At Physiological pH, d(CCG)$_{15}$ Forms a Hairpin Containing Protonated Cytosines and a Distorted Helix†

Adong Yu,‡,§ Maria D. Barron,‖ Rebecca M. Romero,⊥ Mellisa Christy,‡ Barry Gold,▽ Jianli Dai,‡ Donald M. Gray,‖ Ian S. Haworth,⊥ and Michael Mitas*,‡

Department of Biochemistry and Molecular Biology, Oklahoma State University, 246 Noble Research Center, Stillwater, Oklahoma 74078-3035, Program in Molecular and Cell Biology, The University of Texas at Dallas, P.O. Box 830688, Richardson, Texas 75083-0688, Department of Pharmaceutical Sciences, University of Southern California, 1985 Zonal Avenue, Los Angeles, California 90033, and Eppley Institute for Research in Cancer, Department of Pharmaceutical Sciences, University of Nebraska Medical Center, Omaha, Nebraska 68198-6805

Received October 10, 1996; Revised Manuscript Received December 23, 1996⊗

ABSTRACT: To investigate potential structures of d(CGG/CCG)$_n$ that might relate to their biological function and association with triplet repeat expansion diseases (TREDs), the structure of a single-stranded (ss) oligonucleotide containing d(CCG)$_{15}$ [ss(CCG)$_{15}$] was examined by studies of the pH and temperature dependence of electrophoretic mobility, UV absorbance, circular dichroism, chemical modification, and P1 nuclease digestion. ss(CCG)$_{15}$ had an unusually high pK_a (7.7 ± 0.2). At pH 8.5, ss(CCG)$_{15}$ formed a relatively unstable (T_m = 30 °C in 1 mM Na^+) hairpin containing CpG base-pair steps. At pH 7.5, the hairpin contained protonated cytosines but no detectable C·$^+$C base pairs, increased thermal stability (T_m = 37 °C), increased stacking of the CpG base-pair steps, and a single cytosine that was flipped away from the central portion of the helix. Examination of ss(CCG)$_{18}$ and ss(CCG)$_{20}$, which were designed to adopt hairpins containing alternative GpC base-pair steps, revealed hairpins containing CpG base-pair steps, pK_as of ∼8.2 and ∼8.4, respectively, and distorted helices. The results suggest that DNA sequences containing (CCG)$_{n\geq 15}$ adopt hairpin conformations that contain CpG rather than GpC base-pair steps; the mismatched cytosines are protonated at physiological pH but are not H-bonded. We propose that protonation arises from the stacking of two cytosines in the minor groove of a distorted helix.

To aid in the correlation of potential structures of triplet repeat nucleic acids with their function and their propensity to undergo expansion events, a sequence-based classification system of triplet repeats was described (Mitas et al., 1995b). Class I repeats, which were defined by the presence of a GC or CG palindrome, exhibited the lowest rates of slippage synthesis (Schlötterer & Tautz, 1992), had the lowest base-stacking energies, and were associated with nine of the 10 known triplet repeat expansion diseases (TREDs)[1] (Mitas et al., 1995b). Expansions of triplet repeats continue in offspring of affected individuals, resulting in progressive severity of the disease and/or an earlier age of onset, phenomena clinically referred to as "anticipation". Members of the TRED family include fragile X syndrome and Huntington's disease [for a review of TREDs, see Ashley and Warren (1995)].

The six complementary single strands of Class I triplet repeats can potentially form hairpin structures at the lagging strand of the replication fork (Mitas et al., 1995b; Gacy et al., 1995), where expansion events are probably initiated. In support of this possibility, various biophysical and biochemical studies have shown that sequences containing d(CTG)$_n$, d(CGG)$_n$, d(CAG)$_n$, d(GTC)$_n$, and d(GAC)$_n$ form stable hairpins (Mitas et al., 1995a,b; Chen et al., 1995; Gacy et al., 1995; Yu et al., 1995a,b; Smith et al., 1995; Mariappan et al., 1996a,b; Petruska et al., 1996).

Unlike d(CTG)$_n$ or d(CAG)$_n$, which form hairpins containing only GpC base-pair steps, d(CCG)$_n$ sequences can potentially adopt hairpin alignments that contain either CpG or GpC base-pair steps. When the terminal 5′ nucleotide in a d(GCC)$_{5-7}$ sequence is a guanine, a hairpin is formed that contains GpC base-pair steps (Chen et al., 1995; Mariappan et al., 1996a). Consistent with nomenclature describing trinucleotide repeat-containing hairpins (Mitas et al., 1995a), we refer to a hairpin (or duplex structure) containing dCCG repeats and GpC base-pair steps as an **(a)** alignment. In contrast, the duplex structure containing dCCG repeats and CpG base-pair steps is referred to as a **(b)** alignment.

In an **(a)** alignment hairpin containing CCG repeats, the methylatable cytosine of the CpG dinucleotide is mispaired with another cytosine. Studies with the human MTase (Smith et al., 1987; Baker et al., 1991; Smith et al., 1991) and two bacterial cytosine MTases (Klimasauskas & Roberts,

† Studies performed at OSU were supported by the Oklahoma Agricultural Experiment Station at Oklahoma State University. Circular dichroism studies at the University of Texas at Dallas were supported by NSF Grant MCB-9405683, Grant AT-503 from the Robert A. Welch Foundation, and Grant 9741-036 awarded by the Texas Advanced Technology Program to D.M.G.

* Corresponding author: phone (405) 744-6199, fax (405) 744-7799, e-mail mmitas@biochem.okstate.edu.

‡ Oklahoma State University.

§ Present address: Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Ave., P. O. Box 208114, New Haven, CT 06520-8114.

‖ University of Texas at Dallas.

⊥ University of Southern California.

▽ University of Nebraska Medical Center.

⊗ Abstract published in *Advance ACS Abstracts,* March 1, 1997.

[1] Abbreviations: TREDs, triplet repeat expansion diseases; ss, single-stranded; ds, double-stranded; CD, circular dichroism; HA, hydroxylamine; THF-OOH, 2-hydroperoxytetrahydrofuran; DMS, dimethyl sulfate; EMMP, electrophoretic mobility melting profile; T_m, melting temperature; M_{rel}, relative electrophoretic mobility; MTase, methyltransferase.

S0006-2960(96)02541-X CCC: $14.00 © 1997 American Chemical Society

1995; Yang et al., 1995) have shown that when a methylatable cytosine is paired with a base other than guanine, high rates of methylation are observed. Presumably, this is due to the ease with which the methylatable cytosine flips away from the helix, a requirement for methylation at the 5 position [for a review of other DNA reactions catalyzed by base flipping, see Roberts (1995)].

^{1}H NMR studies do not reveal any H-bonds within the C-C mismatches of $(GCC)_{5-7}$ hairpins (Chen et al., 1995; Mariappan et al., 1996a), suggesting that the methylatable cytosine in these hairpins might also be free to flip out of the helix and undergo rapid methylation. In support of this possibility, *in vitro* studies have shown that human MTase rapidly methylates $d(GCC)_{5-7}$ hairpin structures (Chen et al., 1995). It has been proposed that hypermethylation of the triplet repeat region in the *FMR-1* gene is a direct result of formation of hairpins in the **(a)** alignment followed by the action of human MTase (Chen et al., 1995; Laayoun & Smith, 1995).

A completely unexpected and new DNA structure that has recently emerged from studies of triplet repeat sequences is the e-motif formed from $d(CCG)_2$ (Gao et al., 1995). This sequence contains a cytosine at the 5′ terminus, while that characterized by Gupta and colleagues contains a guanine. Gao and colleagues discovered that the duplex structure adopted by $d[(CCG)_2]_2$ was one that contained two CpG base-pair steps rather than a single GpC base-pair step. Surprisingly, the cytosines within the lone C-C mismatch, which was centrally located in the **(b)** alignment duplex, were not stacked within the helix. Instead, the cytosines were extrahelical, pointed away from one another, and symmetrically located in the minor groove.

To understand mechanisms of sequence amplification, gene hypermethylation, and folate-induced chromosomal fragile sites, it is important to determine structures of oligonucleotides containing relatively large numbers of CCG repeats. In this report, we investigated the structural properties of oligonucleotides containing 15, 18, or 20 CCG repeats, the lengths of which are not amenable to high-resolution NMR spectroscopy. We find that all sequences exclusively form hairpins in a **(b)** alignment. However, the hairpins exhibit unusual features: the mismatched cytosines are protonated at a relatively high pH, and the sugar−phosphate backbones are highly distorted. The results described in this report are entirely consistent with a new DNA structure containing multiple stack pairs of extrahelical cytosines.

MATERIALS AND METHODS

Oligonucleotides. With the exception of the oligonucleotide used for circular dichroism studies, all oligonucleotides were synthesized by the OSU Recombinant DNA/Protein Resource Facility on an Applied Biosystems 381A oligonucleotide synthesizer (Foster City, CA) with the trityl group on and purified with oligonucleotide purification cartridges (Cruachem, Glasgow, U.K.). Sequences of oligonucleotides were as follows: $(CCG)_{15}$, GATCC$(CCG)_{15}$GGTACCA; $(GTC)_{15}$, GATCC$(GTC)_{15}$GGTACCA; $(CTG)_{15}$, GATCC-$(CTG)_{15}$GGTACCA; $(CGG)_{20}$, AGCTTGTTAACTG-$(CGG)_{20}$CAG; $(CCG)_{20}$, GATCCTG$(CCG)_{20}$CAGTTAACA; $(GAT)_{15}$, AGCTTCTA$(GAT)_{15}$G.

Plasmids. Plasmids pGTC15 (Yu et al., 1995b), ΔT_{II}GTC15 (Yu et al., 1995b), pCTG15 (Mitas et al., 1995b), and pCCG15 (Mitas et al., 1995a) were previously described. An oligonucleotide containing $d(CGG)_{20}$ (complementary sequences listed above) was cloned into plasmid as previously described (Mitas et al., 1995b) except that vector DNA was Bluescript pSK$^+$ (Stratagene, La Jolla, CA). Plasmid DNAs were isolated from blue bacterial colonies grown on 5-bromo-4-chloro-3-indolyl β-D-galactoside (X-gal)/isopropyl thio-β-D-galactoside (IPTG) plates (Sambrook et al., 1989) and screened for the presence of the oligonucleotide by restriction enzyme digestion with *Hpa*I, which cleaves a sequence immediately adjacent to the triplet repeat region. Two plasmids harboring a new *Hpa*I restriction enzyme site were amplified, purified by cesium chloride gradient centrifugation, and linearized with *Dra*I. The sequences of the triplet repeat regions within the plasmids were determined by the dideoxy chain-termination method with the use of the M13 universal primer. One plasmid contained 20 perfect CCG repeats and was named pCCG20. The other plasmid contained only 18 perfect CCG repeats and was named pCCG18.

pH- and Temperature-Dependent Electrophoretic Analyses: (A) Methylation of pCCG15 and pCCG20. Fifteen-microgram portions of the respective plasmids were incubated at 37 °C for 4 h with *Sss*I CpG methylase (New England Biolabs, Beverly MA) according to manufacturer's instructions. Additional amounts of *S*-adenosylmethionine and *Sss*I CpG methylase were added and incubation was continued overnight. Extent of methylation was determined with the use of the methyl-sensitive restriction enzyme *Hpa*I, which failed to cleave plasmids incubated with *Sss*I CpG methylase (data not shown).

(B) Plasmid Labeling. Labeling of $ss(CCG)_{15}$, $ss(C^mCG)_{15}$, $ss(CCG)_{20}$, $ss(C^mCG)_{20}$, $ss(CTG)_{15}$, and $\Delta T_{II}(GTC)_{15}$ was performed with the use of [α-^{32}P]dCTP, [α-^{32}P]dATP (ICN, Irvine, CA), and Klenow enzyme (New England Biolabs, Beverly, MA) as previously described (Yu et al., 1995b) except that the starting plasmids were pCCG15, methylated pCCG15, pCCG20, methylated pCCG20, pCTG15, and ΔT_{II}CTG15, respectively. Labeling of $ss(GAT)_{15}$ was performed as previously described (Mitas et al., 1995b). Labeling of the pyrimidine-rich triplet repeat sequences resulted in the addition of four nucleotides to the 3′-end (5′AGCT3′).

