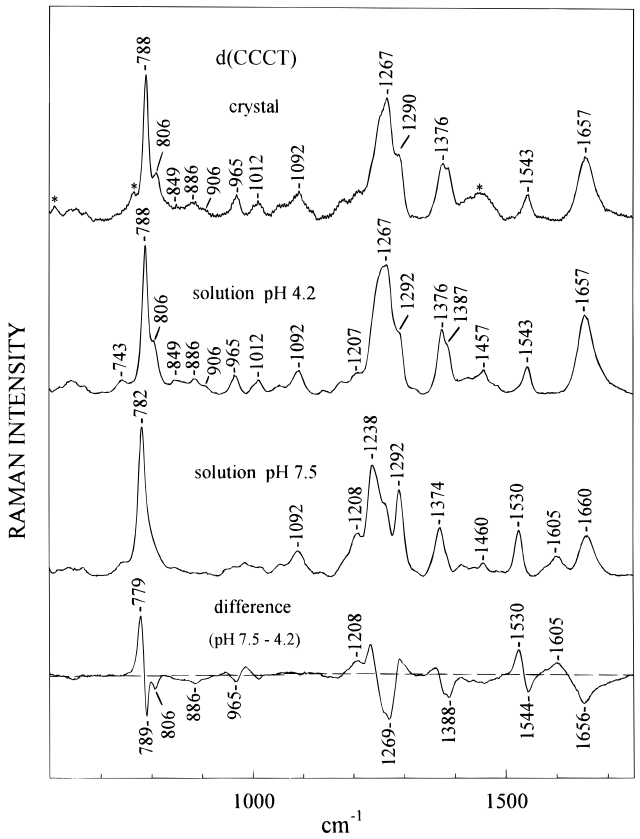

FIGURE 4: Raman spectra in the 600−1750 cm^{-1} region of a crystal of d(CCCT) (top trace) and solutions at pH 4.2 (second trace from top) and pH 7.5 (second trace from bottom). Asterisks in the crystal spectrum indicate Raman bands due to the mother liquor. The difference spectrum between pH 7.5 (minuend) and pH 4.2 (subtrahend) solutions is also shown (bottom trace).

for each oligonucleotide. The Raman bands eliminated near 1530 and 1605 cm^{-1} and those generated near 1388 and 1543 cm^{-1} with decreasing pH signal the protonation of cytosine rings. In addition, the numerous differences observed throughout the 800−1025 cm^{-1} region, particularly the sharpening of otherwise diffuse Raman scattering (800−850 cm^{-1}) to yield a discrete band near 806 cm^{-1}, signal the formation of a relatively well-ordered phosphodiester backbone. In each oligonucleotide, the secondary structure at acidic pH is characterized empirically by Raman markers near 806, 849, 886, 965, and 1010 cm^{-1}. The nature of the low-pH secondary structure which generates these marker bands is next discussed.

3. Raman Markers of the Cytosine Quadruplex in the d(CCCT) Crystal and Application to Solution Structures of d(CCCT) and d(C_8). The Raman spectrum of the d(CCCT) crystal is shown in Figure 4 (top trace), where it is compared with the solution spectrum at pH 4.2 (second trace from top). The d(CCCT) crystal structure (Kang *et al.*, 1994) has been authenticated as that of the four-stranded intercalated complex or so-called *i* motif. The one-to-one correspondence between Raman bands of the crystal and pH 4.2 solution indicates that the structures are essentially identical. This establishes the *i* form as the solution structure of d(CCCT) at pH 4.2. Extension to the Raman signature of d(C_8) (Figure 5) indicates that the cytosine octamer also adopts the intercalated structure in solution at pH 5.

