Published in final edited form as:

Biochemistry. 2010 August 17; 49(32): 6826–6837. doi:10.1021/bi1007782.

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

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 by 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 1-6 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) sequences are 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 5 CGG repeats is expanded to over 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

[†]This work was supported by Brown University.

* Telephone: (401) 863-3590. Fax: (401) 863-9368. sarah_delaney@brown.edu.

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>.

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, and 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 pre-mutation range and are susceptible to expansion over a single familial generation. In fact, a CGG/CCG tract of 90 repeats or more 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 greater 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 to define not only 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 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 Labs) (4.6 × 250 mm) at 90 °C using 100 mM TEAA in acetonitrile:water (99%:1%) (solvent A) and 100 mM TEAA in acetonitrile:water (1%:99%) (solvent B) as the mobile phases (gradient: solvent A was increased from 5 to 25% over 25 min; 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 was increased from 0 to 15% over 35 min; 1.0 mL/min).

DEPC Chemical Probe Analysis

Oligonucleotides were 5'-³²P end-labeled using T4 polynucleotide kinase and following the manufacturer's protocol. In order to obtain the thermodynamically-favored DNA

conformations, oligonucleotides (1 μM) in 10 mM sodium phosphate, 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% diethylpyrocarbonate (DEPC) (v/v) (20 μL 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, 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 phosphorimager.

DMS Methylation Protection Assay

Oligonucleotides were 5'-³²P end-labeled using T4 polynucleotide kinase and following the manufacturer's protocol. Two oligonucleotide samples (375 μL , 0.2 μM) 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, 100 mM KCl, pH 7.5 and the other was in 10 mM sodium phosphate, pH 7.5. To each sample 1 μL of freshly prepared 1:4 DMS:ethanol was added and aliquots of 75 μL were removed after 0, 5, 15, 30 and 45 min and quenched with 20 μL DMS stop solution (1.5 M sodium acetate, 1 M β -mercaptoethanol, 250 $\mu\text{g/mL}$ 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 phosphorimager.

Native Polyacrylamide Gel Electrophoresis

DNA samples (1 μM unless otherwise noted) in 10 mM sodium phosphate, 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 non-denaturing loading buffer (15% ficoll, 0.25% xylene cyanol, 0.25% bromophenol blue) or denaturing loading buffer and were electrophoresed through a 12% native polyacrylamide gel at 450 V (10 V/cm) while at 4 °C. The products were visualized by phosphorimager.

Optical Melting Analysis

Quantification of oligonucleotides was performed at 95 °C using the ϵ_{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 μM and optical melting analysis was performed in 10 mM sodium phosphate, 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 monitoring absorbance 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). Student's t-test was used to determine if the average values are statistically different.

Circular Dichroism

Circular dichroism spectra were obtained at 37 °C with a Jasco J-815 CD spectropolarimeter equipped with a Peltier thermoelectric device. DNA samples had a final concentration of 0.5-2.0 μM in 10 mM sodium phosphate, 100 mM KCl, pH 7.5 and were 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 50 nm/min. All reported spectra represent an average of three scans.

DSC Analysis

Microcalorimetry was performed using a MicroCal VP-DSC. Oligonucleotides (in 0.8 mL) were prepared in 10 mM sodium phosphate, 100 mM KCl, pH 7.5 and 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 was 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 ΔC_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, in order to evaluate the influence a non-zero ΔC_p would have on the reported ΔH , we assumed the maximum possible ΔC_p by setting an upper baseline at the last data point obtained at 95 °C. Indeed, a non-zero ΔC_p decreases the value of ΔH , but only by ~15%.

RESULTS

Design of (CGG)₁₉ DNA Series

Four DNA sequences were designed in order to determine the effect of AGG interruptions on 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 fifteenth repeats.