(C) Electrophoretic Analysis. Electrophoretic mobility melting profiles (EMMPs) were described previously (Yu et al., 1995b). Briefly, to obtain a homogeneous population of labeled double-stranded (ds) DNAs, oligonucleotides liberated from plasmid at 0.7 nM were incubated with 0.5 nM *Escherichia coli* single-stranded DNA binding protein (Lohman & Ferrari, 1994) (a generous gift of Dr. Timothy Lohman, Washington University School of Medicine) in 8% glycerol, 0.2 M NaCl, 18 mM HEPES (pH 7.5−8.5), 1 mM EDTA, and 1 mM DTT at 37 °C for 20 min. To obtain a homogeneous population of labeled ssDNAs, 1.4 pmol of unlabeled synthetic oligonucleotide of the same sequence as the labeled strand (∼0.7 fmol) was added, placed in a 90 °C H_2O bath for 5 min, and placed at 25 °C for 5 min. Labeled DNAs (4×10^4 dpm) were diluted to 10 μL in buffer containing 8% glycerol, 10 mM HEPES, pH 7.5−8.5, and 1 mM EDTA. One microliter of loading dye (50% glycerol and 0.4% bromophenol blue) was added to the DNA samples prior to gel electrophoresis. Electrophoresis was performed in a Hoeffer (San Francisco, CA) SE600 series unit at various

temperatures at 25 mA/gel in 45 mM Tris−borate (pH 8.5) and 1 mM EDTA (TBE buffer). Gel plates were 14 cm (length) × 16 cm (width) × 1.5 mm (thickness). Electrophoresis was stopped when the bromophenol blue marker migrated ∼10 cm. Dried gels were placed between two intensifying screens (Dupont) and exposed to Fuji RX film for 2−5 h. Temperature of the polyacrylamide gels was controlled as previously described (Yu et al., 1995b).

For pH-dependent electrophoretic studies, the same procedure described above was performed at 28 °C at various pH values.

Circular Dichroism Studies. A DNA oligonucleotide of sequence d[GATCC(CCG)$_{15}$GGTACCAAGCT], or ss(CCG)$_{15}$, was purchased from Oligos, Etc. (Wilsonville, OR). ss(CCG)$_{15}$ was dissolved in 2 mM NaCl. An extinction coefficient of 8665 M^{-1} cm^{-1} was calculated from the extinction coefficients of monomers and dimers (Gray et al., 1995) and was used to determine the concentration of stock ss(CCG)$_{15}$ in 2 mM NaCl, at 90 °C. Samples at pH greater than 7 were diluted to be in 50 mM NaCl, 5 mM Tris-HCl, with the pH being adjusted with HCl. Samples at pH 7.0 and below were diluted to be in 50 mM NaCl and 5 mM Na^+ (phosphate), with the pH being adjusted with HCl. All samples were at about 6×10^{-5} M in nucleotide concentration. A separate aliquot of the stock ss(CCG)$_{15}$ was used to make a sample at each individual pH value of 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, and 8.5 ± 0.1. One series of CD spectra as a function of pH was obtained for samples that were heated to 80 °C for 3 min and quick-cooled in ice water to dissociate any self-complex of ss(CCG)$_{15}$ that may have existed in the stock solution. There were no major differences in the spectra obtained for this series compared with two other series of spectra measured on samples without heating.

Absorption spectra were obtained at 20 °C using an Olis-modified Cary 14 spectrophotometer. Circular dichroism (CD) measurements were made in a Jasco J710 spectropolarimeter. CD values are plotted on a molar scale of $\epsilon_L - \epsilon_R$ in units of M^{-1} cm^{-1}, per mole of nucleotide. All spectra were taken at 20 °C.

Chemical Modification with Hydroxylamine or 2-Hydroperoxytetrahydrofuran. Single-stranded (CCG)$_{15}$ or ss(CCG)$_{20}$ was labeled as described above for the EMMP analysis. To separate labeled vector DNA from labeled oligonucleotide, DNAs were subjected to electrophoresis in a 2% agarose gel. Oligonucleotides containing (CCG)$_{15}$ or (CCG)$_{20}$ were excised from the agarose gel and purified with glass beads (Mermaid Kit, Bio101, La Jolla, CA). Hydroxylamine (HA) reactions were performed essentially according to the method of Rubin and Schmid (1980). Briefly, HA was freshly prepared by titrating 4 M hydroxylamine hydrochloride (Aldrich) to the required pH (6.5−8.5) with diethylamine (Sigma). Seven microliters of labeled ss(CCG)$_{15}$ (2.0×10^4 dpm, ∼0.7 fmol) and 1.4 pmol of synthetic oligonucleotide of the same sequence as the labeled strand was heated to 100 °C for 3 min and immediately placed on ice. Various amounts of HA (6−46 μL) were added to the DNA and incubated for 25 min at various temperatures (1−55 °C). NaCl was added to yield a final concentration of 50 mM. Stop solution (0.25 mL) containing 0.3 M sodium acetate (pH 5.2), 0.1 mM EDTA, and tRNA was added to 25 μg/ml, followed by 0.75 mL of cold ethanol. The DNA was precipitated by centrifugation and was resuspended in 100 μL of 1 M piperidine and heated for 30 min at 92 °C to generate strand breaks. After removal of piperidine *in vacuo* (×2), the DNA was resuspended in formamide loading buffer, placed in a boiling water bath for 2 min, immediately chilled on ice, and loaded on a DNA sequencing gel containing 20% polyacrylamide and 8 M urea.

2-Hydroperoxytetrahydrofuran (THF-OOH) was prepared according to the method of Liang et al. (1994). Labeled DNA (2×10^4 dpm., ∼0.7 fmol) and unlabeled synthetic oligonucleotide (1.4 pmol) of the same sequence as the labeled strand were combined in buffer containing 50 mM Tris-HCI, 50 mM NaCl and 10 mM MgCl, in a final reaction volume of 30 μL, placed in a boiling water bath for 3 min, and placed on ice. THF-OOH was added to yield final concentrations of 0.5, 1.0, and 2.0 M. Reactions were incubated for 1 h at various temperatures. The pH of the reactions was either 7.5 or 8.5. In alternative reactions, the amount of THF-OOH was decreased to 0.2 M and the reaction time was increased to 10 h. The results of these studies were indistinguishable from those performed at higher THF-OOH concentrations. Incubations were stopped by cooling the reaction in ice and precipitating the DNA at −70 °C by adding 0.23 mL of stop solution and 0.75 mL of ice-cold EtOH. The precipitated DNA was washed with EtOH and dried *in vacuo*. To generate strand breaks, the THF-OOH-treated DNA was heated at 90 °C for 15 min in 30 μL of 10 mM Tris-HCl (pH 7.0) and 1 mM EDTA. The DNA was precipitated, washed, dried, and treated with piperidine as described above for the HA reactions.

P1 Nuclease Digestion. Single-stranded (CCG)$_{15}$ or ss(CCG)$_{20}$ was labeled as described above for the chemical modification analyses. Unlabeled synthetic oligonucleotide (1.4 pmol) of the same sequence as the labeled strand was added to 4×10^3 dpm (∼0.7 fmol) of 3′ end-labeled DNA, and the sample was placed in a boiling H_2O bath for 3 min and placed on ice for 5 min. P1 nuclease digestions were performed in 50 mM Tris-HCl (pH 6.5−8.5) 10 mM $MgCl_2$, and 50 mM NaCl at various temperatures essentially according to the method of Wohlrab (1992) as previously described (Mitas et al., 1995b) For some reactions, the amount of NaCl was adjusted between 0.0 and 0.4 M.

Dimethyl Sulfate Reactions. Single-stranded (CCG)$_{15}$, ss(CCG)$_{20}$, or ΔT_{II}(GCT)$_{15}$ were labeled as described above for P1 nuclease analysis. Unlabeled synthetic oligonucleotide (1.4 pmol) of the same sequence as the labeled strand was added to 4×10^3 dpm (∼0.7 fmol) of labeled DNA, and the sample was placed in a boiling H_2O bath for 3 min and placed on ice for 5 min. Dimethyl sulfate (DMS) reactions were performed at 37 °C for 3 min essentially according to the method of Maxam and Gilbert (1980) as previously described (Mitas et al., 1995b).

RESULTS

Of the six possible (G + C)-rich single-stranded (ss) repeating triplet sequences, d(CCG)$_n$ contains the highest number of cytosines. Since the pK_a of cytosine is the highest among all the bases (Saenger, 1984), we reasoned that sequences containing d(CCG)$_n$ might exhibit pH-dependent structural transitions. To test this possibility, ss(CCG)$_{15}$ and ss(C^mCG)$_{15}$ were labeled at the 3′-end with the use of Klenow enzyme and examined by polyacrylamide gel electrophoresis at various pH values at a constant temperature (28 °C). The

electrophoretic mobility of each DNA was determined relative to its respective Watson−Crick form. For controls, ss(GAT)$_{15}$, which contains no detectable secondary structure (Mitas et al., 1995b), and ss(CTG)$_{15}$, which forms a stable hairpin at ≤37 °C in low ionic strength (Mitas et al., 1995b, Yu et al., 1995b), were analyzed.

At the five pH values tested, the relative electrophoretic mobility (M_{rel}) of ss(GAT)$_{15}$ was slow ($0.89 \leq M_{rel} \leq 0.87$), while that of ss(CTG)$_{15}$ was fast ($1.16 \leq M_{rel} \leq 1.15$) (Figure 1A). No electrophoretic transitions were detected with ss(GAT)$_{15}$ or ss(CTG)$_{15}$ (Figure 1A) or with ss(ATC)$_{15}$, ss(CAG)$_{15}$, ss(GAC)$_{15}$, or ss(GTC)$_{15}$ (data not shown). These results indicated that the bases within these sequences were not protonated at pH 7.5.

Protonation of ss(CCG)$_{15}$ at a pH above Neutrality. At pH 8.5−7.9, the M_{rel} of intramolecular[2] ss(CCG)$_{15}$ varied between 1.02 and 1.00, values higher than that of random coil ss(GAT)$_{15}$ DNA but lower than that of a hairpin containing paired mismatched bases. This result suggested that although ss(CCG)$_{15}$ contained secondary structure at these pH values, the structure was not as stable as the ss(CTG)$_{15}$ hairpin. The M_{rel} of ss(C^{m}CG)$_{15}$ was similar to that of nonmethylated ss(CCG)$_{15}$ (Figure 1A), suggesting that methylation of the CpG dinucleotide did not result in a significant structural change. At pH 7.7, the M_{rel} of ss(CCG)$_{15}$ and ss(C^{m}CG)$_{15}$ increased to 1.13 and 1.14, respectively, values similar to that of ss(CTG)$_{15}$. The changes in electrophoretic mobilities of the C-rich sequences as a function of pH provided strong evidence that at least some fraction of the cytosines were protonated at pH 7.7, a value well above the pK_a of cytosine. The approximate midpoint in the electrophoretic transition of ss(CCG)$_{15}$ and ss(C^{m}CG)$_{15}$ (i.e., the pK_a) was 7.8. Lowering of the pH to 6.5 did not result in a further increase in the M_{rel} of ss(CCG)$_{15}$ or ss(C^{m}CG)$_{15}$ (data not shown), suggesting that no additional cytosines were protonated between pH 7.7 and 6.5.

