Structural and Thermodynamic Studies of $d(C_3T_4C_3)$ in Acid Solution: Evidence for Formation of the Hemiprotonated CH^+·C Base Pair

Daniel S. Pilch[†,§] and Richard H. Shafer[*,†,‡]

Contribution from the Graduate Group in Biophysics and the Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143

Received October 8, 1992

Abstract: In an effort to probe the structure and stability of the CH^+·C base pair in DNA, we have studied the conformation of the deoxydecanucleotide $d(C_3T_4C_3)$ at pH 5.5, using ^{1}H-NMR and UV absorption spectroscopy. Lowering the pH of a solution containing $d(C_3T_4C_3)$ is accompanied by the induction of six new imino proton NMR resonances in a region (15.4–15.8 ppm) downfield from where either T-H3 or G-H1 Watson–Crick (W-C)-paired imino protons normally resonate. These novel resonances correspond to CH^+-H3 imino protons engaged in hydrogen bonding with the N3 atoms of unprotonated cytosine residues. Since six, and not three, imino proton resonances arise upon lowering of the pH, the base-paired structure formed by $d(C_3T_4C_3)$ appears to be a parallel-stranded, bimolecular duplex $[d(CH^+{}_3T_4CH^+{}_3)\cdot d(C_3T_4C_3)]$, containing six CH^+·C base pairs and an inner loop of eight non-base-paired thymine residues. A nuclear Overhauser enhancement spectroscopy (NOESY) experiment demonstrates dipolar contacts between the CH^+-H3 imino protons and both CH^+-H4 and C-H4 amino protons, confirming the presence of CH^+·C base pairs. Thermal denaturation profiles at differing DNA concentrations are consistent with the existence of a bimolecular duplex at low pH. Thermodynamic analysis of UV melting curves indicates that CH^+·C base pairs are more stable than either Watson-Crick G·C or A·T base pairs.

Introduction

In recent years, many studies have shown that nucleic acid molecules are able to adopt a wide variety of structures and conformations, as dictated by their base sequence and solution conditions.[1-5] Little is currently known about the biological roles that nucleic acid conformation and structure play in various cellular processes, such as gene expression, genetic recombination, and telomere formation. Inter- and intramolecular triple helices, the formation of which generally requires a homopurine·homopyrimidine motif, comprise a family of unusual nucleic acid structures whose potential biological roles and usefulness in applications involving oligonucleotide targeting of duplex DNA in vitro and in vivo are of great current interest. Many of the triple helices which have been studied to date contain homopyrimidine third strands, which require protonation of cytosine residues at their N3 positions in order to form the requisite CH^+·G Hoogsteen base pairs with the homopurine strands of the underlying duplexes.[6-12] While protonation of unmodified cytosine residues requires an acidic pH, protonation of C5-methylated cytosines occurs at neutral pH.[13] Eukaryotic genomes contain numerous stretches of cytosine-rich, homopurine·homopyrimidine sequences that are frequently located near or within regions of functional and/or regulatory importance.[4] These studies have highlighted the importance of investigating the structure and physical chemistry of protonated cytosine residues and the various base pairs or triplets they are able to form.

Early fiber diffraction[14] and thermal denaturation[15] studies on poly(rC) at low pH resulted in a model of a double-helical structure [poly(rCH$^+$)·poly(rC)] with parallel strands and hemiprotonated CH^+·C base pairs (see Figure 1), in which all the bases are in the *anti* conformation. The existence of these base pairs was confirmed by infrared spectroscopic[16] and single crystal[17,18] studies on cytosine derivatives. Circular dichroism (CD)[19-23] and thermal denaturation[24] studies on the low-pH forms of poly(dC)[19-21,24] and poly[d(C-T)][22,23] have also provided evidence for formation of double-helical structures containing CH^+·C base pairs. Upon hemiprotonation of the cytosine residues at 50 mM Na^+, the self-complex of poly(dC) forms with a pK_a of 7.4,[24] while that

* Author to whom correspondence should be addressed.
† Graduate Group in Biophysics.
‡ Department of Pharmaceutical Chemistry.
§ Present address: Laboratoire de Biophysique, Muséum National d'Histoire Naturelle, 43 rue Cuvier, 75005 Paris, France.

(1) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag; New York, 1984; pp 220–349.
(2) Rich, A.; Nordheim, A.; Wang, A. H. J. *Annu. Rev. Biochem.* **1984**, *53*, 791–846.
(3) Wells, R. D.; Harvey, S. C., Eds. *Unusual DNA Structures*; Springer-Verlag; New York, 1988.
(4) Wells, R. D.; Collier, D. A.; Hanvey, J. C.; Shimizu, M.; Wohlrab, F. *FASEB J.* **1988**, *2*, 2939–2949.
(5) Wells, R. D. *J. Biol. Chem.* **1988**, *263*, 1095–1098.
(6) Pilch, D. S.; Brousseau, R.; Shafer, R. H. *Nucleic Acids Res.* **1990**, *18*, 5743–5750.
(7) Rajagopal, P.; Feigon, J. *Biochemistry* **1989**, *28*, 7859–7870.
(8) Santos, C. D. L.; Rosen, M.; Patel, D. *Biochemistry* **1989**, *28*, 7282–7289.
(9) Mirkin, S. M.; Lyamichev, V. I.; Drushlyak, K. N.; Dobrynin, V. N.; Filippov, S. A.; Frank-Kamenetskii, M. D. *Nature (London)* **1987**, *330*, 495–497.
(10) Maher, L. J., III; Wold, B.; Dervan, P. B. *Science* **1989**, *245*, 725–730.
(11) Hanvey, J. C.; Shimizu, M.; Wells, R. D. *Proc. Natl. Acad. Sci. U.S.A* **1988**, *85*, 6292–6296.
(12) Sun, J.S.; François, J. C.; Montenay-Garestier, T.; Saison-Behmoaras, T.; Roig, V.; Thuong, N. T.; Hélène, C. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9198–9202.
(13) Lee, J. S.; Woodsworth, M. L.; Latimer, L. J. P.; Morgan, A. R. *Nucleic Acids Res.* **1984**, *12*, 6603–6614.
(14) Langridge, R. A.; Rich, A. *Nature (London)* **1963**, *198*, 725–728.
(15) Guschlbauer, W. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *57*, 1441–1448.
(16) Borah, B.; Wood, J. L. *J. Mol. Struct.* **1976**, *30*, 13–30.
(17) Marsh, R. E.; Bierstedt, R.; Eichhorn, E. L. *Acta Crystallogr.* **1962**, *15*, 310–316.
(18) Kistenmacher, T. J.; Rossi, M.; Marzilli, L. G. *Biopolymers* **1978**, *17*, 2581–2585.
(19) Gray, D. M.; Bollum, F. J. *Biopolymers* **1974**, *13*, 2087–2102.
(20) Marck, C.; Thiele, D.; Schneider, C.; Guschlbauer, W. *Nucleic Acids Res.* **1978**, *5*, 1979–1996.
(21) Thiele, D.; Marck, C.; Schneider, C.; Guschlbauer, W. *Nucleic Acids Res.* **1978**, *5*, 1997–2012.
(22) Gray, D. M.; Vaughan, M.; Ratliff, R. L.; Hayes, F. N. *Nucleic Acids Res.* **1980**, *8*, 3695–3707.
(23) Brown, D. M.; Gray, D. M.; Patrick, M. H.; Ratliff, R. L. *Biochemistry* **1985**, *24*, 1676–1683.
(24) Inman, R. B. *J. Mol. Biol.* **1964**, *9*, 624–637.