4. Backbone Conformation Markers of the Cytosine Quadruplex. The d(CCCT) crystal structure (Kang *et al.*, 1994) reveals the following distribution of sugar ring conformations among the 12 deoxycytidine residues of the quadruplex: six C3′-*endo*, two C2′-*endo*, and four "envelope" conformers in which C2′, C3′, and C5′ atoms are on the same side of the furanose ring. The crystal structure also shows the four thymidines distributed equally between C2′-*endo* and C3′-*endo* conformations. In addition, different phosphodiester conformations occur along the four chains of the quadruplex. Two CCCT sequences from different strands exhibit a relatively flat and extended chain conformation, while the other two are characterized by phosphodiester groups that are sharply bent. In the phosphodiester stretching region of the Raman spectrum (800−850 cm^{-1}), we anticipate at least two bands corresponding to the two different types of chain conformation in the crystal structure. Curve decomposition analysis (not shown) verifies distinct band components at 806 and 849 cm^{-1} with an intensity ratio of about 2:1. On the basis of the proximity of the former to well-established backbone markers of *A* DNA (807 cm^{-1}) and *A* RNA (813 cm^{-1}) (Thomas & Tsuboi, 1993), we assign the 806 cm^{-1} component to the backbone of d(CCCT) which incorporates the C3′-*endo* nucleosides, i.e., the sharply bent phosphate chain. The 849 cm^{-1} component by default is assigned to phosphodiester groups comprising both C2′-*endo* and envelope conformers of the extended chain. The approximate 2:1 ratio in intensities of Raman bands at 806 and 849 cm^{-1} is not unexpected. A similar intensity difference is observed for the corresponding backbone markers of *A* and *B* DNA (Erfurth *et al.*, 1972; Prescott *et al.*, 1984). We note also that the proposed 849 cm^{-1} marker for nucleosides of d(CCCT) which are not in the *A* conformation is about 10−20 cm^{-1} higher than the frequency of the canonical *B* form marker of double helical DNA, the latter usually observed near 835 ± 5 cm^{-1}. The elevated

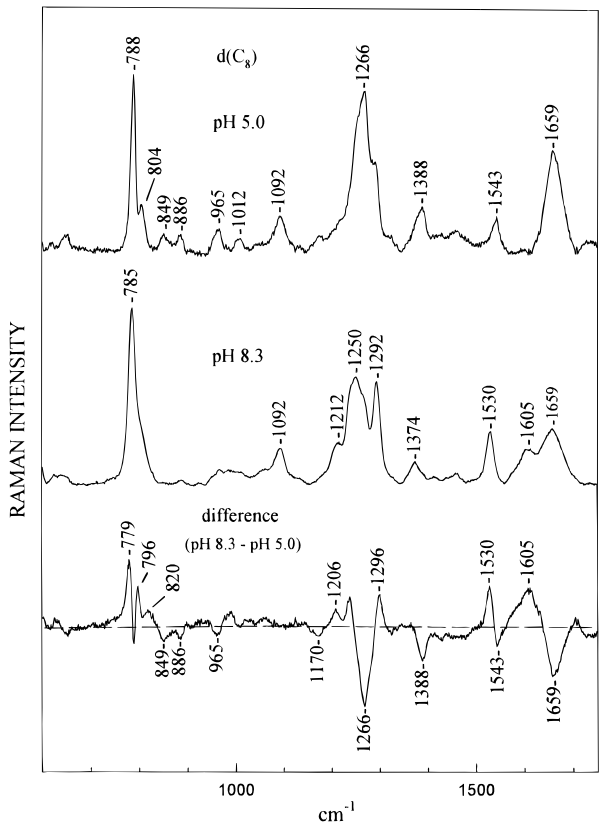

FIGURE 5: Raman spectra (600−1750 cm^{-1}) of d(C_8) at pH 5.0 (top trace) and pH 8.3 (middle trace), and their corresponding difference spectrum (bottom trace). Note the virtual identity of this difference spectrum to that shown in Figure 4 (except for bands in the latter contributed by dT).

Table 2: Raman Conformation Markers of the DNA Backbone[a]

B DNA	*A* DNA	*Z* DNA	*i* DNA
790 ± 5	705 ± 2	745 ± 3	
834 ± 5[b]	807 ± 3[c]		806 ± 2
			849 ± 2
			886 ± 2
			965 ± 2
1092 ± 1	1099 ± 1	1095 ± 2	1092 ± 1

[a] Frequencies are in cm^{-1} units. Data for *A*, *B*, and *Z* DNA are from Raman spectra of crystals and fibers of known structure (Thomas & Wang, 1988, and references therein); data for *i* DNA are from this work. [b] The band center is sensitive to helix geometry (Benevides *et al.*, 1991a). [c] The corresponding band of *A* RNA occurs at 813 cm^{-1}.

frequency in the *i* quadruplex presumably reflects features of helix geometry which differ in detail from those of double-helical *B* DNA. We propose the 806 and 849 cm^{-1} bands as markers of the *i* motif. Table 2 compares Raman conformation markers of the four-stranded intercalated cytosine structure with those of other DNA secondary structures.

