6826 *Biochemistry* **2010,** *49,* 6826–6837
DOI: 10.1021/bi1007782

BIOCHEMISTRY
including biophysical chemistry & molecular biology
*Article*

# AGG Interruptions in $(CGG)_n$ DNA Repeat Tracts Modulate the Structure and Thermodynamics of Non-B Conformations in Vitro†

Daniel A. Jarem, Lauren V. Huckaby, and Sarah Delaney*

*Department of Chemistry, Brown University, Providence, Rhode Island 02912*

*Received May 17, 2010; Revised Manuscript Received July 14, 2010*

ABSTRACT: The trinucleotide repeat sequence CGG/CCG is known to expand in the human genome. This expansion is the primary pathogenic signature of fragile X syndrome, which is the most common form of inherited mental retardation. It has been proposed that formation of non-B conformations by the repetitive sequence contributes to the expansion mechanism. It is also known that the CGG/CCG repeat sequence of healthy individuals, which is not prone to expansion, contains AGG/CCT interruptions every 8−11 CGG/CCG repeats. Using DNA containing 19 or 39 CGG repeats, we have found that both the position and number of interruptions modulate the non-B conformation adopted by the repeat sequence. Analysis by chemical probes revealed larger loops and the presence of bulges for sequences containing interruptions. Additionally, using optical analysis and calorimetry, the effect of these structural changes on the thermodynamic stability of the conformation has been quantified. Notably, changing even one nucleotide, as occurs when CGG is replaced with an AGG interruption, causes a measurable decrease in the stability of the conformation adopted by the repeat sequence. These results provide insight into the role interruptions may play in preventing expansion in vivo and also contribute to our understanding of the relationship between non-B conformations and trinucleotide repeat expansion.

Microsatellites are regions of DNA in which simple sequences of one to six nucleotides are repeated multiple times. Microsatellites comprise ~3% of the human genome and have been shown to have high mutability, leading to both sequence and length polymorphisms (*1*−*4*). Trinucleotide repeat (TNR)[1] sequences make up a class of microsatellites that are generally considered to impact phenotype and have been linked to several genetic diseases (*5*, *6*). TNR sequences have been shown to expand in the human genome, and the proposed mechanisms for the expansion include polymerase slippage during replication or during a DNA repair event (*5*−*12*). In vitro primer extension experiments have shown that an oligonucleotide containing five CGG repeats is expanded to more than 80 repeats when replicated by a purified bacterial or mammalian polymerase (*13*). Formation of non-B conformations by repetitive DNA is also thought to play a critical role in the expansion (*14*). Indeed, repetitive sequences have been shown to adopt conformations such as stem−loop hairpins, quadruplexes, triplexes, and sticky DNA in vitro (*14*−*29*). Recent studies have also linked expanded regions of triplet repeats with epigenetic modifications that result in pathological consequences (*30*, *31*).

One disorder caused by the expansion of a CGG/CCG TNR sequence, fragile X syndrome, is the most common form of inherited mental retardation (*32*). In fragile X syndrome, a CGG/CCG TNR in the 5′-untranslated region of exon 1 of the FMR1 gene expands beyond the healthy length (*33*−*35*). In healthy individuals, the number of repeats varies but falls within the range of 5−55. Furthermore, in healthy individuals, these repeats are interrupted by an AGG/CCT unit every 8−11 repeats (*36*). Repeat lengths within this healthy range that possess AGG/CCT interruptions are stable and are not prone to expansion (*34*).

Repeat lengths of 55−200 are considered to be within a premutation range and are susceptible to expansion over a single familial generation. In fact, a CGG/CCG tract of ≥90 repeats has a nearly 100% risk of expansion to the disease state in a single generation (*37*). The disease state is classified as >200 CGG/CCG repeats. When the number of repeats exceeds this threshold length, affected individuals show hypermethylation of the repeat tract and the FMR1 gene is transcriptionally silenced (*31*, *34*, *38*). Although the function of the FMR1 protein is not fully understood, it is the loss of the protein product that is responsible for the fragile X phenotype.

The importance of AGG/CCT interruptions is emphasized by the observation that an uninterrupted 59-repeat sequence, which would have only a low propensity to expand if interruptions were present, expanded to more than 200 repeats in one generation (*39*). While AGG/CCT interruptions are thought to play a critical role in suppressing the expansion of CGG/CCG repeat sequences, the mechanism by which these interruptions act remains unclear. Therefore, it is important not only to define the conformation adopted by these sequences but also to delineate how AGG interruptions influence the identity and stability of these conformations.

Here, we elucidate the conformation of CGG DNA repeat tracts in vitro both with and without AGG interruptions. We

†This work was supported by Brown University.

*To whom correspondence should be addressed. Telephone: (401) 863-3590. Fax: (401) 863-9368. E-mail: sarah_delaney@brown.edu.

[1]Abbreviations: CD, circular dichroism; DEPC, diethyl pyrocarbonate; DMS, dimethyl sulfate; DMT, dimethoxytrityl; DSC, differential scanning calorimetry; EDTA, ethylenediaminetetraacetic acid; FMR1, fragile X mental retardation 1; PAGE, polyacrylamide gel electrophoresis; TBE, Tris-borate-EDTA; $T_m$, melting temperature; TEAA, triethylammonium acetate; TNR, trinucleotide repeat; Tris, tris(hydroxymethyl)aminomethane.

 Published on Web 07/15/2010 

establish that these interruptions both alter and destabilize the conformation adopted by the repeat DNA. Furthermore, using optical analysis and calorimetry, we provide a quantitative measure of the contribution of AGG interruptions in modulating the stability of non-B DNA conformations.

## EXPERIMENTAL PROCEDURES

*Oligonucleotide Synthesis and Purification*. Oligonucleotides were synthesized using standard phosphoramidite chemistry (*40*) on a BioAutomation DNA/RNA synthesizer. Upon completion of the synthesis, the 5′-dimethoxytrityl (DMT) group was retained to facilitate purification. HPLC purification of these oligonucleotides was performed on a styrene-divinyl benzene reverse phase column (PLRP-S; Polymer Laboratories) (4.6 mm × 250 mm) at 90 °C using 100 mM TEAA in an acetonitrile/water mixture (99:1) (solvent A) and 100 mM TEAA in an acetonitrile/water mixture (1:99) (solvent B) as the mobile phases (gradient, solvent A increased from 5 to 25% over 25 min at a rate of 1.0 mL/min). Following removal of the DMT group by incubation in 80% glacial acetic acid for 12 min at room temperature, the oligonucleotides were subjected to a second round of HPLC purification (gradient, solvent A increased from 0 to 15% over 35 min at a rate of 1.0 mL/min).

*DEPC Chemical Probe Analysis*. Oligonucleotides were 5′-$^{32}$P end-labeled using T4 polynucleotide kinase following the manufacturer's protocol. To obtain the thermodynamically favored DNA conformations, oligonucleotides (1 $\mu$M) in 10 mM sodium phosphate and 100 mM KCl (pH 7.5) were incubated for 5 min at 95 °C and cooled over ~2.5 h to room temperature. Oligonucleotides were then incubated with 5% diethyl pyrocarbonate (DEPC) (v/v) (20 $\mu$L final sample volume) for 30 min at 37 °C. Following incubation with DEPC, the samples were dried in vacuo, treated with 10% piperidine (v/v) for 30 min at 90 °C, and again dried in vacuo. Samples were resuspended in denaturing loading buffer (80% formamide, 0.1% xylene cyanol, and 0.1% bromophenol blue), incubated for 3 min at 95 °C, and electrophoresed through a 14% [for $(CGG)_{19}$ series] or 10% [for $(CGG)_{39}$ series] denaturing polyacrylamide gel. The products were visualized by phosphorimagery.

*DMS Methylation Protection Assay*. Oligonucleotides were 5′-$^{32}$P end-labeled using T4 polynucleotide kinase following the manufacturer's protocol. Two oligonucleotide samples (375 $\mu$L, 0.2 $\mu$M) were incubated for 5 min at 95 °C and cooled over ~2.5 h to room temperature; one sample was in 10 mM sodium phosphate and 100 mM KCl (pH 7.5), and the other was in 10 mM sodium phosphate (pH 7.5). To each sample was added 1 $\mu$L of a freshly prepared 1:4 DMS/ethanol mixture, and 75 $\mu$L aliquots were removed after 0, 5, 15, 30, and 45 min and quenched with 20 $\mu$L of DMS stop solution [1.5 M sodium acetate, 1 M $\beta$-mercaptoethanol, and 250 $\mu$g/mL yeast tRNA (pH 7.0)]. The samples were then twice precipitated with ethanol, dried in vacuo, treated with 10% piperidine (v/v) for 30 min at 90 °C, and again dried in vacuo. Samples were resuspended in denaturing loading buffer, incubated for 3 min at 95 °C, and electrophoresed through a 14% denaturing polyacrylamide gel. The products were visualized by phosphorimagery.

*Native Polyacrylamide Gel Electrophoresis*. DNA samples (1 $\mu$M unless otherwise noted) in 10 mM sodium phosphate and 100 mM KCl (pH 7.5) were heated to 95 °C and either cooled over ~2.5 h to room temperature or flash-cooled in an ice bath. Samples were diluted with nondenaturing loading buffer (15% ficoll, 0.25% xylene cyanol, and 0.25% bromophenol blue) or denaturing loading buffer and electrophoresed through a 12% native polyacrylamide gel at 450 V (10 V/cm) at 4 °C. The products were visualized by phosphorimagery.