Thermal Stability of ss(CCG)$_{15}^+$ Is Modestly Higher Compared to That of ss(CCG)$_{15}$. To investigate the thermal stabilities of the nonprotonated and protonated [ss(CCG)$_{15}^+$] structures of ss(CCG)$_{15}$, electrophoretic mobility melting profiles (EMMPs) were obtained at pH 7.5 and 8.5 ([Na$^+$] ≈ 1 mM; Figure 1B). Between the temperatures of 10 and 66 °C, the M_{rel} of ss(CCG)$_{15}$ at pH 8.5 changed 17.6%. The midpoint of the (gradual) electrophoretic phase transition, which is an estimate of the melting temperature (T_m) of the DNA (Wartell et al., 1990; Ke & Wartell, 1993; Mitas et al., 1995a; Yu et al., 1995a,b), was 30 °C. This T_m value is the lowest among all Class I triplet repeat DNAs (Table 1), indicating that C-C mismatches are the least stable among the four homomismatches. At pH 7.5, the EMMP of ss(CCG)$_{15}$ was steeper and shifted slightly to the right (Figure 1B), indicating that the structure was stabilized modestly by cytosine protonation. An estimate of the T_m of ss(CCG)$_{15}^+$ at pH 7.5 was 37 °C (Table 1).

To further investigate the thermal stabilities of ss(CCG)$_{15}$ and ss(CCG)$_{15}^+$, UV absorbance melting profiles were obtained in 150 mM NaCl at pH 7.5 and 8.5. Similar to the results of the EMMP, the UV absorbance melting profile at pH 7.5 was steeper compared to that at pH 8.5. At the higher pH value, the absorbance of ss(CCG)$_{15}$ changed 9.14% between the temperatures of 4 and 100 °C, whereas at the lower pH value the UV absorbance changed 13.2% (Figure

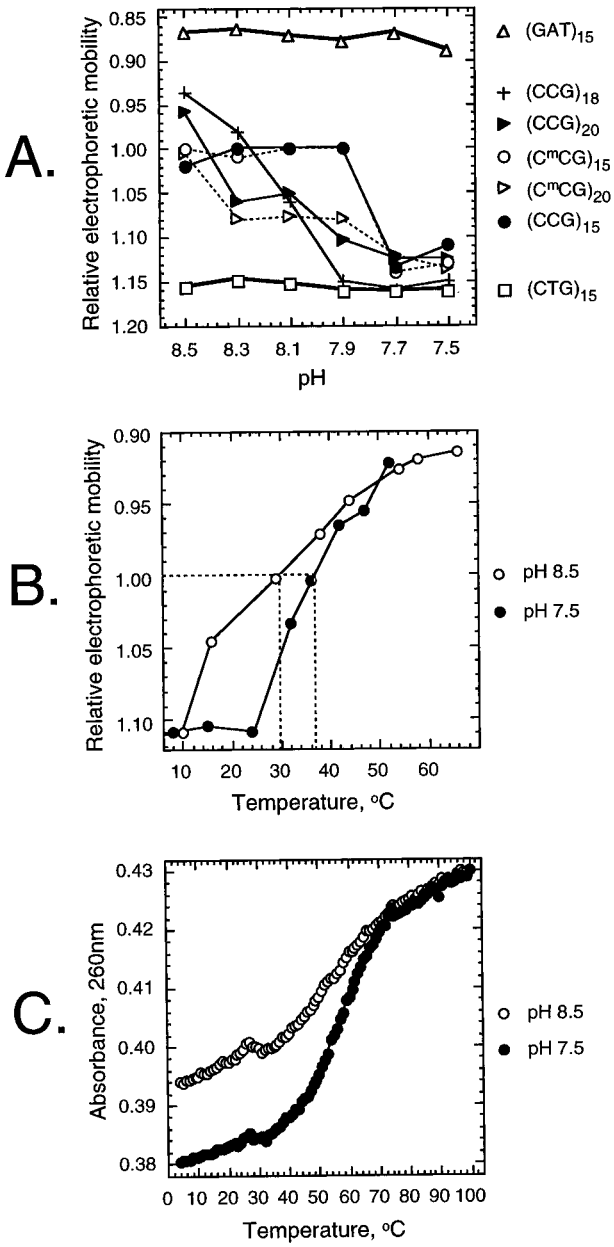

FIGURE 1: Single-stranded (CCG)15 forms two pH-dependent structures. (A) pH-dependent electrophoretic analysis. DNA samples were applied to a native 8% polyacrylamide gel and subjected to electrophoresis at 28 °C at the indicated pH as described in Materials and Methods. The data plotted are the relative electrophoretic mobilites (M_{rel}) of the various ssDNA sequences, which are listed to the right of the figure. M_{rel} of ssDNA = distance ssDNA migrated from origin/distance dsDNA migrated from origin. Data are the mean of two experiments. Except for minor variations due to "smiling" during electrophoresis, the rates of migration of all dsDNAs containing 15 triplet repeats were identical. (B) Temperature-dependent electrophoretic analysis of ss(CCG)$_{15}$. DNA samples were applied to native 8% polyacrylamide gels and subjected to electrophoresis at various temperatures at the indicated pH as described in Materials and Methods. The data plotted are the M_{rel} of ss(CCG)$_{15}$ at the pH indicated to the right of the figure. (C) UV absorbance melting profile of ss(CCG)$_{15}$. The UV absorbance of ss(CCG)$_{15}$ was measured in an Aviv (Lakewood, NJ) Model 14S spectrophotometer at various temperatures at the pH indicated to the right of the figure. The plots represent the mean of two separate experiments. Solution was purged with N_2 prior to absorption measurements and contained 150 mM NaCl, 10 mM Tris HCl, and 1 mM EDTA. Rate of heating was 0.5 °C/min.

[2] The electrophoretic mobility of ss(CCG)$_{15}$ was independent of DNA concentration (70 pM−7 μM).

Table 1: Melting Temperatures of Hairpins Containing 15 Class I Triplet Repeat Sequences[a] in ~1 mM Na^+

$(XXX)_{15}$	pH	T_m (°C)	reference	expansion of triplet repeats associated with the following disease
CGG	8.5	75	Mitas et al., 1995b	fragile X syndrome (Ververk et al., 1991)
GAC	8.5	49	Yu et al., 1995b	none
CTG	8.5	47	Yu et al., 1995a,b	mytotonic dystrophy (Brook et al., 1992; Mahadevan et al., 1992)
CAG	8.5	38	Yu et al., 1995b	Huntington's disease (Huntingon's Disease Collaborative Research Group, 1993)
GTC	8.5	38	Yu et al., 1995a	none
CCG	7.5	37	this study	fragile X syndrome (Ververk et al., 1991)
CCG	8.5	30	this study	fragile X syndrome (Ververk et al., 1991)

[a] Oligonucleotides containing 15 triplet repeats also contained flanking sequences. The sequence of the oligonucleotide containing 15 pyrimidine-rich triplet repeats was GATCC$(XXX)_{15}$GGTACCAAGCT, where XXX = CCG, CTG, and GTC. The sequence of the oligonucleotide containing 15 purine-rich triplet repeats was AGCTTGGTACC$(XXX)_{15}$GGATC, where XXX = CGG, CAG, and GAC.

1C). The increased hypochromicity of ss$(CCG)_{15}^+$ provided direct evidence that the bases of ss$(CCG)_{15}^+$ were better stacked compared to those in ss$(CCG)_{15}$. The T_m values, obtained from first derivatives of the melting profiles, were ~54 °C (a broad peak between 50 and 58 °C was observed) at pH 8.5 and 57.5 °C at 7.5. These results again suggest only a minor increase in thermal stability of ss$(CCG)_{15}^+$ compared to ss$(CCG)_{15}$.

Protonation of cytosine N3 can result in formation of C·$^+$C base pairs (where a center dot designates H-bonds and a superscript plus designates a proton shared between the cytosines) that are arranged in a parallel (Akinrimisi et al., 1963; Inman, 1964; Gehring et al., 1993) or an antiparallel (Brown et al., 1990) double helix. The sequence- and pH-dependent electrophoretic behavior of ss$(CCG)_{15}$ demonstrated that at least some of the cytosines in ss$(CCG)_{15}^+$ were protonated at or near physiologic pH. Sterochemically possible structures of ss$(CCG)_n^+$ are shown in Figure 2. In two theoretical structures, $(CCG)_n$ sequences fold into tetraplexes by virtue of formation of C·$^+$C pairs that are arranged in parallel [stuctures **1(a)** and **1(b)**]. In another type of theoretical structure, mismatched cytosines arranged in ~Watson−Crick hairpin structures in **(a)** or **(b)** alignment hairpins [structures **2(a)** and **2(b)**, respectively] are protonated, allowing formation of two H-bonds between each pair. The duplex structure of the **(a)** alignment contains GpC base-pair steps, whereas the **(b)** alignment contains CpG base-pair steps. In the last theoretical structure, which is an extension of the e-motif recently described by Gao et al. (1995), pairs of extrahelical cytosines, which are formally separated by two C·G base pairs, stack together in the minor groove (structure **3**).

ss$(CCG)_{15}^+$ Contains No Detectable C·$^+$C Base Pairs. When analyzed by circular dichroism (CD), DNA structures that contain C·$^+$C base-pairs arranged in parallel exhibit marked positive bands at 290 nm (Gray et al., 1988). To investigate the possibility that ss$(CCG)_{15}$ contained C·$^+$C base pairs, CD studies were performed at various pH values. The CD values at 290 nm (Figure 3A) and 210 nm (Figure 3B) were plotted as a function of pH. The CD results indicated two spectral transitions; the first transition was between pH 8.5 and pH 6.5, while the second was between about pH 6.5 and 4.5. The first spectroscopic phase transition was consistent with the electrophoretic data.

CD magnitudes of the long-wavelength positive band and the negative band at 210 nm increased concomitantly (Figures 3 and 4A), with little change in the UV absorption spectrum (Figure 4B), as the pH was reduced from 8.5 to 6.5. These spectral features corresponded to an increased stacking and/or base pairing in ss$(CCG)_{15}^+$, and the increased prominence of the band at 210 nm is a feature often associated with the **A** family of nucleic acid conformations (Gray et al., 1992). These results indicated that although the bases in ss$(CCG)_{15}^+$ were better stacked at pH 6.5−7.5, they were not in a normal Watson−Crick arrangement. This result suggested that ss$(CCG)_{15}^+$ did not adopt structures **2(a)** or **2(b)**.

At pH 6.5 there were no CD features associated with the formation of stacked, protonated C·$^+$C base pairs, suggesting that ss$(CCG)_{15}^+$ did not adopt structures **1(a)** or **1(b)**. However, below pH 6.5, the long-wavelength CD and absorption bands underwent characteristic red shifts (as shown in Figure 4), which together with an increase in the magnitude of the long-wavelength positive CD above 285−290 nm provided evidence that some fraction of the cytosines had the spectral characteristics of the hemiprotonated C·$^+$C base pairs of helical poly[d(C·$^+$C)] (Gray et al., 1988, Antao & Gray, 1993). The structure below pH 6.5 might correspond to a tetraplex like structure in which C·$^+$C base pairs form between parallel DNA strands that are adjacent to one another (Figure 2, structure **1**). Since this putative tetraplex structure only forms under nonphysiologic acidic conditions, it will not be discussed in further detail.

