458 *Biochemistry* **2011,** *50,* 458–465
DOI: 10.1021/bi101756e

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Sequence Length Dictates Repeated CAG Folding in Three-Way Junctions†

Natalya N. Degtyareva, Courtney A. Barber, Michael J. Reddish, and Jeffrey T. Petty*

Department of Chemistry, Furman University, Greenville, South Carolina 29613, United States

Received November 2, 2010; Revised Manuscript Received December 10, 2010

ABSTRACT: The etiology of a large class of inherited neurological diseases is founded on hairpin structures adopted by repeated DNA sequences, and this folding is determined by base sequence and DNA context. Using single substitutions of adenine with 2-aminopurine, we show that intrastrand folding in repeated CAG trinucleotides is also determined by the number of repeats. This isomeric analogue has a fluorescence quantum yield that varies strongly with solvent exposure, thereby distinguishing particular DNA motifs. Prior studies demonstrated that $(CAG)_8$ alone favors a stem−loop hairpin, yet the same sequence adopts an open loop conformation in a three-way junction. This comparison suggests that repeat folding is disrupted by base pairing in the duplex arms and by purine-purine mismatches in the repeat stem. However, these perturbations are overcome in longer CAG repeats, as demonstrated by studies of isolated and integrated forms of $(CAG)_{15}$. The oligonucleotide alone forms a symmetrically folded hairpin with looplike properties exhibited by the relatively high emission intensities from a modification in the central eighth repeat and with stemlike properties evident from the relatively low emission intensities from peripheral modifications. Significantly, these hairpin properties are retained when $(CAG)_{15}$ is integrated into a duplex. Intrastrand folding by $(CAG)_{15}$ in the three-way junction contrasts with the open loop adopted by $(CAG)_8$ in the analogous context. This distinction suggests that cooperative interactions in longer repeat tracts overwhelm perturbations to reassert the natural folding propensity. Given that anomalously long repeats are the genetic basis of a large class of inherited neurological diseases, studies with (CAG)-based three-way junctions suggest that their secondary structure is a key factor in the length-dependent manifestation and progression of such diseases.

Genetic instability is determined not only by exogenous factors but also by the properties of DNA itself (*1*, *2*). Such inherent mutagenicity was established through association of fragile X syndrome, spinal and bulbar muscular atrophy, and myotonic dystrophy with abnormally long CNG (N = A, T, G, or C) repeats, and subsequently, ∼30 inherited neurological diseases have been linked with expansion of tri- and tetranucleotide repeats beyond critical thresholds (*3*−*6*). Linkage between these inherited diseases and sequence length is substantiated by its correlation with phenotypic severity and progression. Furthermore, these mutations are dynamically transmitted, as repeat tracts become progressively longer through succeeding generations. Beyond the primary information gleaned from DNA sequence and length, secondary structure is the deeper key to this class of genetic diseases, as self-folded conformations are favored by repeated sequences (*7*). Stem−loop hairpins are favored by CNG repeats and are implicated in repeat expansion via replication, repair, and recombination (*5*). To illustrate, hairpins that form on single-stranded Okazaki fragments can disrupt coordinated synthesis on the template strands, thereby extending the nascent strand on the leading template through repeated pausing and restarting of DNA polymerase (*8*). These self-folded moieties can also preferentially recognize and sequester proteins involved in DNA repair, thereby disrupting normal pathways that maintain the integrity of DNA (*9*, *10*). Thus, establishing the structures of these alternative forms of DNA is necessary for understanding their potential broad-scale biological function.

Within double-stranded DNA, repeated sequences self-associate to produce two distinct structures (*11*). Slipped forms occur when strands have identical lengths of complementary repeats, and the structures are distinguished from canonical duplex DNA by enhanced sensitivity to nucleases and anomalously fast electrophoretic mobility (*12*). Their diverse range of structures is stable to challenging environmental changes, thereby suggesting their biological viability. Slipped intermediates form when opposing strands have different lengths, and the resulting three-way junctions are implicated in DNA expansion during replication (*13*). Within these structures, repeat conformation depends on base sequence, with CTG repeats self associating into hairpins and CAG favoring open, solvent-exposed loops, and this difference may originate in the lower thermodynamic stability of CAG versus CTG repeats (*7*). Our studies show that tract length also dictates conformation.