*Optical Melting Analysis*. Quantification of oligonucleotides was performed at 95 °C using the $\varepsilon_{260}$ values estimated for single-stranded DNA (*41*) and a Beckman Coulter DU800 UV−visible spectrophotometer equipped with a Peltier thermoelectric device. DNA samples had a final concentration of 0.5−2.0 $\mu$M, and optical melting analysis was performed in 10 mM sodium phosphate and 100 mM KCl (pH 7.5). Prior to analysis, samples were incubated for 5 min at 95 °C and cooled to room temperature over ~2.5 h. The samples were then heated at a rate of 1 °C/min from 25 to 95 °C while the absorbance was monitored at 260 nm, held at 95 °C for 5 min, and returned to the starting temperature at a rate of 1 °C/min. The first derivative of the absorbance versus temperature data was obtained and the $T_m$ taken as the maximum in the first-derivative plot. Thermodynamic parameters were extracted from melting profiles using van't Hoff analysis (*42*). A Student's *t* test was used to determine if the average values were statistically different.

*Circular Dichroism*. Circular dichroism spectra were recorded at 37 °C with a Jasco J-815 CD spectropolarimeter equipped with a Peltier thermoelectric device. DNA samples were at a final concentration of 0.5−2.0 $\mu$M in 10 mM sodium phosphate and 100 mM KCl (pH 7.5), incubated for 5 min at 95 °C, and cooled over ~2.5 h to room temperature prior to analysis. Samples were equilibrated at 37 °C for 10 min and then scanned from 320 to 220 nm at a rate of 50 nm/min. All reported spectra represent an average of three scans.

*DSC Analysis*. Microcalorimetry was performed using a MicroCal VP-DSC instrument. Oligonucleotides (in 0.8 mL) were prepared in 10 mM sodium phosphate and 100 mM KCl (pH 7.5), incubated for 5 min at 95 °C, and cooled over ~2.5 h to room temperature. All samples were degassed in vacuo for 12 min at 25 °C prior to analysis by DSC. Data were obtained by continuously monitoring the excess power required to maintain both the sample and the reference cells at the same temperature. The resulting thermograms display excess heat capacity as a function of temperature. Each experiment consisted of a forward scan, in which the temperature was increased from 10 to 95 °C (1.5 °C/min), and a reverse scan, in which the temperature was decreased from 95 to 10 °C (1.0 °C/min). The sample equilibrated for 10 min at 10 and 95 °C between each forward and reverse scan. A total of 10 thermograms were obtained for each DNA sequence. A buffer reference was analyzed using the same procedure described above, and the thermograms were corrected using this background. The thermograms were normalized for concentration, and baseline correction was performed using a systematic linear fit. This type of baseline correction assumes a $\Delta C_p$ value of 0, but with the lack of a baseline at the upper limit (which is a result of the high melting temperatures), it is the only objective way to baseline correct. Nevertheless, to evaluate the influence that a non-zero $\Delta C_p$ would have on the reported $\Delta H$, we assumed the maximum possible $\Delta C_p$ by setting an upper baseline at the last data point obtained at 95 °C. Indeed, a non-zero $\Delta C_p$ decreases the value of $\Delta H$, but only by ~15%.

## RESULTS

*Design of $(CGG)_{19}$ DNA Series*. Four DNA sequences were designed to determine the effect of AGG interruptions on

Table 1: DNA Sequences Used in This Study

| name | nucleotide sequence |
|---|---|
| $(CGG)_{19}$ | 5′-$(CGG)_{19}$-3′ |
| 1AGG-a | 5′-$(CGG)_4$AGG$(CGG)_{14}$-3′ |
| 1AGG-b | 5′-$(CGG)_8$AGG$(CGG)_{10}$-3′ |
| 2AGG | 5′-$(CGG)_4$AGG$(CGG)_9$AGG$(CGG)_4$-3′ |
| $(CGG)_{39}$ | 5′-$(CGG)_{39}$-3′ |
| 4AGG | 5′-$(CGG)_4$AGG$(CGG)_9$AGG$(CGG)_9$ AGG$(CGG)_9$AGG$(CGG)_4$-3′ |

the conformation of a tract of CGG repeats (Table 1). The first sequence in the series is a $(CGG)_{19}$ tract with no AGG interruptions. The second and third sequences each possess a single AGG interruption. In the sequence named 1AGG-a, the interruption replaces the fifth CGG repeat. In the sequence named 1AGG-b, the AGG interruption is more centrally located in the sequence and replaces the ninth CGG repeat. The fourth sequence (2AGG) contains two AGG interruptions, which are located at the fifth and 15th repeats.

*Structural Analysis of the* $(CGG)_{19}$ *Series by Reactivity toward Chemical Probes*. To characterize the conformation(s) formed in solution by the series of $(CGG)_{19}$ sequences, we modified the sequences with diethyl pyrocarbonate (DEPC). DEPC is used commonly as a probe of nucleobase accessibility, as it selectively modifies unpaired purines (A ≫ G) (*43*, *44*). This selectivity is due to the increased solution accessibility of the N-7 position of unpaired purines. While DEPC does not modify purines in well-matched base pairs, it has been shown to be effective in the identification of adenines and guanines in DNA bulges and hairpin loops (*12*, *45*). As seen in Figure 1A, the $(CGG)_{19}$ sequence shows a single region of reactivity toward DEPC, namely, three highly reactive guanines and two guanines with low levels of reactivity. This specific pattern of modification by DEPC suggests that the $(CGG)_{19}$ sequence adopts a stem−loop structure in which the loop contains four bases and the stem consists of G-C base pairs and G·G mismatches. The lower level of modification of the 5′ G in the ninth CGG repeat, relative to the neighboring 3′-G, led us to assign this G as being part of the loop closing base pair instead of being part of the loop. It is possible, however, that the stem−loop structure contains a loop of six bases. Nevertheless, and regardless of whether the loop contains four or six bases, as shown in Figure 1A, the position of the loop dictates that a single G overhang the 3′-end of the stem−loop structure.

Structural characterization by DEPC of the sequences containing a single AGG interruption reveals that the overall conformation of these species is altered relative to $(CGG)_{19}$. The 1AGG-a sequence displays three regions of reactivity toward DEPC in contrast to the one region observed for $(CGG)_{19}$ (Figure 1B). One region of reactivity in 1AGG-a corresponds to a loop. Indeed, the same loop size and location observed for $(CGG)_{19}$ are also observed for 1AGG-a. The other two regions of reactivity correspond to two bulges on either the 5′- or 3′-arm of the stem. These bulges are staggered with respect to one another, and in fact, the AGG interruption is contained within one of these bulges. For 1AGG-b, having the AGG interruption in place of the ninth repeat generates a dramatically different DEPC reactivity profile (Figure 1C). For this sequence, similar to $(CGG)_{19}$, only one region of modification by DEPC is observed. However, in contrast to $(CGG)_{19}$, the reactivity pattern is consistent with a seven-base loop. In addition to the same four bases observed in the loop of $(CGG)_{19}$ and 1AGG-a, the loop of 1AGG-b also contains the AGG interruption. The position of this loop results in a four-base overhang at the 3′-end of the structure. Indeed, although not well-resolved from the unmodified DNA substrate, increased reactivity toward DEPC at the 3′-end of the sequence is observed relative to untreated controls as would be expected for a 3′-overhang.

Lastly, the sequence containing two AGG interruptions was characterized on the basis of its modification by DEPC (Figure 1D). Three regions of reactivity are observed. This reactivity pattern can be described by two different conformations. The first possibility is a stem−loop structure containing a four-base loop and two bulges that are positioned across from one another. This structure would also include an overhang of three bases on the 3′-end. Conversely, the reaction pattern could also describe a boomerang-like structure that consists of two joined stem−loop structures.

It has been widely documented in the literature that G-rich sequences can form quadruplexes. To determine if the reactivity patterns described for the $(CGG)_{19}$ series could be attributed to the formation of an intra- or intermolecular quadruplex, we used a dimethyl sulfate (DMS) methylation protection assay. The unique structure of guanine quadruplexes consists of stacked tetrads where each tetrad is a planar array of four Hoogsteen-bonded guanines (*46*). The formation of guanine quadruplexes is stabilized by monovalent cations (e.g., $K^+$ and $Na^+$) positioned in the center of the structure and coordinated by the electron-rich carbonyl oxygens (*47*, *48*). Quadruplexes can form in an intramolecular fashion from a single strand, from two DNA hairpins, or from four individual strands. DMS methylates N-7 of G; however, in a quadruplex, N-7 is involved in hydrogen bonding and cannot be methylated by DMS. Thus, while guanines in duplex or single-stranded regions of DNA (including bulges, loops, and overhangs) are modified by DMS, guanines in a quadruplex are protected from modification. A control experiment performed with a sequence that has been shown previously to form an intramolecular quadruplex (*49*) illustrates this concept (Supporting Information). In the presence of 100 mM KCl, the control sequence displays protection of guanines from methylation, whereas in the absence of KCl, where the quadruplex structures cannot form, no protection is observed. Conversely, the $(CGG)_{19}$ series shows no such salt-derived protection effects, and thus, quadruplexes are not among our proposed conformations. It is of note that, consistent with the results obtained using the DEPC chemical probe, there are increased levels of methylation at the guanines proposed to be in the loop region.

*Electrophoretic Mobility of* $(CGG)_{19}$ *by Native PAGE*. To improve our understanding of whether the conformations adopted by these repetitive sequences are intra- or intermolecular, $(CGG)_{19}$ was analyzed by native PAGE (Figure 2). As controls for migration of an unstructured single strand and duplex, a 57-mer with mixed sequence and $(CGG)_{19}/(CCG)_{19}$ duplex were used, respectively. When analyzed by native PAGE, each of these control samples migrates as a single species with the unstructured single strand migrating further through the gel matrix than the duplex. Separate samples of the single-stranded control were analyzed, using nondenaturing or denaturing loading buffer, to ensure that it was unstructured.