The additional bands at ≈886 and ≈965 cm^{-1} (Figures 4−6) are also assigned to the ordered DNA backbone of the cytosine quadruplex, because they appear in the oligomers only at low pH. As expected, these bands diminish in intensity with increasing temperature (Figure 6). [Note that bands due to cytosine ring vibrations generally increase in intensity with temperature, a result of base unstacking. This is evident in Figure 6 for the cytosine marker at 788 cm^{-1} and for the complex band near 1250 cm^{-1}, which originates from both cytosine and thymine rings (Table 1).] The band at ≈886 cm^{-1} has a counterpart in Raman spectra of *A* DNA, where it is assigned to coupled C−O and C−C stretching in furanose rings of C3′-*endo* deoxyribose residues (Prescott *et al.*, 1984). A similar weak band has been reported for C3′-*endo* deoxycytidine (Dijkstra *et al.*, 1991). Accordingly, a similar assignment is made here for both d(CCCT) and d(C_8). The band at ≈965 cm^{-1} also has a counterpart in *A* DNA (Prescott *et al.*, 1984). Further, it is enhanced in polarized spectra of *A* DNA fibers when the electric vector of the exciting radiation is parallel to the helix axis (J. M. Benevides and G. J. Thomas, Jr., unpublished results). This is consistent with assignment to C−C and/or C−O bond stretching in the sugar-phosphate backbone. Thus, the band near 965 cm^{-1} in d(CCCT) and d(C_8) is similarly assigned. The temperature dependence of the 965 cm^{-1} band of d(CCCT) is shown in the inset of Figure 6. Finally, the 1012 cm^{-1} band is assigned to a C−O stretching vibration of the DNA backbone. On the basis of the appearance of a similar band in both *A* and *B* forms of DNA (Prescott *et al.*, 1984), the 1012 cm^{-1} band is apparently not conformation sensitive and has been excluded from Table 2. The very weak band at ≈905 cm^{-1}, present only in d(CCCT) (Figure 4), is assigned to the deoxyribose ring of the thymidine residue.

5. Thymine Contributions to Raman Spectra of d(CCCT). The contributions of the thymidine residue to the Raman spectrum of d(CCCT) are revealed in the difference spectra of Figure 7 by appropriate subtractions of spectra of d(C_8) from those of d(CCCT). The left panel of Figure 7 shows the result obtained for the single-stranded form of d(CCCT) (high pH), while the right panel shows the result obtained for the quadruplex (low pH).

Among the salient features of the difference spectra, the band at 1238 cm^{-1} has been assigned to a thymine in-plane ring vibration (Lord & Thomas, 1967; Erfurth & Peticolas, 1975). In the case of *B* DNA, this band exhibits hypochromism greater than any other Raman band (Benevides *et al.*, 1991a). (See also Figure 6.) The comparable intensity of the 1238 cm^{-1} band in the two difference spectra of Figure 7 suggests similar stacking for thymine in both the single- and four-stranded structures. On the other hand, we observe that the thymine marker at 1374 cm^{-1} is significantly more intense at low pH. Because this normal mode involves the exocyclic 5-CH_3 group (Susi & Ard, 1974), and because the Raman intensity is sensitive to the methyl group environment (Benevides *et al.*, 1991b,c), the higher intensity at low pH is interpreted as indicating greater solvent shielding of the thymine methyl group in the quadruplex structure.

The region 1600−1750 cm^{-1} in the difference spectra of Figure 7 is dominated by Raman bands sensitive to hydrogen bonding environments of the carbonyl groups of dT. The principal feature is a broad band centered at 1660 cm^{-1} at high pH, which becomes sharper and shifts to 1654 cm^{-1} at low pH. The frequency decrease and band narrowing are small but consistent with stronger and more uniform hydrogen bonding of dT carbonyl groups in the quadruplex (Kang *et al.*, 1994). The 1654 cm^{-1} marker is also consistent with the dT carbonyl frequency observed for thymine in base pairs of *B* DNA (Thomas & Benevides, 1985).

