Cytosine–Cytosine$^+$ Base Pairing Stabilizes DNA Quadruplexes and Cytosine Methylation Greatly Enhances the Effect†

Charles C. Hardin,* Matthew Corregan, Bernard A. Brown II, and Lori N. Frederick

Department of Biochemistry, North Carolina State University, Raleigh, North Carolina 27695

Received June 18, 1992; Revised Manuscript Received March 17, 1993

ABSTRACT: Previous spectroscopic studies demonstrated that the oligodeoxynucleotide d(CGC G_3 GCG) undergoes a reversible cation-dependent transition between Watson–Crick (WC) hairpin and parallel-stranded "G-DNA" quadruplex structures [Hardin, C. C., Watson, T., Corregan, M., & Bailey, C. (1992) *Biochemistry 31*, 833–841]. The relative stabilities of the structures were assessed as a function of pH, and it was found that the quadruplex was substantially stabilized (ΔT_m = +15 °C) when the pH was shifted from 7.5 to 6 (apparent pK_a = 6.8). In the present study, the effects of different cations and pH on four specific sequence varients were determined to test the proposal that this stabilization is due to C·C$^+$ base pair formation mediated by N3-protonation of cytosine. Characteristically large differences in stability were observed when structures formed by d(TAT G_3 ATA) and d(TAT G_4 ATA) were thermally dissociated at pH 7 in the presence of different cations, verifying that G_n tracts bordered by TAT- and -ATA sequences form quadruplex structures. Imino proton NMR results indicate that the d(m^{5}C G m^{5}C G_3 G m^{5}C G)$_4$ and d(TAT G_4 ATA)$_4$ quadruplex structures are parallel-stranded. It was necessary to increase the K$^+$ concentration from 40 mM to ca. 200 mM to stabilize d(TAT G_3 ATA)$_4$, while the d(TAT G_4 ATA)$_4$ complex was nearly as stable as the quadruplex formed by d(CGC G_3 GCG) under the same conditions. The d(TAT G_4 ATA)$_4$ quadruplex was only slightly stabilized at pH 6 relative to pH 7.5 (ΔT_m = +3 °C), confirming that the unique stabilization that occurs in the pH 6.8 range with [d(CGC G_n GCG)$_4$·ion$_n$] complexes is due to the C residues. The sequence d(m^{5}C G m^{5}C G_3 G m^{5}C G) was found to form a very stable quadruplex in K$^+$ or Ca^{2+}. As with the quadruplex formed by the unmethylated analog, the stability is greatly enhanced when the pH is decreased below about 7.2 (p$K_{a,obs}$ = 6.8). Dissociation kinetic constants and activation energies were determined for quadruplexes formed by d(CGC G_3 GCG), d(m^{5}C G m^{5}C G_3 G m^{5}C G) and d(TAT G_4 ATA). Quantitative comparisons showed that methylation produces a complex that is much more stable at pH 7 in 40 mM Na$^+$ than either of the unmodified structures; the rate-limiting activation energy for dissociation of d(CGC G_3 GCG)$_4$ was 22 kcal mol^{-1} less than for the methylated analog. Statistical analysis of the kinetic data showed that at least three distinct processes occurred with the C- and m^{5}C-containing molecules, while only two different processes could be resolved with the d(TAT G_4 ATA) quadruplex. The results suggest that cytosine methylation stabilizes the complex primarily by producing more favorable entropic (stacking) interactions, not by causing a positive shift in the pK_a for cytosine protonation. To summarize, the following factors can act alone or synergistically to stabilize DNA quadruplexes: (1) moderate increases in the concentrations of K$^+$ or Ca^{2+}, and to a smaller degree Na$^+$, within the physiological ranges, (2) moderate pH decreases through the ranges that occur in eukaryotic nuclei, and (3) methylation of cytosine residues at GCG sites.

Guanosine/cytosine-rich DNA sequences predominate in two important locations within the eukaryotic genome: (i) in highly repetitive sequences at the chromosomal termini called telomeric DNAs , and (ii) in "CpG island" domains within and around the coding and promoter regions of the genes (Bernardi, 1989). These types of sequences differ in cytidine content and strand distribution. Guanosine residues in telomeric DNAs are segregated from cytidine residues, occurring as G_n tracts within specific repetitive sequences that are located exclusively on the 3′-terminal strands (e.g. *Tetrahymena*, T_2G_4; humans, T_2AG_3; *Arabidopsis*, T_3AG_3; *Dictyostelium*, $AG_6GAGAG_6AG_6$). Cytidines are rarely found in the 3′-terminal telomeric strands. In contrast, both G and C typically occur within interspersed motifs that are often on the same strand within "CpG island" sequence domains.

In vitro studies with G-rich telomeric DNA sequences have shown that they can form both intra- and intermolecular quadruplex structures (Henderson et al., 1987; Hardin et al., 1991). The stabilities of these stuctures are uniquely sensitive to changes in the concentrations of important physiological cations such as K$^+$ and Ca$^+$ (Williamson et al., 1989; Sen & Gilbert, 1990; Guschlbauer, 1990; Jin et al., 1990; Hardin et al., 1991, 1992; Raghuraman & Cech, 1990; Zahler et al., 1991). The structures interconvert depending upon solution conditions, including amount and type of cation(s), DNA concentration(s), and temperature. The differential stabilizing effects of different sized ions arises from the fact that O6-carbonyl oxygens from each of the eight guanines that surround the "guest" cation bind to the ion to form a nearly octahedral "cryptand" complex (Sundquist & Klug, 1989).

† These studies were supported by grants from the National Institutes of Health (BRSG RR9223 and GM47431), the North Carolina Biotechnology Center, and the North Carolina Agricultural Research Service.

* To whom correspondence should be addressed.

[1] Abbreviations: C$^+$, N3-protonated cytosine cation; C·C$^+$, hemiprotonated cytosine·cytosine cation base pair; CD, circular dichroism; KP, 20 mM potassium phosphate (pH 7) containing 0.1 mM EDTA; m^{5}C, 5-methylcytosine cation; m^{5}C$^+$, N3-protonated m^{5}C; NaP, 20 mM sodium phosphate (pH 7) containing 0.1 mM EDTA; Tris, tris(hydroxymethyl)aminomethane; WC, Watson–Crick.

0006-2960/93/0432-5870$04.00/0

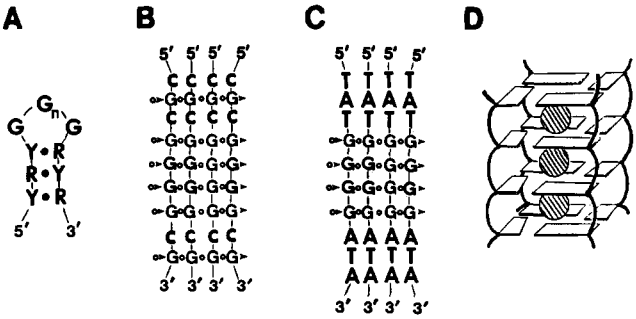

FIGURE 1: Structures formed by the oligonucleotides used in this study. (A) Watson–Crick hairpin, where "Y" and "R" refer to pyrimidine and purine residues, respectively. The subscript n is equal to 1 for the C- and m5C-containing sequences. According to the available evidence, only the C- and m5C-containing oligonucleotides formed appreciable amounts of this structure. (B) Parallel-stranded quadruplex structures formed by d(CGC G_3 GCG) (Hardin et al., 1992) and d(m5C G m5C G_3 G m5C G) (this paper). Symbols: WC base pairs (●); G·G base pairs (○). (C) Parallel-stranded quadruplex structure formed by d(TAT G_4 ATA). (D) Three-dimensional structure of a "generic" (cation·quadruplex) complex. The parallel-stranded quadruplex would represent the four adjacent sets of G·G base-paired residues shown in panels B and C.

The DNA oligonucleotide d(CGC G_3 GCG) can form either *Watson–Crick (WC)*[1] *hairpin* (Figure 1A) or *parallel-stranded quadruplex* (Figure 1B,D) structures under physiological-like conditions (Hardin et al., 1992). Exchange between each of these structures and unpaired strands could be monitered at the same time using CD. It was found that: (i)the hairpin to quadruplex equilibrium is reversible; (ii) the populations can interconvert almost completely by changing the monovalent cation and DNA concentrations; (iii) at a DNA strand concentration of 0.7 mM, the d(CGC G_3 GCG)$_4$ quadruplex is more kinetically and thermodynamically more stable than the hairpin under physiological-like K^+ concentration, pH, and temperature; (iv) K^+ has the optimum ionic radius for complex formation; (v) equivalently sized divalent cations stabilize the quadruplex more effectively than the corresponding monovalent cations; (vi) both forms had an apparent pK_a in the 9–10 range due to dissociation of the guanosine imino protons (T'so, 1974), only the quadruplex had an apparent pK_a in the physiological range (at ca. 6.8).

Hemiprotonated C·C+ base pair formation has been proposed to explain the acid-induced stabilization of complexes composed of both A+C- and G+C-rich DNA fragments (Edwards et al., 1990; Cruse et al., 1983). In each of these cases, complex formation raised the pK_a from a value of ca. 4.2 for the nucleoside to around 7. On the basis of these considerations, it was proposed that the three sets of cytosine residues are constrained in the parallel-stranded quadruplex between the stacked G quartets in a conformation which is conducive to base pair formation (Hardin et al., 1992). Since two hydrogen atoms are incorporated into each stack of C·C+ base pairs in the quadruplex, a slight decrease in pH below 7 has a large stabilizing effect.

The results in this paper verify the role of C·C+ base pair formation in mild acid-induced quadruplex stabilization and show that cytidine methylation greatly stabilizes the structure. This may in part explain the evolutionary selection against cytidine residues in G-rich telomeric DNAs, which must "open up" to the single-stranded form during replication to perform the decoding function. The results also suggest that methylation of GC-rich duplex transcriptional promoter DNAs in "CpG islands" may induce quadruplex formation within heterochromatin domains, resulting in chromosomal condensation. These observations might provide information about a new class of interactions that might occur within or between chromatin "domains" *in vivo*.