The $(CGG)_{19}$ sequence was prepared for electrophoresis in three different ways: (1) no preparation (purified oligonucleotide loaded onto the gel) (Figure 2, lane 6), (2) heated to 95 °C followed by slow cooling (Figure 2, lane 4), and (3) heated to 95 °C followed by flash cooling (Figure 2, lane 5). When analyzed

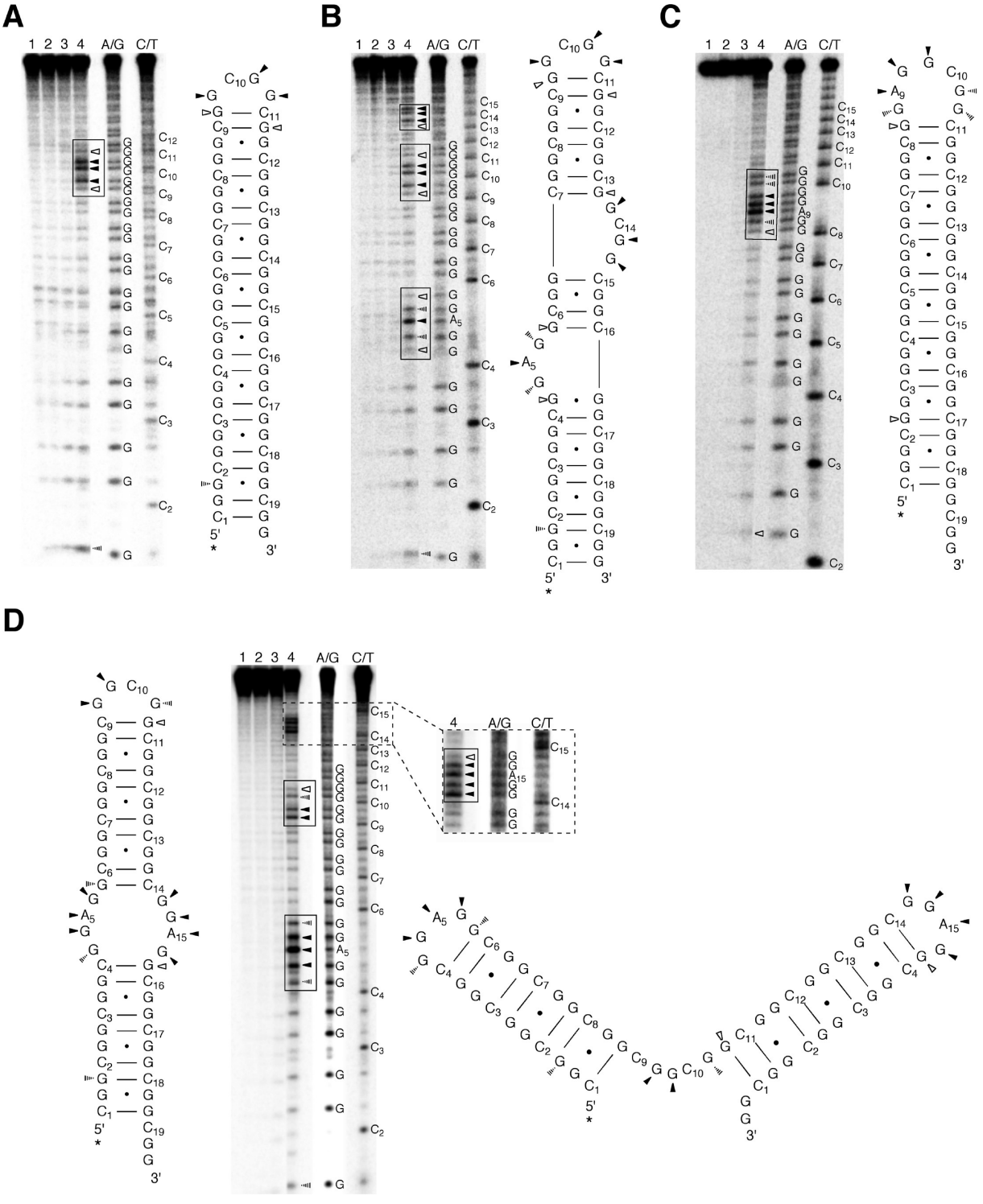


FIGURE 1: Characterization of the $(CGG)_{19}$ series using the chemical probe DEPC. Autoradiograms and schematic representations revealing strand cleavage for (A) $(CGG)_{19}$, (B) 1AGG-a, (C) 1AGG-b, and (D) 2AGG are shown. Conditions: 1 $\mu$M DNA in 10 mM sodium phosphate and 100 mM KCl (pH 7.5). Lanes 1−3 contained DNA alone, DNA cycled through heating methods, and piperidine-treated DNA, respectively. Lane 4 contained DNA incubated for 30 min at 37 °C in the presence of 5% DEPC followed by piperidine treatment. A/G and C/T are Maxam/Gilbert sequencing reactions. The pattern of the arrow at a given site reflects the relative amount of strand cleavage, with the filled arrow being the most reactive, the empty arrow being least reactive, and the striped arrow indicating an intermediate amount of reactivity. The asterisk denotes the location of the $^{32}$P radiolabel.

by native PAGE, the latter two samples migrate as a single species with the same electrophoretic mobility. Notably, this species migrates differently from the unstructured single-stranded and duplex controls. In the absence of any prior sample preparation, the majority of the $(CGG)_{19}$ sample migrates as a single band similar to that observed with the samples that were heated and cooled prior to analysis, but 3% of the sample migrates like the duplex control (Figure 2, lane 6). When the concentration of $(CGG)_{19}$ was 100-fold greater, the amount of this species with a migration similar to the duplex control increased (Figure 2, lane 7).

*Characterization of* $(CGG)_{19}$ *Series by Optical Analysis.* The four sequences were characterized by UV−visible

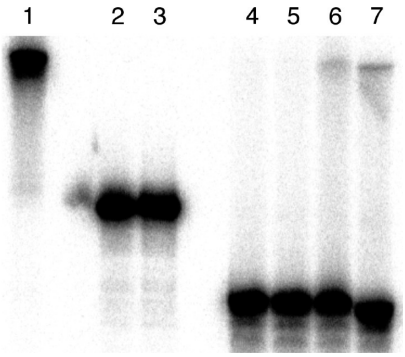


FIGURE 2: Autoradiogram revealing the electrophoretic mobility of DNA through a native polyacrylamide gel. All samples are 1 $\mu$M DNA in 10 mM sodium phosphate and 100 mM KCl (pH 7.5) unless otherwise noted. Lane 1 contained $(CGG)_{19}/(CCG)_{19}$ duplex, lane 2 an unstructured 57-mer single strand in nondenaturing loading buffer, lane 3 the unstructured 57-mer single strand in denaturing loading buffer, lane 4 $(CGG)_{19}$ heated to 95 °C and cooled slowly to room temperature, lane 5 $(CGG)_{19}$ heated to 95 °C and flash-cooled on ice, lane 6 $(CGG)_{19}$ with no treatment prior to loading, and lane 7 100 $\mu$M $(CGG)_{19}$ heated to 95 °C and cooled slowly to room temperature.

spectrophotometry to obtain optical melting profiles. In these optical melting profiles, the absorbance of the DNA is monitored as a function of temperature. All four sequences display a single, sharp transition in which an increase in absorbance is observed at a given temperature (Figure 3A and Supporting Information). Melting temperatures ($T_m$) were determined for the conformation(s) adopted by each sequence and are provided in Table 2. The $T_m$ values obtained for the sequences containing a single AGG interruption are decreased by ~2 °C compared to that of the control sequence, which lacks interruptions. Furthermore, the introduction of a second interruption in 2AGG causes a decrease in the $T_m$ of ~5 °C. A small amount of hysteresis between the melting and annealing profiles was observed for all of the sequences. This hysteresis is observed in the baseline and may occur because the heating and cooling curves are not in thermodynamic equilibrium, implying that the temperature change is faster than the rate at which the conformations relax to a final equilibrium (*50*). Lastly, there is no change in the $T_m$ of $(CGG)_{19}$ over a 10-fold range of concentration (Supporting Information).

Using the profiles generated by optical melting, thermodynamic parameters describing the transition from the structured to the unstructured sequence were obtained by van't Hoff analysis as described by Marky and Breslauer (*42*) (Table 2). Because the sequences are going from structured to unstructured, heat must transfer into the system for melting to occur, resulting in positive thermodynamic parameters. On the basis of the differences between the thermodynamic parameters ($\Delta\Delta H$, $\Delta\Delta G$, and $\Delta\Delta S$) for $(CGG)_{19}$ and the interruption sequences, we are unable to distinguish the $(CGG)_{19}$, 1AGG-a, and 1AGG-b sequences. However, with the introduction of a second AGG interruption, $\Delta\Delta H$, $\Delta\Delta G$, and $\Delta\Delta S$ are negative (Table 2).

Circular dichroism was also employed to characterize the $(CGG)_{19}$ series of sequences. The spectrum obtained for each $(CGG)_{19}$ sequence shows maxima at 240, ~275, and 300 nm and minima at 225, 254, and 290 nm (Figure 4A). With the introduction of AGG interruptions into the sequences, the amplitude of the maximum at ~275 nm increases and the amplitudes for the minima at 225 and 254 nm decrease. For comparison, in Figure 4B are spectra for the $(CGG)_{19}/(CCG)_{19}$ duplex, an unstructured 57-mer single strand, and the quadruplex adopted by *Tetrahymena* telomeric DNA (*51*, *52*).