Toward this goal, the adenine isomer 2-aminopurine is used to develop structural and energetic models of repeated CAG sequences, which are prevalent in many inherited neurological diseases (*14*). Fluorescent nucleobase analogues are powerful tools for assessing DNA structure and function, and 2-aminopurine is widely used for these purposes (*15*, *16*). Single substitutions with this fluorescent adenine analogue do not alter global

†We thank the National Science Foundation (CHE-0718588) for primary support of this work. We are grateful to the National Institutes of Health (R15GM071370), the National Science Foundation (CBET-0853692), the Henry Dreyfus Teacher-Scholar Awards Program, and the National Center for Research Resources (P20 RR-016461) for additional support throughout this work. J.T.P. received partial sabbatical support through matching commitments to a National Science Foundation RII Cooperative Agreement (EPS-0903795).

*To whom correspondence should be addressed. Telephone: (864) 294-2689. Fax: (864) 294-3559. E-mail: jeff.petty@furman.edu.

 Published on Web 12/13/2010

Table 1: Single-Stranded Oligonucleotides

sequence	length[a]	ε^b
5′-CACCATGCCGGTA TTTAAA CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG TACGTA CTGCAGCTCGAGG-3′	95	927000
5′-CACCATGCCGGTA TTTAAA CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG TACGTA CTGCAGCTCGAGG-3′[c]	95	915200
5′-CCTCGAGCTGCAG TACGTA CTGCTG—CTGCTG TTTAAA TACCGGCATGGTG-3′[d]	50	462500
5′-CACCATGCCGGTA TTTAAA CAGCAG—CAGCAG TACGTA CT GCAGCTCGAGG-3′[d]	50	484500
5′-CACCATGCCGGTA TTTAAA CAGCAG—CAGCAG TACGTA CT GCAGCTCGAGG-3′[d]	50	472700
5′-CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG	45	443800
5′-CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG[e]	45	432000
5′-G AAA CAG CAG TTTT CTG CTG TTT C-3′ (DS-CAG)	24	210900
5′-AAA CAG CAG-3′ (SS_β-CAG)	9	86300
5′-CAG CAG CAG-3′ (SS-CAG)	9	78000

[a]Length in bases. [b]Extinction coefficients (M^{-1} cm^{-1}) of the unfolded single strands. [c]Ten variants of this oligonucleotide were used, each with a single substitution of 2-aminopurine, indicated by the underlined bases. [d]The connecting line represents the region corresponding to $(CAG)_{15}$. [e]Six variants of this oligonucleotide were used, each with a single substitution of 2-aminopurine, indicated by the underlined bases.

conformations of CAG-based structures, as demonstrated by electrophoresis, spectroscopic, and energetic studies. Structural motifs are identified from their characteristic fluorescence intensities and thermal stability by utilizing the strong effect of base stacking on the fluorescence quantum yield of 2-aminopurine (*17*, *18*). The studies described in this paper are motivated by earlier studies of $(CAG)_8$ (*19*, *20*). The isolated oligonucleotide behaves as the expected stem−loop hairpin, as supported by intensities in the stem region that are comparable to that of a duplex DNA analogue, and intensities in the central loop are comparable to or exceed those of a single-stranded DNA analogue. However, when this same sequence is integrated into duplex DNA, the hairpin structure is lost, yielding a repeat loop that is highly and uniformly solvent-exposed.

Motivated by how DNA context determines conformation in repeated DNA, we sought to understand how tract length influences secondary structure in slipped intermediate forms of DNA. Utilizing dual structural and energetic perspectives offered by 2-aminopurine substitutions, the secondary structure adopted by $(CAG)_{15}$ in a three-way junction was determined. Our major finding is that this longer 15-repeat sequence retains the inherent hairpin structure of the isolated oligonucleotide when it is incorporated into the three-way junction. This behavior sharply contrasts with the open loop adopted by the analogous $(CAG)_8$ structure, and such length dependence suggests that repeat length dictates secondary structure and potential biological functions of long repeat tracts.