6. Determination of pK_C Values for d(CCCT) and d(C_8). By analogy with pK_C of dCMP (Section 1, above), we define the pK_C for proton dissociation in hemiprotonated *i* quadruplex structures in terms of the equilibrium $CHC^+ = 2C +$

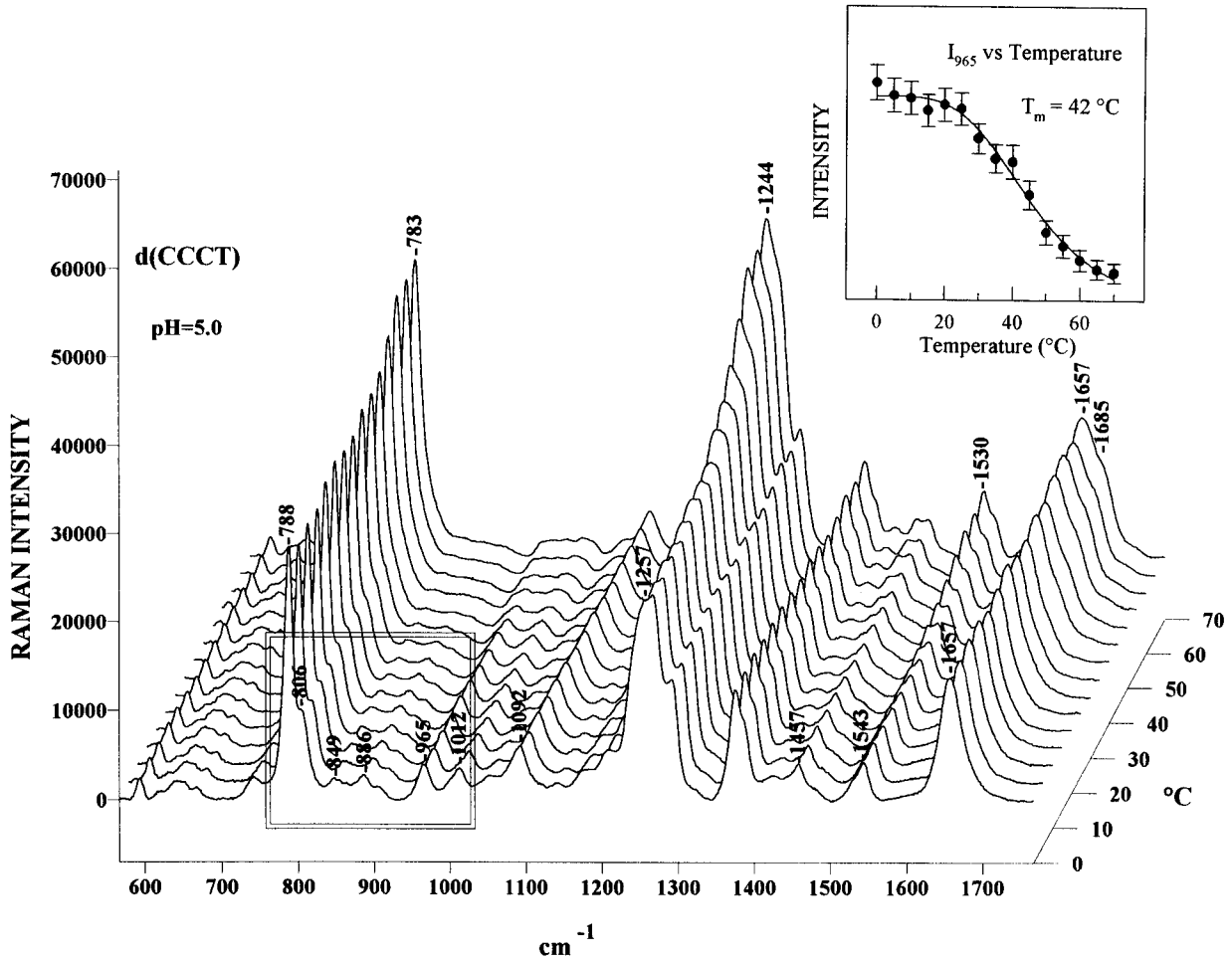

FIGURE 6: Raman spectra (600−1750 cm^{-1}) of d(CCCT) at pH 5.0 and at temperature intervals of 5 °C between 0 and 70 °C. The framed box (left foreground) encloses bands assigned primarily to the sugar−phosphate backbone. The inset (upper right) shows the temperature dependence of the 965 cm^{-1} band from which the apparent melting temperature (T_m = 42 °C) of the quadruplex structure has been determined.

