Influence of Sequence Context and Length on the Structure and Stability of Triplet Repeat DNA Oligomers†

Anthony M. Paiva and Richard D. Sheardy*

Department of Chemistry and Biochemistry, Seton Hall University, South Orange, New Jersey 07079

Received March 23, 2004; Revised Manuscript Received August 30, 2004

ABSTRACT: Genetic expansion diseases have been linked to the properties of triplet repeat DNA sequences during replication. The most common triplet repeats associated with such diseases are CAG, CCG, CGG, and CTG. It has been suggested that gene expansion occurs as a result of hairpin formation of long stretches of these sequences on the leading daughter strand synthesized during DNA replication [Gellibolian, R., Bacolla, A., and Wells, R. D. (1997) *J. Biol. Chem. 272*, 16793−7]. To test the biophysical basis for this model, oligonucleotides of general sequence $(CNG)_n$, where N = A, C, G, or T and n = 4, 5, 10, 15, or 25, were synthesized and characterized by circular dichroism (CD) spectropolarimetry, optical melting studies, and differential scanning calorimetry (DSC). The goal of these studies was to evaluate the influence of sequence context and oligomer length on their secondary structures and stabilities. The results indicate that all single oligomers, even those as short as 12 nucleotides, form stable hairpin structures at 25 °C. Such hairpins are characterized by the presence of N:N mismatched base pairs sandwiched between G:C base pairs in the stems and loops of three to four unpaired bases. Thermodynamic analysis of these structures reveals that their stabilities are influenced by both the sequence of the particular oligomer and its length. Specifically, the stability order of CGG > CTG > CAG > CCG was observed. In addition, longer oligomers were found to be more stable than shorter oligomers of the same sequence. However, a stability plateau above 45 nucleotides suggests that the length dependence reaches a maximum value where the stability of the G:C base pairs can no longer compensate the instability of the N:N mismatches in the stems of the hairpins. The results are discussed in terms of the above model proposed for gene expansion.

Genetic expansion diseases, such as fragile-X syndrome, have been linked to the properties of trinucleotide repeat (also known as triplet repeat) DNA sequences during replication. Of the 64 possible combinations of triplet repeats available within the genetic code, approximately seven have been linked to human disease. Four of these sequences, CTG, CAG, and CCG/CGG have been shown to be involved in the pathogenesis of myotonic dystrophy (DM), Huntington's disease (HD), and fragile-X syndrome (FRAX), respectively (*1*−*3*). In addition, the repeat sequences AAG and CTT have been linked to Freidrich's ataxia. A characteristic common to all triplet expansion diseases is a novel mutational mode termed "dynamic mutation" that leads to an expansion in repeat copy number during replication. Below the disease causation thresholds, (<20 for the CAG repeat and <100 for the CCG/CGG repeat) the number of triplet repeat codons passed from parent to child can remain stable for generations. Once a critical number of repeats have accumulated, an unknown mechanism triggers further expansion. The expansion is currently believed to be the result of abnormal DNA-structure formation that either interferes with downstream mRNA processing or with the ability of the DNA polymerase complex to replicate the template strand. It has been shown that the inherent flexibility of triplet-repeat DNA can lead to an accumulation of superhelicity within triplet repeat DNA tracts (*11*, *12*). These authors suggest the accumulated superhelical strain becomes so great that the DNA_{pol} can no longer proceed with unwinding of the double-stranded DNA. Subsequent cellular events lead to the formation of hairpin structures on the daughter DNA strands and dissociation of the DNA_{pol} complex. If hairpin formation occurs prior to DNA_{pol} reassociation, then the DNA_{pol} will attach to a location upstream (3′ with respect to the parental strand position) of the dissociation point, leading to reiterative synthesis and expansion. Stable hairpin formation will eventually lead to reiterative synthesis, whereas unstable daughter-strand hairpins will eventually reassociate with the parental strand, leading to no expansion in repeat number.

Critical to any discussion about the replication process as it relates to triplet repeat DNA is definition of the structures formed and the stabilities of these structural motifs in terms of the associated thermodynamics. Triplet repeat DNA sequences can spontaneously form stable structures in vitro due to the self-complementarity and inherent strand flexibility that allows alternative hydrogen-bonding arrangements within the resultant DNA secondary structure. The structures formed from the $(CNG)_4$ oligonucleotides, (where N is A, C, T, or G) can be variable and depends on the solution conditions (pH, the ionic strength, and DNA concentration). For example, Zheng and co-workers (*5*) demonstrated that all

† The State of New Jersey Commission on Higher Education provided funds for the acquisition of the differential scanning calorimeter, and Merck, Inc. supported A.P. through their Doctoral Fellowship Program.

* Corresponding author. Phone: (973) 761-9030. Fax: (973) 761-9772. E-mail: sheardri@shu.edu.

10.1021/bi0494368 CCC: $27.50
Published on Web 10/12/2004

CNG triplet repeats form antiparallel hairpins that can readily undergo structural transitions due to increased flexibility within the sequence. On the basis of DEPC modification, Mitas and co-workers (*6*) found that the T:T mismatches in CTG repeats are stacked within the hairpin helix, whereas A:A mismatches in CAG are not well-stacked. CGG repeats have also been shown to form tetraplex hairpins, while CCG can adopt two distinct hairpin conformations. Fry and colleagues observed that CGG repeats could also form stable tetraplexes (*7*), possibly via a foldback hairpin (*8*).

In concert with structural evidence, thermodynamic properties lead to a model-independent understanding of triplet repeat structural stabilities (*9*, *10*). Petruska and co-workers (*9*) determined that the relative order of the 30-mer stability in 20 or 120 mM sodium chloride was CTG > GAC = CAG > GTC. However, the melting temperature (T_m) of the CAG and CTG triplet repeats changed very little as the sequence length changed from 30 to 90 bp. Although the enthalpic contribution was expected to show a 3-fold increase from a 30 to 90 bp sequence, the observed change was only 1.4-fold. On the basis of free energy and van't Hoff enthalpy calculations, Zheng predicted a different stability hierarchy for $(CNG)_4$ repeats (CGG > CAG > CTG > CCG (*5*)). Gacy and McMurray (*10)* find little difference in the thermodynamic stability of triplet repeats, as indicated by the change in enthalpy for the hairpin to single strand transition for $(CTG)_{25}$ and $(CTG)_{10}$ or $(CAG)_{25}$ and $(CAG)_{10}$.

The work of Gellobolian (*11*) and Bacolla (*12*) suggests that the physical properties of the hairpin structures formed will critically impact triplet repeat expansion. The work presented here is the first systematic analysis of the structural and thermodynamic properties that relate triplet repeat tract sequences to their length and base context. We show that the stabilities of the secondary structures formed from $(CNG)_n$ are both sequence context (N) and length (*n*) dependent.