0002-7863/93/1515-2565$04.00/0

of poly[d(C-T)] forms with a pK_a of 6.2.[23] In the latter self-complex, each of the thymine residues is looped out of the helical stack formed by the CH^+·C base pairs.[22,23] Sarma and co-workers have observed formation of a similar structure by the d(CTCTCT) oligonucleotide at low pH.[25]

Initial CD studies by Gray and co-workers on $d(C_4A_4T_4C_4)$ were interpreted in terms of self-association to form an antiparallel duplex containing both CH^+·C and A·T base pairs.[26] Subsequent analysis of other sequences revealed that formation of CH^+·C base pairs was highly sequence dependent.[27] Additional studies on a variety of DNA oligomer sequences led to the conclusion that the structure formed by $d(C_4A_4T_4C_4)$ was multi-stranded and contained both parallel CH^+·C base pairs and antiparallel A·T base pairs.[28]

The structure and physical chemistry of protonated cytosine-containing base pairs and triplets, as well as the range of different structures available to cytosine-rich homopyrimidine sequences under differing solution conditions, are still largely unknown. In the work presented below we describe experiments probing the structure, stability, and thermodynamics of the low-pH form of $d(C_3T_4C_3)$. NMR studies indicate that at low pH, $d(C_3T_4C_3)$ self-associates to form a parallel-stranded, bimolecular duplex $[d(CH^+{}_3T_4CH^+{}_3)\cdot d(C_3T_4C_3)]$, containing six CH^+·C base pairs (see Figure 2, scheme 1). Thermal denaturation experiments are consistent with the presence of a bimolecular duplex at low pH. Thermodynamic analysis of the UV melting curves suggests that CH^+·C base pairs are more stable than either Watson–Crick A·T or G·C base pairs.

Experimental Section

Chemicals and Deoxyoligonucleotides. The $d(C_3T_4C_3)$ oligomer was synthesized and purified as previously reported.[6,29] Following purification, the oligomer was dialyzed against 2 mM sodium phosphate (pH 7.1) and its purity was checked by polyacrylamide gel electrophoresis (PAGE) and ^1H NMR. The concentration of $d(C_3T_4C_3)$ was determined spectrophotometrically, using the previously reported[6] extinction coefficient of $\epsilon_{271} = 8300\ cm^{-1}\ (mol\ of\ base/L)^{-1}$. All DNA concentrations, except where noted otherwise, are reported with respect to strand. All buffer reagents were obtained from Sigma Chemical Co.

One-Dimensional Proton NMR. One-dimensional NMR experiments were carried out on 400 μL of solution containing 4 mM $d(C_3T_4C_3)$. Solution conditions for exchangeable proton NMR measurements were 15 mM sodium phosphate (pH 7.3), 100 mM NaCl, and 90% H_2O/10% D_2O. Aliquots of 1 N HCl were added to the $d(C_3T_4C_3)$ sample until a final pH of 5.5 was achieved, and the proton NMR spectrum was acquired after each addition. The $1\bar{3}3\bar{1}$ solvent suppression pulse sequence[30] was used to acquire the exchangeable proton NMR spectra, with the carrier frequency set on the H_2O resonance. The interval delay, τ, between pulses was 96 μs and the recycle delay was 3 s. The temperature was fixed at 2 °C. Non-exchangeable proton NMR measurements of $d(C_3T_4C_3)$ at pH 5.5 were carried out in D_2O, under buffer conditions identical with those described above. For these measurements, a recycle delay of 3 s was used and the temperature was fixed at 11 °C. All NMR experiments were carried out on a 500-MHz General Electric GN-500 spectrometer, equipped with an Oxford Instruments magnet and a Nicolet 1280 computer.

Two-Dimensional Proton NMR. A two-dimensional NOESY experiment, involving exchangeable proton resonances, was carried out on a 90% H_2O/10% D_2O solution (400 μL) containing 4 mM $d(C_3T_4C_3)$. Buffer conditions were 15 mM sodium phosphate (pH 5.5) and 100 mM NaCl and the temperature was fixed at 2 °C. Phase-sensitive NOESY spectra were acquired in the pure absorption mode.[31] The exciting 90° pulse was replaced by a $1\bar{3}3\bar{1}$ pulse sequence, with the carrier frequency set on the H_2O resonance. A short homospoil pulse was applied at the start of the mixing time to suppress the H_2O signal.[32] A mixing time of 250 ms, an interval delay of 96 μs, and a recycle delay of 3 s were used. The spectra were acquired with a spectral width of 12 048 Hz in both dimensions. Two-dimensional NOESY data sets consisted of 4096 complex points in the t_2 dimension, stored in alternating blocks, with 16 scans for each of 400 increments in the t_1 dimension. The free induction decays (FIDs) of the NOESY spectra were apodized with a 45°-shifted sine-bell function in both dimensions and zero-filled once in each dimension prior to Fourier transformation. The baselines of the spectra were corrected to a sixth-order polynomial in both ω_1 and ω_2.[33,34] Processing of the two-dimensional data sets was performed on a Sun Microsystems Sparc Station 1 computer, using software developed by Drs. S. Manogaran and R. Scheek at the University of California, San Francisco.

Helix-Coil Transitions. The change in UV absorbance at 260 nm was measured as a function of temperature for solutions containing concentrations of $d(C_3T_4C_3)$ ranging from 6 to 176 μM. Buffer conditions for these solutions were 10 mM sodium cacodylate (pH 5.5), 100 mM NaCl, and 0.1 mM EDTA. All DNA solutions were preheated at 70 °C for 5 min and slowly cooled prior to UV analysis. The UV absorbance of all solutions was measured once per minute on a Gilford 2600 spectrophotometer interfaced to an Optima Systems AT-10 computer for data collection and analysis. The temperature was increased at a rate of 0.25 °C/min with a Gilford 2527 thermoprogrammer, for a total of 320 min. The cell path length was either 0.1 or 1.0 cm, depending on the DNA concentration. The cuvette-holding chamber was flushed with N_2 gas for the duration of each experiment. First derivatives were calculated over a window of ±2.5 °C.