Structural Analysis of the (CGG)₁₉ Series by Reactivity Towards Chemical Probes

In order to characterize the conformation(s) formed in solution by the series of (CGG)₁₉ sequences, modification of the sequences by diethylpyrocarbonate (DEPC) was used. 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-match 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 towards DEPC, namely three highly reactive guanines and two guanines with low levels of reactivity. This specific pattern of modification by DEPC suggests that (CGG)₁₉ adopts a stem-loop structure in which the loop contains 4 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 6 bases. Nevertheless, and regardless of whether the loop contains 4 or 6 bases, as shown in Figure 1A the position of the loop dictates that a single G overhangs the 3' end of the stem-loop structure.

Structural characterization by DEPC of the sequences containing a single AGG interruption reveal that the overall conformation of these species is altered relative to (CGG)₁₉. The sequence 1AGG-a displays three regions of reactivity towards 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 is 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 7 base loop. In addition to the same 4 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 4 base overhang at the 3' end of the structure. Indeed, although not well-resolved from the unmodified DNA substrate, increased reactivity towards 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 based on 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 4 base loop and two bulges that are positioned across from one another. This structure would also include an overhang of 3 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. In order to determine if the reactivity patterns described for the (CGG)₁₉ series could be attributed to the formation of an intra- or intermolecular quadruplex a dimethyl sulfate (DMS) methylation protection assay was used. 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⁺, 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 the N-7 of G; however, in a quadruplex the 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 is increased levels of methylation at the guanines proposed to be in the loop region.

Electrophoretic Mobility of (CGG)₁₉ by Native PAGE

To better understand 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 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 non-denaturing or denaturing loading buffer, to assure 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) heating to 95°C followed by slow cooling (Figure 2, lane 4), and (3) heating to 95 °C followed by flash cooling (Figure 2, lane 5). When analyzed 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 similarly to 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 spectrophotometry in order 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 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 in 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 (Table 2) (42). Since the sequences are going from structured to unstructured heat must transfer into the system for melting to occur, resulting in positive thermodynamic parameters. Based on 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 (CGG)₁₉/(CCG)₁₉ duplex, an unstructured 57-mer single strand, and the quadruplex adopted by *Tetrahymena* telomeric DNA (51, 52).

Calorimetric Analysis of (CGG)₁₉ Series

We also employed differential scanning calorimetry (DSC) as a means to characterize 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 degree relative to (CGG)₁₉. Similar to what was observed by UV-visible analysis, the T_m of 2AGG is ~4 degrees lower than the control sequence that lacks interruptions.

Thermodynamic parameters were calculated directly from the DSC thermograms. Notably, for all four sequences the values obtained by DSC for ΔH , ΔG , and ΔS 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

In order 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 by modification with DEPC.

When probed by reactivity towards 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 4 base loop. Consequently, as described for (CGG)₁₉, a single G will overhang at the 3'-end.

Different from (CGG)₃₉, the presence of 4 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/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 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 4

interruptions lowers the T_m by ~2 degrees. 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 elucidated that (CGG)₁₉ forms a stem-loop 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) 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 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 include chemical probing with DEPC/KMnO₄ 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 4 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 hyper-reactivity towards DEPC identified for (CGG)₁₉ suggests a 4 base loop. Although the increased dynamics of a mismatch might be expected to make the bases more accessible to DEPC, very little reactivity towards DEPC was observed for the G•G mismatches in the stem. The marginal reactivity of the G•G mismatches towards DEPC is likely due to G_{syn}•G_{anti} base pairing (29, 45, 55). Rotation about the glycosidic bond, converting from the anti 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 multi-stranded 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, as done for our optical analysis, calorimetry, and chemical probe experiments, no species with a migration similar to the multi-stranded control were observed by native PAGE. However, in the absence of this heat treatment ~3% of the sample migrates similarly to the multi-stranded 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 in order to maintain any G-C base pairs in this stem-loop, bases must overhang the 3' end. The presence of an overhang is supported by the increased reactivity towards DEPC of the guanines at the 3' end of the 1AGG-b sequence.