The Structure of ss$(CCG)_{15}$ at pH 8.5 Is a Hairpin in a (b) Alignment. Our failure to detect C·$^+$C base pairs in $(CCG)_{15}^+$ at pH 7.5 suggested that its structure might be a hairpin rather than a tetraplex. In order to investigate the possibility that ss$(CCG)_{15}$ adopted hairpin conformations at neutral and mildly basic pH, chemical modifications were performed with hydroxylamine (HA) (Singer & Grunberger, 1983; Johnston & Rich, 1985; Johnston, 1992) and 2-hydroperoxytetrahydrofuran (THF-OOH) (Liang et al., 1994). These reagents preferentially react with cytosines that are not base-paired. If ss$(CCG)_{15}$ adopted a hairpin conformation, one of the cytosines within a given CCG triplet should be base-paired to a guanine (i.e., a C·G pair), while the other cytosine should pair with a cytosine. Since it is not known whether the mismatched cytosines in a $(CCG)_n$-containing hairpin are H-bonded, the mismatched cytosines will be designated as a C-C pair. The cytosine of the C·G pair should react poorly with HA or THF-OOH. Therefore, incubation of ss$(CCG)_{15}$ with HA or THF-OOH should reveal which of the two cytosines within a triplet is base-paired to guanine. If ss$(CCG)_{15}$ were to adopt a hairpin in the **(a)** alignment, the cytosine 5′ to the nearest guanine (designated as 5′*C*) should be much more reactive with HA or THF-

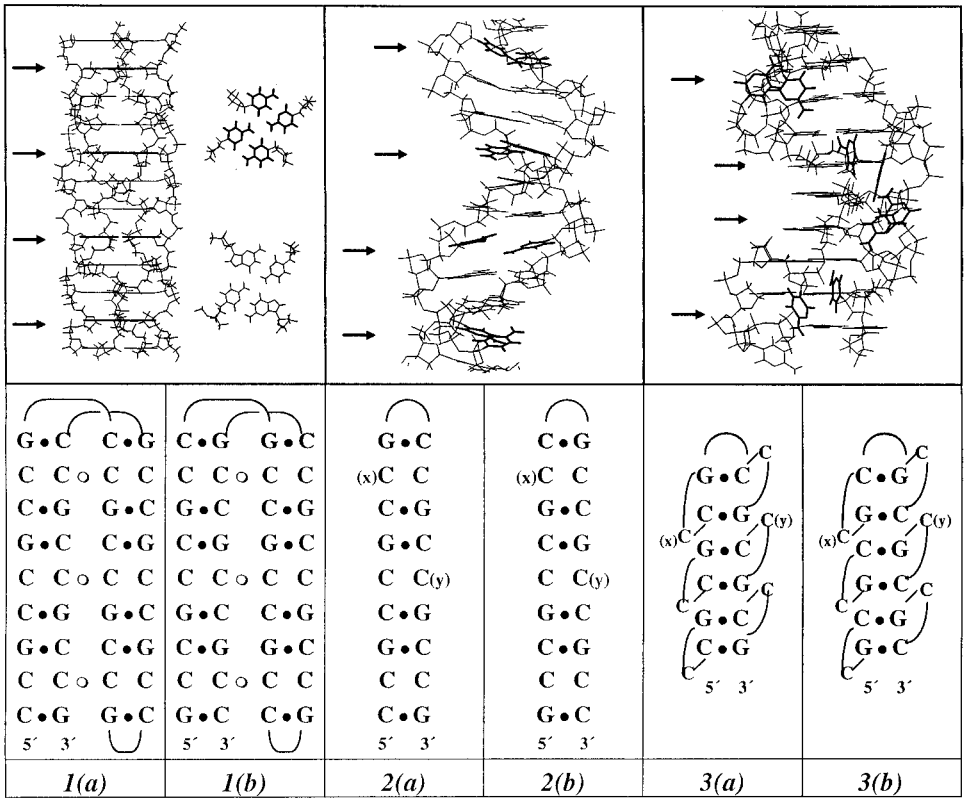

FIGURE 2: Possible conformations adopted by single-stranded $(CCG)_n$. Possible conformations of $d(CCG)_n$ are shown schematically in the lower panels, with the stem regions of each structure shown in all-atom format in the upper panels. Each model structure was built using software developed to allow the flexible construction of DNA conformations (J. Shin, R. Romero, and I. S. Haworth, unpublished work). Structures **1(a)** and **1(b)** are quadruplex conformations, structures **2(a)** and **2(b)** are ~regular B-DNA hairpins, and structures **3(a)** and **3(b)** are extended e-motif hairpins. The **(a)** and **(b)** notation for each structure refers to the alignment of the interacting CCG repeat sequences. In alignment **(a)** a 5′-CCG repeat pairs in an antiparallel manner with a second 5′-CCG repeat, thus forming a GpC base-pair step. In the **(b)** alignment, the 5′-CCG repeat is paired with a 5′-GCC repeat, thus forming a CpG base-pair step. Conformation **1** is a single-stranded quadruplex containing a repeat unit of the one C_4^{2+} and two CGCG quartets, arranged as shown. The quadruplex can be formed by a "folding" of the appropriately aligned hairpin, with the quadruplex loops adopting a "crossover" orientation at the top of the figure. This creates the strand organization necessary to allow two parallel-stranded C + C interactions. Cytosine protonation is necessary to stabilize this conformation. Conformation **2** is a regular B-DNA hairpin conformation containing C-C mismatch pairs (either protonated or nonprotonated). These pairs are clearly unstable, and the result of a minimization from the standard B-DNA conformation is to separate the two cytosine bases, as shown in the center of the stem. We believe that this instability ultimately leads to the development of nonstandard hairpin conformations for this sequence. Conformation **3** is referred to as an extended e-motif, and is based on the original e-motif structure (Gao et al., 1995). The essential element in this structure is the presence of extrahelical cytosine bases formed from cytosines that were originally present as unstable C-C mismatch pairs. In the **(b)** alignment of the extended e-motif, expulsion of the cytosine bases from the body of the helix results in the formation of a pseudo-5′-GC step, which provides the structure with considerable stability, via G-G stacking. Such stacking is not possible in the pseudo-5′-CG step formed in the **(a)** alignment, and consequently we believe the extended e-motif is only plausible for the **(b)** alignment. On the basis of a molecular dynamics simulation (starting from a conformation obtained from the original e-motif; R. Romero, M. Mitas, and I. S. Haworth, to be published), we believe that the extrahelical cytosines, located in the minor groove, fold back toward the 5′-end of the strand and stack with a neighboring extrahelical cytosine, providing further stability. To illustrate this process, two cytosine bases, C(x) and C(y), have been labeled in the regular hairpin conformations shown for structures **2(a)** and **2(b)**. In the extended e-motif conformations, these cytosine bases fold back such that they occupy the minor-groove region of the neighboring 5′-CG step [in structure **3(b)**] and ultimately form a stacked C-C pair. Examples of this stacking interaction can be seen in the all-atom structure in the upper panel of structure **3**.

OOH compared to the cytosine 3′ to the nearest guanine. In contrast, if $ss(CCG)_{15}$ were to adopt a hairpin in the **(b)** alignment, the 3′Cs should be much more reactive with HA compared to the 5′*C*s.

Incubation of $ss(CCG)_{15}$ with HA (Figure 5) or THF-OOH (data not shown) at pH 8.5 revealed high reactivity of the 3′Cs, especially those at the 3′-end of the sequence. These results provided evidence for a hairpin in the **(b)** alignment. The asymmetric nature of the chemical modifications are described in further detail below. Reactivity of the 5′*C*s was not observed, with the exception of C28, a nucleotide located in the presumed loop region of $ss(CCG)_{15}$ (Figure 5). Also, both cytosines in triplet I were nonreactive with HA (Figure 5) or THF-OOH (data not shown), a result consistent with a hairpin in the **(b)** alignment.

Hydroxylamine highly reacted with C55 and C56, two cytosines not part of the triplet repeat region. C55 was opposite to A2 in the hairpin structure of $ss(CCG)_{15}$. Since C55 cannot form H-bonds to A2 at pH 8.5, the HA reactivity of C55 served as a internal control for an unpaired cytosine. The HA reactivity of C55 was comparable to that of C48, a cytosine within a C-C pair. The fact that the HA reactivities of C55 and C48 were similar provided evidence that there were no H-bonds in the C-C mismatches of $ss(CCG)_{15}$ at pH 8.5.

The Sugar−Phosphate Backbone of ss(CCG)$_{15}$ Is Distorted at pH 8.5. One hallmark of a **(b)** alignment duplex containing CCG repeats is distortion of the CpG base-pair steps (Gao et al., 1995). To investigate the possibility that $ss(CCG)_{15}$ formed a **(b)** alignment hairpin containing dis-

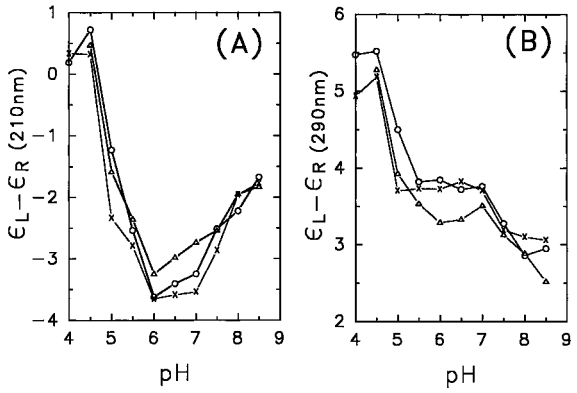

FIGURE 3: Circular dichroism studies indicate two phase transitions of ss(CCG)$_{15}$. Circular dichroism analyses of ss(CCG)$_{15}$ were performed as described in Materials and Methods. Two series of samples were not heat-treated (○, Δ); one series was heat-treated (×) prior to spectroscopic analysis. CD values ($\epsilon_L - \epsilon_R$) at (A) 210 and (B) 290 nm.

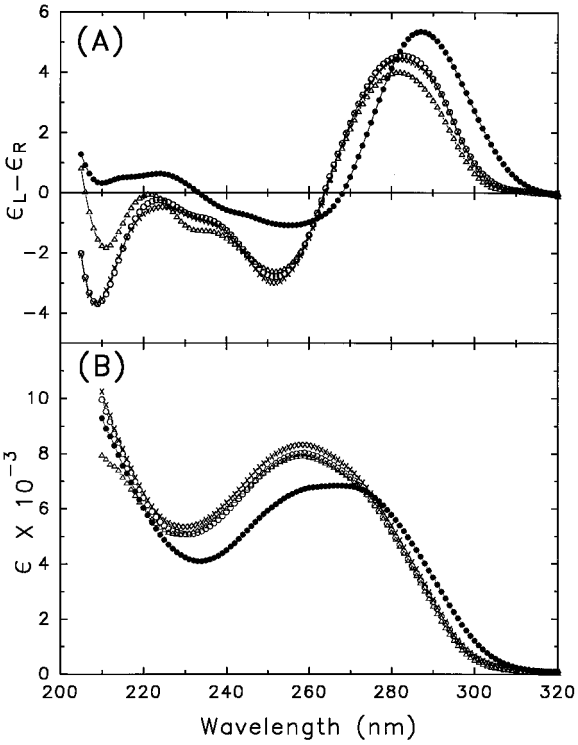

FIGURE 4: Spectral analysis of ss(CCG)$_{15}$. Example CD (A) and absorption (B) spectra of ss(CCG)$_{15}$ taken at pH 8.5 in Tris-HCl (Δ) and, in phosphate buffer at pH 7.0 (×), pH 6.5 (○), and pH 4.5 (●). Spectra were from the heat-treated series; see Materials and Methods.

torted CpG base-pair steps, P1 nuclease reactions were performed at various temperatures in 50 mM NaCl. At pH 8.5 and 25 °C, major sites of P1 nuclease cleavages in the triplet repeat region were the G26−C27, C28−G29, G29−C30, and C31−G32 phosphodiesters (Figure 6). This result indicated that a fold was present near triplet VIII and provided further evidence for a hairpin conformation. At 55 °C, the pattern of P1 nuclease cleavage was more uniform, indicating that the hairpin structure was significantly denatured at this temperature. This result is consistent with UV studies (Figure 1C), which indicated that the T_m of ss(CCG)$_{15}$ in a higher salt concentration (150 mM) was ∼54 °C.