MATERIALS AND METHODS

Oligomer Design, Synthesis, and Purification. To fully evaluate sequence context and length effects on the structures and thermodynamic stabilities of DNA triplet repeat oligomers, we chose to synthesize and characterize oligomers of general sequence $(CNG)_n$, where N = A, C, T, or G and *n* = 4, 5, 10, 15, or 25. All DNA oligomers were synthesized using the phosphoramidite method on an Applied Biosystems Inc. (Foster City, CA) model 380B DNA synthesizer with purification by trityl-selective RPHPLC[1] as previously described (*13*). The β-cyanoethylphosphoramidites and other DNA synthesis reagents were obtained either from Glen Research, Inc. or CruaChem, Inc. After detritylation and a second purification by RPHPLC, the purified DNA oligomers were lyophilized to dryness prior to storage. All DNA solutions were stored as either lyophilized powders or frozen at −20 °C in a solution of 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0. Purity and size analysis of the oligomers was accomplished through a combination of RPHPLC and/or denaturing polyacrylamide gel electrophoresis. The DNA concentration and yield were determined by spectrophotometric absorbance, using an extinction coefficient based upon the summed extinction coefficients of the individual nucleotide bases (*14)*. The evaluated extinction coefficients are available in Table 1S of the Supporting Information

Circular Dichroism Spectropolarimetry (CD). Circular dichroism spectra were recorded at 25 °C with an AVIV Associates CD spectropolarimeter model 62-ADS equipped with a multiple cell turret and Peltier heating/cooling device. All measurements were obtained in a nitrogen environment. The DNA oligomers or duplexes were prepared in standard phosphate buffer (5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0) to final total strand concentration between 1 and 9 μM. The solution was heated to 90−100 °C for 20 min, then slowly cooled to 25 °C over 120 min, and stored at 2−5 °C for 48 h. Prior to all CD measurement the samples were transferred from 2 to 5 °C storage into stoppered cuvettes and equilibrated at 25 °C for 15−20 min. The CD spectrum of the DNA sample was recorded from 320 to 210 nm. All CD data, reported as the average of triplicate scans, were blank-corrected and processed with Aviv DOS-based software for subsequent export into graph-plotting software.

Differential Scanning Calorimeter (DSC). Heat capacity measurements were made with a Nano-II power-compensation differential scanning calorimeter (Calorimetric Sciences Corp.). The calorimeter was interfaced directly to a Windows-based computer with "DSCRun" version 2.1.1 software (Calorimetric Sciences Corp.) for both data collection and DSC control. Samples were either dialyzed or exhaustively washed with a Poly Pak II reversed phase extraction cartridge into the specified buffer solution and then thoroughly degassed and filtered prior to measurement. All solutions for the reference side of the calorimetric measurement were derived from the buffer used to prepare and equilibrate the sample. Each DNA sample was thermally scanned from 10 to 95 °C at 1 °C/min over eight forward and eight reverse scans. The samples were equilibrated for 10 min at the upper temperature (reverse scan) or lower (forward scan) temperature set points between thermal scans. Buffer versus buffer reference scans were performed for every sample if buffer conditions changed (e.g. low salt to high salt) or every three samples if buffering conditions were consistent within a sample series. The raw thermographic data was blank corrected and integrated with "CpCalc" version 2.1 (Calorimetric Sciences Corp.) software. The pre- and post-transition baselines for most samples were linear and enabled integration of the calorimetric enthalpy through the use of a point-to-point linear baseline.

UV−Visible Spectroscopy. Cary model 300E Bio and Cary model 100E spectrophotometers (Varian, Inc.) were used for all spectral measurements. The spectrometers were interfaced to Windows-based computers with Cary/Varian WinUVBio version 2.0 software. Both were equipped with a Peltier thermoelectric heating/cooling block, a multicell-transport device, and a nitrogen-purged sample compartment. The wavelength of maximum absorbance difference between the native form at 10 °C and the denatured form at 90 °C was used for thermal scanning of each oligomer. Oligonucleotide solutions with DNA concentrations ranging from 0.05 to 5 μM in bases in standard phosphate buffer at 115 mM total Na^+ were heated to 90 °C for 20 min and then allowed to cool to 4 °C over 120 min. The samples were equilibrated

[1] Abreviations: CD, circular dichroism; DSC, differential scanning calorimetry; RPHPLC, reverse phase high performance liquid chromatography; T_m, melting temperature.

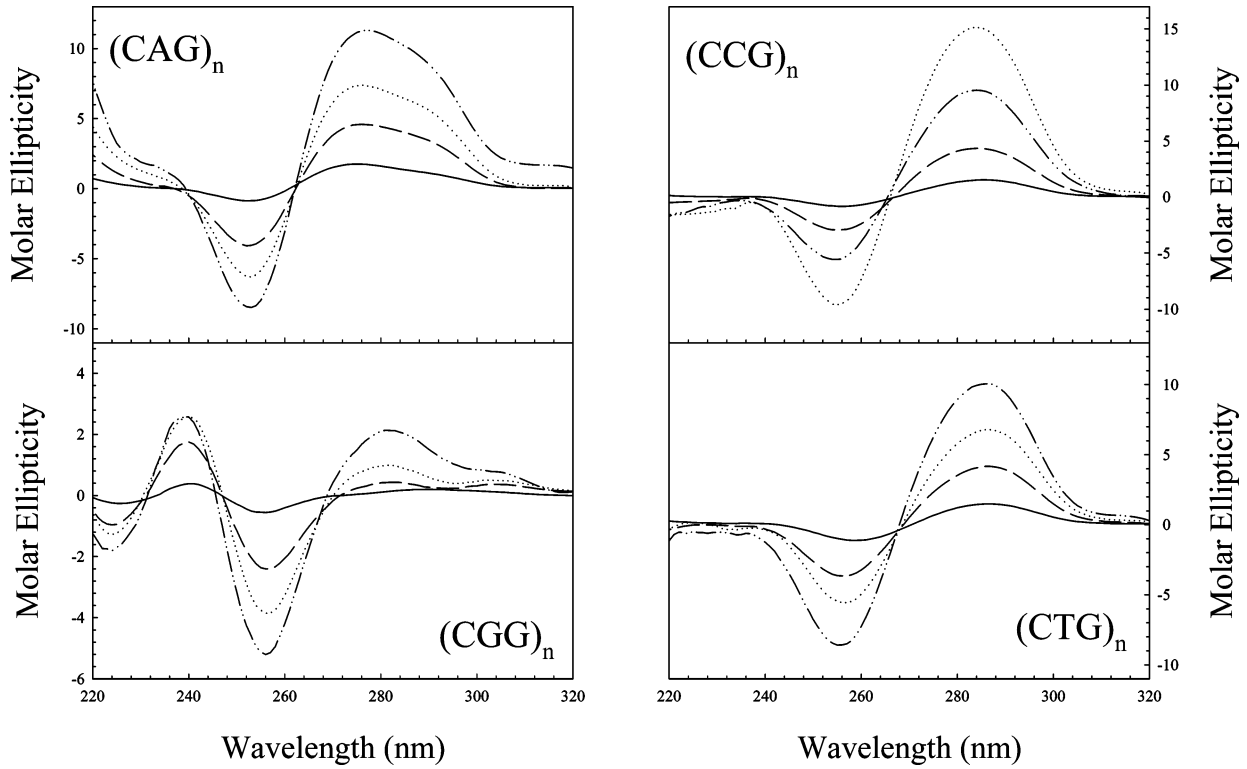

FIGURE 1: Circular dichroism spectra of the individual $(CNG)_n$ oligonucleotides as a function of oligomer length: $n = 4$ (solid), $n = 10$ (dash), $n = 15$ (dot), and $n = 25$ (dash−dot−dot). CD spectra were determined in standard 10 mM phosphate buffer, pH 7.0, 115 mM Na^+ at 25 °C. Molar ellipticities are reported in terms of strand concentration.

at 4 °C for 12−24 h prior to measurement. After equilibration the oligomer solutions were transferred to cuvettes equilibrated at the experimentally relevant start temperature and equilibrated for an additional 20 min. Typically, the temperature of each DNA sample was increased from 10 to 90 °C at approximately 0.5 °C/min over a minimum of five forward and five reverse scans. The melting point values (T_m) for all DNA oligomers were determined with Varian Cary software version 2.0 using a hyperchromicity model to fit the UV-melt data-set to a self-complementary model with $n = 1$.