Thermodynamic Analysis. The NMR results presented below suggest that at low pH, $d(C_3T_4C_3)$ self-associates to form the bimolecular, parallel-stranded duplex $d(CH^+{}_3T_4CH^+{}_3)\cdot d(C_3T_4C_3)$, in which the thymine residues remain unpaired in single-stranded loops (see Figure 2, scheme 1). In the past, studies have shown that duplexes composed of 12 or fewer base pairs melt in a two-state (all-or-none) fashion.[35] This two-state model implies that partially-formed, intermediate structures are not present at any time during the melting process, and therefore they make no thermodynamic contributions. Furthermore, Manzini et al.[36] have shown that approximately one out of six cytosine residues is protonated in single-stranded, oligomeric DNA at pH 5.5, with the remaining cytosines becoming protonated in conjunction with base pairing. Assuming two-state melting behavior and protonation of one out of six cytosine residues in the single-stranded form, the monophasic helix-coil transitions of the $d(CH^+{}_3T_4CH^+{}_3)\cdot d(C_3T_4C_3)$ duplex reflect the following overall equilibrium:

$$d(CH^+{}_3T_4CH^+{}_3)\cdot d(C_3T_4C_3) \rightleftharpoons 2d(C_3T_4CH^+C_2)$$

The equilibrium constant for this reaction may be given by

$$K = \frac{\theta}{2C_t(1-\theta)^2} \tag{1}$$

where C_t is the total concentration of single strands and θ is the fraction of single strands in the duplex form. The value of θ corresponding to the temperature at which $d(absorbance)/d(1/T)$ vs T curves reach their maxima ($T = T_{max}$) has been previously shown[37] to be 0.414.

By the derivation we have previously reported,[6] the variation of T_{max} with total strand concentration may be written as

$$\frac{1}{T_{max}} = \frac{2.3R}{\Delta H^\circ}\log(C_t) + \frac{\Delta S^\circ + 0.504R}{\Delta H^\circ} \tag{2}$$

If the standard enthalpy (ΔH°) and standard entropy (ΔS°) changes are assumed to be temperature independent, their values can be determined from the slopes and y intercepts, respectively, of $1/T_{max}$ vs $\log(C_t)$ plots. All $1/T_{max}$ vs $\log(C_t)$ data were analyzed by linear regression, and the

(25) Sarma, M. H.; Gupta, G.; Sarma, R. H. *FEBS Lett.* **1986**, *205*, 223–229.
(26) Gray, D. M.; Cui, T.; Ratliff, R. L. *Nucleic Acids Res.* **1984**, *12*, 7565–7580.
(27) Edwards, E. L.; Ratliff, R. L.; Gray, D. M. *Biochemistry* **1988**, *27*, 5166–5174.
(28) Edwards, E. L.; Patrick, M. H.; Ratliff, R. L.; Gray, D. M. *Biochemistry* **1990**, *29*, 828–836.
(29) Pilch, D. S.; Levenson, C.; Shafer, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 1942–1946.
(30) Hore, P. J. *J. Magn. Reson.* **1983**, *54*, 539–542.
(31) States, D. J.; Harberhorn, R. A.; Ruben, D. J. *J. Magn. Reson.* **1982**, *48*, 286–292.
(32) Boelens, R.; Scheek, R. M.; Dijkstra, K.; Kaptein, R. *J. Magn. Reson.* **1985**, *62*, 378–386.
(33) Pearson, G. A. *J. Magn. Reson.* **1977**, *27*, 265–272.
(34) Basus, V. J. *J. Magn. Reson.* **1984**, *60*, 138–142.
(35) Puglisi, J. D.; Tinoco, I., Jr. *Methods Enzymol.* **1989**, *180*, 304–325.
(36) Manzini, G.; Xodo, L. E.; Gasparotto, D.; Quadrifoglio, F.; Van Der Marel, G. A.; Van Boom, J. H. *J. Mol. Biol.* **1990**, *213*, 833–843.
(37) Gralla, J.; Crothers, D. M. *J. Mol. Biol.* **1973**, *78*, 301–319.

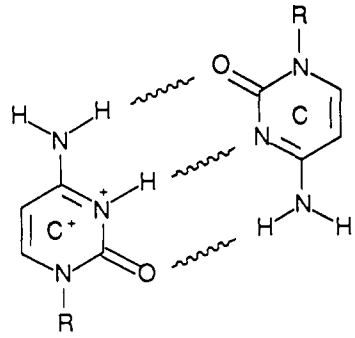
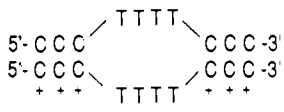
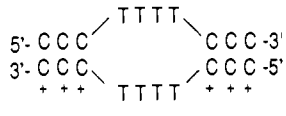


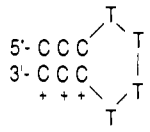
Figure 1. Hemiprotonated CH^+·C base pair, in which the hydrogen bonds are denoted by wavy lines.



SCHEME 1



SCHEME 2



SCHEME 3

Figure 2. Three possible structures that d($C_3T_4C_3$) may adopt at low pH. C with a plus sign denotes a protonated cytosine residue.

errors reported in Table II reflect the magnitudes of the 95% confidence limits as calculated by this analysis.

The standard free energy change ($\Delta G°$) of duplex formation as a function of T_{max} is given by

$$\Delta G°_{T_{max}} = 2.3RT_{max}\log(1.66C_t) \qquad (3)$$

The free energy change and equilibrium binding constant at 25 °C ($\Delta G°_{25}$ and K_{25}, respectively) were determined as previously reported.[6] It should be pointed out that for this particular sequence, the values of $\Delta H°$, $\Delta S°$, and $\Delta G°$ derived in this manner represent the total changes in these thermodynamic parameters associated with duplex formation and therefore include thermodynamic contributions from the protonation of cytosine residues.