MATERIALS AND METHODS

Sample Preparation. Oligonucleotides d(CGC G_3 GCG), d(m5C G m5C G_3 G m5C G), d(TAT G_3 ATA), and d(TAT G_4 ATA) were synthesized using standard phosphoramidite methods and purified by preparative gel electrophoresis as follows. Polyacrylamide gels (20%) were run in 89 mM Tris-borate (pH 8.2) containing 2 mM EDTA and 7 M urea; the two DNA bands (quadruplex = upper; hairpin/single strands = lower) were extracted from the excised gel slices with 10 mM Tris-HCl (pH 8) containing 300 mM NaCl and 1 mM EDTA. Extracts were dialized into 0.1 mM Li_2EDTA (pH 7), lyophilized to dryness, and resuspended into a minimal volume of H_2O. These samples were dialyzed extensively against 0.1 mM Li_2EDTA at ca. 1 mM DNA strand concentration and stored at 4 °C; quadruplexes were stable under these conditions for more than 1 year.

Spectroscopy. Circular dichroism spectra were collected under the solution conditions described in the text and figure legends at controlled sample temperatures on a Jasco J-600 spectropolarimeter (Hardin et al., 1992). Proton NMR data were obtained on a GE Omega system at 500.1 MHz using "1–1 hard pulse" to avoid exciting the solvent HDO resonance. Chemical shifts are reported relative to internal standard 3-(trimethylsilyl)-1-propanesulfonate.

Quadruplex Dissociation Kinetics Measurements. The effect of cytidine methylation on the stability of the quadruplex formed by d(CGC G_3 GCG) was studied by comparing heat-induced dissociation kinetics for the methylated and unmodified structures. Samples of each type of oligonucleotide were prepared for kinetics analysis by diluting the stable quadruplex stock solutions into 20 mM sodium phosphate (NaP), pH 7 at 23 °C. The samples were then placed into a water-jacketed CD cell, and the temperature was "jumped" (within 1–2 s) to values ranging from 50 to 80 °C as data acquisition was begun. Quadruplex dissociation kinetics were observed by CD at 264 nm (Hardin et al., 1992) over 2-, 4-, and/or 6-h periods.

The data were converted to ASCII format on an IBM PS2, transferred to a VAXstation 3600 computer, and analyzed using a version of the exponential multicomponent fitting program DISCRETE (Provencher, 1975; Provencher & Vogel, 1980) which was modified to read digital input. Statistical best fits gave reaction order values (i = 1, 2, or 3 for these reactions) and the corresponding rate constants ($k_{i,obs}$) were determined by residuals analysis using DISCRETE. Dissociation kinetic constants were converted from units of ellipticity change ($\Delta\Delta\epsilon$) per second (mdeg s^{-1}) to first-order rate constants describing the number of moles of quadruplex dissociated per second (s^{-1}) using molar extinction coefficients and absorbance results. Absorbance values were measured for the freshly diluted oligonucleotides in NaP at 23 °C prior to kinetic analysis. Extinction coefficients for the undissociated species were determined by interpolating values for the denatured single-stranded oligonucleotides (obtained after incubation for 1 h at 80 °C) to 23 °C using thermal denaturation ("melt") data.

To obtain activation energy parameters, calculated first-order dissociation rate constants ($k_{1,obs}$) were plotted as a function of inverse temperature. The slope of the resulting Arrhenius plot corresponds to the activation energy (E_A) for the dissociation reaction. First-order kinetic constants were statistically characterized by the standard error in fitting the

ca. 2500 kinetic point files by DISCRETE. Correlation coefficients for the linear regression fits of the Arrhenius data are listed in Table III. Activation energies were used to calculate the "quasithermodynamic" transition state enthalpies ($\Delta H^{\dagger\circ}$) via eq 1 (Eisenberg & Crothers, 1979).

$$E_A = \Delta H^{\dagger\circ} + RT \quad (1)$$

Gibbs free energies ($\Delta G^{\dagger\circ}$) and entropies ($\Delta S^{\dagger\circ}$) for the transitions were calculated via eqs 2–4:

$$k_{1,obs} = (k_B T/h) K^{\dagger} K \quad (2)$$

$$\Delta G^{\dagger\circ} = -RT \ln K^{\dagger} \quad (3)$$

$$\Delta G^{\dagger\circ} = \Delta H^{\dagger\circ} - T\Delta S^{\dagger\circ} \quad (4)$$

where k_B is Boltzmann's constant, h is Planck's constant, $K^{\dagger}$ is the *equilibrium constant for formation of the activated complex*, and K is the "transmission coefficient" which describes the number of complexes that pass over the activation energy barrier without returning (assumed to be 1; Eisenberg & Crothers, 1979).

RESULTS

C·C⁺ Base Pairing Stabilizes the Quadruplex Structure. The CD spectrum of the quadruplex d(CGC G_3 GCG)$_4$ (Figure 1B,D) is characterized by a signature band centered at 264 nm that decreases in magnitude under solution conditions which destabilize the quadruplex (Figure 1B; Hardin et al., 1992). A second CD band centered at 286 nm was shown to correspond to the WC hairpin (Figure 1A). These resolved bands allowed us to monitor the stabilities of the quadruplex and WC hairpin when a mixture of the two forms is present. This is illustrated for d(CGC G_3 GCG), d(m^{5}C G m^{5}C G_3 G m^{5}C G), and d(TAT G_4 ATA) in Figures 2 and 3. Spectra obtained with d(CGC G_3 GCG) samples that are predominantly in the *quadruplex* and *WC hairpin* forms, respectively, are shown in Figure 2A,B. Both total DNA concentration and pH affect the relative amounts and interconversion between these species (Hardin et al., 1992). The results shown in Figure 2A,B were obtained using 7 μM samples in 20 mM NaP (pH 6) buffer. The lower pH stabilized the quadruplex relative to the same experiment at pH 7 (Hardin et al., 1992). At pH 6, a small amount of intact quadruplex d(CGC G_3 GCG)$_4$ was still present despite heating to 80 °C, cooling, and heating *a second time* (Figure 2B). Essentially no quadruplex survived when this process was done at pH 7 (Hardin et al., 1992). These subsequent denaturation steps are referred to as "first melt" and "second melt" here.

Figure 2C shows CD spectra obtained after the *methylated* d(m^{5}C G m^{5}C G_3 G m^{5}C G)$_4$ sample was heated to 80 °C *the second time* at pH 6 in 20 mM NaP. The 264-nm peak only decreased by about 15%. This point is especially interesting because this corresponds to the *second* melt. These results demonstrates that the m^{5}C-containing quadruplex is both kinetically and thermodynamically stable at 80 °C in 40 mM Na^+ at pH 7.5 at a DNA strand concentration of 7 μM. Thus, the methylated complex is stable at a relatively dilute DNA concentration under ionic conditions that do not stabilize the d(CGC G_3 GCG)$_4$ quadruplex.

Figure 3 shows a direct comparison of CD spectra obtained when d(CGC G_3 GCG)$_4$, and structures formed by d(m^{5}C G m^{5}C G_3 G m^{5}C G) and d(TAT G_4 ATA) were denatured at pH 7 in NaP at a strand concentration of 7 μM. Comparison of Figures 2A and 3A shows that the 264-nm CD band in the

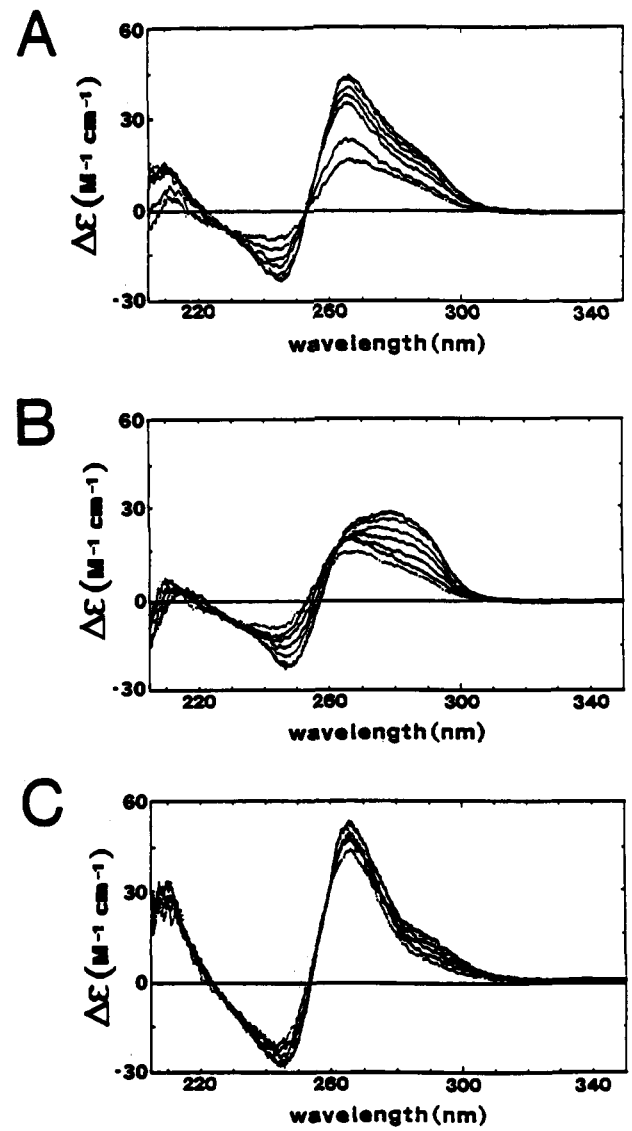

FIGURE 2: CD spectra obtained during thermal denaturation of (A and B) d(CGC G_3 GCG) and (C) d(m^{5}C G m^{5}C G_3 G m^{5}C G) DNAs in NaP (pH 6) buffer. (A) "First melt" obtained with the d(CGC G_3 GCG) sample after dilution to ca. 7 μM (in DNA strand concentration) from a ca. 1 mM stock solution. Spectra were obtained at 23, 30, 40, 50, 60, 70, and 80 °C; the ellipticity values in the 264–300-nm range decreased as the sample temperature was increased. (B) "Second melt" of the same sample used to obtain the results shown in panel A showing a ca. 70% loss of the quadruplex structure, assessed by the change in CD at 264 nm (Hardin et al., 1992), during the "first melt". This structure was replaced by the WC hairpin form upon cooling, as indicated by the increased CD at 282 nm. (C) "Second melt" of the equilibrium mixture of quadruplex and WC hairpin structures formed by d(m^{5}C G m^{5}C G_3 G m^{5}C G). These results demonstrate the higher stability of the m^{5}C-containing quadruplex relative to the unmodified analog.