RESULTS

Length and sequence dependent relationships between the triplet repeat DNA sequences and disease onset have been proposed. The in vivo structural and thermodynamic properties of triplet repeat DNA are here modeled in vitro with oligonucleotides of the $(CNG)_n$ class, where N = A, C, T, or G and $n = 4$, 5, 10, 15 or 25 repeat units. Thus, the major class of triplet repeat DNA sequences, shown to be the causative agents for Huntington's Disease and fragile-X syndrome, are represented at both the non-disease-causing and disease-causing thresholds. The oligomer set not only allows comparisons across different lengths but also between sequence contexts. To date, the structures formed by triplet repeat DNA sequences remains defined for only a limited data set of oligomer lengths and sequence contexts. The structures formed by the triplet repeat sequences studied here are characterized by circular dichroism and optical melting studies.

As can be seen in Figure 1, the CD spectra of the CAG, CCG, and CTG series of oligomers are similar to each other and, with the exception of the CGG oligomers, similar to the CD spectral signature of DNA in the B conformation with peaks centered in the region of 275−280 nm and troughs centered in the region of 240−255 nm. The generally observed increases in ellipticities for both the peaks and troughs are consistent with increasing base-stacking interactions as the length of the oligomers increase. For all samples, the CD spectra remain relatively unchanged with respect to the overall shape and locations of the peak maxima/minima.

The CD spectra for the CGG oligomers (Figure 1) with peaks centered at 303, 282, and 240 nm and two troughs located at 256 and 223 nm for $n = 10-25$ cannot be readily classified by referral to a reference spectrum. The CD spectrum of the 12-mer differs from that of the longer oligomers, with two peaks centered at 292 and 240 nm and two troughs centered at 255 and 225 nm. The overall trends of the spectra suggest that the base-stacking interactions, as indicated by the increase in peak height and the increase in trough depth, becomes larger as the length of the oligonucleotide increases. Overall, the CD spectra of the CGG oligomers resemble a multitude of different CD spectra, none of which form a complete match to the observed CGG spectra. For example, the CD spectra of the CGG oligomers are somewhat reminiscent of a quadruplex DNA with a CD spectral appearance of a trough−peak−trough motif (*15*, *16*).

The CD spectra of the complementary $(CAG)_n/(CTG)_n$ duplexes are consistent with B-form DNA with peaks centered near 274 nm, a slight shoulder appearing at approximately 290 nm, a crossover point occurring at 263 nm for the 12-mer and 30-mer and at 261 nm for all other duplexes, and troughs visible at 254 nm. Figure 2 shows the comparison between the characteristic CD spectrum of the 25-mer with the spectra of the complimentary oligonucleotides. The CD spectra suggest that the structures are forming

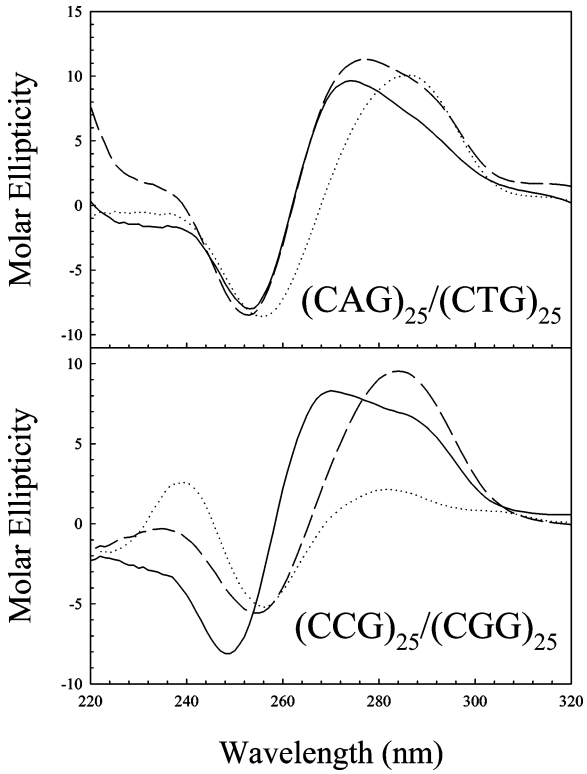

FIGURE 2: Comparison of the CD spectra of the duplex to its component complementary strands for $(CAG)_{25}$ (dash), $(CTG)_{25}$ (dot), and the corresponding duplex $[(CAG)_{25}]:[(CAT)_{25}]$ (solid) in the upper panel and $(CCG)_{25}$ (dash), $(CGG)_{25}$ (dot), and the corresponding duplex $[(CCG)_{25}]:[(CGG)_{25}]$ (solid) in the lower panel. CD spectra were determined in standard 10 mM phosphate buffer, pH 7.0, 115 mM Na^+ at 25 °C. Molar ellipticities are reported in terms of strand concentration.

duplexed B-form DNA structures with full Watson−Crick complementarity between the matched nucleotides.

The CD spectra for the $(CCG)_n/(CGG)_n$ duplexes also suggest B-form DNA with subtle variations observed. For these duplexes there is a major peak centered between 274 and 270 nm with a shoulder between 286 and 284 nm and a trough occurring between 251 nm for the 12-bp duplex to 248 nm for the 75-bp duplex. Figure 2 shows the comparison between the characteristic CD spectrum of the 25-mer with the spectra of the complimentary oligonucleotides.

Comparison of the $(CNG)_{25}$ oligomers with the complementary duplexes (Figure 2) reveals that the peak maxima for all solitary oligomers are red-shifted relative to the complementary duplexes. In addition, the different locations of the maxima and minima, along with the subtle changes in the CD spectra, also indicate that the structures formed by the solitary oligomers differ from the structures formed by the complementary Watson−Crick duplexes.

As noted above, CD spectral data indicate that the CAG, CCG, and CTG oligomers form DNA structures that are similar to B-form DNA and the CGG oligomers form an undefined structure. The triplet repeat DNA sequences under consideration can form two possible structures—either unimolecular hairpin or bimolecular duplex—due to the nearly self-complementary nature of the sequences. Both structures would possess N:N mismatched base pairs and neither structure can be discriminated solely on the basis of CD. To clarify the type of structure formed, the concentration dependence of the melting transition, T_m, was measured via optical melting studies. Such measurements can determine the molecularity of the DNA denaturation (*17*) and, by deduction, eliminate one or more of the possible structural forms under consideration.

Plots of T_m vs ln C_T (where C_T is total strand concentration) for the $(CNG)_n$ series of oligomers, where $n = 4$, 10, or 15 and N = A, C, or T, and for $(CGG)_{10}$ and $(CGG)_{15}$ produced near-zero slopes, indicating unimolecular denaturations (Figure 3). The plot for $(CGG)_4$, however, displayed a high degree of variability that can be interpreted in one of two ways. First, the biphasic nature of the plot, where one-half contains a positive slope and the other half contains a negative slope, suggests that the $(CGG)_4$ oligomer forms two different structures in a concentration-dependent manner. Alternatively, the trend of the melting temperatures could be biased toward large errors due to the small difference between the denatured and native extinction coefficients for this oligomer. Errors made to the pre- and post-transition baselines for plots with shallow slope can have enormous effects on the measured T_m. Both interpretations were equally valid and required additional experiments to rule out the possibility of quadruplex or other structures. However, native gel electrophoresis of all oligomers produced single band migration patterns consistent with unimolecular DNA structures (Supporting Information, Figure 1S). Together, the CD and optical melting studies suggest that the individual oligomers form unimolecular hairpin like structures.

The thermodynamic parameters for the CNG triplet repeat sequences, where N = A, C, G, or T, in conjunction with the complementary triplet repeat duplexes, $(CNG)_n/(CNG)_n$, where N = A/T or C/G, were determined with differential scanning calorimetry in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0, 100 mM NaCl, with DNA concentrations between 20 and 100 μM. A typical DSC thermogram is shown in Figure 4. The reported thermodynamic parameters are expressed as the enthalpy, entropy, and free energy for the hairpin (HP) to single-strand (SS) transition or duplex (DUP) to single-strand (SS) transition. The results of these studies are tabulated in Tables 1 and 2. For the hairpin-forming oligomers (Table 1), the enthalpic contribution to the total free energy of hairpin denaturation, in general, increases with increasing oligomer length. This increase is compensated by a concomitant increase in the entropic contribution to the total free energy of hairpin denaturation as the oligomer becomes longer. Thus, the free energy for the HP to SS transition increases with increasing oligomer length. Overall, the thermodynamic trends indicated that the stability of the $(CNG)_n$ oligomers increases as a function of length and that the increases are enthalpically driven.