MATERIALS AND METHODS

Buffer components (Sigma-Aldrich, St. Louis, MO) and acrylamide (Acros Organics, Geel, Belgium) were used as received. The buffer consisted of 10 mM $HPO_4^{2-}/H_2PO_4^-$ and 50 mM NaCl (pH 7). Oligonucleotides (Integrated DNA Technologies, Coralville, IA) were purified by denaturing 8% PAGE with 7 M urea, 1× TBE buffer at 35 °C, with subsequent visualization of the samples by UV light on a TLC plate. Desired bands were removed from the gel, electroeluted in TBE buffer, and passed through a NAP-10 desalting column (GE Healthcare, Piscataway, NJ). Samples were then lyophilized and resuspended in water. Oligonucleotide concentrations were determined by extrapolating absorbances at 260 nm of the high-temperature post-transition baselines back to 25 °C. Extinction coefficients of single-stranded oligonucleotides were derived from the nearest neighbor approximation (Table 1). For modified oligonucleotides, extinction coefficients were evaluated as the sum of the extinction coefficient without 2-aminopurine plus the relatively small contribution from free 2-aminopurine (1000 M^{-1} cm^{-1} at 260 nm) (*21*). For oligonucleotide and buffer solutions, absorbance was recorded at 260 nm as a function temperature on a Cary 300 spectrometer equipped with a multicell holder (Varian, Palo Alto, CA).

The nomenclature for modified oligonucleotides identifies the position of 2-aminopurine within the primary sequence (Figure 1). Starting from the 5′-terminus, we introduced individual replacements of adenine with 2-aminopurine into the isolated $(CAG)_{15}$ hairpin at the second, fifth, seventh, eighth, eleventh, and fourteenth repeats, as designated by 2-AP_{HP}, 5-AP_{HP}, 7-AP_{HP}, 8-AP_{HP}, 11-AP_{HP}, and 14-AP_{HP}, respectively. For the three-way junction, substitutions denoted as 1-AP_J, 2-AP_J, 5-AP_J, 7-AP_J, 8-AP_J, 11-AP_J, and 14-AP_J were made in the first, second, fifth, seventh, eighth, eleventh, and fourteenth repeats of the central $(CAG)_{15}$, respectively. Modifications labeled α-AP_J, β-AP_J, and α3′-AP_J were also made in the two (CAG) repeats in the duplex region that precedes the junction. Three-way junctions were formed by annealing a 95-base oligonucleotide with a central $(CAG)_{15}$ sequence and flanking 25-base regions that are complementary to a 50-base oligonucleotide (Figure 1A). These two oligonucleotides conserved the duplex arms used for the three-way junction $(CAG)_8$ studies (*20*). We performed annealing by heating equimolar amounts of single strands together to 95 °C for 5 min with slow cooling to room temperature over a period of more than 12 h. Purity was verified by 12% polyacrylmide nondenaturing gel electrophoresis. Formation of annealed three-way junctions was monitored with restriction endonucleases BsaAI and DraI (New England BioLabs, Ipswich, MA) using distinct recognition sites on both sides of the repeat tract. At 0.5 μM in the appropriate buffer (NEB2 from New England BioLabs), DNA samples were mixed with 10−15 units of an enzyme until the total volume reached 20 μL and incubated at 37 °C for at least 3 h. Samples were mixed with SYBR Gold and the gel loading dye and analyzed via native 12% polyacrylamide gel electrophoresis in TBE buffer. Release of a 16 bp fragment confirmed properly annealed strands (Figure S1 of the Supporting Information). Similar thermodynamic properties for the modified and unmodified oligonucleotides support minimal perturbation with the single substitutions with 2-aminopurine (Table S1 of the Supporting Information).

Fluorescence spectra were recorded on a Fluoromax-2 spectrometer (Jobin-Yvon Horiba, Edison, NJ) equipped with DataMax version 3.4 and a Neslab RTE-7 circulating bath. Emission was