Results and Discussion

One-Dimensional Proton NMR. (a) Exchangeable Protons. At low pH, cytosine residues in DNA oligomers become protonated at their N3 positions, thereby obtaining imino protons.[1] Protonated cytosine residues should thus be able to base pair with nonprotonated cytosine residues via three hydrogen bonds. The resulting hemiprotonated CH^+·C base pair is illustrated in Figure 1 and has been previously observed in X-ray crystallographic studies on model compounds.[16-18] One of the three hydrogen bonds in the CH^+·C base pair depicted in Figure 1 involves the CH^+-H3 imino proton. Hence, formation of this hydrogen bond should be accompanied by the induction of an imino proton resonance, which can be observed by NMR.

Assuming thymine residues cannot form base pairs with one another, d($C_3T_4C_3$) strand(s) may associate to form three possible monomeric or dimeric base-paired structures at low pH. These structures are schematically represented in Figure 2. While it is conceivable that at low pH, d($C_3T_4C_3$) may self-associate to form an overlapping, multi-stranded structure (similar to that proposed by Gray and co-workers[28]), containing several double-helical regions of three CH^+·C base pairs each, the comparatively

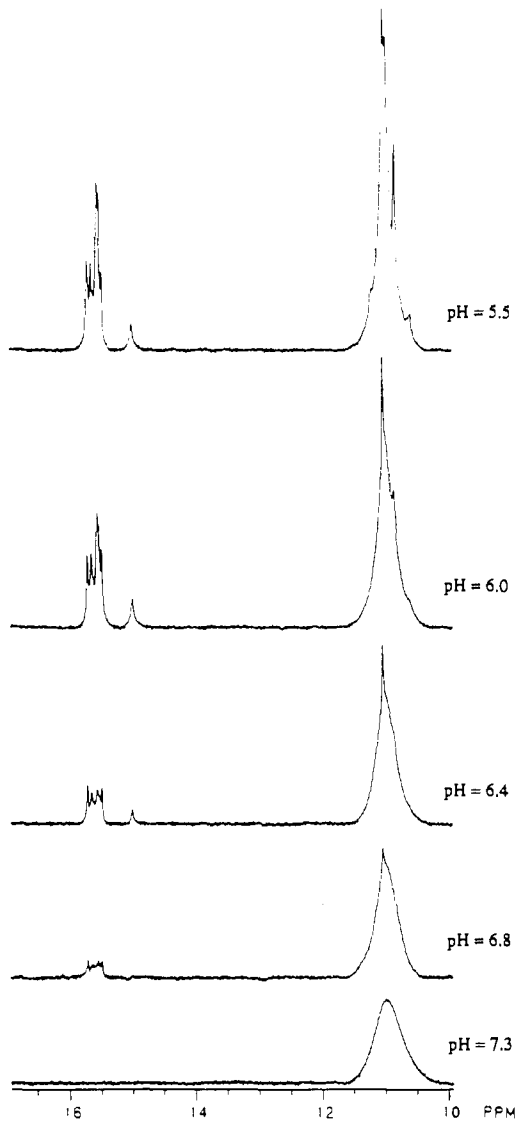


Figure 3. Imino proton NMR spectra of d($C_3T_4C_3$) (4.0 mM in strand) at the indicated values of pH. The titration was carried out in 90% H_2O/10% D_2O, 15 mM sodium phosphate, and 100 mM NaCl at a temperature of 2 °C.

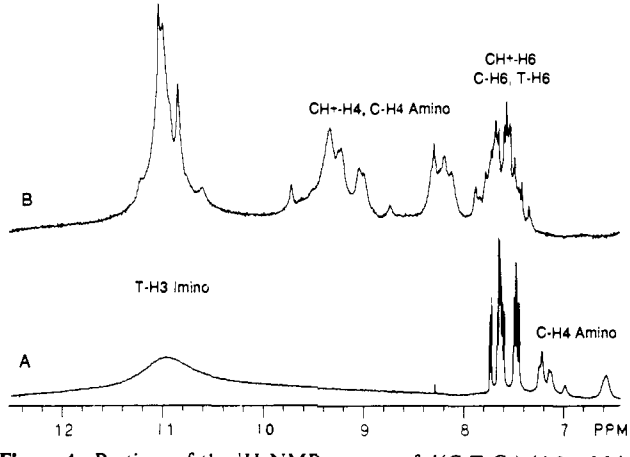


Figure 4. Portions of the ^{1}H-NMR spectra of d($C_3T_4C_3$) (4.0 mM in strand) at pH values of either 7.3 (A) or 5.5 (B). In both cases, solution and temperature conditions are the same as those given for Figure 3.

narrow line widths of the NMR resonances (Figures 3, 4B, and 5) argue against formation of such a structure. If the asymmetric, bimolecular structure (parallel-stranded duplex) presented as scheme 1 is formed upon lowering the pH, one would expect to

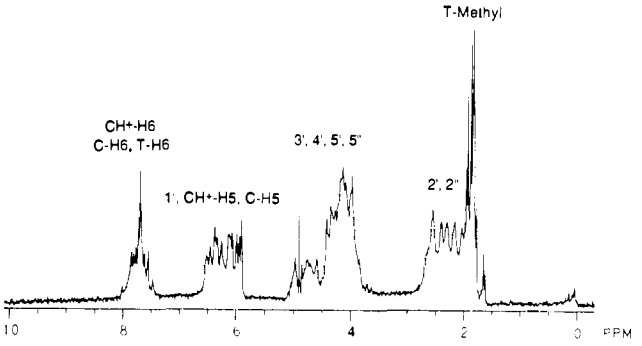


Figure 5. ^{1}H-NMR spectrum of the d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex (2.0 mM) in D_2O. The experiment was carried out in 15 mM sodium phosphate (pH 5.5) and 100 mM NaCl at a temperature of 11 °C.

observe as many as six imino proton resonances at a sufficiently low temperature for the terminal resonances to be observed. In contrast, if either the symmetric, bimolecular structure (antiparallel-stranded duplex) presented as scheme 2 or the monomolecular structure (hairpin duplex) presented as scheme 3 is formed, one would expect to observe at most three imino proton resonances.

Changes in the imino proton spectrum of a d($C_3T_4C_3$) solution upon lowering the pH by titration with HCl are shown in Figure 3. At pH 7.3, no imino proton resonances are present in the region (12–15 ppm) in which the imino proton resonances of double-helical DNA are normally observed, indicating a lack of CH^+·C base pair formation. However, a quite broad resonance band is evident between 10 and 12 ppm. This broad resonance, which is absent in the D_2O NMR spectrum of d($C_3T_4C_3$) (Figure 5), corresponds to the H3 imino protons of the four thymine residues. Other NMR studies[38] probing the structure of various hairpin-containing DNA duplexes, in which runs of thymine residues occur in the looped portions, have shown that the imino protons of these thymines also resonate in the 10–12-ppm region. Furthermore, structural studies by Sarma and co-workers[39] involving single-stranded d($T)_6$ have also indicated the presence of T-H3 imino protor resonances in the absence of any base pairing or regions of duplex formation.