spectrum of d(CGC G_3 GCG) decreased less at higher temperatures for samples at pH 6 versus at pH 7.5. A similar but more pronounced difference is seen when one compares spectra obtained with d(m^{5}C G m^{5}C G_3 G m^{5}C G)$_4$ at pH 6 and 7.5, respectively (Figures 2C and 3B). In contrast, a similar large stabilization was not observed in the spectra of d(TAT G_4 ATA)$_4$ at lower pH (Figure 5C). At pH 7.5, the intensities of the 264-nm bands in the spectra of both d(CGC G_3 GCG)$_4$ and d(TAT G_4 ATA)$_4$ decreased to approximately the same values as the temperature was increased (Figure 3, panels A and C, respectively), demonstrating that the stabilities are of similar magnitude. In contrast, the different surrounding base sequences produced a significant difference in

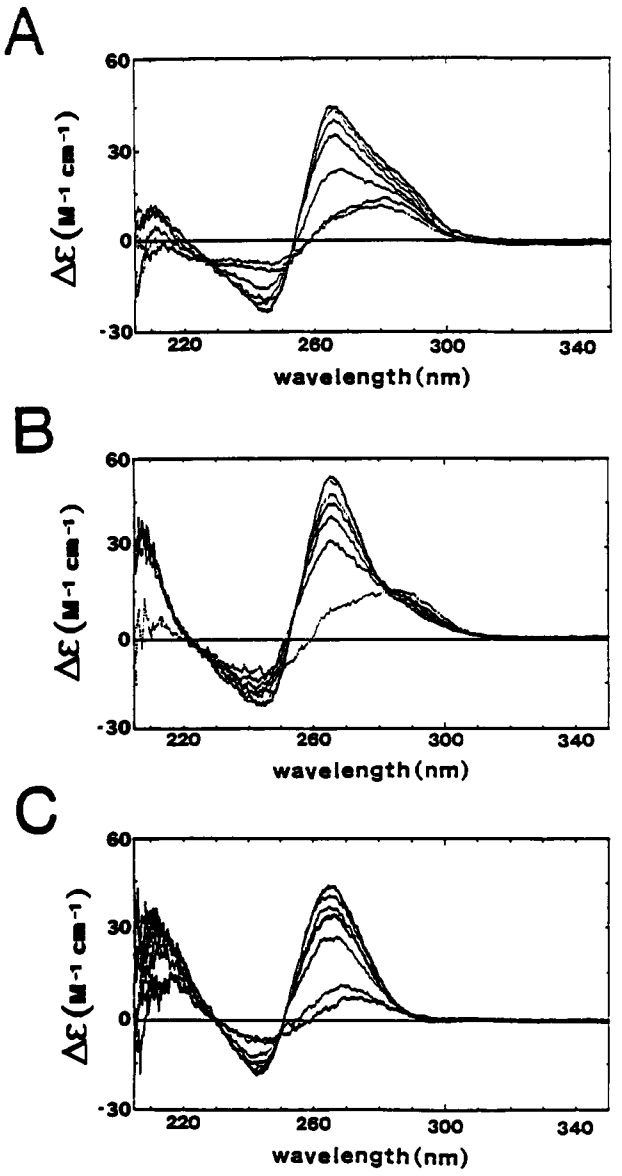

FIGURE 3: Thermal denaturation of the quadruplexes formed by (A) d(CGC G_3 GCG), (B) d(m^5C G m^5C G_3 G m^5C G), and (C) d(TAT G_4 ATA) in NaP buffer at pH 7.5 monitered by CD spectroscopy. Spectra were obtained at 23, 30, 40, 50, 60, 70, and 80 °C; the ellipticity values in the 264–300-nm range decreased as the sample temperature was increased. DNA strand concentrations were ca. 7 μM. Under these conditions, the C- and m^5C-containing oligonucleotides formed WC hairpins after denaturing then cooling the sample to 23 °C. The d(TAT G_4 ATA) sample did not re-form a structure with a CD spectrum that was significantly different from that of the random coil after denaturation to single-strands and then cooling to 23 °C.

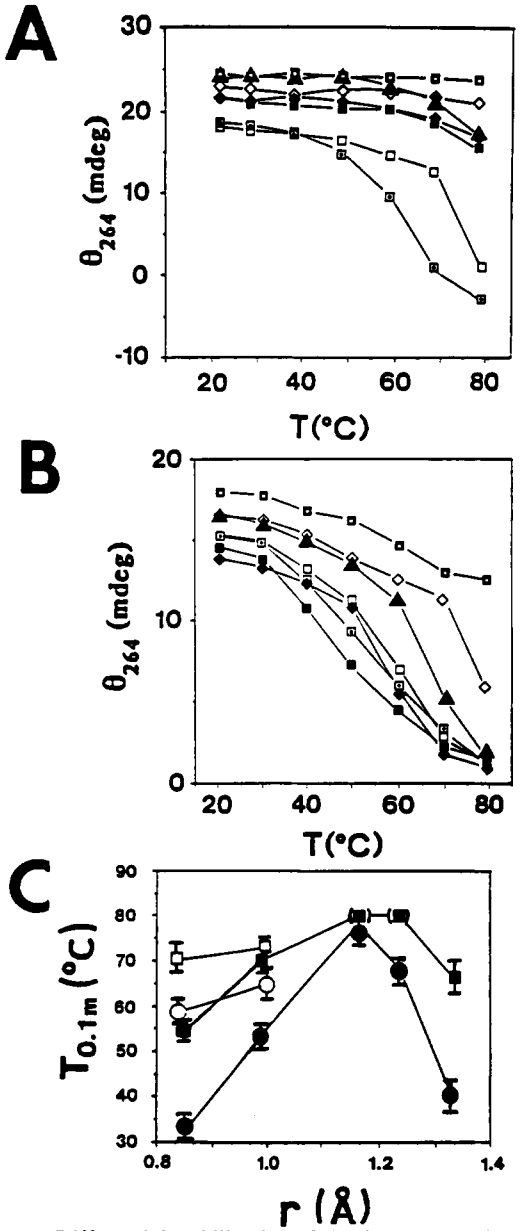

FIGURE 4: Differential stabilization of the G-DNA quadruplex by different ions and sequences. Effect of different ions on the stabilities of (A) d(m^5C G m^5C G_3 G m^5C G) and (B) d(TAT G_4 ATA) containing 10 mM Tris-HCl (pH 7), 0.1 mM EDTA, and each of the following salts at concentrations of 20 mM: LiCl (⊡); NaCl (□); KCl (■); RbCl (◇); CsCl (■); $MgCl_2$ (◆); $CaCl_2$ (▲). DNA strand concentrations were ca. 7 μM. (C) Plots of the temperature corresponding to a 10% denatured quadruplex population ($T_{0.1m}$) versus the van der Waals radius of the corresponding hydrated ion. Symbols: (□, ■) d(m^5C G m^5C G_3 G m^5C G) and (○, ●) d(TAT G_4 ATA).

the stabilities of d(CGC G_3 GCG)$_4$ and d(TAT G_4 ATA)$_4$ at pH 6 (ΔT_m = +11 °C; Figure 5C).

Figure 4 shows the effects of different monoatomic cations on the stabilities of the complexes formed by d(m^5C G m^5C G_3 G m^5C G) and d(TAT G_4 ATA). CD spectra were obtained using samples that contained the oligonucleotide dissolved in 10 mM Tris-HCl (pH 7), 0.1 mM EDTA, and 20 mM cation chloride. As in previous experiments (Hardin et al., 1992), the samples were initially prepared by diluting a ca. 1 mM stock solution of the DNA in 0.1 mM Li_2EDTA to approximately 7 μM. Heat-induced dissociation curves for d(m^5C G m^5C G_3 G m^5C G)$_4$ and d(TAT G_3 ATA)$_4$ complexes formed in the presence of different cations at pH 7 are shown in Figure 4, panels A and B, respectively. As was found with d(CGC G_3 GCG)$_4$ at pH 7, the stabilities of both complexes increased as the cation radius approached that of K^+ from both directions in the group 1A series (Figure 4). This occurs even though this quadruplex cannot reassociate appreciably under these conditions as a result of the metastable nature of the complex after dilution from DNA stock conditions (Hardin et al., 1992). This implies that cations from the external buffer can replace the less tenacious Li^+ ions from the diluted stock complex without complete strand separation. The stability of each type of complex varies depending upon the type of bound cation. These stabilities were compared by plotting the temperatures corresponding to a 10% dissociated quadruplex population (referred to as $T_{0.1m}$) versus the hydrated radius of the ion. This was done because the normal T_m, which correspond to a 50% dissociated quadruplex population, cannot

be determined for many of the ions with d(m^{5}C G m^{5}C G_3 G m^{5}C G)$_4$. These differential stabilization patterns closely resemble each other and the pattern obtained in an analogous experiment with d(CGC G_3 GCG)$_4$ at pH 7; however, $T_{0.1m}$ values for the d(m^{5}C G m^{5}C G_3 G m^{5}C G)$_4$ complex averaged 15–20 °C higher (Hardin et al., 1992). The stabilities depended upon both the size and charge of the cations, and K^+ produced the maximal amount of stabilization with each sequence. Since this large differential cation-dependent stabilization has not been seen with any other type of nucleic acid complex, it is clear that both d(m^{5}C G m^{5}C G_3 G m^{5}C G) and d(TAT G_4 ATA) form quadruplexes. As was noted with d(CGC G_3 GCG)$_4$, divalent cations stabilized these complexes more effectively than physiological monovalent cations of similar size [also see Guschlbauer et al. (1989)]. With all three complexes, the extent of differential stabilization observed as the radius of the divalent cation was increased in going from Mg^{2+} to Ca^{2+} was less dramatic than with equivalently sized divalent cations.

Figure 5 shows the effects of pH changes on the stabilities of d(m^{5}C G m^{5}C G_3 G m^{5}C G)$_4$ and d(TAT G_4 ATA)$_4$ in NaP. While the pattern obtained with d(m^{5}C G m^{5}C G_3 G m^{5}C G)$_4$ was similar to the one obtained with d(CGC G_3 GCG)$_4$ (Hardin et al., 1992), it was distinctly different from the one obtained with d(TAT G_4 ATA)$_4$, which had only a minimal increase in stability as the pH was decreased below 7 (ΔT_m = +3 °C; see Figure 5C). In contrast, the T_m values for d(CGC G_3 GCG)$_4$ and d(m^{5}C G m^{5}C G_3 G m^{5}C G)$_4$ increased by 11–12 °C when the pH was decreased from 7.5 to 6. Thus, although d(TAT G_3 ATA)$_4$ can form a quadruplex, the complex does not exhibit the large stabilization that was observed with both d(CGC G_3 GCG)$_4$ and d(m^{5}C G m^{5}C G_3 G m^{5}C G)$_4$ below pH 7.5. Cytidine-containing quadruplexes are preferentially stabilized relative to non-C-containing quadruplexes as the pH is reduced from 7.5 to 6.5. All of the complexes showed evidence for apparent pK_a values in the 9–10 range (Figure 5C; Hardin et al., 1992), due to dissociation of the guanosine imino protons (T'so, 1974). In contrast, only the C- or m^{5}C-containing quadruplexes had apparent pK_a values in the physiological pH range (at ca. 6.8). Under the conditions used in this work, the cytidine and m^{5}C protonation reactions occurred within the respective complexes with *apparent* pK_a values of ca. 6.8 (Hardin et al., 1992; Figure 5C).