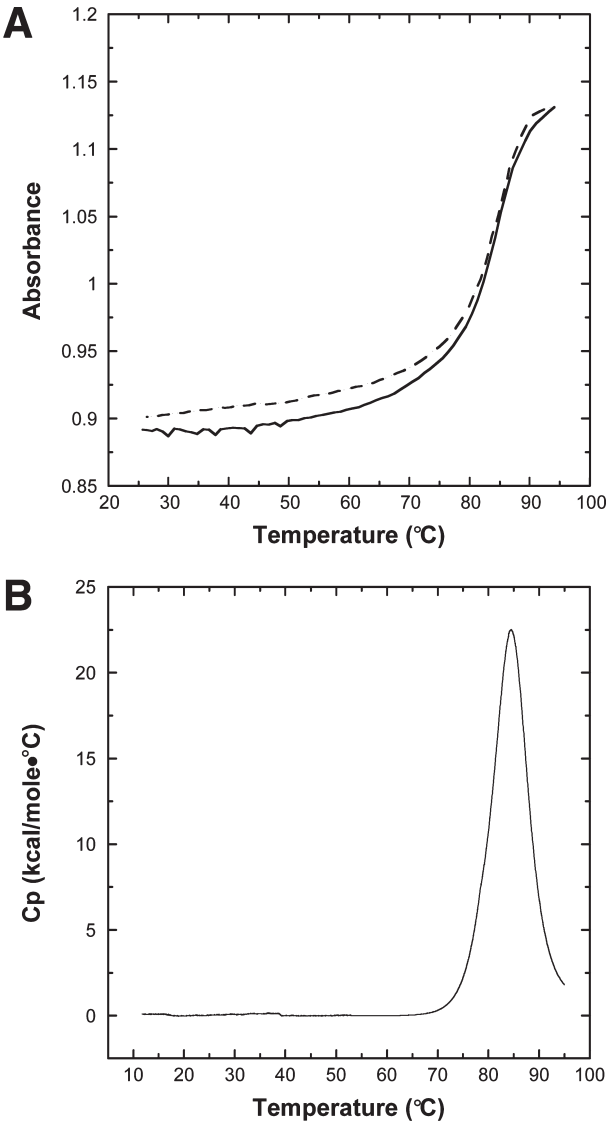


FIGURE 3: (A) Optical and (B) calorimetric analysis of $(CGG)_{19}$ at 1.9 and 72 $\mu$M, respectively, in 10 mM sodium phosphate and 100 mM KCl (pH 7.5). For optical data, the solid line represents data obtained while the sample is heated from 25 to 95 °C and the dotted line represents data obtained while the sample is cooled from 95 to 25 °C.

*Calorimetric Analysis of* $(CGG)_{19}$ *Series*. We also employed differential scanning calorimetry (DSC) as a means of characterizing the transition from the structured to the unstructured sequence. DSC allows for direct measurement of the heat supplied to or released from a system during a melting transition (*42*, *53*). Thermograms obtained by DSC, in which excess heat capacity is plotted as a function of temperature, reveal single, sharp transitions for the $(CGG)_{19}$ series of sequences (Figure 3B and Supporting Information). Melting temperatures obtained from these thermograms are provided in Table 3. The melting temperatures obtained by DSC are similar to those obtained by optical analysis. Indeed, one interruption, in the context of 1AGG-a or 1AGG-b, lowers the $T_m$ of the structure ~1 °C relative to that of $(CGG)_{19}$. Similar to what was observed by UV−visible analysis, the $T_m$ of 2AGG is ~4 °C lower than that of the control sequence that lacks interruptions.

Thermodynamic parameters were calculated directly from the DSC thermograms. Notably, for all four sequences, the values

Table 2: UV−Visible-Derived Thermodynamic Parameters for $(CGG)_{19}$ Repeat Series

| substrate | $T_m$[a] (°C) | $\Delta H$[a,b] (kcal/mol) | $\Delta G$[a,c] (kcal/mol) | $\Delta S$[a,b] (cal mol$^{-1}$ K$^{-1}$) | $\Delta T_m$[a] (°C) | $\Delta\Delta H$ (kcal/mol) | $\Delta\Delta G$ (kcal/mol) | $\Delta\Delta S$ (cal mol$^{-1}$ K$^{-1}$) |
|---|---|---|---|---|---|---|---|---|
| $(CGG)_{19}$ | 84.9 ± 0.6 | 77.7 ± 3.2 | 10.4 ± 0.5 | 217 ± 9 | – | – | – | – |
| 1AGG-a | 83.0 ± 0.6 | 76.3 ± 6.2 | 9.9 ± 0.9 | 214 ± 17 | −1.9 | N/A[d] | N/A[d] | N/A[d] |
| 1AGG-b | 83.1 ± 0.7 | 82.0 ± 2.5 | 10.6 ± 0.4 | 230 ± 7 | −1.8 | N/A[d] | N/A[d] | N/A[d] |
| 2AGG | 80.1 ± 0.7 | 64.7 ± 1.5 | 7.9 ± 0.3 | 183 ± 4 | −4.8 | −13.0 | −2.5 | −34 |

[a]DNA in 10 mM sodium phosphate and 100 mM KCl (pH 7.5). The error represents the standard deviation from a minimum of three experiments. [b]Values derived from van't Hoff analysis as described by Marky and Breslauer (*35*). [c]Values at 37 °C. [d]Not statistically different from the $(CGG)_{19}$ value as determined by a Student's *t* test.

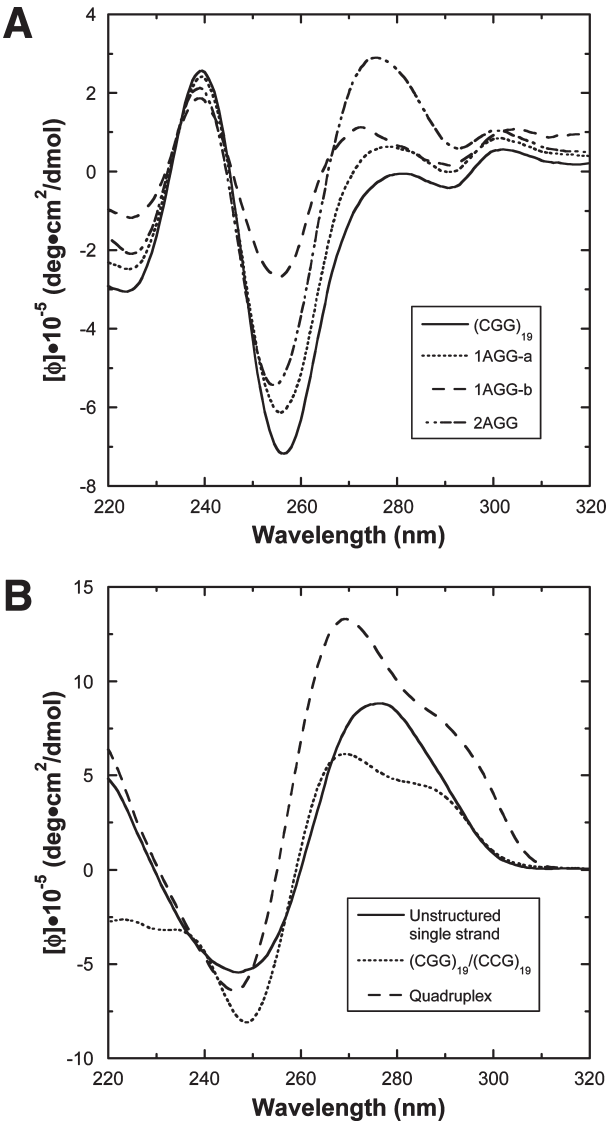


FIGURE 4: Circular dichroism spectra for (A) $(CGG)_{19}$ series and (B) control sequences at 37 °C in 10 mM sodium phosphate and 100 mM KCl (pH 7.5). Each spectrum represents an average of three experiments.

obtained by DSC for $\Delta H$, $\Delta G$, and $\Delta S$ are greater in magnitude than those obtained indirectly from the optical melting profiles. However, as was observed by optical analysis, the trend of $\Delta\Delta H$, $\Delta\Delta G$, and $\Delta\Delta S$ becoming more negative upon the addition of AGG interruptions is upheld.

*$(CGG)_{39}$ Series: Characterization of Longer Repeat Tracts*. To determine if the effect of AGG interruptions observed for the $(CGG)_{19}$ series is upheld as the repeat length increases, two additional sequences were prepared (Table 1). The first sequence is a $(CGG)_{39}$ tract with no interruptions. The second sequence contains four AGG interruptions within the $(CGG)_{39}$ tract (4AGG). The first interruption replaces the fifth CGG repeat, and the remaining interruptions are separated by nine CGG repeats. These longer sequences were analyzed by UV−visible spectrophotometry, CD, and modification with DEPC.

When probed by reactivity toward DEPC, similar to $(CGG)_{19}$, $(CGG)_{39}$ displays only a single region of reactivity (Figure 5A). This sequence also adopts a stem−loop structure. The reactivity is highly concentrated at three guanines, suggestive of a four-base loop. Consequently, as described for $(CGG)_{19}$, a single G will overhang at the 3′-end.

With a structure different from that of $(CGG)_{39}$, the presence of four interruptions in 4AGG leads to five areas of reactivity (Figure 5B). This pattern of reactivity could correspond to one loop and four bulges. The bulges are present as pairs, with each pair having one bulge placed directly opposite its partner. However, this is only one example of the possible conformations that this sequence can form, as DNA modeling programs also predict a number of potential boomerang and hairpin- and bulge-containing structures that are consistent with the same reaction pattern and would have only slightly lower stabilities.

When analyzed by circular dichroism, the $(CGG)_{39}$ series shows the same general spectral profiles as the shorter repeat series (Supporting Information). However, the amplitudes of the maxima and minima are much greater for the $(CGG)_{39}$ series. Moreover, as seen for the $(CGG)_{19}$ sequences, the amplitude of the maximum at 277 nm increases and the amplitude of the minimum at 254 nm decreases upon introduction of the AGG interruptions.

The optical melting profiles each reveal a single, sharp transition similar to those observed with the shorter series of $(CGG)_{19}$ sequences (Supporting Information). The melting temperatures for $(CGG)_{39}$ and 4AGG are reported in Table 4. Interestingly, despite the fact that the sequences in the $(CGG)_{39}$ series are approximately twice as long as those in the $(CGG)_{19}$ series, the melting temperatures are only a few degrees higher than those observed for their shorter counterparts. For $(CGG)_{39}$ versus the 4AGG sequence, the presence of four interruptions lowers the $T_m$ by ~2 °C. Furthermore, as observed for the shorter sequences, with the addition of interruptions the values for $\Delta\Delta H$, $\Delta\Delta G$, and $\Delta\Delta S$ are negative.