Comparison of overall oligomer to oligomer stability shows the following stability order, in terms of the free energy, of the HP to SS transition is (most stable to least stable) CGG > CTG > CAG > CCG. In contrast, the enthalpy trend that shows a reversal in interactions for the CCG and CAG oligomers: CGG > CTG > CCG > CAG. The observed reversal is due most likely to the relatively large contribution made by the entropic term of the $(CCG)_n$ oligomers that serves to reduce the overall stability.

The equilength complementary duplexes of the $(CNG)_n/(CNG)_n$ oligomers, where N = A/T or C/G, were also studied by differential scanning calorimetry (Table 2). The melting

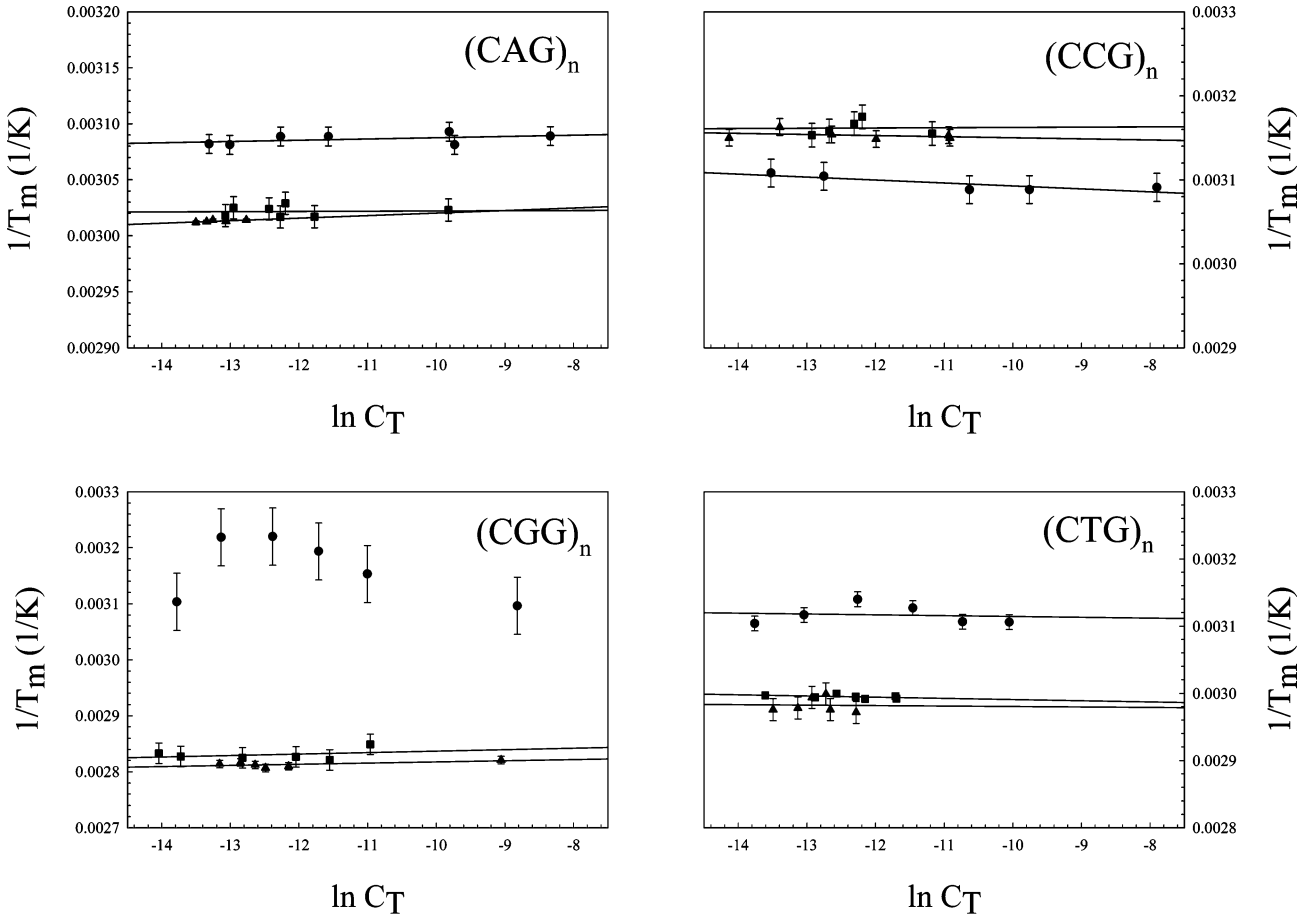

FIGURE 3: Plots of $1/T_m$ vs ln C_T where T_m is the melting temperature determined by optical melting studies and C_T is the total strand concentration for the $(CNG)_n$ oligomers with $n = 4$ (circles), $n = 10$ (squares), and $n = 15$ (triangles). The lines represent least-squares linear fits. Optical melting studies were determined in standard 10 mM phosphate buffer, pH 7.0, 115 mM Na^+.

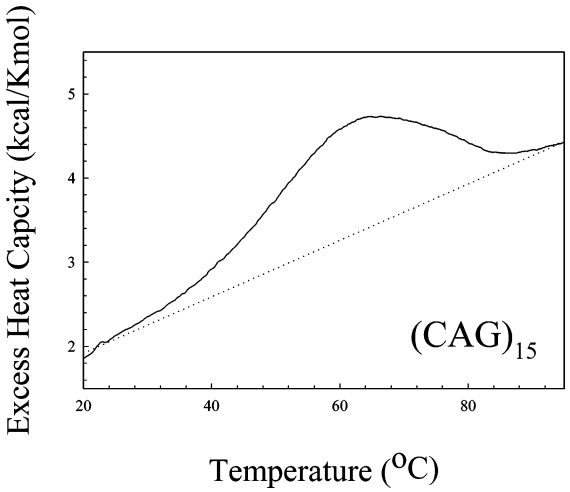

FIGURE 4: Typical differential scanning calorimetric thermogram for $(CAG)_{15}$. The solid line is the thermogram and the dotted line is the determined baseline. Integration of the peak yields ΔH for the transition. All determined thermodynamic parameters are listed in Tables 1 and 2. DSC thermograms were determined in standard 10 mM phosphate buffer, pH 7.0, 115 mM Na^+.

transitions for the CCG/CGG duplex series were too high for accurate measurement by DSC and were therefore not integrated, as in the case of $(CCG)_{10}/(CGG)_{10}$, or not measured, as in the case of $(CCG)_{15}/(CGG)_{15}$ and $(CCG)_{25}/(CGG)_{25}$. The remaining duplex set, the $(CAG)_n/(CTG)_n$ duplexes, were within the realm of accurate measurement and were quantified by DSC with the exception of $(CAG)_{25}/(CTG)_{25}$ oligomer, whose transition temperature was too high for accurate quantitation and was thus not entered into the