The T_m value obtained under the same conditions with d(m^{5}C G m^{5}C G_3 G m^{5}C G)$_4$ was higher than that of d(CGC G_3 GCG)$_4$ by ca. 15 °C (83 vs 68 °C; Figure 5; Hardin et al., 1992). The T_m for dissociation of the complex formed by d(TAT G_4 ATA) (ca. 57 °C) was about 11 °C less than that of d(CGC G_3 GCG)$_4$. Since the complexes do not readily reassociate under these salt and DNA concentration conditions, the equilibria are poised far toward the dissociated species, and the diluted intact quadruplex species is only metastable under the conditions used in the experiments shown in Figure 5 (Hardin et al., 1992). These results show a trend in metastability that decreases in the order d(m^{5}C G m^{5}C G_3 G m^{5}C G)$_4$ > d(CGC G_3 GCG)$_4$ > d(TAT G_4 ATA)$_4$.

Quadruplexes Formed by d(m^{5}C G m^{5}C G_3 G m^{5}C G) and d(TAT G_4 ATA) Are Parallel-Stranded. Imino proton NMR spectra (Figure 6) were obtained with the quadruplexes formed by d(m^{5}C G m^{5}C G_3 G m^{5}C G) and d(TAT G_4 ATA) in 20 mM KP (pH 7) to determine whether these structures are parallel- or antiparallel-stranded. Integration of the major peaks shown in Figure 6A showed that if one assumes that the smaller peak at 11.34 ppm represents four protons, the slightly downfield peak (11.42 ppm) represents eight protons, and the larger upfield set of peaks represents twelve protons (10.97–11.03 ppm). These peaks are in the range observed for other quadruplex DNAs and the spectrum is very similar to results obtained with d(CGC G_3 GCG)$_4$ at 22 °C (Hardin et al., 1992). When the temperature was increased to 37 °C, the 11.42 ppm peak was resolved into two peaks, and the composite peak at 10.97 ppm spread out into three partially resolved peaks (not shown), analogous to the pattern of six partially resolved peaks obtained with d(CGC G_3 GCG)$_4$ at the same temperature (Hardin et al., 1992; unpublished results). In addition, resonances due to a minor species at 13.09 and 13.12 ppm were present in a 1:2 ratio (Figure 6A).

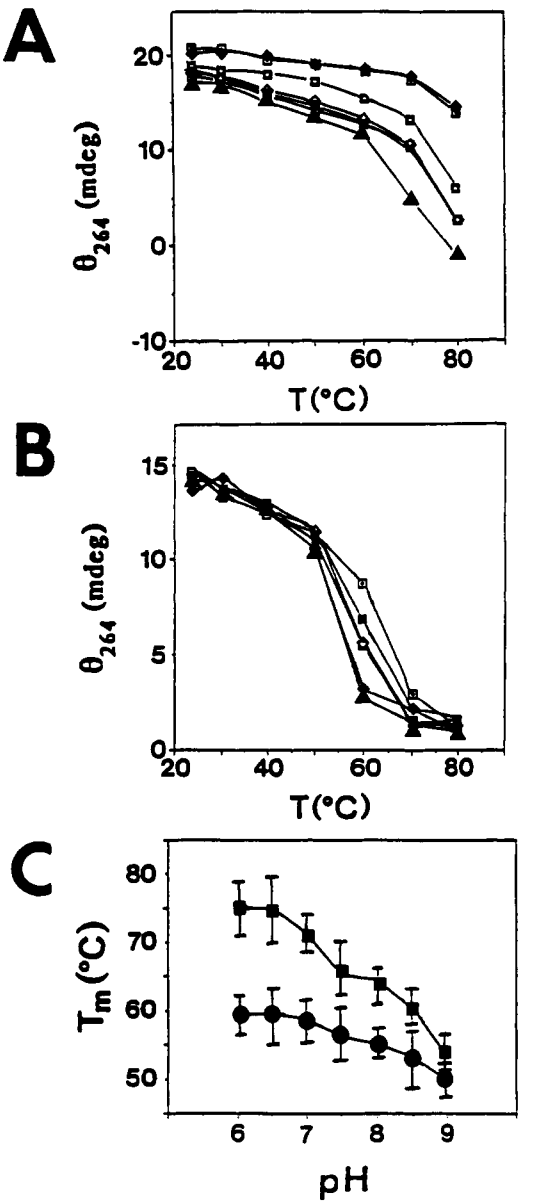

FIGURE 5: Differential stabilization of the G-DNA quadruplex a t different pHs and by different sequences. (A) Effect of pH changes on the stability of d(m^{5}C G m^{5}C G_3 G m^{5}C G) in 20 mM sodium phosphate containing 0.1 mM EDTA. Spectra were obtained at the following pH values: 6 (◆); 6.5 (⊡); 7 (■); 7.5 (◇); 8 (■); 8.5 (□); 9 (▲). (B) Effect of pH changes on the stability of d(TAT G_4 ATA) in 20 mM sodium phosphate containing 0.1 mM EDTA at the following pHs: 6 (⊡); 6.5 (■); 7 (■); 7.5 (◇); 8 (□); 8.5 (◆); 9 (▲). DNA strand concentrations were ca. 7 μM. (C) Temperatures corresponding to the average of the maximum and extrapolated minimum ellipticities (T_m) plotted as a function of the pH values for the 264-nm CD band. Symbols: (◆) d(m^{5}C G m^{5}C G_3 G m^{5}C G) and (⊡) d(TAT G_4 ATA).

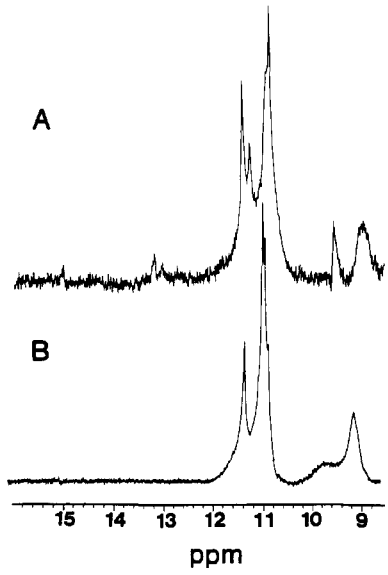

FIGURE 6: Imino ^{1}H NMR spectra of (A) d(m^{5}C G m^{5}C G$_3$ G m^{5}C G)$_4$ and (B) d(TAT G$_4$ ATA)$_4$ in 20 mM potassium phosphate (pH 7) containing 0.1 mM EDTA at 23 °C. Spectra were obtained using the "1–1 hard pulse" solvent supression method in 128 scans. DNA strand concentrations were ca. 0.7 mM.

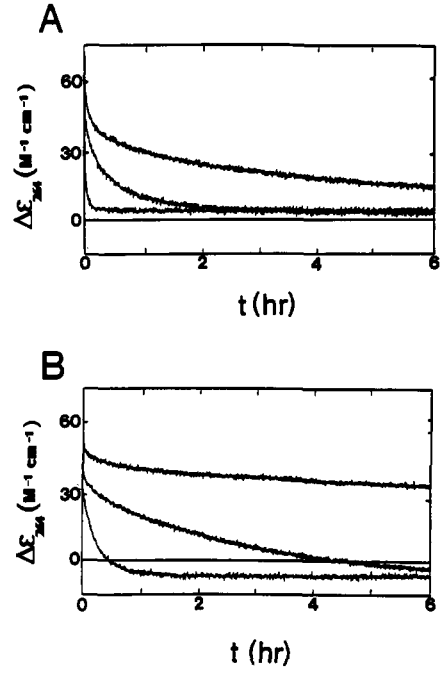

FIGURE 7: Effect of cytosine methylation on the relative stabilities of quadruplexes formed by d(CGC G$_3$ GCG) and d(m^{5}C G m^{5}C G$_3$ G m^{5}C G). Kinetic curves for the quadruplex dissociation reactions are shown for (A) d(CGC G$_3$ GCG)$_4$ in 20 mM sodium phosphate (pH 7) containing 0.1 mM EDTA and (B) d(m^{5}C G m^{5}C G$_3$ G m^{5}C G)$_4$ under the same conditions. DNA strand concentrations were ca. 7 μM. Curves correspond to the circular dichroism values monitered at 264 nm after "jumping" from 23 °C [metastable conditions, see Hardin et al. (1992)] to the following temperatures (listed from top to bottom, i.e., slowest to fastest): 80, 70, and 60 °C.

In light of the CD results discussed above, these peaks can be reasonably attributed to three m^{5}C·G base pairs in the residual WC hairpin population (Hardin et al., 1992). An additional minor peak at 15.0 ppm is likely to be due to C·C$^+$ base pairs (Radhakrishnan et al., 1991). This peak was not as visible with d(CGC G$_3$ GCG)$_4$ under the same conditions, as would be expected given the more stable nature of the methylated quadruplex.

Figure 6B shows the imino proton NMR spectrum obtained with d(TAT G$_4$ ATA)$_4$ in 40 mM K$^+$ at pH 7 and 22 °C. Three peaks and a fourth slightly resolved shoulder can be seen in the 10.7–11.8 ppm range (Figure 6B). Although these peaks were somewhat better resolved as the temperature was increased to 80 °C, they did not separate into additional discrete peaks (not shown).