For 1AGG-a the reactivity towards 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 as 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}•G_{syn} (56) and G_{syn}•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 sequence 2AGG incorporates two interruptions with this spacing. Three areas of reaction towards 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 has been previously described for RNA (57) and is 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 to gain 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 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 when comparing the CD spectrum obtained for (CGG)₁₉ to that of (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 nm and 275 nm. While we observe a significant difference in this ratio for (CGG)₁₉, with the maxima at 240 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 length of the sequences, this difference in CD spectra may result from the different buffers, ionic strength, and the acquisition temperatures used in the two experiments. Indeed, the conformation adopted by (CGG)_n sequences have been shown to be very sensitive to salt concentration, identity of cation, 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 ΔH , ΔG , and ΔS 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. Since these intermediates are not accounted for 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 accounted for 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 ($n = 14, 15, 16, 18, 20$) (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.

Since the thermodynamic data obtained by optical analysis are underestimates we will discuss in detail only the DSC results. When considering the results obtained by DSC there is a correlation between the number of AGG interruptions and thermal stability (T_m). One interruption lowers the $T_m \sim 1$ degree and two interruptions lower the $T_m \sim 4$ degrees. This decrease in T_m is consistent with the patterns of reactivity towards 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 , ΔH of the AGG-interrupted structures is also decreased in magnitude relative to (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 that the structural changes revealed by DEPC modification are a result of 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 1AGG-a and 1AGG-b, $\Delta\Delta H$ for this sequence is more negative and indicates an even further disruption of hydrogen bonding and/or stacking interactions.

In addition to ΔH decreasing with the introduction of AGG interruptions, a decrease in ΔS 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 higher 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 pre-mutation length. Prior to this work, the conformations adopted by sequences with lengths approaching the pre-mutation 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 similarly to 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, as 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 5 areas of reactivity towards DEPC identified for 4AGG could correspond to a stem-loop structure with 4 bulges. There could also be any number of Y-branched/bulge combination structures that would show this reaction pattern towards DEPC. These structures are predicted by mfold (58) to have similar thermodynamic stabilities.

The impact of 4 AGG interruptions on the thermodynamics of the (CGG)₃₉ series follows the same trend as was identified for interruptions in the (CGG)₁₉ series. Changes in enthalpy, entropy, and free energy associated with melting of 4AGG are lower than those for the uninterrupted sequence. However, it is important to note that the thermodynamic parameters reported are those obtained by 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 this data using nano-DSC, which requires significantly less material.

An interesting observation is the lack of effect the increase in size from 19 to 39 repeats has on melting temperature. This plateau effect that occurs for T_m values as the size of the stem-

loop increases has been observed previously for TNR stem-loops 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 of well-matched duplexes containing 12 to 45 base pairs do not plateau with increasing 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 magnitude of both increase with the larger sequence, an important feature that would have been missed if one simply compared melting temperatures. The larger ΔH indicates that there is more hydrogen bonding/base stacking occurring with a larger sequence, while the larger ΔS implies that the ground state has less disorder.

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 pre-mutation range. r(CGG)_n (n = 19, 23, 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-loops 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 1 or 2 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 coworkers 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 guards 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 loops 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 decrease 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge the Brown University EPSCoR Proteomics Facility for the use of the DSC and CD instrumentation (supported by NSF/EPSCoR grant 0554548, Rhode Island Science and Technology Advisory Council grant and NIH NCRR 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.

Abbreviations

CD	circular dichroism
DEPC	diethylpyrocarbonate
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

REFERENCES

1. Kelkar Y, Tyekucheva S, Chiaromonte F, Makova K. The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome Res.* 2008; 18:30. [PubMed: 18032720]
2. Pumpernik D, Oblak B, Borštnik B. Replication slippage versus point mutation rates in short tandem repeats of the human genome. *Mol. Genet. Genomics.* 2008; 279:53–61. [PubMed: 17926066]
3. Madsen B, Villesen P, Wiuf C. Short tandem repeats in human exons: a target for disease mutations. *BMC genomics.* 2008; 9:410. [PubMed: 18789129]
4. Tóth G, Gáspári Z, Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* 2000; 10:967. [PubMed: 10899146]
5. Kozlowski P, De Mezer M, Krzyzosiak W. Trinucleotide repeats in human genome and exome. *Nucleic Acids Res.* 2010; 38:4027–4039. [PubMed: 20215431]
6. Gatchel J, Zoghbi H. Diseases of unstable repeat expansion: mechanisms and common principles. *Nat. Rev. Genet.* 2005; 6:743–755. [PubMed: 16205714]