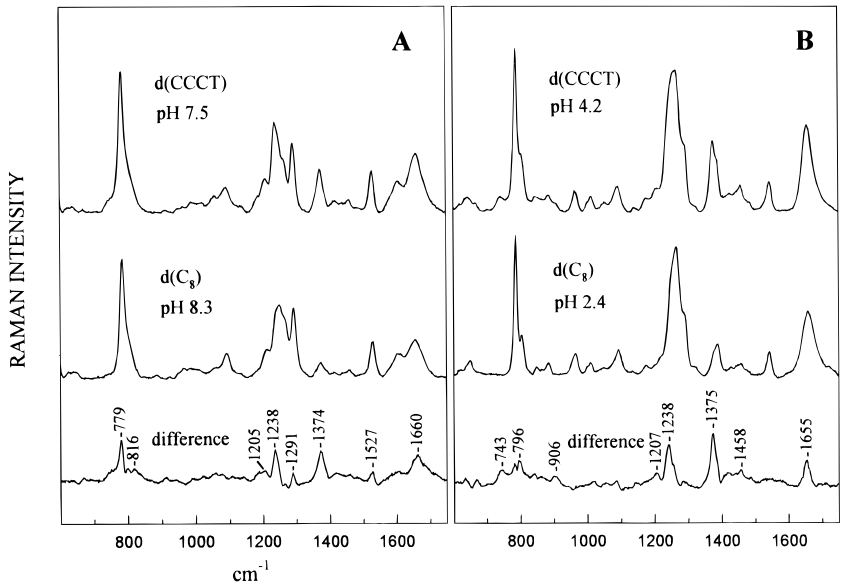

FIGURE 7: Comparison of solution Raman spectra (600−1750 cm^{-1}) of d(CCCT) and d(C_8). Data are compared for structures containing neutral cytosines (left panel) and hemiprotonated cytosine pairs (right panel). In each panel, the computed difference spectrum (bottom trace) indicates the Raman contribution of thymidine to the d(CCCT) spectrum. (See text.)

H^+, where CHC^+ represents the hemiprotonated base pair. (Note also that the hemiprotonated species may be further protonated to yield 2 equiv of CH^+.)

Determination of pK_C values for d(CCCT) and d(C_8), requires a different spectrophotometric approach than employed for 5′-dCMP. In the mononucleotide, markers of unprotonated and protonated dC are observed at 1530 and 1546 cm^{-1}, respectively, and these exhibit nearly equivalent intensities at the midpoint of the titration (Figure 2). However, this is not the case for d(CCCT) because the

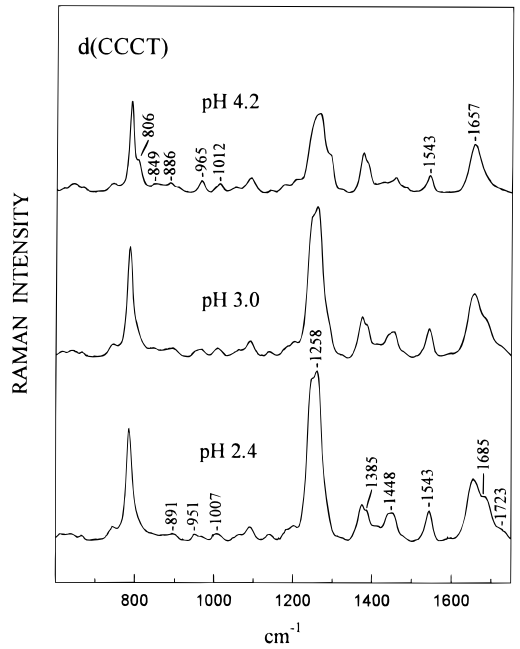

FIGURE 8: Raman spectra (600−1750 cm^{-1}) of d(CCCT) at pH 4.2 (top trace), pH 3.0 (middle trace), and pH 2.4 (bottom trace).

hemiprotonated base pair $(C{\cdot}C)^+$ is structurally distinct from a simple mixture of unprotonated and protonated bases. Thus, the $(C{\cdot}C)^+$ species of the d(CCCT) quadruplex exhibits a single peak at 1543 cm^{-1} (Figure 4) in lieu of a pair of bands near 1530 and 1546 cm^{-1}. Nevertheless, upon thermal disruption of the quadruplex, a marker diagnostic of unprotonated cytosine is generated near 1530 cm^{-1} and that of protonated cytosine is retained at 1543 cm^{-1} (Figure 6), as expected for quadruplex denaturation. Note that the occurrence of a Raman singlet for $(C{\cdot}C)^+$ at 1543 cm^{-1}, in lieu of a doublet, in effect extends the NMR determined lower limit of the imino proton shuttling rate from $\approx 8 \times 10^4$ s^{-1} (Leroy *et al.*, 1993) to $\approx 5 \times 10^{14}$ s^{-1}. Figure 4 also shows that the d(CCCT) quadruplex exhibits no Raman band that can be considered a counterpart to the 1605 cm^{-1} marker of unprotonated cytosine.