Cytosine Methylation Further Stabilizes the Quadruplex: Dissociation Kinetics Measurements. The kinetics for heat-induced dissociation of the quadruplexes at pH 7 in 40 mM Na$^+$ were measured by using the large change in CD at 264 nm to moniter the time-course of the reaction. Kinetic traces for the unmethylated and modified structures in NaP (pH 7) obtained at the same three temperatures are shown in Figure 7, panels A and B, respectively. Observed kinetic constant values ($k_{i,obs}$) were determined from the data using the multicomponent exponential fitting program DISCRETE (Table I). The subscript i refers to the kinetic order of the corresponding reaction step; "obs" emphasizes that the kinetics are observed by CD at 264 nm. DISCRETE calculates the residuals curve, which corresponds to the measured data subtracted from the fit curve and then determines the exponential reaction order that best describes the results according to statistical analysis. The results indicate that the complexes formed by both C- and m^{5}C-containing oligonucleotides dissociate with apparent third-order kinetics (Table I). Exceptions occurred at higher temperatures, where processes with lifetimes of less than 5 min (Table II) were not resolved, apparently because there were only a limited number of meaningful data points in the early range of the kinetic profiles. Process 1 is the *rate-limiting step* that controls the rate of the overall dissociation process observed by CD. The reaction may require three steps, but they *do not* necessarily occur in the same temporal order as the numerical subscripts.

The logarithms of the observed first order kinetic constants (ln $k_{1,obs}$) were plotted as a function of inverse temperature to produce the Arrhenius plot shown in Figure 8. Activation energies for the first-order processes were determined from the slopes of these plots. Results obtained in this manner with the quadruplexes formed by d(m^{5}C G m^{5}C G$_3$ G m^{5}C G), d(CGC G$_3$ GCG), and d(TAT G$_4$ ATA) are listed in Table III. Statistical characteristics of the data are summarized as standard errors for the $k_{i,obs}$ values in Table I and as correlation coefficients for the Arrhenius plots shown in Figure 8. The difference in activation energies for dissociation of the methylated and unmodified C-containing quadruplex structures is 22 kcal mol^{-1}.

Quantitative "transition-state" enthalpies ($\Delta H^{\dagger\circ}$), entropies ($\Delta S^{\dagger\circ}$), and free energies ($\Delta G^{\dagger\circ}$) for these reactions were calculated for three different temperatures. These results and the corresponding $k_{1,obs}$ values are listed for comparative purposes in Table III. The molecules are listed in Table III according to decreasing T_m values. Dissociation of the methylated quadruplex has a higher activation energy and proceeds less spontaneously than with d(CGC G$_3$ GCG)$_4$. The dissociation rates and $\Delta G^{\dagger\circ}$ values demonstrate that the stabilities of the complexes decrease in the following order: d(m^{5}C G m^{5}C G$_3$ G m^{5}C G)$_4$ > d(CGC G$_3$ GCG)$_4$ > d(TAT G$_4$ ATA)$_4$. The enthalpy for dissociating the d(CGC G$_3$ GCG)$_4$ complex is 22 kcal mol^{-1} less than obtained with d(m^{5}C G m^{5}C G$_3$ G m^{5}C G)$_4$. Activation entropy contributions to $\Delta G^{\dagger\circ}$ that oppose reaching the transition state are smaller for the unmethylated complex than for the modified structure. For example, at 60 °C $T\Delta S^{\dagger\circ}$ is 29.3 kcal mol^{-1} for d(m^{5}C G m^{5}C G$_3$ G m^{5}C G)$_4$ and 8.1 kcal mol^{-1} for d(CGC G$_3$ GCG)$_4$. Thus, all of the energetic contributions favor

Table I: Kinetic Parameters for the Dissociation of Quadruplex DNAs Formed by d(CGC G_3 GCG), d(m^{5}C G m^{5}C G_3 G m^{5}C G), and d(TAT G_4 ATA) in 20 mM Sodium Phosphate at pH 7

T (°C)	components[b]	A_1[a]	k_1[a]	A_2	k_2	A_3	k_3
			(I) d(CGC G_3 GCG)$_4$				
60	3	8.39 ± 0.09[c]	7.12 ± 0.56 × 10^{-5}	3.35 ±0.21	5.35 ± 0.68 × 10^{-4}	8.57 ± 0.46	7.31 ± 0.65 × 10^{-3}
65	3	7.14 ± 0.12	8.80 ± 0.54 × 10^{-5}	5.01 ± 0.19	6.08 ± 0.42 × 10^{-4}	9.91 ± 0.42	1.38 ± 0.09 × 10^{-2}
70	2	6.65 ± 0.38	2.88 ± 0.11 × 10^{-4}	8.10 ± 0.70	1.52 ± 0.20 × 10^{-3}		
75	2	4.03 ± 0.55	8.66 ± 0.68 × 10^{-4}	10.05 ± 1.10	4.25 ± 0.65 × 10^{-3}		
80	2	0.85 ± 0.20	1.05 ± 0.22 × 10^{-3}	14.63 ± 0.15	1.09 ± 0.10 × 10^{-2}		
			(II) d(m^{5}C G m^{5}C G_3 G m^{5}C G)$_4$				
60	3	6.10 ± 1.50	2.42 ± 0.93 × 10^{-5}	2.15 ± 0.11	5.40 ± 0.57 × 10^{-4}	2.26 ± 0.22	1.10 ± 0.19 × 10^{-2}
65	3	10.10 ± 0.28	4.15 ± 0.25 × 10^{-5}	2.03 ± 0.12	8.61 ± 0.96 × 10^{-4}	4.32 ± 0.37	1.61 ± 0.24 × 10^{-2}
70	1	13.21 ± 0.27	2.20 ± 0.45 × 10^{-4}				
75	1	14.59 ± 0.69	7.18 ± 0.29 × 10^{-4}				
80	1	10.91 ± 0.47	3.35 ± 0.16 × 10^{-3}				
			(III) d(TAT G_4 ATA)$_4$				
50	2	4.42 ± 0.14	6.94 ± 0.41 × 10^{-5}	0.27 ± 0.10	6.76 ± 4.3 × 10^{-3}		
55	2	6.48 ± 0.26	1.19 ± 0.13 × 10^{-4}	0.91 ± 0.42	5.54 ± 2.1 × 10^{-4}		
60	2	4.19 ± 0.63	2.88 ± 0.27 × 10^{-4}	4.08 ± 0.61	8.50 ± 1.0 × 10^{-4}		
65	2	3.92 ± 0.63	1.00 ± 0.07 × 10^{-3}	4.78 ± 0.58	3.24 ± 0.48 × 10^{-3}		
70	1	8.17 ± 0.29	5.83 ± 0.20 × 10^{-3}				

[a] Dissociation kinetic amplitudes (A_i) and constants ($k_{d,i}$) corresponding to component i obtained as the best fit to the experimental CD data using DISCRETE (Provencher, 1976). Units for the first-order kinetic constants are s^{-1}. [b] The number of components was determined by DISCRETE to be the optimum number based on statistical analysis of the residuals (data – fit). [c] Standard errors in the fits; all error values are of the same exponential order as the corresponding data.

Table II: Comparison of Kinetic Lifetimes for the Dissociation of Quadruplex DNAs in 20 mM Sodium Phosphate at pH 7

	τ_1[a]		τ_2		τ_3	
molecule	60 °C	70 °C	60 °C	65 °C	60 °C	65 °C
d(CGC G_3 GCG)	3.8 h	56 min	31 min	27 min	2.3 min	1.2 min
d(m^{5}C G m^{5}C G_3 G m^{5}C G)	11.6 h	1.3 h	31 min	19 min	1.5 min	1.0 min
d(TAT G_4 ATA)	56 min	2.9 min	19 min	2.5 min	*b*	*b*

[a] Dissociation kinetic lifetimes (τ_i) corresponding to component i determined from k_i values as $\tau_i = 1/k_i$. [b] The d(TAT G_4 ATA)$_4$ complex dissociates in only two statistically resolved kinetic processes.

preferential dissociation of the unmethylated complex relative to the methylated analog. Entropic factors that oppose transition state formation ($T\Delta S^{\dagger\circ}$ = 21.3 kcal mol^{-1} for d(TAT G_4 ATA)$_4$ at 60 °C) produced a smaller contribution to $\Delta G^{\dagger\circ}$ than in the reaction with d(CGC G_3 GCG)$_4$ (8.1 kcal mol^{-1}).

DISCUSSION

In this work we sought to understand how flanking cytidine residues stabilize quadruplexes formed by oligonucleotides of the sequence motif d(YRY G_n RYR) at slightly acidic pH ("R" refers to purine residues and "Y" to pyrimidines). We replaced the cytidine residues in the most thoroughly characterized molecule in this series, d(CGC G_3 GCG) (Hardin et al., 1992), with either 5-methylcytidines or thymidines. In addition, we reasoned that it would be useful to retain the possibility that they might also form WC hairpin structures (Figure 1A). Thus, we characterized the effects of different ions and pH changes on the stabilities of the quadruplexes formed by d(m^{5}C G m^{5}C G_3 G m^{5}C G), d(TAT G_3 ATA) and d(TAT G_4 ATA). We demonstrated previously that the quadruplex formed by d(CGC G_3 GCG) is *parallel*-stranded (Figure 1B; Hardin et al., 1992). Since the 3′-terminal sequence, -GCG, extends the number of contiguous guanosines by one residue (*i.e.*, G_3G), the stability of the quadruplex formed by d(CGC G_3 GCG) should be better approximated by that of d(TAT G_4 ATA)$_4$ than that of d(TAT G_3 ATA)$_4$. After diluting d(TAT G_3 ATA)$_4$ to 1.8 μM, it was necessary to raise the K^+ concentration to ca. 200 mM to obtain a quadruplex with a stability comparable to that of 1.8 μM

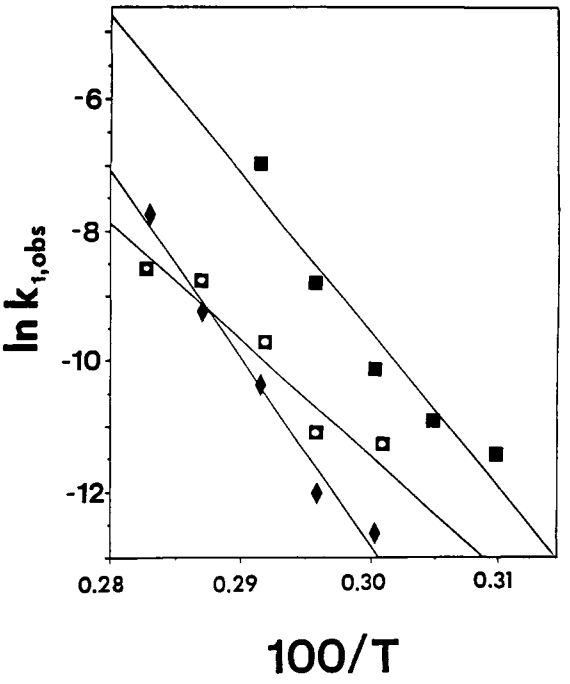

FIGURE 8: Arrhenius plots showing the effects of temperature changes on the kinetic constants for the quadruplex dissociation reactions involving either d(CGC G_3 GCG)$_4$ (⧈), d(m^{5}C G m^{5}C G_3 G m^{5}C G)$_4$ (◆), or d(TAT G_4 ATA)$_4$ (■) in 20 mM sodium phosphate (pH 7) containing 0.1 mM EDTA. Standard errors obtained in fitting the kinetic data to exponential decay curves by DISCRETE are given in Table I. Standard errors in linear regression analysis of the Arrhenius data are given in Table III.

d(CGC G_3 GCG)$_4$ in 40 mM Na^+. In contrast, d(TAT G_4 ATA)$_4$ was only moderately less stable than the complex formed by d(CGC G_3 GCG) in the presence of 40 mM Na^+ at pH 7.5. Therefore, d(TAT G_4ATA) was used for the comparisons in these studies.