7. Liu Y, Prasad R, Beard W, Hou E, Horton J, McMurray C, Wilson S. Coordination between Polymerase β and FEN1 can modulate CAG repeat expansion. *J. Biol. Chem.* 2009; 284:28352. [PubMed: 19674974]
8. López Castel A, Cleary JD, Pearson CE. Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol.* 2010; 11:165–170. [PubMed: 20177394]
9. Kovtun I, McMurray C. Features of trinucleotide repeat instability in vivo. *Cell Res.* 2008; 18:198–213. [PubMed: 18166978]
10. Garber K, Smith K, Reines D, Warren S. Transcription, translation and fragile X syndrome. *Curr. Opin. Genet. Dev.* 2006; 16:270–275. [PubMed: 16647847]
11. Pearson CE, Nichol Edamura K, Cleary JD. Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet.* 2005; 6:729–742. [PubMed: 16205713]
12. Mitas M. Trinucleotide repeats associated with human disease. *Nucleic Acids Res.* 1997; 25:2245–2253. [PubMed: 9171073]
13. Ji J, Clegg N, Peterson K, Jackson A, Laird C, Loeb L. In vitro expansion of GGC:GCC repeats: Identification of the preferred strand of expansion. *Nucleic Acids Res.* 1996; 24:2835. [PubMed: 8759019]
14. Wells RD. Non-B DNA conformations, mutagenesis and disease. *Trends Biochem. Sci.* 2007; 32:271–278. [PubMed: 17493823]
15. Renciuik D, Zemánek M, Kejnovská I, Vorlíčková M. Quadruplex-forming properties of FRAXA (CGG) repeats interrupted by (AGG) triplets. *Biochimie.* 2009; 91:416–422. [PubMed: 19028545]
16. Amrane S, Mergny J. Length and pH-dependent energetics of (CCG)_n and (CGG)_n trinucleotide repeats. *Biochimie.* 2006; 88:1125–1134. [PubMed: 16690198]
17. Paiva A, Sheardy R. The influence of sequence context and length on the kinetics of DNA duplex formation from complementary hairpins possessing (CNG) repeats. *J. Am. Chem. Soc.* 2005; 127:5581–5585. [PubMed: 15826196]
18. Paiva A, Sheardy R. Influence of Sequence Context and Length on the Structure and Stability of Triplet Repeat DNA Oligomers. *Biochemistry.* 2004; 43:14218–14227. [PubMed: 15518572]
19. Sinden R, Potaman V, Oussatcheva E, Pearson C, Lyubchenko Y, Shlyakhtenko L. Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *J. Biosciences.* 2002; 27:53–65.
20. Pearson CE, Tam M, Wang Y-H, Montgomery SE, Dar AC, Cleary JD, Nichol K. Slipped-strand DNAs formed by long (CAG)•(CTG) repeats: slipped-out repeats and slip-out junctions. *Nucleic Acids Research.* 2002; 30:4534–4547. [PubMed: 12384601]
21. Weisman-Shomer P, Cohen E, Fry M. 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.* 2000; 28:1535. [PubMed: 10710419]
22. Pearson CE, Wang YH, Griffith JD, Sinden RR. Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)_n•(CAG)_n repeats from the myotonic dystrophy locus. *Nucleic Acids Research.* 1998; 26:816–823. [PubMed: 9443975]
23. Mariappan S, Catasti P, Chen X, Ratliff R, Moyzis R, Bradbury E, Gupta G. Solution structures of the individual single strands of the fragile X DNA triplets (GCC)_n•(GGC)_n. *Nucleic Acids Res.* 1996; 24:784. [PubMed: 8604324]
24. Pearson CE, Sinden RR. Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry.* 1996; 35:5041–5053. [PubMed: 8664297]
25. Chen X, Mariappan S, Catasti P, Ratliff R, Moyzis R, Laayoun A, Smith S, Bradbury E, Gupta G. 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.* 1995; 92:5199. [PubMed: 7761473]
26. Gacy A, Goellner G, Jurani N, Macura S, McMurray C. Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell.* 1995; 81:533. [PubMed: 7758107]
27. Mitas M, Yu A, Dill J, Haworth I. The trinucleotide repeat sequence d(CGG)₁₅ forms a heat-stable hairpin containing G_{syn}•G_{anti} base pairs. *Biochemistry.* 1995; 34:12803–12811. [PubMed: 7548035]