We note that the markers of the d(CCCT) quadruplex at 1543 and 1657 cm^{-1} (Figure 4) are close in frequency to markers of protonated cytosine (1546 and 1660 cm^{-1}, respectively). A similar situation occurs for $d(C_8)$ (Figure 5). Evidently, the vibrational coupling in the hemiprotonated cytosine ring is rather similar to that of the fully protonated cytosine ring. Moreover, these $(C{\cdot}C)^+$ marker bands typically exhibit half the intensity anticipated for corresponding bands of fully protonated cytosines. Figure 8 shows that at pH 3.0, the Raman band intensities at 1543 and 1657 cm^{-1} have increased with respect to those in the pH 4.2 spectrum. The intensity of the 1543 cm^{-1} band continues to increase as the pH is further lowered, eventually reaching the limiting intensity of the marker of fully protonated cytosine. Similar doublings of intensity are observed for bands at 1258, 1689, and 1723 cm^{-1}. Accordingly, pH titration of the Raman spectrum allows the pK_C value of a cytosine quadruplex to be computed in either of two ways: (i) on the basis of elimination of a definitive marker band of unprotonated cytosine (e.g., the 1530 cm^{-1} band) or (ii) on the basis of 50% diminution of the intensity of a marker band of protonated cytosine (e.g., the 1543 cm^{-1} band). By either method, we obtain for d(CCCT), p$K_C = 4.5 \pm 0.2$ at 10 °C. Similarly, we find for $d(C_8)$, p$K_C = 5.8 \pm 0.3$ at 10 °C.

The difference in pK_C values between the quadruplexes of d(CCCT) and $d(C_8)$ corresponds to a stabilization in the free energy of the octamer (16 base pairs) relative to the tetramer (six base pairs) of 1.7 kcal/mol at 10 °C (Cantor & Schimmel, 1980). The observed pK_C difference suggests an upper limit of ≈0.17 kcal for the stabilization free energy per mole $(C{\cdot}C)^+$ base pairs. This approximation neglects end effects which are expected to destabilize the tetramer relative to the octamer.

Upon further lowering the solution pH, the hemiprotonated base pairs of the cytosine quadruplex are expected to become fully protonated, leading to dissociation of the four strands. Figure 8 shows, for example, that the d(CCCT) quadruplex is essentially completely converted to fully protonated cytosines at pH 2.4. This is indicated both by the intensities of the markers of protonated cytosine (1258, 1385, 1448, 1543, 1685, and 1723 cm^{-1}) and by the elimination of backbone markers of the *i* motif (806, 849, 886, and 965 cm^{-1}). Interestingly, the $d(C_8)$ quadruplex does not achieve full protonation of cytosines at pH 2.4 and Raman markers diagnostic of the *i* form remain prominent (Figure 7, right panel, middle trace). Thus, the $d(C_8)$ quadruplex exhibits *both high-pH and low-pH stabilization* vis-à-vis d(CCCT).

7. Thermal Denaturation of d(CCCT). The inset to Figure 6 shows the temperature dependence of the intensity of the 965 cm^{-1} Raman band of d(CCCT), from which the apparent median melting temperature, $T_m = 42$ °C, is determined. The 965 cm^{-1} band is well suited for this purpose, owing to its dependence on the backbone conformation of the quadruplex and its freedom from significant overlap by other Raman bands. The T_m of 42 °C for d(CCCT) is also in agreement with the value determined for d(TCCC) by NMR spectroscopy (Leroy *et al.*, 1993). For $d(C_8)$, on the other hand, we find $T_m = 80 \pm 2$ °C (data not shown), consistent with the value reported by Leroy and co-workers (1993).

CONCLUSIONS

The nucleic acid *i* motif, which results from antiparallel intercalation of two parallel-stranded duplexes containing hemiprotonated cytosine base pairs $[(C{\cdot}C)^+]$, is characterized by a unique Raman signature. This signature can be exploited to monitor the formation of four-stranded intercalated cytosine structures in the presence of other nucleic acid conformations, a capability not shared by other spectroscopic methods. Raman spectroscopy has been applied here to monitor both thermostability (T_m) and the extent of cytosine protonation (pK_C) in *i* motif quadruplexes of d(CCCT) and $d(C_8)$.