The 5′ *C*pG 3′ phosphodiester linkages in triplets X−XIV of ss(CCG)$_{15}$ were also cleaved to a minor extent (Figure

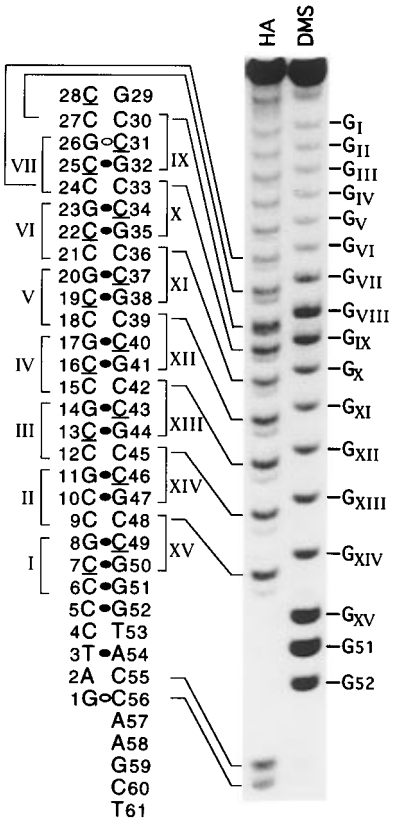

FIGURE 5: Chemical modification with hydroxylamine at pH 8.5 reveals a hairpin in a **(b)** alignment. Single-stranded (CCG)$_{15}$ was incubated with hydroxylamine (HA) at 37 °C, pH 8.5, in 50 mM Na^+ as described in Materials and Methods and applied to a 20% polyacrylamide gel containing 8 M urea. The concentration of HA in the reaction mixture was 6.3 M. The deduced hairpin structure of ss(CCG)$_{15}$ is shown on the left. The ^{32}P labels in the oligonucleotide are 5′ to A58 and C60. Roman numerals indicate triplet repeat numbers. Conventional Arabic numerals indicate the position of a given nucleotide with respect to the 5′-end. The positions of the reactive Cs within the deduced hairpin structure are shown. Dimethyl sulfate (21 mM) reactions were performed as described in Materials and Methods.

6). This result was significant since minor cleavages of the phosphodiesters in the stem regions of oligonucleotides containing all other class I triplet repeat sequences have not been observed at pH 7.5 and 37 °C (Mitas et al., 1995a,b; Yu et al., 1995a,b). However, what is more important is that the cleaved phosphodiesters were within the CpG base-pair steps and not the C-C mismatches. The ability of P1 nuclease to cleave the phosphodiester linkage within two C·G base pairs provides direct evidence that the sugar−phosphate backbone was distorted.

The major site of P1 nuclease cleavage at the 3′ terminus of ss(CCG)$_{15}$ was the C56−A57 phosphodiester (Figure 6). This result was in contrast to P1 nuclease cleavage of ss(CTG)$_{15}$ and ss(GTC)$_{15}$, where the major site of cleavage in both of these oligonucleotides was the C55−C56 phosphodiester (Yu et al., 1995b). In ss(CTG)$_{15}$ and ss(GTC)$_{15}$, the base-pairing arrangements at the termini are exactly the same as would be observed in an **(a)** alignment hairpin of ss(CCG)$_{15}$. Therefore, if ss(CCG)$_{15}$ were to adopt an **(a)** alignment, the major site of cleavage at the 3′ terminus should be the C55−C56 phosphodiester. The finding that the major site of P1 cleavage at the 3′ terminus of ss(CCG)$_{15}$ is the C56−A57 phosphodiester further supports a **(b)** alignment.

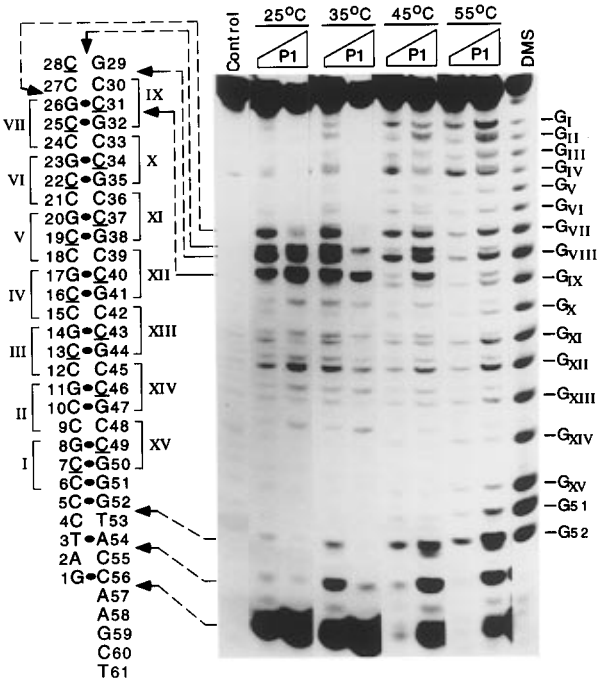

FIGURE 6: P1 nuclease digestion of ss(CCG)$_{15}$ at pH 8.5. Single-stranded (CCG)$_{15}$ was incubated with P1 nuclease for 5 min as described in Materials and Methods at the indicated temperatures. Buffer contained 50 mM Na^+. Reaction products were applied to a 20% polyacrylamide gel containing 8 M urea. The amounts of P1 nuclease added at the respective temperatures were as follows: 25−45 °C, 0.116 (left lane) and 0.31 unit (right lane); 55 °C, 0.035 (left lane) and 0.116 unit (right lane). Control lane was incubated at 25 °C in the absence of P1 nuclease. Dimethyl sulfate reactions were performed as described in Materials and Methods. Roman numerals represent triplet repeat numbers. Conventional Arabic numerals indicated the position of a given nucleotide with respect to the 5′-end. Arrows indicate sites of P1 nuclease cleavage in the ss(CCG)$_{15}$ hairpin.

P1 Nuclease Reactions with a Control $\Delta T_{II}(GTC)_{15}$ Sequence Indicate a Lack of Cleavage in the Hairpin Stem. Detection of P1 nuclease susceptibility at the CpG base-pair step in the ss(CCG)$_{15}$ hairpin might have been due to experimental (e.g., high pH) rather than structural conditions. To provide a control for P1 nuclease studies performed at pH 8.5, a sequence variant of ss(GTC)$_{15}$ was analyzed (Figure 7). The hairpin formed by ss(GTC)$_{15}$ is similar to that formed by ss(CCG)$_{15}$ at pH 8.5 in a number of respects. First, both hairpins are in **(b)** alignments and contain CpG base-pair steps. Second, the thermal stabilities of the two hairpins are low relative to ss(CTG)$_{15}$ (Table 1), and third, the mismatched bases in the ss(CCG)$_{15}$ and ss(GTC)$_{15}$ hairpins are pyrimidines. The sequence variant $\Delta T_{II}(GTC)_{15}$ is comparable to ss(GTC)$_{15}$ except that the thymine in triplet II is missing. In the hairpin structure of $\Delta T_{II}(GTC)_{15}$, the extrahelical thymine in triplet XIV does not have a nucleotide with which it can pair, and hence, it is readily oxidized by $KMnO_4$ (Yu et al., 1995b).

At pH 8.5 and 50 mM Na^+, P1 nuclease did not cleave the stem region of $\Delta T_{II}(GTC)_{15}$ (including the region corresponding to the extrahelical thymine) at 25 and 35 °C (Figure 7), indicating that the sugar−phosphate backbone was not distorted. These results indicated that cleavages of the CpG dinucleotides in the ss(CCG)$_{15}$ hairpin at pH 8.5 were due to major distortion of the sugar−phosphate backbone.

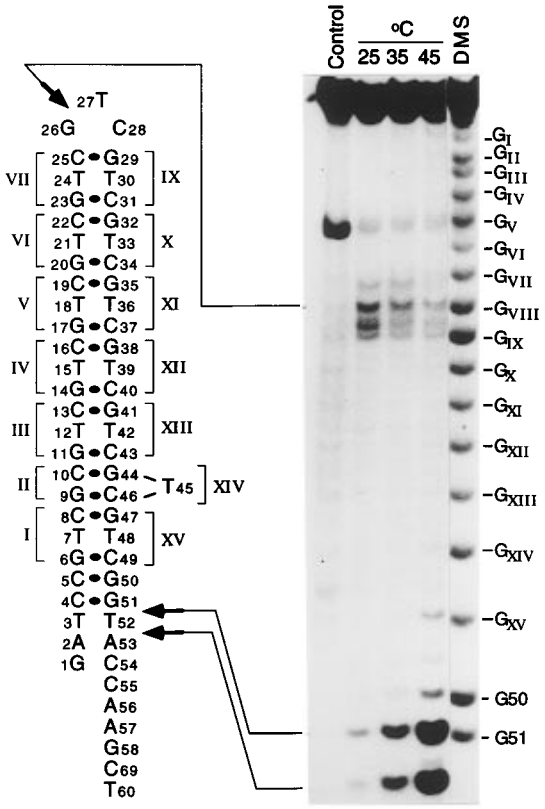

FIGURE 7: P1 nuclease digestion of $\Delta T_{II}(GTC)_{15}$ at pH 8.5. Single-stranded $\Delta T_{II}(GTC)_{15}$ was incubated with P1 nuclease at the indicated temperatures as described in the legend to Figure 6. The amounts of P1 nuclease added to each of the reactions was 0.10 unit. Control lane was incubated at 25 °C in the absence of P1 nuclease. The signal in the control lane that corresponds to G_V in the marker lane is a contaminant of dsDNA. Quantitative analysis of the control lane with a Storm Molecular Dynamics PhosphorImager (Sunnyvale, CA) revealed that the contaminant represented <5% of the total DNA.

*The Structure of ss(CCG)$_{15}^+$ at pH 7.5 Is a Hairpin in a **(b)** Alignment.* The results of electrophoretic mobility (Figure 1A), UV absorbance (Figure 1C), and CD studies (Figure 3) indicated that the bases of ss(CCG)$_{15}$ were more highly stacked at pH 6.5−7.5. In order to investigate the base-pairing arrangement below neutral pH, HA reactions at various temperatures (15−55 °C) were performed in 50 mM Na^+ at pH 7.5. We were particularly interested to determine whether the hairpin at pH 7.5 might have switched to an **(a)** alignment. In this alignment, it is possible to form antiparallel C·$^+$C pairs similar to those described by Brown et al. (1990) [Figure 2, structure **2(a)**].

Similar to results obtained at pH 8.5 (Figure 5), reactions performed at or below 45 °C revealed high reactivity of the 3′Cs, indicating that ss(CCG)$_{15}$ formed a hairpin in the **(b)** alignment (Figure 8). At ≤ 35 °C, all cytosines in triplets VIII and IX reacted with HA, indicating that the bases in the loop region of ss(CCG)$_{15}$ were not paired. Identical results were obtained at 37 °C at pH 7.5 with THF-OOH (data not shown) or with HA at pH 6.5, 15−35 °C (50 mM Na^+) or 37 °C (pH 6.5), 600 mM NaCl (data not shown). At 55 °C, near-uniform reactivities of adjacent cytosines in ss(CCG)$_{15}$ were observed, indicating that all C·G base pairs were melted at this temperature (Figure 8). The result is consistent with UV absorbance studies (Figure 1C), which indicated that the melting temperature of ss(CCG)$_{15}^+$ in 150 mM NaCl was 57.5 °C.

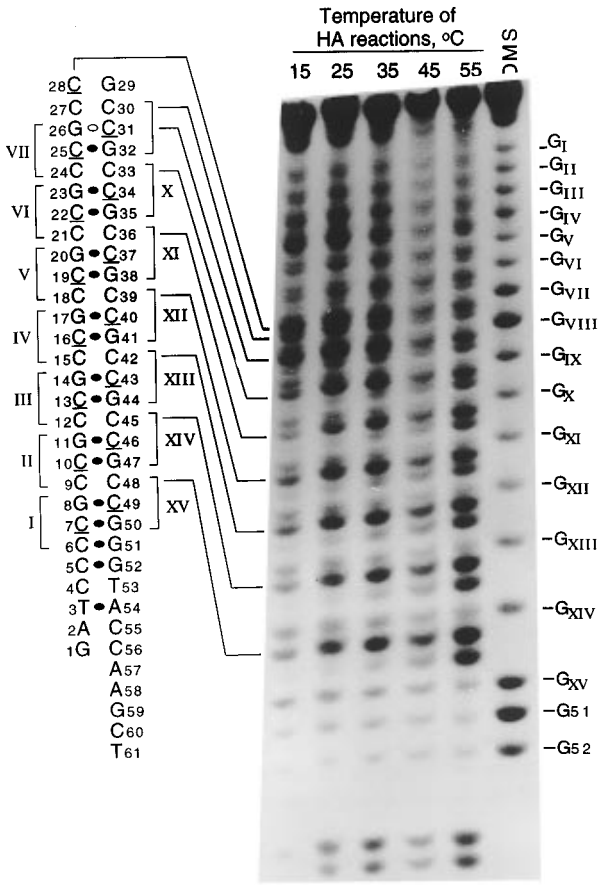

FIGURE 8: Hydroxylamine modification of ss(CCG)$_{15}$ at pH 7.5. Single-stranded (CCG)$_{15}$ was incubated with HA at the indicated temperatures in buffer containing 50 mM Na^+, pH 7.5, as described according to the legend to Figure 5. Concentrations of HA in the reactions were as follows: 35, 45, and 55 °C, 3.03 M; 15 and 25 °C, 5.68 M. Control lane was incubated at 25 °C in the absence of HA. The deduced hairpin structure of ss(CCG)$_{15}$ is shown on the left. Dimethyl sulfate (21 mM) reactions were performed as described in Materials and Methods.