Table 1: DSC-Derived Thermodynamic Parameters for the $(CNG)_n$ Oligomers[a]

oligomer	length (nucleotides)	ΔH (kcal/mol)	ΔS (cal/kmol)	$\Delta G_{37°C}$ (kcal/mol)
$(CAG)_4$	12	5.5 ± 0.2	16.4 ± 0.2	0.4 ± 0.02
$(CAG)_5$	15	16.9 ± 0.7	49.8 ± 2.3	1.5 ± 0.07
$(CAG)_{10}$	30	44.7 ± 1.9	132 ± 5.0	3.4 ± 0.2
$(CAG)_{15}$	45	56.4 ± 1.7	168 ± 5.9	4.3 ± 0.2
$(CAG)_{25}$	75	114 ± 4.1	339 ± 12	9.0 ± 0.4
$(CCG)_4$	12	5.7 ± 0.2	17.8 ± 0.8	0.2 ± 0.01
$(CCG)_5$	15	8.7 ± 0.4	26.5 ± 1.3	0.5 ± 0.03
$(CCG)_{10}$	30	57.0 ± 2.3	177 ± 7.1	2.4 ± 0.1
$(CCG)_{15}$	45	76.6 ± 2.7	235 ± 8.2	3.5 ± 0.1
$(CCG)_{25}$	75	102 ± 3.8	315 ± 12	4.0 ± 0.2
$(CGG)_4$	12	23.2 ± 1.2	70.3 ± 3.5	1.4 ± 0.07
$(CGG)_5$	15	9.5 ± 0.4	28.5 ± 1.2	0.7 ± 0.03
$(CGG)_{10}$	30	92.8 ± 3.2	261 ± 9.1	11.9 ± 0.4
$(CGG)_{15}$	45	148 ± 6.2	413 ± 17	19.6 ± 0.8
$(CGG)_{25}$	75	148 ± 6.7	411 ± 18	20.5 ± 0.9
$(CTG)_4$	12	1.9 ± 0.1	6.0 ± 0.3	0.0 ± 0.2
$(CTG)_5$	15	23.4 ± 0.9	73.8 ± 2.8	0.6 ± 0.02
$(CTG)_{10}$	30	77.1 ± 3.5	228 ± 10	6.4 ± 0.3
$(CTG)_{15}$	45	126 ± 4.4	371 ± 13	10.9 ± 0.4
$(CTG)_{25}$	75	165 ± 5.9	486 ± 17	14.5 ± 0.5

[a] The thermodynamic parameters of the $(CNG)_n$ oligomers, derived from differential scanning calorimetry measurements, are shown below as a function of oligomer length. For all thermodynamic parameters listed above, the assumed direction is the hairpin to single-strand transition.

Table 2: DSC-Derived Thermodynamic Parameters for the $(CNG)_n/(CNG)_n$ Duplexes[a]

duplex	length (base pairs)	ΔH (kcal/mol)	ΔS (cal/kmol)	$\Delta G_{37°C}$ (kcal/mol)
$(CAG)_4/(CTG)_4$	12	52.2 ± 2.3	155 ± 6.9	4.0 ± 0.02
$(CAG)_{10}/(CTG)_{10}$	30	138 ± 5.2	386 ± 15	18.5 ± 0.7
$(CAG)_{15}/(CTG)_{15}$	45	220 ± 9.5	609 ± 26	31.5 ± 1.4
$(CAG)_{25}/(CTG)_{25}$	75	b	b	b
$(CCG)_4/(CGG)_4$	12	51.2 ± 2.5	145 ± 7.2	6.2 ± 0.3
$(CCG)_{10}/(CGG)_{10}$	30	*c*	*c*	*c*
$(CCG)_{15}/(CGG)_{15}$	45	*b*	*b*	*b*
$(CCG)_{25}/(CGG)_{25}$	75	*b*	*b*	*b*

[a] The thermodynamic parameters of the $(CNG)_n/(CNG)_n$ duplexes, derived from differential scanning calorimetry measurements, are shown as a function of duplex length and were adjusted for duplex concentration. For all thermodynamic parameters listed above, the assumed direction is the duplex to single-strand transition. [b] Samples with an apparent melting transition that was near or above 95 °C. [c] Samples with incomplete transitions, where the endpoint of the thermogram could not be clearly defined.

data set for analysis. The enthalpy for the $(CAG)_4/(CTG)_4$ through $(CAG)_{15}/(CTG)_{15}$ samples became more favorable with increasing length and increased from 50 kcal/mol for the 12-bp duplex to 220 kcal/mol for the 45-bp duplex. The calorimetrically derived entropy for these duplexes became less favorable with increasing duplex length, going from 155 cal/kmol for the 12-bp duplex to 610 cal/kmol for the 45-bp duplex. For these duplexes, the free energy of the duplex to single strand transition increases with increase duplex length.

DISCUSSION

There has been much discussion about the nature of the structures formed by triple repeat DNA sequences. Because of the diverse solution conditions employed by the different research laboratories and the lack of a complete database encompassing both sequence and length comparison, we undertook the task of characterizing the structures formed by triplet repeat DNA under physiological conditions. The goal for such a systematic thermodynamic study is similar to that of the structural studies: to define the database of thermodynamic properties and to determine if a correlation exists between stability and either the length or the sequence context of triplet repeat DNA.

Triplet Repeat Oligomers Form Unimolecular Hairpin Structures Similar to B-Form DNA. Analysis of the CD spectra shows that the individual oligomers form stable DNA secondary structures even when as short as 12 nucleotides. Further, the overall nature of these structures changes very little within a sequence class, regardless of the length of the oligomer under consideration. While there are minor changes in the height of a peak or trough, indicative of changes in stacking interactions, or changes in the location of the maxima/minima, indicating changes in hydration, the overall structures remain similar. With the exception of the $(CGG)_n$ oligomers, all CD spectra bear a striking resemblance to each other and to the idealized CD spectra of B-form DNA. As previously noted, the CD spectral signals are not exactly the same as B-form DNA due possibly to two reasons. First, triplet repeat DNA sequences have an unusually high GC content that ranges from 66% G/C for CAG and CTG to 100% G/C for CCG or CGG. It has been shown by Johnson that DNA sequence content can alter the observed CD spectrum of a DNA molecule from that of an idealized B-form CD spectrum (*18*). Second, the presence of mismatches located every third nucleotide can serve to alter the CD spectrum by introducing structural changes that affect the nucleotide interactions.

While the CD spectra indicate that the $(CNG)_n$ oligomers form stable structures, there are two possible conformers that can be formed from the DNA sequences under consideration, either a unimolecular hairpin or a bimolecular duplex, where every third nucleotide in the stem of the hairpin or in the duplex is mismatched. Both the hairpin and duplex structures should produce similar CD spectra, due to the structural similarity between the duplex and the hairpin stem. The loop region of the hairpin, however, should not significantly contribute to the CD spectrum, due to the absence of base-stacking interactions and the small size of the loop relative to the stem region.

Due to the spectral similarity between these two structural classes of DNA, there can be no assignment of structure without ancillary evidence. For this reason, the concentration dependence of the melting transition was undertaken. The DNA concentration dependence of the melting transition for all $(CNG)_n$ oligomers (where $n = 4$, 10, or 15), except $(CGG)_4$, were essentially independent of DNA concentration, indicating that all oligomers form structures that melt as unimolecular units.

To further define the type of structures formed, the oligomers were also examined with native gel electrophoresis. All solitary sequences migrate as single bands with mobilities higher than the corresponding complementary duplex. On the basis of native gel electrophoresis alone, the data indicate that all oligomers form unimolecular structures that are likely to be hairpins. However, as demanded by the nature of the triplet repeat sequence, the hairpins do not maintain full complementarity, due to the mismatch located at every third nucleotide within the sequence. Taken together, the data indicate that the $(CNG)_n$ oligomers with N = C, A, or T, under solution conditions specified form hairpin structures with variable stem and loop regions, as depicted in Figure 5.