Raman spectra indicate that the quadruplex of the d(CCCT) crystal structure is conserved in aqueous solution, despite the fact that C3′-*endo* deoxynucleosides are rarely encountered for aqueous nucleic acids. In the case of previously examined DNA crystal structures, C3′-*endo* nucleoside conformers are invariably converted in solution to the more energetically favorable C2′-*endo* conformation (Benevides *et al.*, 1984, 1986, 1988). Stabilization of the *i* motif by cytosine base pairing and stacking, estimated here as <0.17 kcal/mol of $(C{\cdot}C)^+$, is apparently sufficient to compensate for the incorporation of C3′-*endo* sugars in the backbone of the solution quadruplex.

Raman bands of the 750−950 cm^{-1} interval provide a particularly sensitive measure of DNA backbone geometry (Thomas & Tsuboi, 1993). Because this region of the Raman spectrum of d(CCCT) is highly conserved between crystal and solution states, we conclude further that details of the backbone geometry of the d(CCCT) crystal structure are preserved in the solution structure. In the d(CCCT) crystal, the four phosphodiester strands are not conformationally identical, leading to an asymmetric quadruplex (Kang *et al.*, 1994). This asymmetry is apparently retained in the solution structure. A different conclusion has been reached from NMR studies of solution structures of oligonucleotides of the type d(TC_n) (Gehring *et al.*, 1993), which may reflect a property of the different sequences involved or may represent the effect of time averaging of structures on the slower time scale of NMR spectroscopy.

We have tentatively interpreted the Raman markers of the thymine residue of d(CCCT), obtained by spectral difference, as consistent with T·T base pairs and thymine base stacking in the solution structure. Further studies will be required to confirm or reject this hypothesis.

The Raman signature of hemiprotonated cytosine base pairs is distinct from the signatures of unprotonated and protonated cytosines. Specifically, $(C{\cdot}C)^+$ lacks marker bands of unprotonated cytosine, and the marker bands of protonated cytosine are diminished in intensity by approximately 50%. The latter are also slightly shifted in frequency. Therefore, for any DNA structure, the degree of cytosine protonation can be determined from the ratio of Raman band intensities using appropriate markers of protonated and unprotonated cytosines. This should be of value in assessing the extent of cytosine protonation or hemiprotonation in duplex, triplex, and quadruplex structures of DNA.

ACKNOWLEDGMENT

We thank Mr. Andrew Armstrong for technical assistance.

REFERENCES

Benevides, J. M., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., Rich, A., & Thomas, G. J., Jr. (1984) *Nucleic Acids Res. 12*, 5913−5925.

Benevides, J. M., Wang, A. H.-J., Rich, A., Kyogoku, Y., van der Marel, G. A., van Boom, J. H., & Thomas, G. J., Jr. (1986) *Biochemistry 25*, 41−50.

Benevides, J. M., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., & Thomas, G. J., Jr. (1988) *Biochemistry 27*, 931−938.

Benevides, J. M., Stow, P. L., Ilag, L. L., Incardona, N. L., & Thomas, G. J., Jr. (1991a) *Biochemistry 30*, 4855−4863.

Benevides, J. M., Weiss, M. A., & Thomas, G. J., Jr. (1991b) *Biochemistry 30*, 4381−4388.

Benevides, J. M., Weiss, M. A., & Thomas, G. J., Jr. (1991c) *Biochemistry 30*, 5955−5963.

Benevides, J. M., Tsuboi, M., & Thomas, G. J., Jr. (1993) *J. Am. Chem Soc. 115*, 5351−5359.

Benevides, J. M., Weiss, M. A., & Thomas, G. J., Jr. (1994) *J. Biol. Chem. 269*, 10869−10878.

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

Borah, B., & Wood, J. L. (1976) *J. Mol. Struct. 30*, 13−30.

Cantor, C. R., & Schimmel, P. A. (1980) *Biophysical Chemistry, Part 1: The Conformation of Biological Macromolecules*, p 45, W. H. Freeman, San Francisco, CA.

Cheng, Y.-K., & Pettitt, B. M. (1992) *Prog. Biophys. Mol. Biol. 58*, 225−257.

Chou, C. H., & Thomas, G. J., Jr. (1977) *Biopolymers 16*, 765−789.