Guanosine-rich DNAs that contain cytidines form distinctly different types of quadruplexes than those that do not. The CD thermal denaturation results indicate qualitatively that the terminal sequence differences (CGC–GCG; TAT–ATA) do not cause large differences in the stabilities of the quadruplexes at pH 7.5, suggesting that the common central

Table III: Transition-State Thermodynamic Parameters for the Dissociation of Quadruplex DNAs Formed by the Oligonucleotides d(CGC G_3 GCG), d(m^5C G m^5C G_3 G m^5C G), and d(TAT G_4 ATA) in 20 mM Sodium Phosphate at pH 7.0

DNA	k_1 (s^{-1})	E_A (kcal mol^{-1})	$\Delta G^{\dagger\circ}$ (kcal mol^{-1})	$\Delta H^{\dagger\circ}$ (kcal mol^{-1})	$\Delta S^{\dagger\circ}$ (cal mol^{-1} K^{-1})
		60 °C			
GCG G_3 GCG	11.7×10^{-6}	35.9	27.09	35.2	24.3
m^5C analog	3.3×10^{-6}	57.9	27.93	57.2	88.0
TAT G_4 ATA	32.0×10^{-6}	48.3	26.43	47.6	64.0
		65 °C			
CGC G_3 GCG	14.9×10^{-6}	35.9	27.35	35.2	23.1
m^5C analog	6.0×10^{-6}	57.9	27.97	57.2	86.4
TAT G_4 ATA	96.4×10^{-6}	48.3	26.06	47.6	63.6
		70 °C			
CGC G_3 GCG	55.8×10^{-6}	35.9	26.86	35.2	24.2
m^5C analog	32.0×10^{-6}	57.9	27.24	57.2	87.5
TAT G_4 ATA	429×10^{-6}	48.3	25.47	47.6	64.4

[a] Activation energies obtained as the slope of the Arrhenius plot using the $k_{1,obs}$ values listed in Table I. Linear regression correlation coefficients (R^2) were 0.948 for d(CGC G_3 GCG), 0.984 for d(m^5 C G m^5 C G_3 G m^5 C G), and 0.941 for d(TAT G_4 ATA). [b] Transition-state quasi-thermodynamic parameters, defined as characterizing the thermodynamics for the pathway from the reactants to the "activated state" (Eisenberg & Crothers, 1979).

stack of four G quartets provides much of the stability under these conditions. This loss of dramatic stabilization upon changing the the sequences that flank the central quadruplex is consistent with the proposal that C·C+ base pairs provide a significant amount of stability at or below the pK_a in parallel-stranded quadruplexes that contain GpCpG tracts (Hardin et al., 1992). By providing both *positive* and *negative* examples of stabilization relative to the C-containing quadruplex, the results obtained with d(m^5C G m^5C G_3 G m^5C G)$_4$ and d(TAT G_4 ATA)$_4$, respectively, show how the surrounding sequence motif can modulate the stability of the parallel-stranded quadruplex.

When d(TAT G_4 ATA)$_4$ and d(TAT G_3 ATA)$_4$ are heat denatured in 40 mM K^+ and 200 mM K^+, respectively, and then cooled, the CD spectra do not change significantly relative to the 80 °C spectrum in the 250–300-nm range. These results imply that, unlike d(CGC G_3 GCG), d(TAT G_n ATA) sequences do not form stable WC hairpin structures when melted to single strands and then cooled. Thus, molecules involved in the quadruplex to single-strand equilibrium are probably not depleted by a competing single-strand to WC hairpin equilibrium, as occurs with d(CGC G_3 GCG)$_4$ (Hardin et al., 1992). With d(CGC G_3 GCG), the hairpin must be denatured then renatured under quadruplex-prone conditions (at high DNA and/or K^+ concentration) to offset the competition with the WC hairpin population for the common single-stranded equilibrium partner.

Cytidines can form hemiprotonated C·C+ base pairs at moderately acidic to neutral pH in a variety of structures (Cruse et al., 1983; Edwards et al., 1990; Bratek-Wiewiorowska et al., 1990; Haner & Dervan, 1990; SantaLucia et al., 1991), including the C-rich strand of telomeric DNA sequences (Lyamichev et al., 1989; Ahmed & Henderson, 1992). Since the contribution of the stacked G-quartet domain to the overall stability of the complex is obscured by the contribution(s) of surrounding C·C+ base pairing interactions, the d(TAT G_n ATA)$_4$ complex is probably a better model for an "isolated" quadruplex system than the complex formed by d(CGC G_n GCG) or its methylated analog. Given the sequence differences, these results confirm the original postulate that the anomolous stabilization is the result of pH-dependent hemiprotonated C·C+ base pair formation (Hardin et al., 1992). Thus, slight changes in pH in the physiological range (Ling, 1984; Guo & Cole, 1989; Barasch et al., 1991) can significantly modulate the stability of both methylated and unmodified C-containing quadruplexes.

Structural Characterization. If the d(m^5C G m^5C G_3 G m^5C G)$_4$ complex contains two sets of antiparallel strands, the NMR spectrum should contain clearly resolved peaks in the 12–13 ppm range corresponding to Watson–Crick m^5C·G imino protons in the terminal (m^5C G m^5C)·(G m^5C G) duplex tracts. If the four strands in the complex are all aligned in the same direction, the spectra should contain six peaks corresponding to the six sets of electronically equivalent G imino hydrogen atoms. Since no major peaks were seen downfield of 12 ppm, the complex is *parallel-stranded.*

The minor peak(s) seen at 15 ppm probably correspond to m^5C·m^5C+ base-paired protons (Radhakrishnan et al., 1991). These results imply that only a small population of proton-containing base pairs are present at pH 7.0, about 0.2 units above the apparent pK_a measured by CD. Since on the order of 1% of the base pairs remain protonated, the intrinsic pK_a for m^5C·m^5C deprotonation in quadruplex DNA observed by NMR is about 1 or 2 units below pH 7. These data are consistent with triplex DNA studies reported by Singleton and Dervan (1992), who pointed out that the pK_a determined by ^{1}H NMR is the intrinsic pK_a of a single atom and not the value for the global (triplex/duplex, single-strand) association–dissociation equilibrium. Global pK_a values of 6.8–7.0 were obtained for triplex DNAs based upon absorbance "melt" and triplex "lifetime" measurements, while intrinsic pK_as of 5.5–5.7 were reported for cytosine protonation (Singleton & Dervan, 1992). ^{1}H NMR temperature- and pH-dependence data obtained with d(CGC G_4 GCG)$_4$ showed that the quadruplex "tightens up" in a trend that levels off at ca. pH 5.8 (H. Li, B. A. Brown II, M. Corregan, and C. C. Hardin, unpublished results).

The imino proton NMR spectrum obtained with d(TAT G_4 ATA)$_4$ in 40 mM K^+ at pH 7 (Figure 6B) contained four resolved peaks in the G·G imino proton range at 22 °C. As in the situation described above, if the strands are antiparallel one should see well-resolved resonances in the 14 ppm range due to A·T imino protons with intensities on the same order of those of G-quartet base pairs in the 10.5–12.0 ppm range. Since the downfield region is clearly devoid of such peaks, one can reasonably conclude that d(TAT G_4 ATA)$_4$ is parallel-stranded. These results are consistent with the interpretation that the complex has four G quartets, each containing four axially equivalent imino protons, as expected for a parallel-stranded structure. This number of peaks would be less likely for an antiparallel-stranded structure (Smith & Feigon, 1992). Resonances due to T·T base pairs might also be expected to occur; however, it is likely that these interactions would be relatively weak and thus exhibit reduced peak intensities due to exchange broadening (Cheong & Moore, 1992).

Insights from Kinetics Studies. The quadruplex to single-strand equilibrium is driven to quadruplex by increasing the d(CGC G_3GCG) strand concentration and/or the K^+ concentration. The kinetic barrier to quadruplex formation is crossed by heating the sample to 80 °C, after adding sufficient KCl, and then renaturing the complex by cooling to 23 °C (Hardin et al., 1992). Analogous results were also obtained with both d(m^5C G m^5C G_3 G m^5C G)$_4$ and d(TAT G_4 ATA)$_4$ (results not shown). In each case the equilibrium was found to be mechanistically reversible under conditions similar to

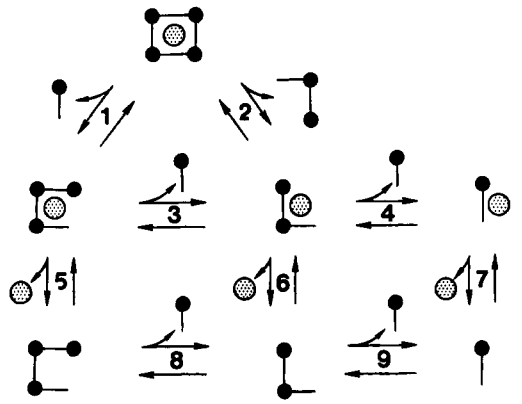

FIGURE 9: Experimentally consistent quadruplex dissociation pathways. Dissociated strands are represented by solid circles, unpaired bases and base pairs are indicated by the connected lines, and stippled circles represent cations. Individual putative kinetic steps are numbered only to facilitate discussion.

those described for the d(CGC G_3 GCG)$_4$ system. Since the equilibrium was not practically reversible as a function of temperature in 20 mM NaP (pH 7) at a DNA strand concentration of ca. 7 μM, true thermodynamic energies were not measured. Therefore, we determined the *relative stabilities* of the quadruplexes by measuring *dissociation kinetic constants* under the same set of conditions (ca. 40 mM Na^+, pH 7, ca. 7 μM DNA strands). In the context of the quadruplex dissociation reaction, "transition state" parameters describe the energy requirements for transformation of a solvated quadruplex to the transient highly strained "activated" species (Piszkiewicz, 1977; Eisenberg & Crothers, 1979). In the case of the first-order quadruplex dissociation reaction, this "activated state" is poised at the top of the rate-limiting energy barrier that must be crossed to allow spontaneous formation of separated solvated strands and liberated "guest" ions. All of the energetic factors that control dissociation favor the unmethylated complex relative to the methylated structure. The difference in activation energies for dissociation of the methylated and unmodified C-containing quadruplex structures (22 kcal mol^{-1}) demonstrates that substantially more energy is required to dissociate the methylated quadruplex.