As previously noted, the $(CGG)_n$ oligomers display unusual CD spectra that could not be readily assigned to a particular structural class of DNA. Consequently, determination of the $(CGG)_n$ oligomer structure relied upon electrophoretic and optical melting studies. The $(CGG)_n$ oligomers do not form higher order structures such as quadruplexes or slip structures within the limits of the electrophoretic measurement. All oligomers migrate as single bands on an electrophoretic gel under native conditions. In addition, no multiphasic transitions were observed by UV-thermal melting techniques (data not shown). Finally, for $(CGG)_{10}$ and $(CGG)_{15}$, the slopes of the concentration-dependence T_m curves were linear with essentially zero slopes. With the $(CGG)_4$ oligomer, however, the trend of the concentration dependence plot was not clearly defined. Despite this data set, the bulk of the experimental results are consistent with $(CGG)_n$ oligomers forming unimolecular hairpin structures. The unusual CD spectra observed for the $(CGG)_n$ oligomers can possibly be the result of reorientation of the guanine nucleotides within the mismatch regions of the oligomer hairpin structure.

Complementary Oligomers Form Duplexes. This study also shows that there are no unusual CD spectral properties

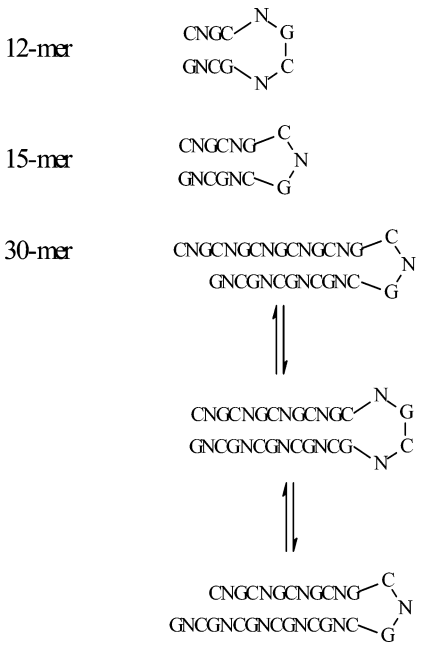

FIGURE 5: Proposed secondary structures for the $(CNG)_n$ oligomers for N = A, C, or T. The 12-mers form minihairpin-like structures with stems of three G:C base pairs and one N:N mismatch with loops of four bases. The hairpins formed from the 15-mers have stems possessing four G:C base pairs and two N:N mismatched bases with loops of three bases. The most stable hairpins for the 30-mers would possess stems of nine G:C base pairs and four N:N mismatched with loops of four bases. It is likely, however, that the longer hairpins could also form reasonably stable "slipped" structures, where one strand of the stem "slides" three bases at a time over the complimentary strand. For the 30-mer, the resultant hairpin formed by sliding only three base pairs would have a stem of eight G:C base pairs with four N:N mismatches, a loop of three bases, and an overhang (5′ or 3′) of three unpaired bases. Slippage by another three bases would generate a hairpin with a stem containing six G:C base pairs, three N:N mismatches, a loop of bases, and an overhang (5′ or 3′) of unpaired six bases. Slipped structures should, in general, be less stable than blunt-ended structures. Similar structures can also be generated for the 45-mers.

associated with the duplexes formed from complementary strands of triplet repeat DNA sequences. On the basis of CD spectral data and native gel electrophoresis, the complementary strands anneal to form B-form duplex DNA structures. Examination of the oligomers by native gel electrophoresis shows that complementary mixtures of $(CAG)_n/(CTG)_n$ or $(CCG)_n/(CGG)_n$ form structures that migrate with the expected size as the corresponding duplex DNA structure.

Proposed Structures. Together, the results indicate that each oligomeric sequence forms a unimolecular hairpin with a stem region similar to B-form DNA. Due to the alignment of the triplet repeat unit within such a structure, it is impossible to achieve full complementarity between the "sense" and "antisense" legs of the hairpin. Every third position within the primary structure of the stem contains a mismatched nucleotide. Because this structural feature is unlike B-form DNA, where all nucleotides within the sequence of the sense strand maintain full complementarity with the anti-sense strand, this structural form is 'B-like'. While there are many properties in common between the B-form and B-like DNA, the location of a mismatch at every third nucleotide introduces a localized distortion or "bulge" to the backbone of the DNA helix. This distortion imparts to the hairpin molecule unique thermodynamic and structural properties. As was previously shown, the hairpins are less stable than the fully base-paired duplex of similar length. The size of the loop is variable and can range from a minimum of three nucleotides within the loop (the minimum number required to achieve a turn) to larger sizes that are dependent upon the overall stability of the hairpin structure. Loop structures within DNA molecules can have a destabilizing influence on the thermodynamic stability of the molecule, due to torsional strain, in the case of small loops ($n = 2$), or exposure of hydrophobic groups, in the case of large loops ($n = 5$ or more) (*19*, *20*). Because the loop region can contribute negatively to the thermodynamic stability of the DNA structure, the size of the loop will vary according to the prevailing thermodynamics of the hairpin. A balance must be struck between relaxation of loop strain and exposure of hydrophobic groups within the larger thermodynamic "picture." The most probable structure formed will be the thermodynamically most stable structure that maximizes the hydrogen bonding and stacking interactions between the sense and antisense legs of the hairpin. Such a structure is the blunt-ended hairpin (Figure 5). Slip-structures, where the alignment shifts in the 5′ (N + 1 slip) or 3′ (N − 1 slip) direction by one repeat unit are also theoretically possible but not thermodynamically favored beyond an N + 1 or N − 1 slip due to the decreased stability. It is quite possible that in solution the oligomers are in rapid equilibrium between the blunt-ended, N + 1 and N − 1 structures, with the predominant form being the blunt-ended alignment. For all oligomers, only a single band was observed by native gel electrophoresis, suggesting only one predominant form of the hairpin.

Sequence Context Comparisons. Comparison of the calorimetrically derived free energies of the hairpin (HP) to single strand (SS) transitions across the $(CNG)_n$ triplet repeat sequences shows a stability order, ranking from most to least stable, of CGG > CTG > CAG > CCG. These calorimetric trends were similar to those discovered by Peyret, who found the order of mismatch stability for N:N mismatches at 37 °C to be G:G > T:T > A:A > C:C (*21*).

Structural Basis for Observed Sequence Context Stability Profiles. Because of the sequence redundancy of the triplet repeat DNA, the stability of the molecules is predicted to be dependent on the prevailing thermodynamic behavior of the N:N mismatches. The observed stability ranking, CGG > CTG > CAG > CTG, is consistent with what is known from a structural perspective for the triplet repeat sequences. Both Lane (*22*) and Zheng (*5*) studied the CGG triplet repeat sequence by solution NMR and found that the nucleotide base of the G:G mismatch can reorient about the glycosidic bond. The guanine base on the sense strand was in the anti conformation about the glycosidic bond, whereas the guanine base on the antisense strand was in the syn conformation. They also found rapid orientation exchange between opposite guanine mismatches. The reorientation of guanine within the G:G mismatch allows the formation of Hoogstein hydrogen bonds between the N-1 of the *anti*-guanine and the O-6 of the *syn*-guanine and between the C-2 amino of the *anti*-guanine and the N-7 of the *syn*-guanine. This reorientation allows two hydrogen bonds to form between the mismatches and adds to the overall stability of the DNA oligomer that otherwise would be destabilized by a G:G mismatch. The solution NMR of the T:T mismatches within the CTG triplet

repeat class were studied by Arnold et al., who found that the mismatched thymines maintain an anti orientation about the glycosidic bond (*23*). However, they also found that the T:T mismatches are capable of forming "wobble" base-pairs between O-4 of the sense strand and N-3 of the antisense strand and between N-5 of the sense strand and O-2 of the antisense strand. Like the G:G mismatch, the T:T mismatch confers additional stability to the oligomer through the formation of two hydrogen bonds and base-stacking interactions. Arnold et al. also studied the A:A mismatches in the CAG sequence context and found, like the T:T mismatches, rapidly interchanging hydrogen-bonded structures between the bases of two *anti*-adenines (*23*). There was only one hydrogen bond proposed that occurred between the C-4 amino of the sense strand and the N-3 of the antisense strand, unlike the G:G and T:T mismatches. While it is likely that the A:A mismatches had more base-stacking interactions than the T:T mismatches, overall the hydrogen bonding was sufficiently reduced to lower the stability ranking of the A:A mismatches below that observed for the T:T mismatch. Finally, the C:C mismatch, also studied by Arnold et al., shows a propensity for wobble-pair formation, much like that observed for the T:T mismatch (*22*). Like the T:T mismatch, both bases of the nucleotide are in the anti conformation. There is a single hydrogen bond formed between N-3 of the sense strand and the C-4 amino of the antisense strand.