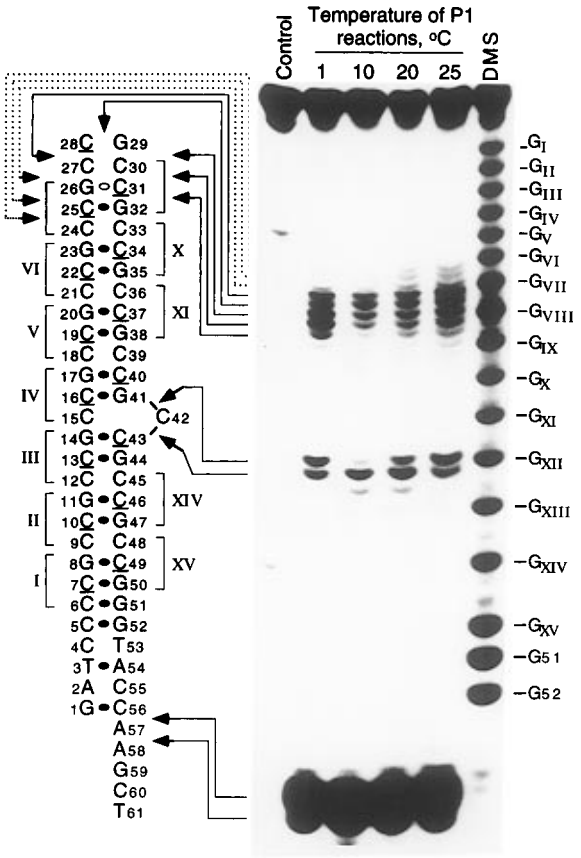

FIGURE 9: P1 nuclease digestion of ss(CCG)$_{15}$ at pH 6.5. Single-stranded (CCG)$_{15}$ was incubated with P1 nuclease as described according to the legend to Figure 5 at the indicated temperatures. Buffer contained 50 mM Na^+, pH 6.5. The amount of P1 nuclease added per reaction was 3.5×10^{-2} unit. Incubation times are as follows: control, 5 h; 1 °C, 5 h; 10 °C, 25 min; 20 °C, 15 min; and 25 °C, 9 min. Control lane was incubated at 25 °C in the absence of P1 nuclease. Dimethyl sulfate reactions were performed as described in Materials and Methods.

Base Stacking and Base Flipping at pH 6.5. To further investigate the structure of ss(CCG)$_{15}^+$ at or slightly below neutral pH, P1 nuclease digestions were performed at pH 6.5 and 7.5. Since the results at both pH values were indistinguishable, only those obtained at pH 6.5 will be described. Incubation of ss(CCG)$_{15}$ with P1 nuclease at pH 6.5 resulted in extensive cleavage of the loop region (Figure 9). This result was in complete agreement with the HA results (Figure 8), which indicated that C31 was not extensively paired to G26. Substantial cleavages of the G41−C42 and C42−C43 phosphodiester linkages in the center of the helix stem were also observed. These phosphodiesters flank a cytosine (C42) of a C-C mismatch. Little or no cleavages of the C40−G41 or C43−G44 phosphodiester linkages were observed (Figure 9), indicating that C42 and only C42 was extruded or flipped away from the 3′ portion of the helix. Cleavages of the G41−C42 and C42−C43 phosphodiesters were also observed over a wide range of salt concentrations (0−400 mM NaCl) and over a range of temperatures (37−57 °C) at pH 7.5 (data not shown), indicating that the flipped cytosine configuration was heat- and salt-stable. Little or no cleavage of the G14−C15 and C15−C16 phosphodiesters was observed, indicating that C15, the base opposite that of C42 in the formal hairpin, was not flipped out of the helix.

In contrast to the P1 nuclease results at pH 8.5 (Figure 6), minor cleavages of all phosphodiesters between triplets IX and XII were observed at pH 6.5, most prominently at 25 °C (Figure 9). Cleavages of these phosphodiesters were also observed at pH 6.0, 7.0, and 7.5 (data not shown). These results provide evidence that the sugar−phosphate backbone of ss(CCG)$_{15}^+$ was distorted.

*(CCG)$_n$-Containing Sequences Designed to Form Hairpins in the **(a)** Alignment Form Hairpins in the Alternative **(b)** Alignment.* The results presented above demonstrate that ss(CCG)$_{15}$ and ss(CCG)$_{15}^+$ formed hairpins in the **(b)** alignment. One possibility was that ss(CCG)$_{15}$ folded into a **(b)** alignment because of the influence of the loop structure and/or the sequences in the non-triplet repeat region. To address this possibility, we analyzed oligonucleotides with the following features: (i) a greater number of CCG repeats (18 and 20), which diminished the influence of loop structures and the non-triplet repeat regions, (ii) terminal sequences that were designed to force the triplet repeat region to fold into an **(a)** hairpin alignment (i.e., four consecutive Watson−Crick base pairs are formed), and (iii) an **(a)** alignment that contained a four- (rather than a three-) nucleotide loop. Energy minimizations predicted that the former loop size was more stable than the latter (I.S.H., and R.M.R., unpublished results).

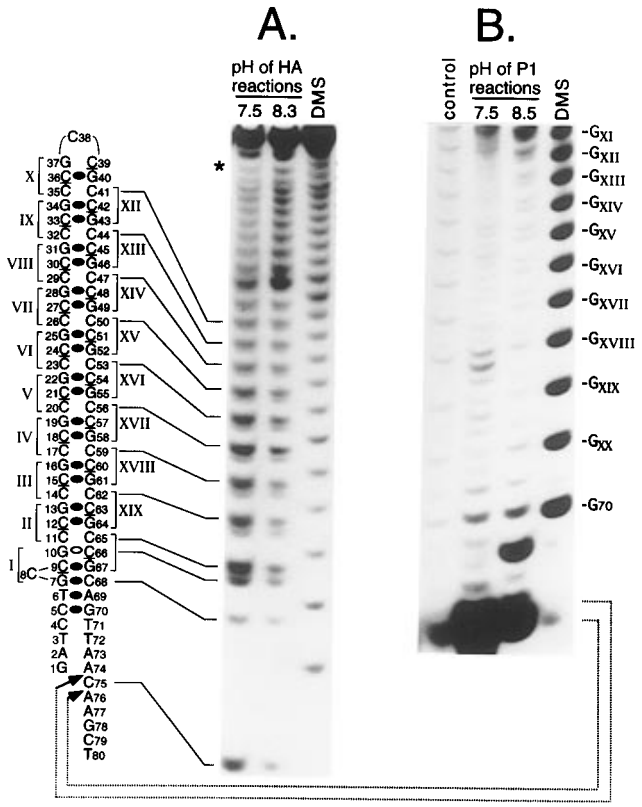

FIGURE 10: Analysis of ss(CCG)$_{20}$. (A) Hydroxylamine modification. Single-stranded (CCG)$_{20}$ was incubated with hydroxylamine (HA) at 37 °C in 50 mM Na^+ as described according to the legend to Figure 4. The concentration of HA in each reaction was 6.3 M. Times of incubation were 25 min (pH 7.5) and 40 min (pH 8.3). The deduced hairpin structure of ss(CCG)$_{15}$ is shown on the left. The ^{32}P labels in the oligonucleotide are 5′ to A77 and C79. The band corresponding to C8 is marked with an asterisk to the left of the figure. Dimethyl sulfate (21 mM) reactions were performed as described in Materials and Methods. (B) P1 nuclease analysis. Single-stranded (CCG)$_{20}$ was incubated with P1 nuclease for 5 min at 37 °C (pH as indicated in the figure) as described according to the legend to Figure 5. The amounts of P1 nuclease added were 0.10 (pH 7.5) and 0.31 (pH 8.5) unit. Time of incubations were 10 min (pH 7.5) and 27 min (pH 8.5). Dimethyl sulfate reactions were performed as described in Materials and Methods.

To investigate the structures of ss(CCG)$_{18}$ and ss(CCG)$_{20}$, their electrophoretic mobilities were analyzed as a function of pH (Figure 1A). Similar to ss(CCG)$_{15}$, the M_{rel} of ss(CCG)$_{18}$, ss(CCG)$_{20}$, and ss(C^{m}CG)$_{20}$ increased as the pH was lowered from 8.5 to 7.5. For ss(CCG)$_{18}$, a sharp transition occurred between pH 8.5 and 7.9, indicating that the pK_a of the sequence was ~8.2. For ss(CCG)$_{20}$ and ss(C^{m}CG)$_{20}$, sharp transitions occurred between pH 8.5 and 8.3, indicating that the pK_a of the sequences was ~8.4. A gradual transition occurred between pH 8.3 and 7.5. These results indicate that a fraction of the cytosines in ss(CCG)$_{18}$ and ss(CCG)$_{20}$ were protonated above pH 8.0.

To investigate the base-pairing arrangement(s) of ss(CCG)$_{18}$ and ss(CCG)$_{20}$, chemical modifications were performed with HA at pH 7.5 or 8.3 at 37 °C, and 50 mM Na^+. Since the results of the HA reactions indicated that the structures of ss(CCG)$_{18}$ and ss(CCG)$_{20}$ were identical (with the exception of the number of nucleotides in the stem), only the results of ss(CCG)$_{20}$ will be presented. Analysis of ss(CCG)$_{20}$ (Figure 10A) revealed that the 3′Cs in triplets XII−XIX were more reactive compared to the 5′*C*s, providing evidence for a hairpin in the **(b)** alignment. However, the reactivities of the cytosines in the non-triplet repeat region were not consistent with the **(b)** alignment hairpin. For example, C66, which was predicted to form a base pair with G10 in the **(b)** alignment (and hence poorly react with HA), was nearly as reactive as C65. This result suggested that C66 only formed "weak" H-bonds with some base, presumably G10. Further, C68, which was predicted to pair with C8 (and hence react well with HA), exhibited HA reactivity that was lower compared to that of C65, a cytosine within a C-C mismatch. This result suggested that C68 formed H-bonds with a guanine, presumably G7. Therefore, the results suggested that the bases in the non-triplet repeat region were partially paired according to an **(a)** alignment, while those in the triplet repeat region were paired according to a **(b)** alignment. These apparently conflicting data can be reconciled if one of the cytosines in triplet I becomes extrahelical (Figure 10A).

The identity of the potential extrahelical base was determined by discerning which of the two cytosines in triplet I reacted with HA. The position of the band at triplet I (marked with an asterisk in Figure 10A) in the reactions performed at pH 7.5 and 8.3 was consistent with reactivity of C8 (and not C9), indicating that C8 might be extrahelical. The extrahelical configuration of this base would allow for formation of three base pairs in the non-triplet repeat region and a **(b)** hairpin alignment of triplets II−XIX.

To further investigate the hairpin structures of ss(CCG)$_{20}$ and ss(CCG)$_{20}^+$, P1 nuclease studies were performed at pH 7.5 or 8.5 at 37 °C and 50 mM Na^+. At pH 7.5, the major sites of P1 nuclease cleavage at the 3′-end were the A74−C75 and C75−A76 phosphodiesters (Figure 10B), a result consistent with the hairpin structure shown in Figure 10. Similar to results obtained with ss(CCG)$_{15}$ at pH 8.5, minor cleavages of the phosphodiesters within the CpG dinucleotide were observed (Figure 10B), indicating that the sugar−phosphate backbone was distorted. At pH 7.5, minor cleavages of all phosphodiesters in the helix region were observed. Significant cleavages of the G61−C62 and C62−C63 phosphodiester linkages were also observed. These phosphodiesters flank a cytosine (C62) of a C-C mismatch, suggesting that C62 was partially flipped away from the helix. However, unlike C42 of ss(CCG)$_{15}$, C62 of ss(CCG)$_{20}$ was not located in the center of the helix but rather toward the 3′ terminus.