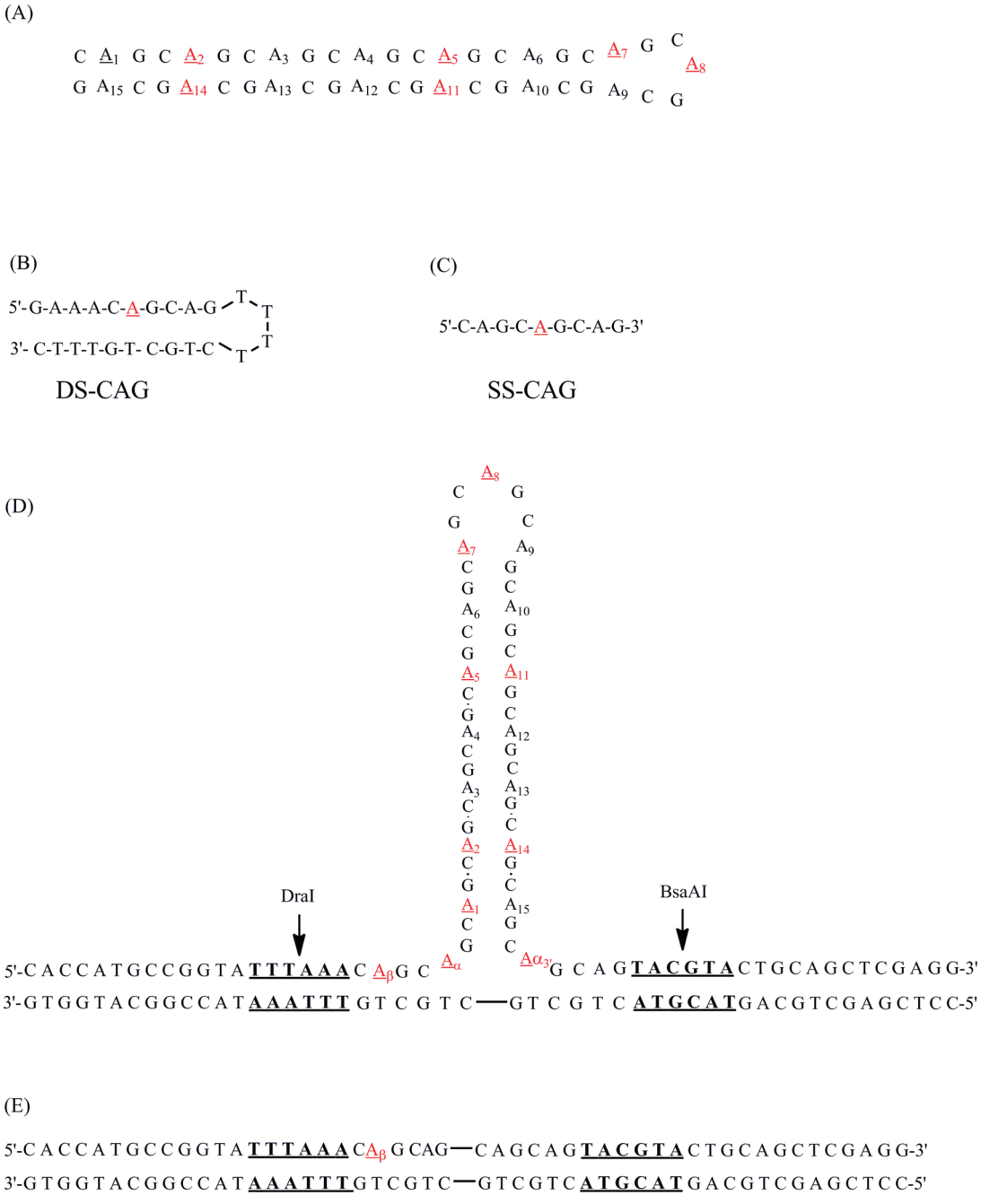

FIGURE 1: Structures for the hairpin formed by $(CAG)_{15}$ (A) along with the double-stranded (DS-CAG) (B) and single-stranded (SS-CAG) (C) structural references. In three-way junction formed with $(CAG)_{15}$, the DraI and BsaAI restriction enzyme sites are shown in bold and underlined (D). The duplex that comprises the three-way junction includes connecting lines to represent the position of the abstracted repeat sequence (E). For the isolated and integrated forms of $(CAG)_{15}$, positions of 2-aminopurine substitution are underlined and enumerated.

collected at 370 nm using excitation at 307 nm, and the temperature was varied by 0.3 °C with a 1 min equilibration at each temperature (Figure S2 of the Supporting Information). Thermal denaturation curves accounted for temperature-dependent changes in the emission of 2-aminopurine using the single-stranded control SS-CAG (Figure 1C) (*18*, *19*). Absorbance melting curves were determined at a rate of 0.1 °C/min using 1 μM concentrations of $(CAG)_{15}$ and 0.6 μM concentrations of the three-way junctions (Figure S3 of the Supporting Information). For denaturation profiles exhibiting monophasic profiles, standard van't Hoff analysis was used to determine the enthalpy and entropy changes. For biphasic transitions, derivative profiles were deconvoluted into Gaussian peaks, from which the thermodynamic information was extracted (*22*). For denaturation of the *i*th domain, the enthalpy change (ΔH_i) at the melting temperature (T_m) is