Cohn, W. E. (1955) *The Nucleic Acids* (Chargaff, E., & Davidson, J. N., Eds.) Vol. 1, p 227, Academic Press, New York.

Dijkstra, S., Benevides, J. M., & Thomas, G. J., Jr. (1991) *J. Mol. Struct. 242*, 283−301.

Erfurth, S. C., & Peticolas, W. L. (1975) *Biopolymers 14*, 247−264.

Erfurth, S. C., Kiser, E. R., Kiser, J., & Peticolas, W. L. (1972) *Proc. Natl Acad. Sci. U.S.A. 69* 938−941.

Frank-Kamenetskii, M. D., & Mirkin, S. M. (1995) *Annu. Rev. Biochem. 64*, 65−95.

Gehring, K., Leroy, J.-L., & Guéron, M. (1993) *Nature 363*, 561−565.

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

Hartman, K. A., Lord, R. C., & Thomas, G. J., Jr. (1973) in *Physicochemical Properties of Nucleic Acids* (Duchesne, J., Ed.) Vol. 2, pp 1−89, Academic Press, New York.

Hurst, R. O., Marko, A. M., & Butler, G. C. (1953) *J. Biol. Chem. 204*, 847−856.

Inman, R. B. (1964) *J. Mol. Biol. 9*, 624−637.

Jaishree, T. N., & Wang, A. H.-J. (1993) *Nucleic Acids Res. 21*, 3839−3844.

Jordan, D. O. (1960) *The Chemistry of Nucleic Acids*, pp 134−139, Butterworth, Washington, D. C.

Kang, C. H., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R., & Rich, A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11636−11640.

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

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

Leroy, J. L., & Guéron, M. (1995) *Structure 3*, 101−120.

Leroy, J. L., Gehring, K., Kettani, A., & Guéron, M. (1993) *Biochemistry 32*, 6019−6031.

Leroy, J. L., Guéron, M., Mergny, J. L., & Helene, C. (1994) *Nucleic Acids Res. 22*, 1600−1606.

Liu, K., Miles, H. T., Frazier, J., & Sasisekharan, V. (1993) *Biochemistry 32*, 11802−11809.

Lord, R. C., & Thomas, G. J., Jr. (1967) *Spectrochim. Acta 23A*, 2551−2591.

Manzini, G., Yathindra, N., & Xodo, L. E. (1994) *Nucleic Acids Res. 22*, 4634−4640.

Miura, T., & Thomas, G. J., Jr. (1994) *Biochemistry 33*, 7848−7856.

Miura, T., & Thomas, G. J., Jr. (1995) *Biochemistry 34*, 9645−9653.

Miura, T., Benevides, J. M., & Thomas, G. J., Jr. (1995) *J. Mol. Biol. 248*, 233−238.

O'Connor, T., Johnson, C., & Scovell, W. (1976) *Biochim. Biophys. Acta 447*, 484−494.

Prescott, B., Steinmetz, W., & Thomas, G. J., Jr. (1984) *Biopolymers 23*, 235−256.

Robinson, H., van der Marel, G. A., van Boom, J. H., & Wang, A. H.-J. (1992) *Biochemistry 31*, 10510−10517.

Robinson, H., & Wang, A. H.-J. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 5224−5228.

Sen, D., & Gilbert, W. (1990) *Nature 344*, 410−414.

Susi, H., & Ard, J. S. (1974) *Biochim. Biophys. Acta 312*, 311−322.

Steitz, T. A. (1990) *Q. Rev. Biophys. 23*, 205−280.

Thomas, G. J., Jr., & Benevides, J. M. (1985) *Biopolymers 24*, 1101−1105.

Thomas, G. J., Jr., & Wang, A. H.-J. (1988) *Nucleic Acids Mol. Biol. 2*, 1−30.

Thomas, G. J., Jr., & Tsuboi, M. (1993) *Adv. Biophys. Chem. 3*, 1−70.

Thomas, G. J., Jr., Benevides, J. M., Overman, S. A., Ueda, T., Ushizawa, K., Saitoh, M., & Tsuboi, M. (1995) *Biophys. J. 68*, 1073−1088.

Wang, Y., & Patel, D. J. (1994) *J. Mol. Biol. 242*, 508−526.

Williamson, J. R. (1994) *Annu. Rev. Biophys. Biophys. Chem. 23*, 703−730.

BI9529420