DISCUSSION

We have provided evidence that ss(CCG)$_{15}$ forms a hairpin in the **(b)** alignment, whereby the cytosines 3′ to the nearest guanine were mismatched with one another. Chemical modification experiments revealed that the **(b)** alignment was stable in a variety of salt concentrations (0−600 mM NaCl) and pH values (6.0−8.5) (Figures 5 and 9; data not shown). Although we did not investigate the structures of ss(CCG)$_{18}$ and ss(CCG)$_{20}$ as thoroughly as ss(CCG)$_{15}$, the results described in this report indicate that ss(CCG)$_{18}$ (Figure 1A and data not shown) and ss(CCG)$_{20}$ (Figures 1A and 10) also adopted a hairpin in the **(b)** alignment.

In theory, ss(CCG)$_{15,18,20}$ could have adopted the alternative **(a)** alignment, whereby the cytosines 5′ to the nearest guanine were mismatched with one another. The **(a)** alignment contains GpC base-pair steps, whereas the **(b)** alignment contains CpG base-pair steps. The stacking energies of GpC

and CpG base-pair steps have been calculated to be −14.59 and −9.69 kcal/mol, respectively (Saenger, 1984). Therefore, if one assumes that the interactions between the mismatched Cs in **(a)** or **(b)** alignments are identical, one would also assume that an **(a)** alignment should be more stable than a **(b)** alignment. Our failure to detect an **(a)** alignment of ss(CCG)$_{15,18,20}$ suggests that when the number of repeats is ≥15, the interactions of the mismatched Cs in a **(b)** alignment are more stable compared to the interactions of the mismatched Cs in a **(a)** alignment. When the number of repeats is small, the loop structure and/or end effects may favor an **(a)** alignment, as previously observed in ^{1}H NMR studies for the sequences d(GCC)$_{5-7}$ (Chen et al., 1995; Mariappan et al., 1996a).

pH-dependent Stabilization of ss(CCG)$_{15}$ Is Not Due to the Formation of Parallel or Antiparallel C·$^+$C Pairs. The midpoint in the transition of the pH-dependent M_{rel} profile of ss(CCG)$_{15}$ was pH 7.8 (Figure 1A). A similar midpoint was observed from CD studies conducted at 290 nm (pH 7.5; Figure 3B), providing additional evidence for protonation of ss(CCG)$_{15}$ at physiological pH. In further support of a pH-dependent structural transition of ss(CCG)$_{15}$ near pH 7.5, P1 nuclease experiments indicated prominent cleavages of the G41−C42 and C42−C43 phosphodiester bonds up to pH 8.0 (Figure 9 and data not shown). Therefore, on the basis of three independent analyses, we assign a pK_a value of 7.7 ± 0.2 to ss(CCG)$_{15}$. Interestingly, although spectroscopic data indicated further protonation of ss(CCG)$_{15}$ between 7.5 and 7.0 (Figure 3), electrophoretic data did not (Figure 1A). We suspect that this result was because the M_{rel} of ss(CCG)$_{15}$ is governed by charge, as well as structure. Without any structural change, protonation of ss(CCG)$_{15}$ alone would result in a *decrease* in its M_{rel} and not an increase. Therefore, the increase in the M_{rel} of ss(CCG)$_{15}$ at pH 7.7 relative to pH 7.9 must have been due to a large structural change. Between pH 7.7 and 7.0, structural changes were probably offset by the addition of positive charges such that there was no net gain in the M_{rel} of ss(CCG)$_{15}$.

Since C·G base pairs are not stabilized by cytosine protonation, we concluded that the protonation in ss(CCG)$_{15}$ was limited to the C-C mismatches. We are aware of three potential types of cytosine−cytosine interactions that are stabilized by protonation. The first and classic example is the hemiprotonated C·$^+$C pair arranged in parallel that was first described in 1963 (Akinrimisi et al., 1963; Langridge & Rich, 1963). Since C·$^+$C base pairs have been observed at pH values as high as neutrality (Lavelle & Fresco, 1995), we initially speculated that ss(CCG)$_{15}^+$ might contain C·$^+$C pairs within an intramolecular tetraplex (Figure 2A). However, several results rule out the possibility that ss(CCG)$_{15}^+$, ss(CCG)$_{18}^+$, or ss(CCG)$_{20}^+$ contained C·$^+$C pairs arranged in parallel. First, results of CD studies provided no indication of C·$^+$C pairs in ss(CCG)$_{15}$ until the pH was reduced below pH 6.0 (Figures 3 and 4). Second, the failure of HA to react with C16 or C43 of ss(CCG)$_{15}^+$ and C21 or C57 of ss(CCG)$_{20}^+$ (Figure 10A) provided evidence that these cytosines were not involved in a hairpin fold, which is required for parallel arrangement of cytosines. Third, the results of P1 nuclease digestion were not consistent with a hairpin fold in ss(CCG)$_{18}^+$ (data not shown) or ss(CCG)$_{20}^+$ (Figure 10B). Fourth, DMS was able to alkylate the N7s of ss(CCG)$_{15}^+$ (see DMS lanes in Figures 8 and 9), a result that was not consistent with a tetraplex containing CGCG tetrads. Fifth, the results of the UV absorbance melting profile performed at pH 7.5 indicated a single structural transition of ss(CCG)$_{15}^+$ (Figure 1C). This result was most consistent with a hairpin → random coil transition, and not a tetraplex → hairpin → random coil transition. A single structural transition of ss(CCG)$_{15}^+$ in 50 mM Na$^+$ was further verified by HA modification (Figure 8) and by P1 nuclease digestion experiments conducted at pH 6.5 (Figure 9 and data not shown). In both sets of experiments, the data reveal a hairpin at lower temperatures (≤45 °C) and a random coil at higher temperatures (≥50 °C). The results provided no evidence for an intermediate structure.

The second type of a protonated C-C pair is one arranged in an antiparallel orientation (Brown et al., 1990). Again, there are several lines of evidence that rule out the possibility that ss(CCG)$_{15}^+$ or ss(CCG)$_{20}^+$ contained C·$^+$C pairs arranged in antiparallel. First, the pK_a of the antiparallel C·$^+$C pair is 5.25 (Brown et al., 1990), well below the pK_a of ss(CCG)$_{15}$ or ss(CCG)$_{20}$. Second, ^{1}H NMR studies have provided direct evidence that C-C mismatches in a **(b)** alignment do not interact with one another but instead are extrahelical (Gao et al., 1995). Third, the high HA reactivity of the cytosines 3′ to the nearest guanine in ss(CCG)$_{15}^+$ (Figure 5) and ss(CCG)$_{20}^+$ (Figure 10A) suggest a lack of H-bonds to N4 of these cytosines. Fourth, an antiparallel C·$^+$C pair is similar in structure to the T·T pair observed in d(CTG)$_n$ hairpins (Smith et al., 1995; Marriapan, 1996b); both base pairs are pyrimidine-pyrimidine mismatches that involve H-bonds formed between N3 and O2. The results of P1 nuclease studies performed with ss(CCG)$_{15}^+$ (Figures 9 and 10B), ss(CCG)$_{18}^+$ (data not shown), and ss(CCG)$_{20}^+$ (Figure 10B) are not consistent with a helix that contains pyrimidine−pyrimidine base pairs such as those observed in ss(CTG)$_{15}$ or ss(GTC)$_{15}$ (Mitas et al., 1995a; Yu et al., 1995a,b).

C-C$^+$ Pairs Stacked in the Minor Groove: A New DNA Structure. A third type of a protonated C-C pair (C-C$^+$) is theoretical in nature (Figure 2C) and results from extension of the e-motif structure (Gao, 1995). The C-C$^+$ pairs in this extended e-motif are not base pairs *per se*, since they are not stabilized by H-bonds and do not lie in the same plane. Rather, the stability of the pairs arises from stacking interactions within the minor groove of a rather distorted helix [Figure 2, structure **3(b)**]. The experimental data described in this report are completely consistent with this new DNA structure, in which the interacting cytosines are formally separated by two C·G base pairs. At pH ≥ 7.9, the mismatched cytosines might be pushed toward the minor groove but do not fully develop into C-C$^+$ stack pairs.

A striking feature of ss(CCG)$_{15}^+$ was the flipping of a single cytosine away from the central portion of the hairpin stem (Figure 9). In the absence of other data provided in this report, it could appear that the flipped cytosine provided evidence for a tetraplex structure, which must contain a fold in the middle of the hairpin stem. However, several lines of evidence, including those described above, suggest that the flipped cytosine in ss(CCG)$_{15}^+$ was a manifestation of the stress imposed upon the sugar−phosphate backbone by the extrahelical C-C stack pairs, and not due to a tetraplex arrangement. First, although cleavages of the G41−C42 and C42−C43 phosphodiester bonds were pH-dependent (Figures 6 and 9), significant cleavage of the G41−C42 phosphodi-

ester of ss$(CCG)_{15}$ was achieved at pH 8.5 by lowering the temperature of the reaction to 1 or 10 °C (data not shown). Minor cleavage of the G41−C42 phosphodiester was also observed at 25 °C (Figure 6). This result suggested that the central portion of the helix was significantly deformed in the absence of cytosine protonation, a result that was not consistent with a tetraplex structure. One interpretation is that reduction in temperature allowed for the extrahelical cytosines within a stack pair to move closer toward one another, thereby adding strain to the sugar−phosphate backbone. We also note that the UV absorbance melting profile conducted at pH 8.5 (Figure 1C) indicated a minor structural perturbation at ∼25 °C, perhaps indicating reorientation of the mispaired cytosines at this temperature.

Second, molecular dynamics simulations of a d$(CCG)_{11}$ hairpin have been performed starting from a conformation in an extended e-motif (R.M.R. and I.S.H., unpublished results). The simulations predict a flipping of the central cytosine away from the core of the helix. Third, evidence for a flipped cytosine was not found in the central portion of the hairpin stem of ss$(CCG)_{20}^+$ (Figure 10B), a DNA sequence predicted to form a more stable tetraplex compared to ss$(CCG)_{15}^+$. Instead, minor but significant cleavages of the phosophodiesters flanking C62, a nucleotide in triplet XIX involved in a C-C mismatch, were observed at pH 7.5 but not pH 8.5. The fact that C62 of ss$(CCG)_{20}$ was not located in the center of the helix stem suggests that nuclease cleavage around this nucleotide results from a structural deformation (such as relief of helical strain imposed by the C-C^+ mismatches) that was not a tetraplex. It is possible that the reason cleavages of the C62 phosphodiesters [in ss$(CCG)_{20}^+$] were not as pronounced as the C42 phosphodiesters [in ss$(CCG)_{15}^+$] was that the additional helical length of ss$(CCG)_{20}$ allowed for greater dissipation of the backbone strain imposed by the C-C^+ stack pairs. The amount of helical strain may be inversely related to the pK_a of the sequence. In support of this possibility, the pK_as of ss$(CCG)_{18}$ and ss$(CCG)_{20}$ were ∼8.2 and ∼8.4. To our knowledge, these are the highest values ever reported for pK_a of a cytosine.

Third, an asymmetric pattern of HA reactivity was observed at pH 8.5 (Figure 5) and 7.5 (Figure 8), such that the cytosines at the 5′-end of the ss$(CCG)_{15}$ hairpin were more reactive compared to those at the 3′-end. These results suggest that at the two pH values, the overall structural conformations were similar, a result that is not consistent with a tetraplex structure at the lower pH. Interestingly, the side of the hairpin that was more susceptible to modification by HA was also the side that contained the flipped cytosine. The nature of the structural asymmetry is not known.