## DISCUSSION

In this work, we have determined that $(CGG)_{19}$ forms a stem−loop structure that possesses considerable hydrogen bond and base stacking interactions despite the presence of G·G mismatches. Previously, $(CGG)_n$ ($n$ = 4−12, 14−16, 18, 20, and 25)

Table 3: DSC-Derived Thermodynamic Parameters for $(CGG)_{19}$ Repeat Series

| substrate | $T_m$[a] (°C) | $\Delta H$[a,b] (kcal/mol) | $\Delta G$[a,c] (kcal/mol) | $\Delta S$[a,b] (cal mol$^{-1}$ K$^{-1}$) | $\Delta T_m$[a] (°C) | $\Delta\Delta H$ (kcal/mol) | $\Delta\Delta G$ (kcal/mol) | $\Delta\Delta S$ (cal mol$^{-1}$ K$^{-1}$) |
|---|---|---|---|---|---|---|---|---|
| $(CGG)_{19}$ | 84.4 ± 0.1 | 234 ± 4 | 29.4 ± 0.5 | 661 ± 11 | – | – | – | – |
| 1AGG-a | 82.9 ± 0.2 | 227 ± 4 | 27.8 ± 0.5 | 644 ± 10 | −1.5 | −7 | −1.6 | −17 |
| 1AGG-b | 83.5 ± 0.3 | 222 ± 6 | 27.2 ± 0.6 | 628 ± 18 | −0.9 | −12 | −2.2 | −33 |
| 2AGG | 80.3 ± 0.2 | 206 ± 5 | 24.6 ± 0.5 | 584 ± 13 | −4.1 | −28 | −4.8 | −77 |

[a]DNA in 10 mM sodium phosphate and 100 mM KCl (pH 7.5). The error represents the standard deviation from a minimum of three experiments. [b]Values derived directly by integration of the excess heat capacity curve (*35*). [c]Values at 37 °C.

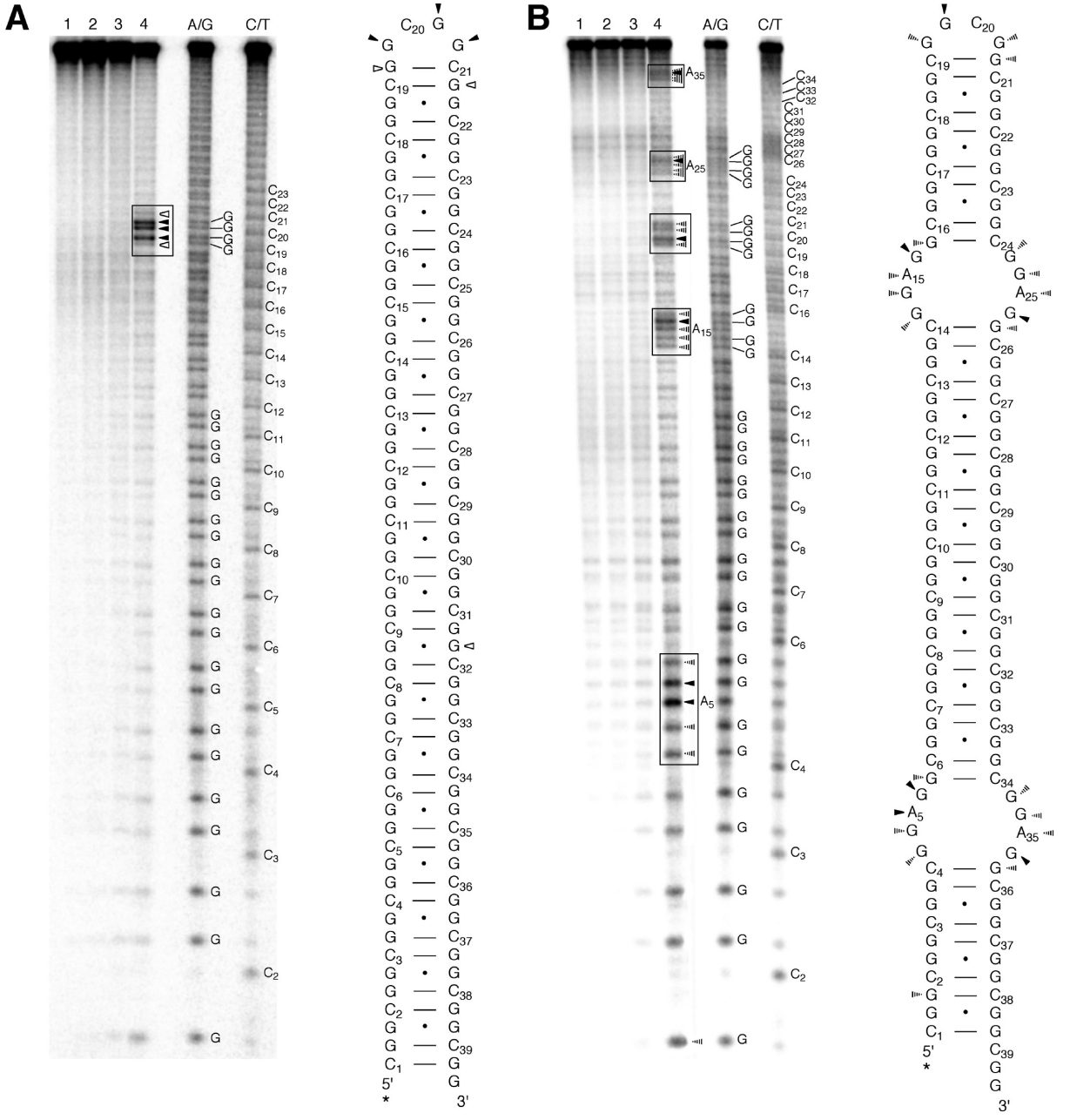


FIGURE 5: Characterization of the $(CGG)_{39}$ series using the chemical probe DEPC. Autoradiograms and schematic representations revealing strand cleavage for (A) $(CGG)_{39}$ and (B) 4AGG are shown. Conditions: 1 $\mu$M DNA in 10 mM sodium phosphate and 100 mM KCl (pH 7.5). Lanes 1−3 contained DNA alone, DNA cycled through heating methods, and piperidine-treated DNA, respectively. Lane 4 contained DNA incubated for 30 min at 37 °C in the presence of 5% DEPC followed by piperidine treatment. A/G and C/T are Maxam/Gilbert sequencing reactions. The pattern of the arrow at a given site reflects the relative amount of strand cleavage, with the filled arrow being the most reactive, the empty arrow being the least reactive, and the striped arrow indicating an intermediate amount of reactivity. The asterisk represents the location of the $^{32}$P radiolabel.

DNA sequences have been analyzed using calorimetric analysis, optical analysis, and native PAGE (*15*−*18*, *29*, *54*). While the results of such analyses provide support for the formation of intramolecular secondary structure, they do not provide a

Table 4: UV−Visible-Derived Thermodynamic Parameters for $(CGG)_{39}$ Repeat Series

| substrate | $T_m$[a] (°C) | $\Delta H$[a,b] (kcal/mol) | $\Delta G$[a,c] (kcal/mol) | $\Delta S$[a,b] (cal mol$^{-1}$ K$^{-1}$) | $\Delta T_m$[a] (°C) | $\Delta\Delta H$ (kcal/mol) | $\Delta\Delta G$ (kcal/mol) | $\Delta\Delta S$ (cal mol$^{-1}$ K$^{-1}$) |
|---|---|---|---|---|---|---|---|---|
| $(CGG)_{39}$ | 85.6 ± 0.2 | 110 ± 4 | 15.0 ± 0.6 | 309 ± 12 | − | − | − | − |
| 4AGG | 83.4 ± 0.3 | 91.9 ± 1.4 | 12.0 ± 0.2 | 258 ± 4 | −2.2 | −18.1 | −3.0 | −51 |

[a]DNA in 10 mM sodium phosphate and 100 mM KCl (pH 7.5). The error represents the standard deviation from a minimum of three experiments. [b]Values derived from van't Hoff analysis as described by Marky and Breslauer (*35*). [c]Values at 37 °C.

molecular-level picture of the conformation. Studies by NMR demonstrated that $(CGG)_3$ forms a homoduplex and also characterized the hydrogen bonding at G·G mismatches, but sequences of this length are too short to form intramolecular structures (*29*). Structural characterization of longer sequences includes chemical probing with DEPC and $KMnO_4$ and digestion by mung bean nuclease to derive the conformation of $(CGG)_{11}$ (*28*) and $(CGG)_{15}$ (*16*), respectively. A stem−loop structure with a four-base loop and either a single-stranded overhang or bulge was reported in both cases.

Our use of DEPC, a small molecule chemical probe of nucleobase accessibility, has provided evidence that $(CGG)_{19}$ also forms a stem−loop conformation. The single region of hyperreactivity toward DEPC identified for $(CGG)_{19}$ suggests a four-base loop. Although the increased dynamics of a mismatch might be expected to make the bases more accessible to DEPC, very little reactivity toward DEPC was observed for the G·G mismatches in the stem. The marginal reactivity of the G·G mismatches toward DEPC is likely due to $G_{syn}\cdot G_{anti}$ base pairing (*29*, *45*, *55*). Rotation about the glycosidic bond, converting from the anti conformation to the syn conformation, of one of the guanines involved in a G·G mismatch makes the Hoogsteen edge available for base pairing. This indicates that on the time scale of the DEPC experiments the mismatches are intrahelical and not otherwise extruded from the stem.

Given the possibility for repetitive CGG sequences to form a homoduplex (*29*), namely, a duplex formed by two strands of $(CGG)_n$, we considered whether this intermolecular structure was forming for $(CGG)_{19}$. The concentration-independent melting temperatures we observe are consistent with an intramolecularly folded structure; a multistrand structure would display a concentration-dependent change (*42*). Indeed, concentration-independent melting temperatures have been reported previously for $(CGG)_n$ ($n$ = 10−12, 14−16, 18, 20, and 25) (*16*, *18*). Furthermore, when $(CGG)_{19}$ was heat denatured prior to analysis, which was done for our optical analysis, calorimetry, and chemical probe experiments, no species with a migration similar to that of the multistrand control were observed by native PAGE. However, in the absence of this heat treatment, ~3% of the sample migrates like the multistrand control. The amount of this species increases with DNA concentration and is consistent with a homoduplex. Taken together, the $T_m$ and native PAGE results support the formation of an intramolecular structure by $(CGG)_{19}$ under our experimental conditions.