28. Nadel Y, Weisman-Shomer P, Fry M. The Fragile X Syndrome single strand d(CGG) nucleotide repeats readily fold back to form unimolecular hairpin structures. *J. Biol. Chem.* 1995; 270:28970. [PubMed: 7499428]
29. Zheng M, Huang X, Smith G, Yang X, Gao X. Genetically unstable CXG repeats are structurally dynamic and have a high propensity for folding. An NMR and UV spectroscopic study. *J. Mol. Biol.* 1996; 264:323–336. [PubMed: 8951379]
30. Al-Mahdawi S, Pinto RM, Ismail O, Varshney D, Lymperi S, Sandi C, Trabzuni D, Pook M. The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues. *Hum. Mol. Genet.* 2008; 17:735–746. [PubMed: 18045775]
31. Robertson KD. DNA methylation and human disease. *Nat. Rev. Genet.* 2005; 6:597–610. [PubMed: 16136652]
32. Murray J, Cuckle H, Taylor G, Hewison J. Screening for fragile X syndrome. *Health Technol. Assess.* 1. 1997
33. Zarnescu D, Shan G, Warren S, Jin P. Come FLY with us: toward understanding fragile X syndrome. *Genes Brain Behav.* 2005; 4:385–392. [PubMed: 16098136]
34. O'Donnell W, Warren S. A decade of molecular studies of Fragile X Syndrome. *Annu. Rev. Neurosci.* 2002; 25:315–338. [PubMed: 12052912]
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, Warren S. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell.* 1991; 65:905–914. [PubMed: 1710175]
36. Kunst CB, Warren ST. Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell.* 1994; 77:853–861. [PubMed: 7911740]
37. Nolin S, Brown W, Glicksman A, Houck J, Gargano A, Sullivan A, Biancalana V, Brøndum-Nielsen K, Hjalgrim H, Holinski-Feder E. Expansion of the fragile X CGG repeat in females with premutation or intermediate alleles. *Am. J. Hum. Genet.* 2003; 72:454–464. [PubMed: 12529854]
38. Kass S, Pruss D, Wolffe A. How does DNA methylation repress transcription? *Trends Genet.* 1997; 13:444–449. [PubMed: 9385841]
39. Nolin S, Lewis F, Ye L, Houck G Jr, Glicksman A, Limprasert P, Li S, Zhong N, Ashley A, Feingold E, Sherman S, Brown T. Familial transmission of the FMR1 CGG repeat. *Am. J. Hum. Genet.* 1996; 59:1252–1261. [PubMed: 8940270]
40. Beaucage SL, Caruthers MH. Synthetic strategies and parameters involved in the synthesis of oligodeoxyribonucleotides according to the phosphoramidite method. *Curr. Protoc. Nucleic Acid Chem.* 2000;3.3.1–3.3.20.
41. Warshaw MM, Tinoco I Jr. Optical properties of sixteen dinucleoside phosphates. *J. Mol. Biol.* 1966; 20:29–38. [PubMed: 5970660]
42. Marky LA, Breslauer KJ. Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers.* 1987; 26:1601–1620. [PubMed: 3663875]
43. Leonard N, McDonald J, Henderson R, Reichmann M. Reaction of diethyl pyrocarbonate with nucleic acid components. Adenosine. *Biochemistry.* 1971; 10:3335–3342. [PubMed: 5118618]
44. Vincze A, Henderson R, McDonald J, Leonard N. Reaction of diethyl pyrocarbonate with nucleic acid components. Bases and nucleosides derived from guanine, cytosine, and uracil. *J. Am. Chem. Soc.* 1973; 95:2677–2682. [PubMed: 4694530]
45. Huertas D, Bellolell L, Casasnovas J, Coll M, Azorín F. Alternating d(GA)_n DNA sequences form antiparallel stranded homoduplexes stabilized by the formation of G•A base pairs. *EMBO J.* 1993; 12:4029–4038. [PubMed: 8404869]
46. Williamson JR, Raghuraman MK, Cech TR. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell.* 1989; 59:871–880. [PubMed: 2590943]
47. Miura T, Thomas GJ. Structural polymorphism of telomere DNA: interquadruplex and duplex-quadruplex conversions probed by Raman spectroscopy. *Biochemistry.* 1994; 33:7848–7856. [PubMed: 8011648]