$$\Delta H_i = 4RT_m^2 \times h_i/\Delta A_i$$

where R is the gas constant, h_i is the peak amplitude, and ΔA_i is the peak area. Entropy changes accounted for the concentration dependence using

$$\Delta S_i = -R\ln(c_0/2) + \Delta H_i/T_m$$

where c_0 is the initial strand concentration (*23*). Average thermodynamic parameters and standard deviations were derived from a minimum of three thermal denaturation profiles that were determined using separate samples.

For acrylamide studies, standard Stern−Volmer analysis was an adequate model for determining the static quenching

Table 2: Thermodynamic Data for Thermal Denaturation and Acrylamide Quenching Constants for $(CAG)_{15}$ Sequences[a]

position[b]	$T_{m,abs}$ (°C)	ΔH_{abs} (kcal/mol)	ΔS_{abs} (cal mol^{-1} K^{-1})	$T_{m,fluor}$ (°C)	ΔH_{fluor} (kcal/mol)	ΔS_{fluor} (cal mol^{-1} K^{-1})	K_q (M^{-1})[f]
2-AP_{HP}	58.4 ± 0.3	35.9 ± 0.9	108 ± 3	57.7 ± 0.3	38.2 ± 1.8	116 ± 5	3.7 ± 0.4
5-AP_{HP}	58.9 ± 0.6	38.1 ± 2.9	114 ± 9	58.1 ± 0.3	41.8 ± 1.6	127 ± 5	5.8 ± 0.6
7-AP_{HP}	57.4 ± 0.8	39.2 ± 0.7	118 ± 2	58.7 ± 1.0	41.2 ± 4.6	124 ± 14	6.8 ± 0.6
8-AP_{HP}	57.8 ± 0.4	39.0 ± 0.6	118 ± 2	ND[e]	ND[e]	ND[e]	8.0 ± 0.5
11-AP_{HP}	59.4 ± 1.0	35.9 ± 0.9	108 ± 3	58.4 ± 0.8	39.2 ± 2.3	119 ± 7	5.5 ± 0.4
14-AP_{HP}	57.4 ± 0.2	39.4 ± 0.4	119 ± 1	57.5 ± 0.8	42.7 ± 1.4	129 ± 4	4.5 ± 0.7
average[c]	58.2 ± 0.8	37.9 ± 1.6	114 ± 5	58.0 ± 0.5	40.6 ± 1.9	123 ± 5	
$(CAG)_{15}$[d]	58.0 ± 0.3	38.7 ± 1.8	117 ± 5				

[a]Subscripts abs and fluor indicate measurements using absorbance at 260 nm and fluorescence using a λ_{ex} of 307 nm and a λ_{em} of 370 nm. Standard deviations are derived from a minimum of three measurements. Melting temperatures (T_m), enthalpy changes (ΔH), and entropy changes (ΔS) were derived from absorbance and fluorescence changes. Quenching constants (K_q) were derived from acrylamide quenching of 2-aminopurine fluorescence. [b]See Figure 1A for the location of modifications in $(CAG)_{15}$. [c]Average values for the absorbance and fluorescence measurements for the six variants listed. [d]Thermodynamic data derived for the unmodified $(CAG)_{15}$ oligonucleotide. [e]Values not determined because of a poorly defined denaturation profile. [f]Quenching constants for the double-stranded (DS-CAG) and single-stranded standards (SS-CAG) are 4.2 ± 1.7 and 9.6 ± 1.0 M^{-1}, respectively.

constants (*24*). Acrylamide concentrations up to 0.1 M were used. Oligonucleotide concentrations up to 0.3−0.5 μM were used, and average quenching constants with standard deviations were calculated from at least three replicate experiments.