Length Comparisons. The calorimetrically derived thermodynamic parameters were compared to examine the "macroscopic" (total length) and "microscopic" (sequence) trends in the thermodynamic parameters. Oligomers as short as 12 nucleotides long formed stable structures, and as the length of the oligomer increases, the observed thermodynamic stability also increases. These observations are consistent with increased base stacking and hydrogen-bond interactions as a function of oligomer length. The enthalpic contribution to the total free energy of duplex formation is dependent upon hydrogen-bond interactions, base-stacking interactions, and charge stabilization that cannot be factored into a theoretical enthalpy calculation, because no context-relevant data sets exist. On the basis of the structural considerations above and the enthalpy values listed in Table 1, it appears that the $(CGG)_n$ and $(CTG)_n$ oligomers possess considerable hydrogen-bond and base-stacking interactions, even with the presence of unpaired mismatches or complementary nucleotides within the stem of the hairpin. The $(CCG)_n$ oligomers, which have the lowest stability of all oligomers examined, have enthalpies higher than those measured for the $(CAG)_n$ oligomers. This suggests fewer base-stacking and hydrogen-bond interactions for both the $(CCG)_n$ and the $(CAG)_n$ oligomers.

Examination of the entropies for the hairpin to single-strand transition for the two data sets shows that the $(CGG)_n$ oligomers have higher entropy values (more negative entropy in going from single-strand to hairpin) than the $(CAG)_n$ oligomers. The entropy for the hairpin to single strand transition gives a relative measure of the disorder of the ground state of the oligomer and enables a ranking of the oligomers in terms of the disorder and/or flexibility of the oligomer. The entropy of this process is the sum of the conformational flex of the molecule, the conformational freedom of the individual nucleotides, and the localized melting/reannealing of the base pairs (i.e., "breathing" of the molecule) that includes the end-fraying of both the stem and loop regions of the hairpin. Assuming that the single-stranded states are isoenergetic across all oligomers of the same length and contain the maximum allowable disorder, oligomers with a high amount of disorder will have smaller entropy values relative to the oligomers with a low amount of disorder. Relative to the $(CGG)_n$ oligomers, the $(CAG)_n$ and the $(CCG)_n$ oligomers have a more disordered structure in the ground state at 37 °C.

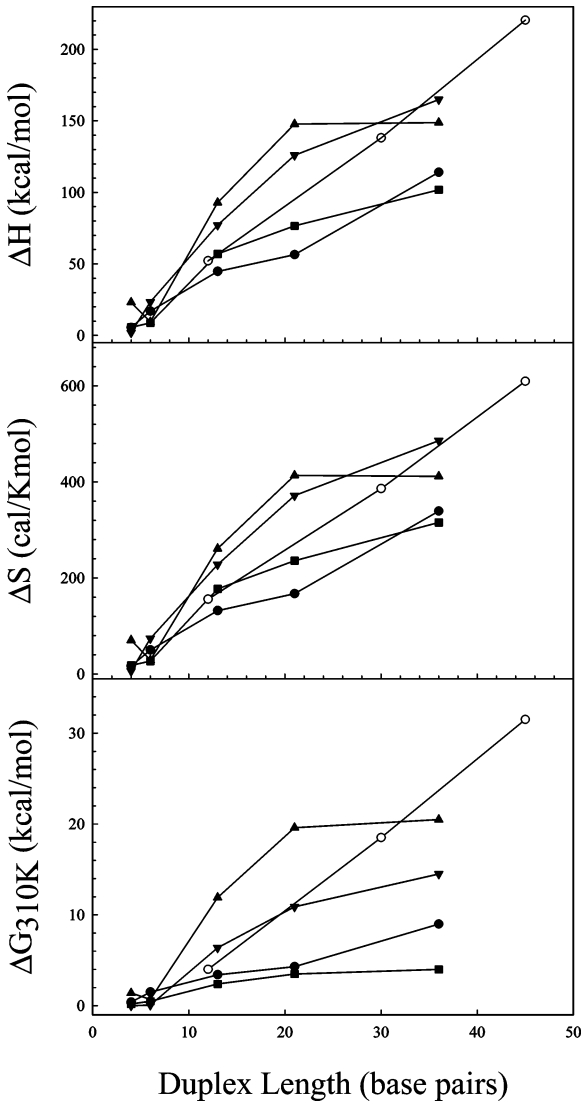

FIGURE 6: Plots of Δ*H*, Δ*S*, and Δ*G* (calculated at 310 K) as a function of duplex length in bases for the $(CNG)_n$ oligomers, where N = A (circles), N = C (squares), N = G (triangles), and N = T (upside down triangles), and for the CAG/CTG duplex (open circles). For the oligomers the number of bases in the duplex was determined from the stem length of the blunt-ended hairpin (see Figure 5) and includes both G:C base pairs and N:N mismatches.

The Stability Plateau and Triplet Repeat DNA Expansion. The nature of the observed stability, enthalpy, and entropy trends for all solitary oligomers examined was not a linear function of the oligomer length but rather appeared to reach a plateau at approximately 45 nucleotides (Figure 6). This work shows, for the first time, a thermodynamic relationship between triplet repeat DNA length and stability. While the stability of the component oligomers tended to plateau, the duplex oligomers displayed an additive behavior where the stability increased in a near-linear fashion over the range of

lengths examined. A plateau in the DNA stability plot that occurs below the duplex stability plot suggests that the triplet repeat oligomers adopt conformations with stabilities suited for expansion. Furthermore, the plateau in the thermodynamic data trends suggests that the stability of the hairpin is compromised as the length of the hairpin increases; i.e., the stability gained by increasing the number of G:C base pairs can no longer overcompensate for the instability of additional mismatches. The instability of the oligomer as the length increases makes the molecule ideally suited for a slip-type of expansion mechanism. In such a mechanism, the daughter DNA strand undergoes localized denaturation that allows extensive breathing or bulge formation to occur within the stem of the hairpin. Under these conditions, the localized denatured region is free to reanneal to a segment of DNA upstream or downstream of the melting position, due to sequence redundancy. From the data tables previously shown, the lower the stability of the oligomer under consideration, the greater the entropy of the DNA segment and the more likely this process will occur due to a combination of low enthalpy and high entropy. Therefore, the plateau in the thermodynamic data suggests that expansion in the DNA sequence beyond 45 nucleotides in length is the inevitable outcome of an inherent thermodynamic instability.