48. Hardin CC, Henderson E, Watson T, Prosser JK. Monovalent cation induced structural transitions in telomeric DNAs: G-DNA folding intermediates. *Biochemistry*. 1991; 30:4460–4472. [PubMed: 2021636]
49. Han H, Hurley L, Salazar M. A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic Acids Res*. 1999; 27:537–542. [PubMed: 9862977]
50. Risitano A, Fox KR. Stability of intramolecular DNA quadruplexes: comparison with DNA duplexes. *Biochemistry*. 2003; 42:6507–6513. [PubMed: 12767234]
51. Guo Q, Lu M, Kallenbach NR. Effect of thymine tract length on the structure and stability of model telomeric sequences. *Biochemistry*. 1993; 32:3596–3603. [PubMed: 8466901]
52. Víglaský V, Bauer L, Tluczková K. Structural features of intra- and intermolecular G-quadruplexes derived from telomeric repeats. *Biochemistry*. 2010; 49:2110–2120. [PubMed: 20143878]
53. Chalikian T, Völker J, Plum G, Breslauer K. 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.* 1999; 96:7853–7858. [PubMed: 10393911]
54. Fojtik P, Kejnovska I, Vorlickova M. The guanine-rich fragile X chromosome repeats are reluctant to form tetraplexes. *Nucleic Acids Res*. 2004; 32:298–306. [PubMed: 14718550]
55. Lane A, Peck B. Conformational flexibility in DNA duplexes containing single G-G mismatches. *Eur. J. Biochem*. 2008; 230:1073–1087. [PubMed: 7601138]
56. Leonard G, Booth E, Brown T. Structural and thermodynamic studies on the adenine.guanine mismatch in B-DNA. *Nucleic Acids Res*. 1990; 18:5617–5623. [PubMed: 2216754]
57. Napierala M, Michalowski D, De Mezer M, Krzyzosiak W. Facile FMR1 mRNA structure regulation by interruptions in CGG repeats. *Nucleic Acids Res*. 2005; 33:451–463. [PubMed: 15659577]
58. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 2003; 31:3406–3415. [PubMed: 12824337]
59. Qin Y, Rezler E, Gokhale V, Sun D, Hurley L. 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*. 2007; 35:7698–7713. [PubMed: 17984069]
60. Zhou J, Yuan G, Liu J, Zhan C. Formation and stability of G-quadruplexes self-assembled from guanine-rich strands. *Chem.-Eur. J*. 2006; 13:945–949. [PubMed: 17036297]
61. Völker J, Plum G, Klump H, Breslauer K. Energetic coupling between clustered lesions modulated by intervening triplet repeat bulge loops: Allosteric implications for DNA repair and triplet repeat expansion. *Biopolymers*. 2009; 93:355–369. [PubMed: 19890964]
62. Ludwig A, Raske C, Tassone F, Garcia-Arocena D, Hershey J, Hagerman P. Translation of the FMR 1 mRNA is not influenced by AGG interruptions. *Nucleic Acids Res*. 2009; 37:6896–6904. [PubMed: 19752155]

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 are 1 μ M DNA in 10 mM sodium phosphate, 100 mM KCl, pH 7.5. Lanes 1, 2 and 3 contain DNA alone, DNA cycled through heating methods, and piperidine-treated DNA, respectively. Lane 4 contains DNA incubated for 30 min at 37 $^{\circ}$ 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 solid arrow being the most reactive,

open arrow being least reactive, and the striped arrow indicating an intermediate amount of reactivity. The asterisk represents the location of the ^{32}P -radiolabel.