*Stability of the **(b)** alignment hairpin: relevance to recognition by the human MTase. In vitro* studies have provided evidence that the human MTase is capable of methylating hairpins formed from $(CCG)_{15}$ (Smith et al., 1994) and $(GCC)_{5-7,11}$ (Chen et al., 1995). The rates of methylation of these sequences increased as the number of repeats increased. Previous ^{1}H NMR studies have shown that $(GCC)_{5-7}$ fold into **(a)** alignment hairpins (Chen et al., 1995), while the present study has shown that $(CCG)_{15,18,20}$ fold into **(b)** alignment hairpins. The base-pair alignment adopted by the $(GCC)_{11}$ sequence is not known.

It is possible that the high rate of *in vitro* methylation of d$(CCG)_{15}$ observed in the previous study might be due to one of two factors. First, d$(CCG)_{15}$ may exist in solution in a number of conformations, of which the **(b)** alignment hairpin is the predominant form. The **(b)** alignment hairpin might be ignored by the human MTase, while a minor population of **(a)** alignment hairpins might be rapidly methylated as a consequence of the increased length of the hairpin stem. In support of this possibility, it has been shown that optimal rates of methylation by the human MTase are achieved when DNA fragment sizes are at least 20 bp (Laayoun & Smith, 1995), which is the approximate helix length of d$(CCG)_{15}$. The second and more intriguing possibility is that due to the distorted nature of sugar−phosphate backbone, the **(b)** alignment hairpin may be an excellent substrate for the MTase. In support of this possibility, cytosines within extrahelical CpG dinucleotides (i.e., an extremely distorted helix) are rapidly methylated by the human MTase (Laayoun & Smith, 1995).

Potential Relationship Between **(b)** *Alignment Hairpins, Slipped Structures, and Fragile Sites.* Recently, Peason and Sinden (1996) provided evidence that reduplexing of d$(CCG)_n$•$(CGG)_n$ repeats resulted in formation of alternative secondary structures, which were best understood as slipped structures stabilized by single-stranded hairpins. These structures not only may play a role in expansion events but also may play a role in the genesis of chromosomal fragile sites, which appear as gaps in metaphase chromosomes and can be induced under a variety of culture conditions (Sutherland & Hecht, 1985). Those fragile sites induced by folic acid deprivation have been localized to d$(CGG/CCG)_n$ sequences (Ashley & Warren, 1995) methylated at CpG dinucleotides (Hornstra et al., 1993). It is not known whether methylation plays a direct role by inducing the formation of unusual DNA structures or plays an indirect role by slowing DNA replication.

Prior to the isolation of the *FMR-1* gene, Sutherland et al. (1985) hypothesized that a folate-sensitive fragile site was an amplified polypurine/polypyrimidine tract which did not package during mitosis. Failure of the DNA to package could be due to the presence of single-strand DNA gaps (Ledbetter et al., 1986) or some type of DNA structure that prevented packaging, or both. In some cases, the chromosome physically breaks at the fragile site (Sutherland, 1983), suggesting that d$(CGG/CCG)_n$ sequences might be cleaved by an endonuclease.

In this study, we have provided evidence that the hairpin structure formed by d$(CCG)_{n\geq 15}$ is the only one among class I triplet repeat sequences that contains a helix which can be cleaved by an endonuclease (Figures 6 and 9). The results of pH-dependent electrophoretic studies (Figure 1A) provided strong evidence that the structures of ss$(CCG)_{15}$ and ss$(C^mCG)_{15}$ were similar, suggesting that the helix of a methylated **(b)** alignment hairpin is also distorted. Further studies are required to determine whether folate-sensitive fragile sites result from hairpin structures in the **(b)** alignment, result from tetraplex structures formed by d$(CGG)_n$ (Fry & Loeb, 1994; Mitas et al., 1995a; Usdin & Woodford, 1995; Kettani et al., 1995), or result from other factors such as nucleosome exclusion (Wang & Griffith, 1996).

ACKNOWLEDGMENT

We thank Dr. Xiaolian Gao (University of Houston) for providing NMR coordinates of the e-motif. We thank Drs.

Steven S. Smith (City of Hope National Medical Center) and Goutam Gupta (Los Alamos National Laboratory) for helpful discussions. We thank Dr. Ulrich Melcher (Oklahoma State University) for critical review of the manuscript, Ms. Janet Rogers and Ms. SueAnn Hudiburg of the OSU Recombinant DNA/Protein Resource Facility for DNA sequence analysis and oligonucleotide preparation, and Dr. Wayne Fish (Phillips Petroleum, Bartlesville, OK) for the use of the AV/IV spectrophotometer. We are grateful to Mr. Christopher L. Clark (Molecular and Cell Biology Program, The University of Texas at Dallas) for the use of his program to obtain the first derivative of melting profiles. The program was based on the "quintic-sexic" sliding 13-point first-derivative procedure of Savitzky and Golay (1964).

REFERENCES

Akinrimisi, E. O., Sander, C., & Ts'o, P. O. P. (1963) *Biochemistry 2*, 340−344.

Antao, V. P., & Gray, D. M. (1993) *J. Biomol. Struct. Dyn. 10*, 819−839.

Ashley, J. C. T., & Warren, S. T. (1995) *Annu. Rev. Genet. 29*, 703−728.

Baker, D. J., Kan, J. L. C., & Smith, S. S. (1991) *Gene 74*, 207−210.

Brook, J. D., McCurrash, A. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J.-P., Hudson, T., Sohn, R., Zemelman, B., Snell, R. G., Rundle, S. A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P. S., Shaw, D. J., & Housman, D. E. (1992) *Cell 78*, 799−808.

Brown, T., Leonard, G. A., Booth, E. D., & Kneale, G. (1990) *J. Mol. Biol. 212*, 437−440.

Chen, X., Mariappan, S. V. S., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoun, A., Smith, S. S., Bradbury, E. M., & Gupta, G. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 5199−5203.

Fry, M., & Loeb, L. A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 4950−4954.

Gacy, A. M., Goeliner, G., Juranic, N., Macura, S., & McMurray, C. T. (1995) *Cell 81*, 533−540.

Gao, L., Huang, X., Smith, K. G., Zheng, M., & Liu, H. (1995) *J. Am. Chem. Soc. 114*, 8883−8884.

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

Gray, D. M., Ratliff, R. L., Antao, V. P., & Gray, C. W. (1988) in *Structure and Expression, Vol. 2: DNA and Its Drug complexes* (Sarma, M. H., & Sarma, R. H., Eds.) pp 147−166, Adenine Press, Guilderland, NY.

Gray, D. M., Ratliff, R. L., & Vaughan, M. R. (1992) *Meth. Enzym. 211*, 389−406.

Gray, D. M., Hung, S.-H., & Johnson, K. H. (1995) *Methods Enzymol. 246*, 19−34.

Hornstra, I. K., Nelson, D. L., Warren, S. T., & Yang, T. P. (1993) *Hum. Mol. Genet. 2*, 1659−1665.

Huntingon's Disease Collaborative Research Group. (1993) *Cell 72*, 971−983.

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

Johnston, B. H. (1992) *Methods Enzymol. 212*, 180−194.

Johnston, B. H., & Rich, A. (1985) *Cell 42*, 713−724.

Ke, S. H., & Wartell, R. M. (1993) *Nucleic Acids Res. 21*, 5137−5143.

Kettani, A., Kumar, R. A., & Patel, D. J. (1995) *J. Mol. Biol. 254*, 638−656.

Klimasauskas, S., & Roberts, R. J. (1995) *Nucleic Acids Res. 23*, 1388−1395.

Laayoun, A., & Smith, S. S. (1995) *Nucleic Acids Res. 23*, 1584−1589.

Langridge, R., & Rich, A. (1963) *Nature 198*, 726−728.

Lavelle, L., & Fresco, J. (1995) *Nucleic Acids Res. 23*, 2692−2705.

Ledbetter, D. H., Ledbetter, S. A., & Nussbaum, R. (1986) *Nature 324*, 161−163.

Liang, G., Gannett, P., Shi, X., Zhang, Y., Chen, F.-X., & Gold, B. (1994) *J. Am. Chem. Soc. 116*, 1131−1132.

Lohman, T. M., & Ferrari, M. E. (1994) *Annu. Rev. Biochem. 63*, 527−570.

Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'Hoy, K., Leblond, S., Earle-Macdonald, J., De Jong, P. J., Wieringa, B., & Korneluk, R. G. (1992) *Science 255*, 1253−1255.

Mariappan, S. V. S., Catasti, P., Chen, X., Ratcliff, R., Moyzis, R. K., Bradbury, E. M., & Gupta, G. (1996a) *Nucleic Acids Res. 24*, 784−792.

Mariappan, S. V. S., Garcia, A., & Gupta, G. (1996b) *Nucleic Acids Res. 24*, 775−783.

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499−560.

Mitas, M., Yu, A., Dill, J., & Haworth, I. S. (1995a) *Biochemistry 34*, 12803−12811.

Mitas, M., Yu, A., Dill, J., Kamp, T. J., Chambers, E. J., & Haworth, I. S. (1995b) *Nucleic Acids Res. 23*, 1050−1059.

Pearson, C. E., & Sinden, R. R. (1996) *Biochemistry 35*, 5041−5053.

Petruska, J., Arnheim, N., & Goodman, M. F. (1996) *Nucleic Acids Res. 24*, 1992−1998.

Roberts, R. J. (1995) *Cell 82*, 9−12.

Rubin, C. M., & Schmid, C. W. (1980) *Nucleic Acids Res. 8*, 4613−4619.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning*, 2nd ed., Cold Spring Harbor Press, Cold Spring Harbor, NY.

Savitzky, A., & Golay, M. J. E. (1964) *Anal. Chem. 36*, 1627−1639.

Schlötterer, C., & Tautz, D. (1992) *Nucleic Acids Res. 20*, 211−215.

Singer, B., & Grunberger, D. (1983) *Molecular biology of mutagens and carcinogens*, Plenum Press, New York.

Smith, G. K., Jie, J., Fox, G. E., & Gao, X. (1995) *Nucleic Acids Res. 23*, 4303−4311.

Smith, S. S., Hardy, T. A., & Baker, D. J. (1987) *Nucleic Acids Res. 15*, 6899−6917.

Smith, S. S., Kan, J. L. C., Baker, D. J., Kaplan, B. E., & Dembek, P. (1991) *J. Mol. Biol. 217*, 39−51.

Smith, S. S., Lingeman, R. G., & Kaplan, B. E. (1992) *Biochemistry 31*, 850−854.

Smith, S. S., Layoun, A., Lingeman, R. G., Baker, D. J., & Riley, J. (1994) *J. Mol. Biol. 243*, 143−151.

Sutherland, G. R., & Hecht, F. (1985) *Fragile sites on human chromosomes*, Vol. 13, Oxford University Press, New York.

Sutherland, G. R., Baker, E., & Fratini, A. (1985) *Am. J. Med. Genet. 22*, 110−122.

Usdin, K., & Woodford, K. J. (1995) *Nucleic Acids Res. 23*, 4202−4209.

Ververk, A. J. J. H., Pieretti, M., Sutcliffe, J. S., Fu, Y.-H., Kuhl, D. P. A., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F., Eussen, B. E., van Ommen, G.-J. B., Blonden, L. A. J., Riggins, G. J., Chastain, J. L., Kunst, C. B., Galjaard, H., Caskey, C. T., Nelson, D. L., Oostra, B. A., & Warren, S. T. (1991) *Cell 65*, 905−914.

Wang, Y.-H., & Griffith, J. (1996) *J. Biol. Chem. 271*, 22937−22940.

Wartell, R. M., Hosseini, S. H., & Moran, C. P. J. (1990) *Nucleic Acids Res. 18*, 2699−2705.

Wohlrab, F. (1992) *Methods Enzymol. 212B*, 294−301.

Yu, A., Dill, J., & Mitas, M. (1995a) *Nucleic Acids Res. 23*, 4055−4057.

Yu, A., Dill, J., Wirth, S. S., Huang, G., Lee, V. H., Haworth, I. S., & Mitas, M. (1995b) *Nucleic Acids Res. 23*, 2706−2714.

BI9625410