Having determined the conformation adopted by the sequence that lacks interruptions, we next used chemical probes to examine the effect of AGG interruptions on the structure of the CGG repeat sequence. We find that these interruptions have an effect on both the conformation adopted by the sequence and the relative stability of the conformation. For 1AGG-b, where the AGG interruption is centrally located in the sequence, the loop size increases relative to $(CGG)_{19}$ and the interruption is incorporated into the loop. It is also of note that to maintain any G-C base pairs in this stem−loop structure, bases must overhang the 3′-end. The presence of an overhang is supported by the increased reactivity toward DEPC of the guanines at the 3′-end of the 1AGG-b sequence.

For 1AGG-a, the reactivity toward DEPC in repeats 9 and 10 that was identified for $(CGG)_{19}$ is conserved, implying that the loop region is of the same size and at the same position. Thus, in contrast to incorporating the interruption into the loop which occurs with 1AGG-b, a bulge is used to accommodate the AGG interruption. A second bulge is observed on the opposing arm of the stem. These two bulges do not lie across from one another but instead are staggered. These staggered bulges maintain G-C base pairing in the stem and also prevent the need for a 3′-overhang.

For the $(CGG)_{19}$ stem−loop structure, each G·G mismatch is flanked by two well-paired G-C base pairs. However, with the introduction of an AGG interruption, A replaces C, and there are two mismatches in a row (A·G and G·G). Although $A_{anti}\cdot G_{syn}$ (*56*) and $G_{syn}\cdot G_{anti}$ (*29*, *45*, *55*) base pairing has been reported when present independently in a duplex, when these two mismatches are adjacent to one another an increase in dynamics may facilitate the formation of the larger loop and bulges observed for 1AGG-b and 1AGG-a, respectively.

The most common genotype in the FMR1 gene of healthy individuals contains AGG interruptions spaced 9−11 repeats apart (*36*). The 2AGG sequence incorporates two interruptions with this spacing. Three areas of reaction toward DEPC were observed, which is similar to the reactivity observed for 1AGG-a. This pattern of modification by DEPC is expected if, in addition to the loop, two bulges were positioned directly across from one another and there is a 3′-end overhang. However, as previously described for RNA (*57*) and as supported by structural predictions using mfold (*58*), a Y-type or boomerang structure would also be consistent with the observed DEPC reaction pattern. While we cannot definitively assign the structure of 2AGG, we can conclude that the introduction of two interruptions yields a conformation that is unique from both the uninterrupted $(CGG)_{19}$ and the sequences containing a single interruption.

Sequences containing runs of two sequential guanines have been reported to form a quadruplex from two or four strands (*15*, *21*, *59*, *60*). Of particular relevance to the $(CGG)_{19}$ sequences, it has been shown that two stem−loop structures can associate and form a quadruplex (*21*). Although our $T_m$ and native PAGE analysis indicate that a single-stranded structure is formed, we employed DMS, another chemical probe of nucleobase accessibility, to examine the possibility of quadruplex formation by these G-rich sequences. The lack of protection from modification by DMS for $(CGG)_{19}$, 1AGG-a, 1AGG-b, or 2AGG demonstrates that these sequences do not form quadruplexes under these conditions. Rather, our chemical probe experiments support the formation of stem−loop structures by the $(CGG)_{19}$ series.

Although chemical probe analysis is a powerful tool for gaining an understanding of the conformations adopted by DNA sequences in solution, we used complementary methods to corroborate these findings and to provide a quantitative measure of the destabilizing effect of AGG interruptions. CD spectra for stem−loop structures formed by other TNR sequences [$(CAG)_n$, $(CTG)_n$, and $(CCG)_n$] include a signature profile similar to that of the B-DNA duplex, i.e., a maximum at 275−280 nm and a minimum at 240−255 nm (*18*). However, the CD spectra observed for the $(CGG)_{19}$ series do not share this signature. It has been reported previously that the CD spectra for sequences with high G-C content, and for structures containing mismatches, are often perturbed (*15*, *18*). Indeed, similarities are observed upon comparison of the CD spectrum obtained for $(CGG)_{19}$ to that for $(CGG)_{15}$ and $(CGG)_{25}$ reported previously by Paiva and Sheardy (*18*). The wavelengths at which maxima and minima occur are comparable; however, one notable difference is the ratio of the maxima at 240 and 275 nm. While we observe a significant difference in this ratio for $(CGG)_{19}$, with the maxima at 240 nm being greater, the ratio is nearly comparable in the spectra reported previously for $(CGG)_{15}$ and $(CGG)_{25}$ (*18*). This result suggests that there are differences in stacking and/or overall conformation between the sequences studied in this work versus those of Paiva and Sheardy. In addition to the different lengths of the sequences, this difference in CD spectra may result from the different buffers, ionic strengths, and acquisition temperatures used in the two experiments. Indeed, the conformations adopted by $(CGG)_n$ sequences have been shown to be very sensitive to salt concentration, cation identity, temperature, and pH (*15*, *16*, *18*, *27*, *54*).

Upon the introduction of AGG interruptions, the changes we observe in the intensity of the maxima and minima in the CD spectra are consistent with a change in the extent of base stacking interactions. Indeed, structural changes that influence the extent of base stacking have been shown previously to affect CD spectra (*15*, *18*, *53*, *59*).

Using both optical analysis and DSC, we have quantified the extent to which AGG interruptions modulate the stability of the conformation adopted by CGG repeat sequences. Using van't Hoff analysis, thermodynamic parameters describing the transition from structured to unstructured sequence were extracted from the optical melting profiles. Using DSC, the same parameters were measured directly. The two methods result in values for $\Delta H$, $\Delta G$, and $\Delta S$ that are different from one another; the values obtained by DSC are much larger. This discrepancy between thermodynamic parameters derived from optical analysis versus calorimetry arises when the two-state assumption used during van't Hoff analysis fails. The two-state assumption fails when stable intermediates populate the transition from structured to unstructured sequence. Because these intermediates are not included in van't Hoff analysis, the result is an underestimate of the heat associated with a melting transition (*42*). The heat of the transition is measured directly in DSC, and any intermediates are included and identified in the experiment. Because of the difference in values obtained by van't Hoff analysis and DSC, we know that for the $(CGG)_{19}$ series the two-state model does not hold and intermediates populate the transition from structured to unstructured sequence (*42*). This difference is not observed for all TNR sequences. For example, $(CAG)_6$ has been shown to abide by the two-state model (*61*). However, it has been shown previously for $(CGG)_n$ ($n$ = 14, 15, 16, 18, and 20) (*16*) that this model does indeed fail for larger CGG repeat tracts. Regardless, the trend observed in the thermodynamic parameters obtained by both methods is upheld, solidifying the notion that introducing AGG interruptions has an effect on the stability of these repeat tracts.

Because the thermodynamic data obtained by optical analysis are underestimates, we will discuss in detail only the DSC results. Upon consideration of the results obtained by DSC, there is a correlation between the number of AGG interruptions and thermal stability ($T_m$). One interruption lowers the $T_m$ ~1 °C, and two interruptions lower the $T_m$ ~4 °C. This decrease in $T_m$ is consistent with the patterns of reactivity toward DEPC across the series which indicate that the structure of $(CGG)_{19}$ is perturbed by the addition of AGG interruptions. Along with a decrease in $T_m$, $\Delta H$ of the AGG-interrupted structures is also decreased in magnitude relative to that of $(CGG)_{19}$. This enthalpic contribution to the total free energy is dependent upon hydrogen bonding and base stacking (*42*), both of which have been altered when considering the proposed structures adopted by the sequences containing AGG interruptions relative to $(CGG)_{19}$. A negative $\Delta\Delta H$ for 1AGG-a and 1AGG-b supports the possibility that the structural changes revealed by DEPC modification result from fewer hydrogen bonds and/or reduced base−base stacking due to the larger loop or bulges. Thereby, less heat is required for melting the conformations adopted by 1AGG-a and 1-AGG-b. The same holds true for 2AAG. Relative to those of 1AGG-a and 1AGG-b, the $\Delta\Delta H$ for this sequence is more negative and indicates an even further disruption of hydrogen bonding and/or stacking interactions.

In addition to $\Delta H$ decreasing with the introduction of AGG interruptions, a decrease in $\Delta S$ is also observed across the $(CGG)_{19}$ series. A decrease in the entropy associated with the transition from structured to unstructured sequence is a result of the structured form having more disorder when interruptions are present. The structures predicted by DEPC modification include bulges and loops, both of which would result in an increase in disorder. Structures with more disorder will display smaller changes in entropy upon melting assuming that all the unstructured single-stranded states are isoenergetic across the $(CGG)_{19}$ series (*42*, *53*).

The free energy required for melting also decreases as AGG interruptions are introduced into the sequence. The combination of negative values for both $\Delta\Delta H$ and $\Delta\Delta S$ with the incorporation of AGG interruptions means that the enthalpic contribution that reduces the free energy required for melting is somewhat moderated by the entropic contribution that increases the free energy. The overall result, however, is still a more negative $\Delta\Delta G$ with an increasing number of interruptions and a structure that is less thermodynamically stable.

It is noteworthy that AGG interruptions occur in the genome of healthy individuals, and in these individuals, the CGG repeat sequence is not prone to expansion (*36*). Indeed, a length of 19 CGG repeats falls within the healthy range. The $(CGG)_{39}$ sequence, while still within the healthy range, is closer to the premutation length. Prior to this work, the conformations adopted by sequences with lengths approaching the premutation range for fragile X syndrome had not been examined. This lack of information is likely due to the difficulties associated with synthesizing and purifying sequences of this length. In addition to length, these sequences have a G-C content of 100%, and this can further complicate synthesis and purification. Here we have synthesized and purified $(CGG)_{39}$ and a corresponding sequence containing four AGG interruptions (4AGG).