Large positive $\Delta S^{\dagger\circ}$ values (Table III) infer that the reactions are multimolecular and that the molecules must become properly oriented to pass from the reactant configuration to the transition state (Piszkiewicz, 1977). This is consistent with nearly simultaneous dissociation of a number of electrostatic interactions (e.g., ion–carbonyl oxygen attractive forces, hydrogen bonds), as occurs when the base pairs and GO6–ion contacts in the complex are severed to yield the products of steps 1 or 2 in Figure 9.

Kinetic experiments were also done in the pH 6–9 range to quantitatively determine the effects of pH on the energetics of the quadruplex dissociation reaction (results not shown). When the quadruplex formed by d(CGC G_3 GCG) dissociates in NaP, the activation energies for the rate-limiting kinetic process decreased with pH as follows: 36.2 kcal mol^{-1} (pH 6), 35.9 kcal mol^{-1} (pH 7), 18.4 kcal mol^{-1} (pH 8), and 17.0 kcal mol^{-1} (pH 9). These results indicate that as much as *half* of the stability of the complex can be attributed to the C·C$^+$ base-pairing interactions at neutral pH!

Two different mechanisms can be proposed to explain how cytidine methylation stabilizes the quadruplex. The first would involve an increase in the pK_a when the base is methylated, which would stabilize base pairing. When cytosine pyranoside is methylated at C5, the pK_a increases by 0.25 pH units (Fox & Shugar, 1952). The pK_a increases by 0.8 units when uridine is methylated to form thymidine; the pK_a values of the corresponding pyranosides differ by 0.5 units. If this trend is quantitatively similar with the cytidines (Sowers et al., 1987a), one would expect that pK_a for 5-methylcytidine would be 0.4 units higher than that of cytidine. In the quadruplex this would theoretically raise the pK_a to 7.2, suggesting that it could shift a large part of the quadruplex population to the much more stable C·C$^+$ base-paired form at pH 7. The second potential explanation for methylation-induced stabilization is based on the fact that methylation of uracil to form thymidine increases the polarizability of the base (Miller & Savchik, 1979), which has been cited as likely to make an important contribution to base–base stacking energies in nucleic acids via dipole/induced dipole (dispersion) forces (Sowers et al., 1987b).

The results presented in this paper provide two types of evidence that favor the second mechanism (dipole/induced dipole forces enhance stacking energies). First, the apparent pK_a for quadruplex dissociation does not shift noticably when the pH profiles for the methylated and unmodified complexes are compared [Figure 5C versus Figure 6C in Hardin et al. (1992)]. Although a slight shift (<0.2 units) may be present, it is not likely to be enough to account for the large difference in stabilities. Secondly, there is a large difference in the relative contributions of enthalpic and entropic factors to the stabilities of the complexes. The ratio of entropic contributions to enthalpic contributions at 60 °C increases from 23% in the case of d(CGC G_3 GCG)$_4$ to 51% with d(m^{5}C G m^{5}C G_3 G m^{5}C G)$_4$ (Table III). Since stacking interactions are characterized by entropic thermodynamic contributions (Saenger, 1984) and dipole–dipole interactions make a dominant contribution to stacking energies, we conclude that cytidine methylation stabilizes the quadruplex primarily by increasing favorable stacking forces.

Comparing d(CGC G_3 GCG)$_4$ and d(TAT G_4 ATA)$_4$, there is an interesting sequence-dependent difference in the balance between enthalpic and entropic influences on quadruplex dissociation. Entropic factors that opposed transition state formation were less obstructive in the reaction with d(CGC G_3 GCG)$_4$ [$T\Delta S^{\dagger\circ}$ at 60 °C of 8.1 kcal mol^{-1} vs 21.3 kcal mol^{-1} for d(TAT G_4 ATA)$_4$]. This produced an unusual situation in which the activation energy (E_A) required to drive the first-order reaction with d(TAT G_4 ATA)$_4$ was higher than for d(CGC G_3 GCG)$_4$, yet the reaction occurred *more* spontaneously (lower $\Delta G^{\dagger\circ}$, faster rate) (Tables II and III)! Therefore, relative to the enthalpic energy that must be expended to reach the activated state in the reaction with d(TAT G_4 ATA)$_4$, more entropic "ordering" energy is necessary than with d(CGC G_3 GCG)$_4$. This is an example of "entropy–enthalpy compensation" and illustrates the complex relationship that exists between E_A, $\Delta G^{\dagger\circ}$, $\Delta H^{\dagger\circ}$, and $\Delta S^{\dagger\circ}$. The excess entropic energy may be required because d(TAT G_4 ATA)$_4$ has an additional favorable purine–purine stacking interactions (G_3G-A) which are not present within the analogous G_3G-C linkage in the latter molecule. The different number of 3′-terminal residues (-ATA vs -CG) may also contribute to the difference.

Figure 9 shows a reasonable set of strand-separation pathways that could lead to quadruplex dissociation. The reaction scheme was simplified to only include steps for dissociation of either a lone strand, a G·G base-paired duplex unit, or all of the coordinated ions *en mass*. In this mechanism, the "transition state" is a transient energized complex that must form prior to loss of the electrostatic and hydrogen-

bonding interactions and subsequent solvation of the liberated strands. These donor and acceptor groups are shown as ball and stick sites viewed down the center of the quadruplex in Figure 9. The model assumes that the observed CD changes occur when a strand (step 1) or a base-paired duplex unit (step 2) gains enough momentum to "zip loose" from the complex in an apparent first-order dissociation process. In descriptive terms, the dissociating strand or duplex would leave behind a rapidly solvating ion-binding pocket and unpaired hydrogen bond donor and acceptor groups on the remaining triplex or duplex structures, respectively. Results from electrospray ionization mass spectrometry studies indicate that the quadruplex structure d(CGC G_4 GCG)$_4$ is stable enough to survive the electrospray process and suggest that a residual duplex population can be observed at mass/charge = m_{dimer} − 11 (i.e., the 11^- anion) under more destabilizing conditions (Goodlett et al., 1993).

When a metastable complex in 0.1 mM Li_2EDTA (pH 7) is diluted into a solution that contains a more stabilizing cation, the quadruplex denatures with a stability characteristic of the newly introduced cation. This indicates that dynamic fluctuations in the structure of the intact complex can allow ions to leach in and out of the complex by either passing between the strands or via the fraying ends of nonproductively denaturing complexes. A quadruplex could theoretically exist without internally bound cations; however, given the probable thermodynamic instability of such structures, it is likely that the departing ion would be quickly replaced by a new cation. Otherwise, it is likely that the quadruplex would dissociate due to repulsive forces between the unneutralized cluster of electronegitive guanine carbonyl oxygens that line the cation binding site.

The weak peak at 15.0 ppm (Figure 6A) can reasonably be assigned to the trapped imino protons in the remaining population of m^5C·m^5C^+ base pairs. The 15.0 ppm peak was not detected in the NMR spectrum of d(CGC G_3 GCG)$_4$ under the same conditions (20 mM KP at pH 7) (Hardin et al., 1992). This indicates that C·C+ base pairs are less stable than m^5C·m^5C^+ base pairs, in agreement with CD kinetics results (Table III). Subintegral m^5C·m^5C^+ imino NMR peaks indicate that relatively stable structures can form even when the terminal C·C+, m^5C·m^5C^+ base pairs and G quartet base-pairing interactions "fray" apart prior to detachment of an entire strand or duplex unit from the quadruplex (Figure 6; Hardin et al., 1992; unpublished results). Since base-stacking interactions would be perturbed by these reactions, a separate distinguishable kinetic step might be detected by CD. Dissociation of the quadruplexes formed by both d(CGC G_3 GCG) and d(m^5C G m^5C G_3 G m^5C G) could be resolved into three distinct processes, yet dissociation of d(TAT G_4ATA)$_4$ was always best characterized by two steps. These results could be explained by a pathway in which the overall dissociation rate is measurably affected by "fraying" of the terminal G quartet(s) and C·C+ base-pairing interactions. Since C·C+ base pairing is pH-dependent, the ability of this population of quadruplexes to directly affect kinetic processes should decrease strongly as the pH is raised above the pK_a (ca. 6.8). CD kinetics studies showed that dissociation of d(CGC G_3 GCG)$_4$ is more likely to have second-order kinetics than third-order kinetics at or above pH 8.0 (unpublished results). Given the sequence differences, these results indicate that C·C+ and m^5C·m^5C^+ dissociation subreactions are probably "activated" during the fast step (at pH 7), which cannot occur with d(TAT G_4 ATA)$_4$ (Tables II and III).

Potential Biological Implications. Between 60% and 90% of the CpG dinucleotides in vertebrate genomes are methylated at C5 on the cytosine ring (Bird, 1986; Bernardi, 1989; Tazi & Bird, 1990). The degree of methylation can directly affect both chromatin condensation and transcriptional repression (Bernardi, 1989). Because RNA polymerases cannot transcribe methylated DNA sequences as effectively, cytosine methylation has long been thought to partly determine the transcriptional viability of a given gene. The results in this paper suggest that DNAs in semicondensed chromatin may interact under the control of intracellular ion and/or pH gradients to form reversible quadruplex-mediated "cross-links". A plausible mechanism for how cytosine methylation stabilizes these putative domains is also suggested. This mechanism might switch distinct structural units between different functional states and thereby regulate changes in higher order chromatin structure. Physiological metastasis may be under the control of cell cycle- or developmentally-programmed methylase or signal-transduction pathways (Hardin et al., 1992). In theory, proper metastasis would support proper gene expression, incipient chromatin condensation, recombination, and probably other genetic events. Ion ports would presumably be key controlling agents (DeFelice et al., 1993; Mazzanti et al., 1993).