Control NMR studies (data not shown) on deoxythymidine 5′-monophosphate (dTMP) in 90% H_2O at pH 7.1 and a temperature of 2 °C also reveal the presence of a very broad resonance line between 10 and 12 ppm, corresponding to the H3 imino proton. Furthermore, the line width of this resonance decreases with decreasing pH. Thus, the intrinsic solvent exchange rate of the H3 imino proton of dTMP (at pH 7.1 and T = 2 °C), whose hydrodynamic radius and hence correlation time (τ_c) are much smaller and shorter, respectively, than that of d($C_3T_4C_3$), is sufficiently slow so as to render it detectable by NMR. It is therefore likely that the broad resonance between 10 and 12 ppm in the NMR spectrum of d($C_3T_4C_3$) at pH 7.3 (Figure 3) reflects the intrinsic solvent exchange properties of the four T-H3 imino protons within the single-stranded oligomer.

Lowering the pH induces significant changes in the imino proton resonance pattern of d($C_3T_4C_3$) (Figure 3). Six new resonances arise between 15.4 and 15.8 ppm. Although one of the six new resonances (at 15.61 ppm) is not easily observable in the one-dimensional spectrum at pH 5.5 (Figure 3, top) due to partial overlapping with that at 15.63 ppm, it is much more evident in the NOESY plot in Figure 6A. These new resonances correspond to the CH^+-H3 imino protons engaged in hydrogen bonding with C–N3 atoms. It is interesting to note that the chemical shifts of these CH^+–H3 imino protons are markedly downfield from the region where W–C base-paired thymine and guanine imino protons

(38) Germann, M. W.; Kalisch, B. W.; Lundberg, P.; Vogel, H. J.; van de Sande, J. H. *Nucleic Acids Res.* **1990**, *18*, 1489–1498.
(39) Sarma, M. H.; Umemoto, K.; Gupta, G.; Luo, J.; Sarma, R. H. *J. Biomol. Struct. Dyn.* **1990**, *8*, 461–482.

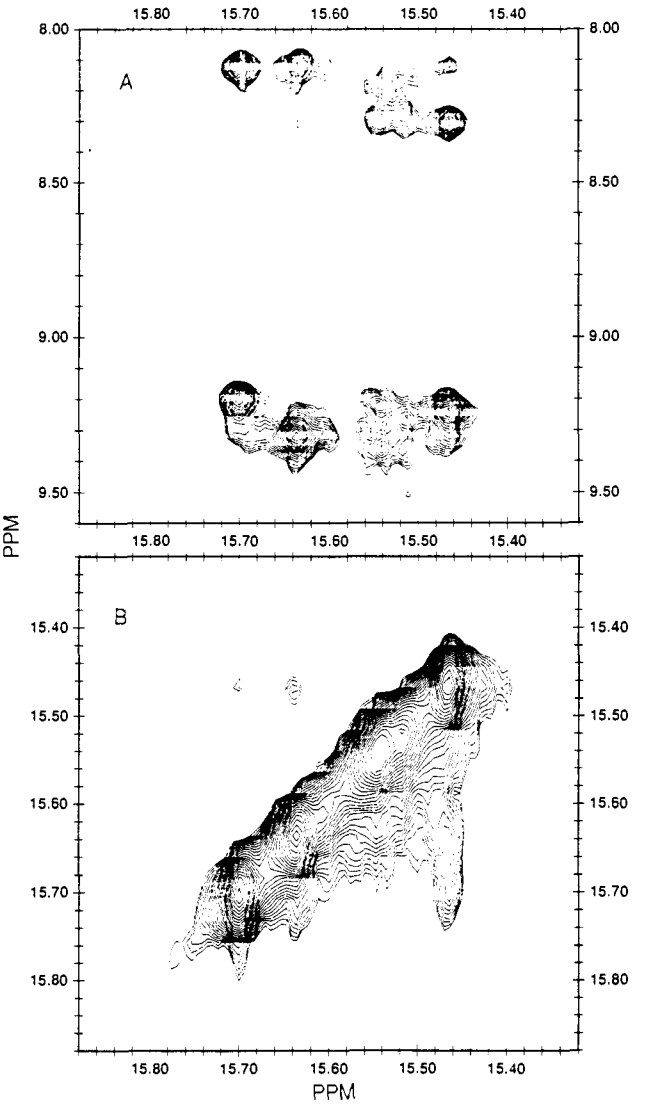


Figure 6. Portions of a NOESY spectrum (250 ms mixing time) of the d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex (2.0 mM), showing contacts between CH^+–H3 imino and either both CH^+–H4 and C–H4 amino (A) or other CH^+–H3 imino proton resonances (B). Solution and temperature conditions are the same as those given for Figure 3.

normally resonate (12–14.5 ppm). This distinction may prove to be a useful signature for detecting the presence of CH^+·C base pair formation. These results differ from those reported by Sarma and co-workers[25] on d(CTCTCT) at low pH, in which a resonance at 11.10 ppm was attributed to CH^+–H3 imino protons engaged in CH^+·C base pairing.

The presence of six, and not three, imino proton resonances suggests that at low pH, d($C_3T_4C_3$) adopts the parallel-stranded duplex conformation, [d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$)], presented as scheme 1 in Figure 2. This result also suggests that a mixture of the two structures presented as schemes 1 and 3 in Figure 2, which would result in as many as nine imino proton resonances, is not present under these conditions. Finally, PAGE studies on d($C_3T_4C_3$) at pH 5.4 in native polyacrylamide gels (data not shown) reveal only a single band. This is consistent with the presence of only one structural species and implies that a mixture of structures such as schemes 2 and 3 is not present, as these structures would have substantially different gel mobilities and give rise to two bands.

A very weak resonance line at 15.14 ppm also arises upon decreasing the pH to 5.5 (Figure 3). The proton to which this resonance corresponds is not clear. It showed no contacts to any other proton resonances in the NOESY experiment discussed

below. However, this lack of observed connectivity may simply be due to the weak intensity of the resonance line. Furthermore, in contrast to all the other resonance lines, the width and intensity of this resonance line do not change as the pH is lowered below 6.0, suggesting that it does not reflect a CH^+–H3 imino proton.