We found that, with respect to the ability of AGG interruptions to modulate the structure and stability of the repeat sequence, the $(CGG)_{39}$ series behaves like the $(CGG)_{19}$ series. Furthermore, even with a sequence that is more than twice the length, modification by DEPC reveals a single region of reactivity for $(CGG)_{39}$. The reactivity is consistent with a stem−loop structure, which was observed for $(CGG)_{19}$. This result was unexpected because mfold (*58*) predicts a number of thermodynamically stable structures that include branched motifs with more than one loop, but no evidence for these structures is observed.

Again, similar to the $(CGG)_{19}$ series, the introduction of AGG interruptions alters the conformation adopted by $(CGG)_{39}$. The five areas of reactivity toward DEPC identified for 4AGG could correspond to a stem−loop structure with four bulges. There could also be any number of Y-branched/bulge combination structures that would show this reaction pattern toward DEPC. These structures are predicted by mfold (*58*) to have similar thermodynamic stabilities.

The impact of four AGG interruptions on the thermodynamics of the $(CGG)_{39}$ series follows the same trend that was identified for interruptions in the $(CGG)_{19}$ series. Changes in enthalpy, entropy, and free energy associated with melting of 4AGG are smaller than those for the uninterrupted sequence. However, it is important to note that the thermodynamic parameters reported are those obtained from optical melting studies. Technical issues, including the low yield of the syntheses of $(CGG)_{39}$ and 4AGG, in addition to the relatively large amount of material required for calorimetry, prevented us from analyzing these sequences by DSC. It may be possible to obtain these data using nano-DSC, which requires significantly less material.

An interesting observation is the lack of an effect of the increase in size from 19 to 39 repeats on melting temperature. This plateau effect that occurs for $T_m$ values as the size of the stem−loop structure increases has been observed previously for TNR stem−loop structures and is not fully understood; however, it has been proposed that the stabilizing effect of increasing the number of base pairs is moderated by the instability of the additional mismatches (*18*). This theory is supported by the fact that the $T_m$ values of well-matched duplexes containing 12−45 bp do not plateau with an increase in length (*18*). Indeed, although the $T_m$ values obtained by optical analysis for $(CGG)_{19}$ and $(CGG)_{39}$ are nearly identical, the changes associated with enthalpy and entropy are markedly different; the magnitudes of both increase with the larger sequence, an important feature that would have been missed if one simply compared melting temperatures. The larger $\Delta H$ indicates that there is more hydrogen bonding or base stacking occurring with a larger sequence, while the larger $\Delta S$ implies that the ground state is less disordered.

Although in the disease state the FMR1 gene is transcriptionally silenced, it has been proposed that defects in FMR1 mRNA metabolism might be responsible for the different phenotypes observed for individuals with repeat lengths within the premutation range. $r(CGG)_n$ ($n$ = 19, 23, and 28) repeat sequences present in the natural sequence context of the 5′-UTR FMR1 mRNA were structurally characterized using nuclease digestions (*57*). The mRNA CGG repeat sequences adopted stem−loop structures similar to those described here for DNA. When AGG interruptions were present, branched structures with multiple stem−loop structures were observed. However, using both in vitro and in vivo methods, it was determined that the translational efficiency of CGG repeat mRNA was not influenced by one or two AGG interruptions (*62*). This result suggests that the protective role the AGG interruptions play in preventing expansion may lie at the DNA level rather than at the mRNA level.

Fry and co-workers studied the effect of AGG interruptions on the processing of CGG repeat DNA by an enzyme relevant to DNA replication, namely, the human Werner syndrome DNA helicase (*21*). Two DNA stem−loop structures containing CGG repeats were incubated under conditions that favor formation of intermolecular quadruplexes. Upon the insertion of an AGG interruption, unwinding of the stem−loop DNA conformation by the DNA helicase was accelerated (*21*). Therefore, AGG interruptions are able to modulate the ability of a helicase to process the repeat tract.

During replication, DNA polymerase, in concert with the rest of the replicative machinery, copies a region of duplex DNA. First, DNA helicase unwinds the parent duplex, and each single strand is used as a template by polymerase. As DNA synthesis progresses, polymerase can dissociate from and reassociate at the DNA replication fork. Reassociation at the appropriate position relies on the general requisite that both the parent strand template and the nascent daughter strand remain unstructured and, thus, not folded intramolecularly. Single-stranded DNA binding proteins that maintain the single-stranded nature of the unwound DNA guard this structural requirement. If a stem−loop structure was to form in the daughter strand of leading strand synthesis, the position of the nascent strand would slip with respect to the parent strand. If the formation of these stem−loop structures were to occur faster than the binding of the single-stranded DNA binding proteins, thus allowing these structures to persist, replication would continue with these stem−loop structures embedded in the daughter strand. Following another round of replication, a helicase would unwind the stem−loop structure, and DNA polymerase would replicate the full length of the structure, resulting in an expansion of the TNR sequence.

Here we have shown that AGG interruptions decrease the stability of the structures formed by CGG repeat sequences and may hinder their ability to persist during replication. Thus, AGG interruptions may play a significant role in distinguishing sequences that are stable and do not expand from those that are prone to expansion. Moreover, the ability of AGG interruptions to disrupt non-B DNA structure may also be important during a DNA repair event. For example, following the removal of a modified base by a DNA glycosylase, downstream proteins, including DNA polymerase, complete the steps for DNA repair. It has recently been shown that during long-patch base excision repair the formation of stem−loop structures by a repeat sequence can lead to expansion (*7*). Our results provide insight into the role interruptions may play in preventing expansion in vivo and also contribute to our understanding of the relationship between non-B conformations and trinucleotide repeat expansion.

## ACKNOWLEDGMENT

We acknowledge the Brown University EPSCoR Proteomics Facility for the use of the DSC and CD instrumentation (supported by NSF/EPSCoR Grant 0554548, a Rhode Island Science and Technology Advisory Council grant, and National Center for Research Resources Grant 1S10RR020923-01A1), as well as Dr. James Clifton for technical support. We also thank Ms. Amalia Ávila Figueroa and Ms. Nicole Wilson for helpful discussions.

## SUPPORTING INFORMATION AVAILABLE

Optical melting profiles for the $(CGG)_{19}$ series, optical analysis for $(CGG)_{19}$ concentration dependence, DSC thermograms for $(CGG)_{19}$ series, autoradiogram revealing modification by DMS for quadruplex control sequence and $(CGG)_{19}$ series, optical melting profiles for $(CGG)_{39}$ series, and CD spectra for $(CGG)_{39}$ series. This material is available free of charge via the Internet at http://pubs.acs.org.