How the Triplet Repeat DNA Sequences Expand—A Biophysical View. At the cellular level, the expansion of triplet DNA begins during replication. Replication is a complex process that involves a multitude of DNA-processing enzymes, DNA synthesis proteins, and other proteins that maintain the single-stranded nature of the unwound DNA. Throughout this process the parent DNA strand is first unwound and separated into single strands through the combined actions of DNA helicase and other processing enzymes collectively referred to as the DNA_{pol} complex. It is during the unwinding process that the expansion of triplet repeat DNA begins. The unwinding process introduces an extraordinary amount of superhelical torsion about the DNA duplex downstream from the replication fork such that the DNA_{pol} complex can no longer process the parent DNA strand (*11*, *12*). The DNA_{pol} complex then dissociates from the parent DNA and allows, through a process that has not been well-characterized, the formation of hairpin structures on the leading daughter DNA strand. The redundant nature of the triplet repeat sequences allows a hairpin formation process that is faster than the binding kinetics of the single-strand binding proteins, whose function is to maintain any single-strand region of DNA as an uncomplemented single-strand. Hairpin formation is accompanied by a positional shift of the 3′-terminus of the daughter DNA strand such that a region of complementary parental DNA that once base-paired with the daughter strand is now exposed. Once formed, the hairpin structures have two possible fates that dictate the downstream disease pathogenesis. After release of the superhelical torsion and hairpin formation, the DNA_{pol} complex will reassociate to the parent DNA strand. The location of the DNA_{pol} complex reassociation will be dependent upon both the thermodynamic stability and the reassociation kinetics of the hairpins. If the hairpin remains as a quasistable species, then the DNA_{pol} complex reassociates at a location that is downstream with respect to the parental strand where the complex dissociated. Upon resumption of DNA synthesis, the DNA_{pol} complex reiteratively synthesizes a previously templated region of DNA, which leads to an expansion in the triplet repeat copy number. If end-fraying of the hairpin occurs and the free 3′-terminus of the hairpin nucleates with the parent DNA strand, then it is probable that the daughter strand will realign and form a duplex DNA or other structure closely resembling duplex DNA. In such a situation, the reassociation of the DNA_{pol} complex occurs at a location at or near where the complex dissociated. Upon resumption of synthesis, there is no expansion in the triplet repeat sequence copy number.

On the basis of the data presented within this study, the $(CNG)_n$ triplet repeat DNA sequences present a multifaceted dependence on the thermodynamic properties that depends on both the sequence and the length of the hairpin under consideration. Sequences with G:G mismatches within the stem region of the hairpin also have the highest stability of all oligomer sequences, due to the formation of two hydrogen bonds and extensive base-stacking interactions. Sequences with C:C mismatches, on the other hand, have the least stability, due to the formation of only one hydrogen bond in the mismatch and limited base-stacking interactions. Considering the similarity between the (CNG) sequences—a 5′-CG-3′ on the sense strand base paired to a 5′-CG-3′ on antisense strand—it is the thermodynamic properties of the N:N mismatches that in large part govern the stability of the hairpin structures.

In summation, the probability for expansion of triplet repeat DNA sequences is dependent on whether the hairpin structure formed during replication remains prior to the reassociation of the DNA_{pol} complex. The results presented here suggests that the probability for the stable hairpin formation is increased with increasing length of the hairpin sequence and sequence context (G > T > A > C). The stability of the hairpin appears to be driven by the stability of the mismatch and reaches a plateau at approximately 45 nucleotides.

ACKNOWLEDGMENT

The authors thank Prof. George Turner for his critical reading of the manuscript.

SUPPORTING INFORMATION AVAILABLE

A table of the calculated extinction coefficients (Table 1S) and a figure of the PAGE electrophorograms of selected oligomers and their respective duplexes (Figure 1S). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Mandel, J.-L. (1994) Trinucleotide diseases on the rise. *Nat. Genet. 4*, 453−455.
2. 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.
3. Reddy, P. S., and Housman, D. E. (1997) The complex pathology of trinucleotide repeats. *Curr. Opin. Cell Biol. 9*, 364−372.
4. Gacy, A. M., Geollner, G., Juranic, N., Macura, S., and McMurray, C. T. (1995) Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell 81*, 553−540.
5. Zheng, M., Huang, X., Smith, G. K., 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*, 323−326.

6. Mitas, M. (1997) Trinucleotide repeats associated with human disease. *Nucleic Acids Res. 25*, 2245−2253.
7. Fry, M., and Loeb, A. L. (1994) The fragile X syndrome of $d(CGG)_n$ nucleotide repeats form a stable tetrahelical structure. *Proc. Natl. Acad. Sci. U.S.A. 33*, 4950−4954.
8. Nadel, Y., Weisman-Shomer, P., and Fry, M. (1995) The fragile X syndrome single strand d(CGG)n nucleotide repeats readily fold back to form unimolecular hairpin structures. *J. Biol. Chem. 270*, 28970−28977.
9. Petruska, J., Arnheim, N., and Goodman, M. F. (1996) Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNAS triplet repeats associated with neurological diseases. *Nucleic Acids Res. 24*, 1992−1998.
10. Gacy, A. M., and McMurray, C. T. (1998) Influence of hairpins on template reannealing at trinucleotide repeat duplexes: A model for slipped DNA. *Biochemistry 37*, 9426−9434.
11. Gellibolian, R., Bacolla, A., and Wells, R. D. (1997) Triplet repeat instability and DNA topology: An expansion model based on statistical mechanics. *J. Biol. Chem. 272*, 16793−16797.
12. Bacolla, A., Gellibolian, R., Shimizu, M, Amirhaeri, S., Kang, S., Ohshima, K., Larson, J. E., Harvey, S. C., Stollar, B. D., and Wells, R. D. (1997) Flexible DNA: Genetically unstable CTG-.CAG and CGG.CCG from human hereditary neuromuscular disease genes. *J. Biol. Chem. 272*, 16783−16792.
13. Sheardy, R. D., Levine, N., Marotta, S., Suh, D., and Chaires, J. B. (1994) A Thermodynamic investigation of the melting of B−Z junction forming DNA oligomers. *Biochemistry 33*, 1385−1391.
14. Fassman, G., Ed. (1975) CRC Handbook of Biochemistry and Molecular Biology, 3rd ed., pp 589, CRC Press, Boca Raton, FL.
15. Sheardy, R. D. (1991) Monitoring conformational transition in synthetic DNA olgomers using circular dichroism. *Spectroscopy 6*, 14−17.
16. Marotta, S. P., Tamburri, P. A., and Sheardy, R. D. (1996) Sequence and environmental effects on the self-assembly of DNA oligomers possessing $G_xT_2G_x$ segments. *Biochemistry 35*, 10484−10492.
17. Marky, L. A., and Breslauer, K. J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers 26*, 1601−1620.
18. Johnson, W. C. (1994) CD of Nucleic Acids. In *CD of Nucleic Acids, Principles and Applications* (Nakanishi, K., Berova, N, and Woody, R. W., Eds.), pp 523−540, VCH Publishers, New York.
19. Amaratunga, M., Snowden-Ifft, E., Wemmer, D. E., and Benight, A. S. (1992) Studies of DNA dumbbells. II. Construction and characterization of DNA dumbbells with a 16 base-pair duplex stem and T_n end loops ($n = 2, 3, 4, 6, 8, 10, 14$). *Biopolymers 32*, 865−879.
20. Rentzeperis, D., Alessi, K., and Marky, L. A. (1993) Thermodynamics of DNA hairpins: Contribution of loop size to hairpin stability and ethidium binding. *Nucleic Acids Res. 21*, 2683−2689.
21. Peyret, N., Seneviratne, P. A., Allawi, H. T., and SantaLucia, J., Jr. (1999) Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. *Biochemistry 38*, 3468−3477.
22. Lane, A. N., and Peck, B. (1995) Conformational flexibility in DNA duplexes containing single G:G mismatches. *Eur. J. Biochem. 230*, 1073−1087.
23. Arnold, F. H., Wolk, S., Cruz, P., and Tinoco, I., Jr. (1987) Structure, dynamics, and thermodynamics of mismatched DNA oligonucleotide duplexes $d(CCCAGGG)_2$ and $d(CCCTGGG)_2$. *Biochemistry 26*, 4068−4075.

BI0494368