In addition to the induction of new imino proton resonances, lowering the pH is accompanied by a sharpening and an increase in intensity of the T–H3 imino proton resonances between 10 and 12 ppm (Figure 3). This result indicates that the intrinsic solvent exchange rate of the T–H3 protons becomes progressively slower with decreasing pH. Aside from the direct contribution of the pH change, the observed line sharpening may also be due in part to formation of the d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex structure. The restraints imposed on solvent accessibility of the thymine residues within the internal loop of the d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex may lead to an additional decrease in solvent exchange rates of the T–H3 imino protons. Assuming the parallel-stranded d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex is formed in conjunction with decreasing pH, eight rather than four T–H3 imino protons should be resonating in the 10–12-ppm region. Consistent with this assumption, five T–H3 resonance lines are present at pH 5.5 (Figure 3, top), two of which have intensities that are significantly higher than the others and probably reflect more than one imino proton. It should be pointed out that the NOESY experiment on d($C_3T_4C_3$) at pH 5.5 in 90% H_2O discussed below showed no evidence of NOE contacts between any of the thymine H3 imino proton resonances and those of the CH^+ residues. This result suggests that formation of T·T base pairs, which has been previously observed in an antiparallel DNA duplex by Kouchakdjian et al.,[40] does not occur in the parallel-stranded, d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex.

The aromatic and amino proton regions of the H_2O-NMR spectra of d($C_3T_4C_3$) at either pH 7.3 or 5.5 are shown in Figure 4, spectra A and B, respectively. At pH 7.3 (Figure 4A), very narrow resonance lines, corresponding to the H6 protons of cytosine and thymine, are present between 7.4 and 7.8 ppm. The narrowness of these aromatic proton resonances at pH 7.3 indicates that d($C_3T_4C_3$) exists in a single-stranded state under these conditions. Weaker and broader resonance lines are evident between 6.6 and 7.4 ppm at pH 7.3 (Figure 4A). These resonances reflect the cytosine amino protons of single-stranded d($C_3T_4C_3$), as they are not observable in D_2O.

At pH 5.5 (Figure 4B), new resonances, which are absent at pH 7.3 (Figure 4A), are present between 7.8 and 9.8 ppm. These resonances are not observable in D_2O at pH 5.5 (Figure 5), indicating that they correspond to exchangeable protons. Furthermore, they occur in a chemical shift region similar to that frequently observed for amino proton resonances in duplex DNA and upfield from that in which either the T–H3 or CH^+–H3 imino protons are resonating. Hence, these broad resonance lines most likely reflect cytosine amino protons engaged in hydrogen bonding and are indicative of CH^+·C base pair formation. In addition, the resonance lines corresponding to the H6 aromatic protons (7.4–7.8 ppm) are considerably broader at pH 5.5 than at pH 7.3 (compare spectra A and B in Figure 4). This result is also indicative of duplex or base pair formation. Additional evidence confirming the complete propagation of the duplex at pH 5.5 is provided by the disappearance of the amino proton resonances of single-stranded d($C_3T_4C_3$) (6.6–7.4 ppm) as the pH changes from 7.3 to 5.5 (compare spectra A and B in Figure 4). The sharp resonance line at 8.24 ppm in both spectra in Figure 4 is an artifact of the NMR spectrometer.

(b) Nonexchangeable Protons. The nonexchangeable proton NMR spectrum of the d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex at pH 5.5 in D_2O is shown in Figure 5. Resonances between 8 and 10 ppm are absent, thus confirming that the resonance lines observed in this region of the spectrum acquired in 90% H_2O (Figure 4B) reflect exchangeable, amino protons. More than four narrow and intense resonance lines corresponding to thymine methyl protons are present between 1.5 and 2.0 ppm. These resonances have significantly narrower line widths than are normally observed for methyl protons in duplex DNA. Similarly, some of the H6 aromatic proton resonances are narrower than others and are likely to reflect TH6 aromatic protons (Figures 4B and 5). These latter two results indicate that the runs of thymine residues exist in single-stranded states at pH 5.5 and are consistent with formation of the asymmetric, parallel-stranded d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex, presented as scheme 1 in Figure 2.

Two-Dimensional Proton NMR. In the model for the CH^+·C base pair presented in Figure 1, the CH^+–H3 imino protons should be close in space, and hence dipolar-coupled, to both the C–H4 and CH^+–H4 amino protons within the same base pair. In an effort to demonstrate these spatial connectivities, a NOESY experiment (250 ms mixing time) was carried out on the d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex in 90% H_2O at pH 5.5. Two portions of the resulting NOESY spectrum showing contacts between imino and either amino or other imino protons are presented in Figure 6, spectra A and B, respectively. Each of the CH^+–H3 imino proton resonances show either two, three, or four resolvable cross-peaks to resonances in the amino proton region (8–10 ppm) of the spectrum (Figure 6A). These cross-peaks reflect connectivities between the CH^+–H3 imino protons and the hydrogen- and non-hydrogen-bonded amino protons of both the CH^+ and C base-paired residues.

Three of the six CH^+–H3 proton resonances (at 15.52, 15.54, and 15.61 ppm) show much weaker contacts to those of the amino protons than the other three. Two of these three imino proton resonances have significantly higher peak intensities than the three showing the strong amino contacts, while the third has a lower peak intensity (Figure 3, top). The peak intensities of the three imino proton resonances showing strong amino proton contacts (at 15.47, 15.63, and 15.70 ppm) are all similar. Furthermore, imino proton–imino proton connectivities, as shown in Figure 6B, demonstrate that these three resonances reflect H3 imino protons of three contiguous CH^+ residues, and they are therefore at one end of the duplex. Although the other three imino proton resonances should also show contacts to each other, these contacts are not observed, as they reside too close to the diagonal. These results suggest that a structural asymmetry exists within the d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex. Fiber diffraction studies[14] on poly(rC) at a pH of approximately 5.0 have proposed formation of a double-stranded poly(rCH^+)·poly(rC) duplex, in which the strands are associated in a parallel orientation. This unusual orientation may be characteristic of all duplexes containing solely or some CH^+·C base pairs. In such a case, the 5′ end of the parallel-stranded duplex may have certain structural dissimilarities with respect to the 3′ end. Others have suggested that the structure formed by poly(rC) at low pH is a single-stranded helix with an A conformation.[41] However, the possible existence of the double-stranded helix is not ruled out by these studies.

Contacts between CH^+–H3 imino and either T–H6 or C–H6 aromatic protons were not observed (data not shown). The chemical shifts of the CH^+–H3 imino proton resonances, as well as those of the resolvable amino proton resonances to which they are coupled, are summarized in Table I. Additional two-dimensional NMR studies on the d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex in both 90% H_2O and D_2O are currently underway, in an effort to gain further insight into its structural characteristics.