## REFERENCES


1. Kelkar, Y., Tyekucheva, S., Chiaromonte, F., and Makova, K. (2008) The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome Res. 18*, 30.
2. Pumpernik, D., Oblak, B., and Borštnik, B. (2008) Replication slippage versus point mutation rates in short tandem repeats of the human genome. *Mol. Genet. Genomics 279*, 53–61.
3. Madsen, B., Villesen, P., and Wiuf, C. (2008) Short tandem repeats in human exons: A target for disease mutations. *BMC Genomics 9*, 410.
4. Tóth, G., Gáspári, Z., and Jurka, J. (2000) Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Res. 10*, 967.
5. Kozlowski, P., De Mezer, M., and Krzyzosiak, W. (2010) Trinucleotide repeats in human genome and exome. *Nucleic Acids Res. 38*, 4027–4039.
6. Gatchel, J., and Zoghbi, H. (2005) Diseases of unstable repeat expansion: Mechanisms and common principles. *Nat. Rev. Genet. 6*, 743–755.
7. Liu, Y., Prasad, R., Beard, W., Hou, E., Horton, J., McMurray, C., and Wilson, S. (2009) Coordination between polymerase $\beta$ and FEN1 can modulate CAG repeat expansion. *J. Biol. Chem. 284*, 28352.
8. López Castel, A., Cleary, J. D., and Pearson, C. E. (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol. 11*, 165–170.
9. Kovtun, I., and McMurray, C. (2008) Features of trinucleotide repeat instability in vivo. *Cell Res. 18*, 198–213.
10. Garber, K., Smith, K., Reines, D., and Warren, S. (2006) Transcription, translation and fragile X syndrome. *Curr. Opin. Genet. Dev. 16*, 270–275.
11. Pearson, C. E., Nichol Edamura, K., and Cleary, J. D. (2005) Repeat instability: Mechanisms of dynamic mutations. *Nat. Rev. Genet. 6*, 729–742.
12. Mitas, M. (1997) Trinucleotide repeats associated with human disease. *Nucleic Acids Res. 25*, 2245–2253.
13. Ji, J., Clegg, N., Peterson, K., Jackson, A., Laird, C., and Loeb, L. (1996) In vitro expansion of GGC:GCC repeats: Identification of the preferred strand of expansion. *Nucleic Acids Res. 24*, 2835.
14. Wells, R. D. (2007) Non-B DNA conformations, mutagenesis and disease. *Trends Biochem. Sci. 32*, 271–278.
15. Renciuk, D., Zemánek, M., Kejnovská, I., and Vorlícková, M. (2009) Quadruplex-forming properties of FRAXA (CGG) repeats interrupted by (AGG) triplets. *Biochimie 91*, 416–422.
16. Amrane, S., and Mergny, J. (2006) Length and pH-dependent energetics of $(CCG)_n$ and $(CGG)_n$ trinucleotide repeats. *Biochimie 88*, 1125–1134.
17. Paiva, A., and Sheardy, R. (2005) The influence of sequence context and length on the kinetics of DNA duplex formation from complementary hairpins possessing (CNG) repeats. *J. Am. Chem. Soc. 127*, 5581–5585.
18. Paiva, A., and Sheardy, R. (2004) Influence of Sequence Context and Length on the Structure and Stability of Triplet Repeat DNA Oligomers. *Biochemistry 43*, 14218–14227.
19. Sinden, R., Potaman, V., Oussatcheva, E., Pearson, C., Lyubchenko, Y., and Shlyakhtenko, L. (2002) Triplet repeat DNA structures and human genetic disease: Dynamic mutations from dynamic DNA. *J. Biosci. 27*, 53–65.
20. Pearson, C. E., Tam, M., Wang, Y.-H., Montgomery, S. E., Dar, A. C., Cleary, J. D., and Nichol, K. (2002) Slipped-strand DNAs formed by long (CAG)·(CTG) repeats: Slipped-out repeats and slip-out junctions. *Nucleic Acids Res. 30*, 4534–4547.
21. Weisman-Shomer, P., Cohen, E., and Fry, M. (2000) Interruption of the fragile X syndrome expanded sequence $d(CGG)_n$ by interspersed d(AGG) trinucleotides diminishes the formation and stability of $d(CGG)_n$ tetrahelical structures. *Nucleic Acids Res. 28*, 1535.
22. Pearson, C. E., Wang, Y. H., Griffith, J. D., and Sinden, R. R. (1998) Structural analysis of slipped-strand DNA (S-DNA) formed in $(CTG)_n$·$(CAG)_n$ repeats from the myotonic dystrophy locus. *Nucleic Acids Res. 26*, 816–823.
23. Mariappan, S., Catasti, P., Chen, X., Ratliff, R., Moyzis, R., Bradbury, E., and Gupta, G. (1996) Solution structures of the individual single strands of the fragile X DNA triplets $(GCC)_n$·$(GGC)_n$. *Nucleic Acids Res. 24*, 784.
24. Pearson, C. E., and Sinden, R. R. (1996) Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry 35*, 5041–5053.
25. Chen, X., Mariappan, S., Catasti, P., Ratliff, R., Moyzis, R., Laayoun, A., Smith, S., Bradbury, E., and Gupta, G. (1995) Hairpins are formed by the single DNA strands of the fragile X triplet repeats: Structure and biological implications. *Proc. Natl. Acad. Sci. U.S.A. 92*, 5199.
26. Gacy, A., Goellner, G., Juranić, N., Macura, S., and McMurray, C. (1995) Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell 81*, 533.
27. Mitas, M., Yu, A., Dill, J., and Haworth, I. (1995) The trinucleotide repeat sequence $d(CGG)_{15}$ forms a heat-stable hairpin containing $G_{syn}$·$G_{anti}$ base pairs. *Biochemistry 34*, 12803–12811.
28. Nadel, Y., Weisman-Shomer, P., and Fry, M. (1995) The Fragile X Syndrome single strand d(CGG) nucleotide repeats readily fold back to form unimolecular hairpin structures. *J. Biol. Chem. 270*, 28970.
29. Zheng, M., Huang, X., Smith, G., Yang, X., and Gao, X. (1996) Genetically unstable CXG repeats are structurally dynamic and have a high propensity for folding. An NMR and UV spectroscopic study. *J. Mol. Biol. 264*, 323–336.
30. Al-Mahdawi, S., Pinto, R. M., Ismail, O., Varshney, D., Lymperi, S., Sandi, C., Trabzuni, D., and Pook, M. (2008) The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues. *Hum. Mol. Genet. 17*, 735–746.
31. Robertson, K. D. (2005) DNA methylation and human disease. *Nat. Rev. Genet. 6*, 597–610.
32. Murray, J., Cuckle, H., Taylor, G., and Hewison, J. (1997) Screening for fragile X syndrome. *Health Technology Assessment 1*, No. 4.
33. Zarnescu, D., Shan, G., Warren, S., and Jin, P. (2005) Come FLY with us: Toward understanding fragile X syndrome. *Genes, Brain Behav. 4*, 385–392.
34. O'Donnell, W., and Warren, S. (2002) A decade of molecular studies of Fragile X Syndrome. *Annu. Rev. Neurosci. 25*, 315–338.
35. Verkerk, A., Pieretti, M., Sutcliffe, J., Fu, Y., Kuhl, D., Pizzuti, A., Reiner, O., Richards, S., Victoria, M., Zhang, F., Eussen, B., van Ommen, G., Bionden, L., Riggins, G., Chastain, J., Kunst, C., Galjaard, H., Caskey, C., Nelson, D., Oostra, B., and Warren, S. (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell 65*, 905–914.
36. Kunst, C. B., and Warren, S. T. (1994) Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell 77*, 853–861.
37. Nolin, S., Brown, W., Glicksman, A., Houck, J., Gargano, A., Sullivan, A., Biancalana, V., Bröndum-Nielsen, K., Hjalgrim, H., and Holinski-Feder, E. (2003) Expansion of the fragile X CGG repeat in females with premutation or intermediate alleles. *Am. J. Hum. Genet. 72*, 454–464.
38. Kass, S., Pruss, D., and Wolffe, A. (1997) How does DNA methylation repress transcription? *Trends Genet. 13*, 444–449.
39. Nolin, S., Lewis, F., Ye, L., Houck, G., Jr., Glicksman, A., Limprasert, P., Li, S., Zhong, N., Ashley, A., Feingold, E., Sherman, S., and Brown, T. (1996) Familial transmission of the FMR1 CGG repeat. *Am. J. Hum. Genet. 59*, 1252–1261.
40. Beaucage, S. L., and Caruthers, M. H. (2000) Synthetic strategies and parameters involved in the synthesis of oligodeoxyribonucleotides according to the phosphoramidite method. Current Protocols in Nucleic Acid Chemistry, pp 3.3.1−3.3.20, Wiley-Interscience, New York.
41. Warshaw, M. M., and Tinoco, I., Jr. (1966) Optical properties of sixteen dinucleoside phosphates. *J. Mol. Biol. 20*, 29–38.
42. Marky, L. A., and Breslauer, K. J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers 26*, 1601–1620.
43. Leonard, N., McDonald, J., Henderson, R., and Reichmann, M. (1971) Reaction of diethyl pyrocarbonate with nucleic acid components. Adenosine. *Biochemistry 10*, 3335–3342.
44. Vincze, A., Henderson, R., McDonald, J., and Leonard, N. (1973) Reaction of diethyl pyrocarbonate with nucleic acid components. Bases and nucleosides derived from guanine, cytosine, and uracil. *J. Am. Chem. Soc. 95*, 2677–2682.
45. Huertas, D., Bellsolell, L., Casasnovas, J., Coll, M., and Azorín, F. (1993) Alternating $d(GA)_n$ DNA sequences form antiparallel

stranded homoduplexes stabilized by the formation of G·A base pairs. *EMBO J. 12*, 4029–4038.
46. Williamson, J. R., Raghuraman, M. K., and Cech, T. R. (1989) Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell 59*, 871–880.
47. Miura, T., and Thomas, G. J. (1994) Structural polymorphism of telomere DNA: Interquadruplex and duplex-quadruplex conversions probed by Raman spectroscopy. *Biochemistry 33*, 7848–7856.
48. Hardin, C. C., Henderson, E., Watson, T., and Prosser, J. K. (1991) Monovalent cation induced structural transitions in telomeric DNAs: G-DNA folding intermediates. *Biochemistry 30*, 4460–4472.
49. Han, H., Hurley, L., and Salazar, M. (1999) A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic Acids Res. 27*, 537–542.
50. Risitano, A., and Fox, K. R. (2003) Stability of intramolecular DNA quadruplexes: Comparison with DNA duplexes. *Biochemistry 42*, 6507–6513.
51. Guo, Q., Lu, M., and Kallenbach, N. R. (1993) Effect of thymine tract length on the structure and stability of model telomeric sequences. *Biochemistry 32*, 3596–3603.
52. Víglaský, V., Bauer, L., and Tlucková, K. (2010) Structural features of intra- and intermolecular G-quadruplexes derived from telomeric repeats. *Biochemistry 49*, 2110–2120.
53. Chalikian, T., Völker, J., Plum, G., and Breslauer, K. (1999) A more unified picture for the thermodynamics of nucleic acid duplex melting: A characterization by calorimetric and volumetric techniques. *Proc. Natl. Acad. Sci. U.S.A. 96*, 7853–7858.
54. Fojtik, P., Kejnovska, I., and Vorlickova, M. (2004) The guanine-rich fragile X chromosome repeats are reluctant to form tetraplexes. *Nucleic Acids Res. 32*, 298–306.
55. Lane, A., and Peck, B. (2008) Conformational flexibility in DNA duplexes containing single G·G mismatches. *Eur. J. Biochem. 230*, 1073–1087.
56. Leonard, G., Booth, E., and Brown, T. (1990) Structural and thermodynamic studies on the adenine·guanine mismatch in B-DNA. *Nucleic Acids Res. 18*, 5617–5623.
57. Napierala, M., Michalowski, D., De Mezer, M., and Krzyzosiak, W. (2005) Facile FMR1 mRNA structure regulation by interruptions in CGG repeats. *Nucleic Acids Res. 33*, 451–463.
58. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res. 31*, 3406–3415.
59. Qin, Y., Rezler, E., Gokhale, V., Sun, D., and Hurley, L. (2007) Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP 4. *Nucleic Acids Res. 35*, 7698–7713.
60. Zhou, J., Yuan, G., Liu, J., and Zhan, C. (2006) Formation and stability of G-quadruplexes self-assembled from guanine-rich strands. *Chem.—Eur. J. 13*, 945–949.
61. Völker, J., Plum, G., Klump, H., and Breslauer, K. (2009) Energetic coupling between clustered lesions modulated by intervening triplet repeat bulge loops: Allosteric implications for DNA repair and triplet repeat expansion. *Biopolymers 93*, 355–369.
62. Ludwig, A., Raske, C., Tassone, F., Garcia-Arocena, D., Hershey, J., and Hagerman, P. (2009) Translation of the FMR 1 mRNA is not influenced by AGG interruptions. *Nucleic Acids Res. 37*, 6896–6904.