Figure 2.

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

Figure 3. (A) Optical and (B) calorimetric analysis of (CGG)₁₉ at 1.9 μ M and 72 μ M, respectively, in 10 mM sodium phosphate, 100 mM KCl, pH 7.5. For optical data solid line represents data obtained while heating sample from 25 to 95 °C and dotted line represent data obtained while cooling sample from 95 to 25 °C.

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

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 are 1 μ M DNA in 10 mM sodium phosphate, 100 mM KCl, pH 7.5. Lanes 1, 2 and 3 contain DNA alone, DNA cycled through heating methods, and piperidine-treated DNA, respectively. Lane 4 contains DNA incubated for 30 min at 37 $^{\circ}$ 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 solid arrow being the most reactive, open arrow being least

reactive, and the striped arrow indicating an intermediate amount of reactivity. The asterisk represents the location of the ^{32}P -radiolabel.

Table 1

DNA Sequences Used in this Study

Name	Nucleotide Sequence
(CGG) ₁₉	5'-(CGG) ₁₉ -3'
1AGG-a	5'-(CGG) ₄ <u>AGG</u> (CGG) ₁₄ -3'
1AGG-b	5'-(CGG) ₈ <u>AGG</u> (CGG) ₁₀ -3'
2AGG	5'-(CGG) ₄ <u>AGG</u> (CGG) ₉ <u>AGG</u> (CGG) ₄ -3'
(CGG) ₃₉	5'-(CGG) ₃₉ -3'
4AGG	5'-(CGG) ₄ <u>AGG</u> (CGG) ₉ <u>AGG</u> (CGG) ₉ <u>AGG</u> (CGG) ₉ <u>AGG</u> (CGG) ₄ -3'

Table 2

UV-Visible-Derived Thermodynamic Parameters for (CGG)₁₉ Repeat Series

Substrate	T _m ^a (°C)	ΔH ^{a,b} (kcal/mole)	ΔG ^{a,c} (kcal/mole)	ΔS ^{a,b} (cal/(mole•K))	ΔT _m ^a (°C)	ΔΔH (kcal/mole)	ΔΔG (kcal/mole)	ΔΔS (cal/(mole•K))
(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

^aDNA in 10 mM sodium phosphate, 100 mM KCl, pH 7.5 buffer. Error represents standard deviation from a minimum of three experiments.

^bValues derived from van't Hoff analysis as described by Marky and Breslauer (35).

^cValues at 37 °C.

^dNot statistically different from (CGG)₁₉ as determined by Student's t-test.

Table 3

DSC-Derived Thermodynamic Parameters for (CGG)₁₉ Repeat Series

Substrate	T _m ^a (°C)	ΔH ^{a,b} (kcal/mole)	ΔG ^{a,c} (kcal/mole)	ΔS ^{a,b} (cal/(mole•K))	ΔT _m ^a (°C)	ΔΔH (kcal/mole)	ΔΔG (kcal/mole)	ΔΔS (cal/(mole•K))
(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

^aDNA in 10 mM sodium phosphate, 100 mM KCl, pH 7.5 buffer. Error represents standard deviation from a minimum of three experiments.

^bValues derived directly by integration of the excess heat capacity curve (35).

^cValues at 37 °C.

Table 4

UV-Visible-Derived Thermodynamic Parameters for (CGG)₃₉ Repeat Series

Substrate	T _m ^a (°C)	ΔH ^{a,b} (kcal/mole)	ΔG ^{a,c} (kcal/mole)	ΔS ^{a,b} (cal/(mole•K))	ΔT _m ^a (°C)	ΔΔH (kcal/mole)	ΔΔG (kcal/mole)	ΔΔS (cal/(mole•K))
(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

^aDNA in 10 mM sodium phosphate, 100 mM KCl, pH 7.5 buffer. Error represents standard deviation from a minimum of three experiments.

^bValues derived from van't Hoff analysis as described by Marky and Breslauer (55).

^cValues at 37 °C.