Helix-Coil Transitions. Another method by which the molecularity of the structure formed by d($C_3T_4C_3$) at pH 5.5 can

(40) Kouchakdjian, M.; Li, B. F. L.; Swann, P. F.; Patel, D. J. *J. Mol. Biol.* **1988**, *202*, 139–155.

(41) Arnott, S.; Chandrasekharan, R.; Marttila, C. M. *Biochem. J.* **1974**, *141*, 537–538.

Table I. Resolvable Chemical Shifts of the Exchangeable Proton Resonances of the Cytosine Residues in the $d(CH^+_3T_4CH^+_3)\cdot d(C_3T_4C_3)$ Duplex for $CH^+\cdot C$ Base Pairing[a]

imino CH^+–H3	amino C–H4, CH^+–H4
15.47	8.10, 8.29, 9.25, 9.35
15.52	8.15, 8.28, 9.28, 9.37
15.54	8.17, 8.29, 9.23, 9.38
15.61	8.12, 9.32
15.63	8.10, 9.30, 9.39
15.70	8.11, 9.21, 9.31

[a] Buffer conditions are 15 mM sodium phosphate (pH 5.5) and 100 mM NaCl. Temperature is fixed at 2 °C. All chemical shifts are in units of ppm.

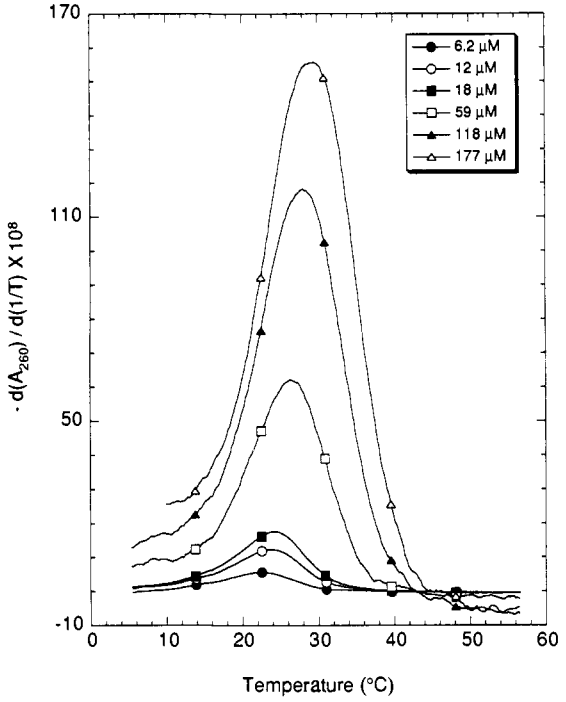


Figure 7. First-derivative plots of helix-coil transitions at 260 nm for the indicated strand concentrations of the $d(CH^+_3T_4CH^+_3)\cdot d(C_3T_4C_3)$ duplex in 10 mM sodium cacodylate (pH 5.5), 100 mM NaCl, and 0.1 mM EDTA.

be established is by conducting thermal denaturation studies at differing concentrations of $d(C_3T_4C_3)$. If the structure formed by $d(C_3T_4C_3)$ is a bimolecular duplex (Figure 2, schemes 1 and 2), the T_{max} values of these helix-coil transitions will increase with increasing DNA concentration. In contrast, the T_{max} values of these transitions will not change with increasing DNA concentration, if the structure formed is a monomolecular, hairpin duplex (Figure 2, scheme 3). First-derivative plots of the thermal denaturation profiles for differing concentrations of the duplex formed by $d(C_3T_4C_3)$ at pH 5.5 are shown in Figure 7. The T_{max} values clearly increase with increasing DNA concentration, and they are thus consistent with formation of a bimolecular duplex. Although this result does not distinguish between formation of the parallel or antiparallel bimolecular duplex (Figure 2, schemes 1 and 2, respectively), it is consistent with the findings of the NMR studies discussed above. Furthermore, the melting profiles are monophasic throughout the range of DNA concentrations, suggesting that a mixture of bi- and monomolecular duplexes is not present. If such a mixture were present, biphasic transitions would likely arise as the DNA concentration increases, since the T_{max} values of the transitions of the bimolecular duplex would increase, while those of the monomolecular duplex would remain unchanged. This result is also in agreement with the NMR and PAGE results discussed above.

Thermodynamic Analysis. A plot of $1/T_{max}$ vs $\log(C_t)$ for the $d(CH^+_3T_4CH^+_3)\cdot d(C_3T_4C_3)$ duplex-coil transitions at pH 5.5 is

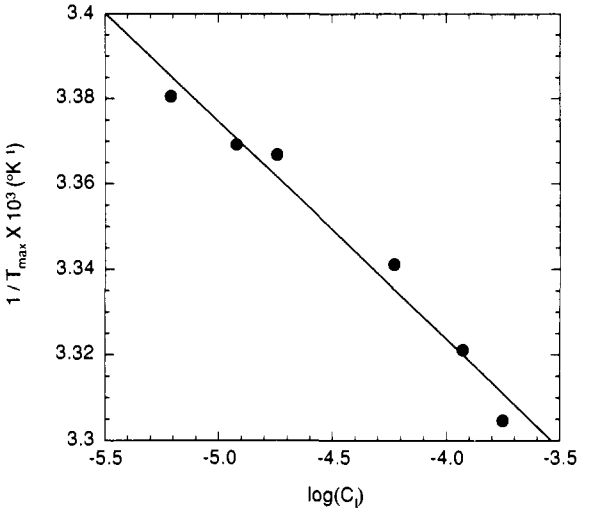


Figure 8. Plot of $1/T_{max}$ vs $\log(C_t)$ for the $d(CH^+_3T_4CH^+_3)\cdot d(C_3T_4C_3)$ duplex-coil transitions. The solution conditions are the same as those given for Figure 7.

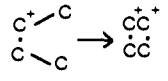
Table II. Thermodynamic Parameters for Formation of the $d(CH^+_3T_4CH^+_3)\cdot d(C_3T_4C_3)$ Duplex[a]

$\Delta H°$ (kcal/mol)[b,c]	$\Delta S°$ (eu)[b,c]	$\Delta G°_{25}$ (kcal/mol)[b,c]	K_{25} (M^{-1})
−89.4 ± 16.8	−280 ± 56	−6.0 ± 0.5	2.5×10^4

[a] Solution conditions are the same as those given for Figure 7. [b] Refers to moles of duplex formed. [c] Errors reflect the magnitudes of the 95% confidence limits from linear regression analysis.

presented in Figure 8. The thermodynamic parameters and equilibrium binding constant derived from this plot are listed in Table II. These thermodynamic properties describing the association of two strands of $d(C_3T_4C_3)$ can be analyzed in order to obtain an approximate value for the stability of the $CH^+\cdot C$ base pair.