Intracellular pH varies depending upon cell type and the stage in the cell cycle. For example, the pH in *Physarum* cells shifts from 5.9 during interphase to 6.8 during mitosis (Gerson, 1978). It was suggested that DNA synthesis might be stimulated by decondensation of chromatin as a consequence of the increased pH (Guo & Cole, 1989). In contrast, cultured cancer or virus-transformed cells have a less pronounced pH optimum than normal cells and usually have a more acidic intracellular pH (Gerson, 1978). For example, normal rat liver cells have a pH of 7.4, while rat and mouse hepatomas have pH values of 7.0 and 6.7, respectively. The acidosis in cancerous cells may produce the physiological changes in cancerous cells (Racker, 1972). It is interesting that these physiologically important effects titrate over the same pH range as the C·C+-mediated quadruplex stabilization/destabilization phenomenon. In theory, the lower pH in cancerous cells could cause "improper" pairing or condensation interactions within or between chromatin domains and thus lead to faulty chromosomal interaction(s), resulting in pathological function(s). In addition to other manifestations, transformed cells might suffer from improper cell cycle control because the more acidic pH alters the functional state of the chromatin structure.

It has been proposed that ion-induced quadruplex formation could be involved in telomere–telomere interactions, chromosomal pairing, recombination, and possibly transcriptional repression (Sen & Gilbert, 1988; Boeke, 1991; Hardin et al., 1991, 1992 and references therein; Murchie & Lilley, 1992). Recent studies have shown that physiological malfunctions are correlated with altered G-rich telomeric DNA sequences and impaired telomere maintainence (Yu et al., 1990; Biessmann & Mason, 1992). Faulty telomeric maintainence can increase the susceptibility of chromosomes to mutations by "healing", and thus sealing into place, a mutant sequence. These mutations could conceivably be acquired via "faulty" recombination, transposition, viral action, etc. In essence, the G-richness might produce an unusually "sticky" multi-strand-prone DNA, with the ability to form C·G, G·G, and C+·C+ (m^5C^+·m^5C^+) base pairs, intrinsically aided by cations (including H^+). These entities may then lead to successful illegitimate recombination events. This mechanism might

thus cause stable inheritance of genetic diseases associated with chromosomal truncation (Wilkie et al., 1990), translocation events and subsequent protooncogene activation (Bonetta et al., 1990; Smith et al., 1989; Murchie & Lilley, 1992), or spontaneous breakage at "fragile" sites (Heitz et al., 1991; Rennie, 1992). These pathways could conceivably lead to problematically "imprinted" changes in gene expression, which might in some cases produce the associated physiological defects.

CONCLUSIONS

CD and NMR studies demonstrate that cytosine methylation stabilizes a m^5C-containing quadruplex relative to the unmodified parent complex at pH 7 in physiological-like salt solutions. Both types of structures are highly stabilized at or below similar apparent pK_a values of 6.8. Proton NMR spectra obtained with the m^5C-containing complex at pH 7 show a small downfield peak with a fairly unique chemical shift (15 ppm), which has previously been assigned to imino protons in $C{\cdot}C^+$ base pairs in WC duplex and triplex structures. This peak is less apparent in the spectrum of the unmethylated complex, indicating that $C{\cdot}C^+$ base pairs are more stable in the modified structure. Replacing the flanking G+C, or m^5C+G, tracts with A+T sequences produces a loss of the large stabilization that occurs with C- or m^5C-containing quadruplexes at or below pH 6.8. By providing examples of both negative and positive modulation of the $C{\cdot}C^+$-dependent stabilization phenomenon, these studies demonstrate that flanking sequences can affect quadruplex *stabilities* and *dissociation timescales* . This suggests that *if* quadruplex interactions are involved in genetic processes *in vivo*, they might exchange with other types of structures on a variety of time scales which are strongly dependent upon the immediate solution environment.

ACKNOWLEDGMENT

We thank Ling Xia for early experimental help, Patricia Minchew for writing the CD data conversion program and Dr. David Zarling for encouragement and supplying m^5C phosphoramidites.

REFERENCES

Ahmed, S., & Henderson, E. H. (1992) *Nucleic Acids Res. 20*, 507–511.
Barasch, J., Kiss, B., Prince, A., Saiman, L., Gruenert, D., & Al-Awqati, Q. (1991) *Nature 352*, 70–73.
Bernardi, G. (1989) *Annu. Rev. Genet. 23*, 637–661.
Biessman, H., & Mason, J. M. (1992) *Adv. Genet.* (in press).
Bird, A. P. (1986) *Nature 321*, 209–213.
Boeke, J. D. (1990) *Cell 61*, 193–195.
Bonetta, L., Kuehn, S. E., Huang, A., Law, D. J., Kalikin, L. M., Koi, M., Reeve, A. E., Brownstein, B. H., Yeger, H., Williams, B. R. G., & Feinberg, A. P. (1990) *Science 250*, 994–997.
Bratek-Wieworowska, M. D., Popenda, M., Malinowska, N., & Wieworowska, M. (1990) *J. Mol. Struct. 237*, 123–137.
Cheong, C., & Moore, P. B. (1992) *Biochemistry 31*, 8406–8414.
Cruse, W. B. T., Egert, E., Kennard, O., Sala, G. B., Salisbury, S. A., & Viswamitra, M. A. (1983) *Biochemistry 22*, 1833–1839.
DeFelice, L. J., Dale, B., Kyozuka, K., Santella, L., & Tosti, E. (1993) *Biophys. J. 63*, A330.
Edwards, E. L., Patrick, M. H., Ratliff, R. L., & Gray, D. M. (1990) *Biochemistry 29*, 828–836.
Eisenberg, D., & Crothers, D. M. (1979) *Physical Chemistry with Applications to the Life Sciences*, Benjamin-Cummings, Menlo Park, CA.
Fox, J. J., & Shugar, D. (1952) *Biochim. Biophys. Acta 9*, 369–384.
Gerson, D. F. (1978) in *Cell Cycle Regulation* (Jeter, J. R., Jr., Cameron, R., Padilla, G. M., & Zimmerman, A. M., Eds.) pp 105–131, Academic Press, Orlando, FL.
Goodlett, D. R., Camp, D. G., II, Hardin, C. C., Corregan, M., & Smith, R. D. (1993) *Biol. Mass. Spectrom. 22*, 181–183.
Guo, X., & Cole, R. D. (1989) *J. Biol. Chem. 264*, 11653–11657.
Guschlbauer, W., Chantot, J.-F., & Thiele, D. (1990) *J. Biomol. Struct. Dyn. 8*, 491–511.
Haner, R., & Dervan, P. B. (1990) *Biochemistry 29*, 9761–9765.
Hardin, C. C., Henderson, E., Watson, T., & Prosser, J. K. (1991) *Biochemistry 30*, 4460–4472.
Hardin, C. C., Watson, T., Corregan, M., & Bailey, C. (1992) *Biochemistry 31*, 833–841.
Heitz, D., Rousseau, F., Devys, D., Saccone, S., Abderrahim, H., Le Paslier, D., Cohen, D., Vincent, A., Toniolo, D., Della Valle, G., Johnson, S., Schlessinger, D., Oberle, I., & Mandel, J. L. (1991) *Science 251*, 1236–1239.
Henderson, E., Hardin, C. C., Wolk, S. K., Tinoco, I., Jr., & Blackburn, E. H. (1987) *Cell 51*, 899–908.
Jin, R., Breslauer, K. J., Jones, R. A., & Gaffney, B. L. (1990) *Science 250*, 543–546.
Ling, G. N. (1984) in *Search of the Physical Basis of Life*, Plenum Press, New York.
Lyamichev, V. I., Mirkin, S. M., Danilevskaya, O. N., Voloshin, O. N., Balatskaya, S. V., Dobrynin, V. N., Filippov, S. A., & Frank-Kamenetskii, M. D. (1989) *Nature 339*, 634–637.
Mazzanti, M., Innocenti, B., & Rigatelli, M. (1993) *Biophys. J. 63*, A236.
Miller, K. J., & Savchik, J. A. (1979) *J. Am. Chem. Soc. 101*, 7206–7213.
Murchie, A. I. H., & Lilley, D. M. J. (1992) *Nucleic Acids Res. 20*, 49–53.
Piszkiewicz, D. (1977) *Kinetics of Chemical, and Enzyme-Catalyzed Reactions*, Oxford University Press.
Provencher, S. W. (1975) *Biophys. J. 16*, 27–41.
Provencher, S. W., & Vogel, R. H. (1980) *Math. Biosci. 50*, 251–262.
Radhakrishnan, I., Gao, X., de los Santos, C., Live, D., & Patel, D. (1991) *Biochemistry 30*, 9022–9030.
Raguhuraman, M. K., & Cech, T. R. (1990) *Nucleic Acids Res. 18*, 4543–4552.
Racker, E. (1972) *Am. Sci. 60*, 56.
Rennie, J. (1992) *Sci. Am.*, May issue, p 92.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
SantaLucia, J., Kierzek, R., & Turner, D. H. (1991) *Biochemistry 30*, 8242–8251.
Sen, D., & Gilbert, W. (1988) *Nature 334*, 364–366.
Sen, D., & Gilbert, W. (1990) *Nature 344*, 410–414.
Singleton, S. F., & Dervan, P. B. (1992) *Biochemistry 31*, 10995–11003.
Smith, F. S., & Feigon, J. (1992) *Nature 356*, 164–168.
Smith, S. S., Baker, D. J., & Jardines, L. A. (1989) *Biochem. Biophys. Res. Commun. 160*, 1397–1402.
Sowers, L. C., Shaw, B. R., Veigl, M. L., & Sedwick, M. D. (1987a) *Mutat. Res. 177*, 201–218.
Sowers, L. C., Shaw, B. R., & Sedwick, W. D. (1987b) *Biochem. Biophys. Res. Commun. 148*, 790–794.
Sundquist , W. I., & Klug, A. (1989) *Nature 342*, 825–829.
Tazi, J., & Bird, A. (1990) *Cell 60*, 909–920.
Ts'o, P. O. P. (1974) in *Basic Principles of Nucleic Acid Chemistry*, Vol I, Academic Press, New York.
Wilkie, A. O. M., Lamb, J., Harris, P. C., Finney, R. D., & Higgs, D. R. (1990) *Nature 346*, 868–871.
Williamson, J. R., Raghuraman, M. K., & Cech, T. R. (1989) *Cell 59*, 871–880.
Yu, G.-L., Bradley, J. D., Attardi, L. D., & Blackburn, E. H. (1990) *Nature 344*, 126–132.
Zahler, A. M., Williamson, J. R., Cech, T. R., & Prescott, D. M. (1991) *Nature 350*, 718–720.