Gralla and Crothers[37] have measured the thermodynamic stability of internal loops in RNA oligonucleotide duplexes, ranging from 2 to 6 unpaired cytosine bases, bordered on either side by a G·C base pair. These data were extrapolated to larger loop sizes, with the result that formation of a loop composed of 8 unpaired bases (four on each strand) would entail a change in free energy at 25 °C (ΔG_{loop}) of ~+3.0 kcal/mol. In addition, Breslauer and co-workers[42] have indicated that at 25 °C, the free energy change for closure of the first base pair of DNA oligonucleotide duplexes ($\Delta G_{nucleation}$) containing G·C base pairs may be approximated by +5.0 kcal/mol. Assuming values of ΔG_{loop} are similar in both RNA and DNA, and independent of the base composition of the loop, an estimate of the $\Delta G°_{growth}$, the change in free energy for addition of a $CH^+\cdot C$ base pair to a pre-existing $CH^+\cdot C$ base pair according to the reaction



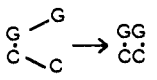
may be estimated by (−6.0–5.0–3.0)/4 = −3.5 kcal/mol.

While the above analysis clearly involves several assumptions, and hence should be considered as subject to some uncertainty, there are data which support the validity of some of the approximations used. For example, Aboul-ela et al.[43] have determined the stability of an internal loop composed of 2 unpaired cytosines bound by A·T base pairs in a DNA duplex to be 2.0 kcal/mol, which compares favorably with the value of 1.8 kcal/

(42) Breslauer, K. J.; Frank, R.; Blöcker, H.; Marky, L. A. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 3746–3750.
(43) Aboul-ela, F.; Koh, D.; Martin, F. H.; Tinoco, I., Jr. *Nucleic Acids Res.* **1985**, *13*, 4811–4824.

mol calculated by Gralla and Crothers[37] for a loop of 2 cytosines bound by A·U base pairs in an RNA duplex. Hence the free energy changes for loop formation in RNA and DNA duplexes are similar. Also, Aboul-ela et al. reported than when the cytosines in the loop were replaced with thymines, the value of ΔG_{loop} was observed to be 1.5 kcal/mol. While this represents a slightly greater dependence of loop stability on base composition, the differences are not large.

We can compare our results for ΔG_{growth} obtained above for CH^+·C base pairs with data reported by Breslauer and coworkers[42] on standard Watson–Crick base pairs. Thus, for adding a G·C base pair to a pre-existing one, i.e. for the reaction represented by



the corresponding ΔG_{growth} value is −3.1 kcal/mol. This represents a slightly smaller stability than our result for the CH^+·C base pair. Also, as expected, the CH^+·C base pair stability is greater than that of A·T base pairs. We can expect that these differences would be enhanced if our measurements were made at higher salt (1 M NaCl) corresponding to those of Breslauer et al.,[42] while they would be diminished if our measurements were carried out at higher pH. The thermodynamic contributions from nearest-neighbor interactions involving CH^+ residues are as yet unknown. Additional studies on oligonucleotides containing all the other possible CH^+–X (X = A, T, G) nearest neighbors will be necessary to evaluate these thermodynamics.

The $\Delta G°_{25}$ for formation of the d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex is 3.1–7.2 kcal/mol less negative (less favorable) than that which we have previously demonstrated for formation of either the $d(A)_{10}$·$d(T)_{10}$ or the d($G_3A_4G_3$)·d($C_3T_4C_3$) duplex at pH 5.5 in either 50 mM $MgCl_2$ or 2.0 M NaCl.[6,44] An obvious contributor to this difference is the absence of four base pairs in the former duplex. However, despite this reduced overall stability, the enthalpy change for formation of the d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex is more negative (more favorable) than that for formation of the other duplexes. Hence, formation of the d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex is associated with a large entropic cost ($\Delta S°$ = −280 eu), which is the main contributing factor to the comparatively unfavorable $\Delta G°_{25}$. Solvent effects may contribute, in part, to this large entropic barrier. Water molecules, which would normally be released upon base pair formation, may remain clustered around the thymine nucleotides in the single-stranded, looped regions of the d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex.

Electrostatic effects are likely to contribute to the reduced stability of the d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex as well. Positive charges on adjacent CH^+ residues are likely to be in close enough proximity to exert repulsive forces against one another. These forces would tend to compensate attractive, stacking interactions that might occur between neighboring CH^+ residues. However, despite the presence of adjacent positive charges, $\Delta H°$ is quite favorable. Manzini et al.[36] have indicated that one out of six cytosine residues in oligomeric DNA are protonated at pH 5.5, which suggests that protonation of the additional four cytosine residues occurs in conjunction with base pairing. Furthermore, they have estimated that protonation of a cytosine residue is associated with a $\Delta H°$ of −2.8 kcal/mol. Protonation of cytosine residues may therefore contribute as much as −11.2 kcal/mol (≈13%) to the overall $\Delta H°$ of duplex formation. This potential enthalpic contribution might account, in part, for the favorable overall $\Delta H°$.

Conclusions

One- and two-dimensional NMR studies indicate that at low pH, d($C_3T_4C_3$) self-associates to form a parallel-stranded, bimolecular duplex [d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$)], containing six CH^+·C base pairs and an internal loop of eight non-base-paired thymine residues (see Figure 2, scheme 1). T_{max} values of the thermal denaturation profiles for d($C_3T_4C_3$) at low pH increase with increasing DNA concentration, consistent with formation of a bimolecular duplex. Thermodynamic analysis of thermal denaturation profiles for the d($CH^+_3T_4CH^+_3$)·d($C_3T_4C_3$) duplex suggests that propagation of CH^+·C base pairs is more favorable than W–C A·T or G·C base pairs.

Acknowledgment. This research was supported by U.S. Public Health Service Grant CA27343, awarded by the National Cancer Institute, and the University of California Cancer Research Coordinating Committee. D.S.P. is supported, in part, by Training Grant GM08284, awarded by the National Institute for General Medical Sciences, Department of Health and Human Services. We thank Dr. Corey Levenson of the Cetus Corporation, Emeryville, CA, for help with the synthesis and purification of d($C_3T_4C_3$).

Note Added in Proof: We have recently become aware of the publication by Luo et al.[45] on the acid-induced parallel duplex of d(A5C5) in which they observe imino protons from hemiprotonated CH^+·C base pairs in the same chemical shift region as this work.

(44) Pilch, D. S. Ph.D. Thesis, University of California, San Francisco, 1991; pp 8–71.

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