RESULTS

$(CAG)_{15}$ Hairpin. Six variants with 2-aminopurine establish the propensity of $(CAG)_{15}$ to fold into a stem−loop hairpin (*7*). A minimal degree of structural perturbation by this adenine isomer is supported by similar circular dichroism spectra of modified and unmodified oligonucleotides (Figure S4 of the Supporting Information). The spectra are characteristic of B-form DNA, suggesting a duplex stem that is stabilized by the high proportion of G/C pairs (*25*). Base stacking in the folded structure is also indicated by hyperchromic absorbance changes at high temperatures. Derived thermodynamic parameters do not vary with 2-aminopurine modification, with the unmodified hairpin having a ΔH of 38.7 ± 1.8 kcal/mol and a ΔS of 117 ± 5 cal mol^{-1} K^{-1} at 58.0 ± 0.3 °C and the six variants having an average ΔH of 37.9 ± 1.6 kcal/mol and a ΔS of 114 ± 5 cal mol^{-1} K^{-1} at 58.2 ± 0.8 °C (Table 2).

The secondary structure of $(CAG)_{15}$ was established through the solvent accessibility of individually substituted 2-aminopurines, and two types of references relate emission intensities with particular structural moieties. First, a hairpin oligonucleotide with a 2-aminopurine/thymine pair in the stem defines a lower limit of solvent exposure [DS-CAG (Figure 1B)]. In this arrangement, fluorescence is quenched through base stacking, although stacking is less efficient when compared with that of the corresponding adenine/thymine pair (*26*). Second, a short nine-base $(CAG)_3$ oligonucleotide with a centrally placed 2-aminopurine defines an upper limit of solvent exposure [SS-CAG (Figure 1C)]. This length precludes folding and base stacking, thereby enhancing the fluorescence quantum yield of 2-aminopurine. In addition, this sequence reflects the environment in the repeat tracts, in which repeats with 2-aminopurine are flanked by CAG trinucleotides. This strand exhibits no tendency to self-associate on the basis of fluorescence studies conducted at higher and lower concentrations and similar temperature-dependent fluorescence changes for this short oligonucleotide and free 2-aminopurine (*19*). Fluorescence intensities are as follows: DS-CAG < 14-AP_{HP} ~ 2-AP_{HP} ~ 5-AP_{HP} ~ 11-AP_{HP} < 7-AP_{HP} < 8-AP_{HP} < SS-CAG (where the extreme intensities are associated with the double-stranded and single-stranded references, respectively).

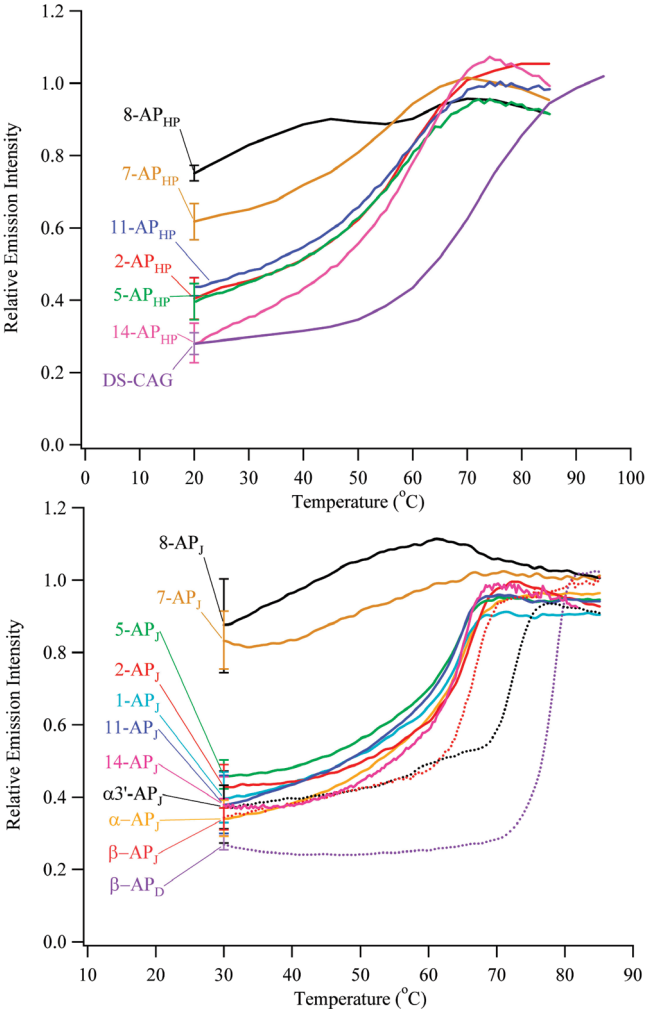

FIGURE 2: Emission intensities of the $(CAG)_{15}$ hairpin (top) and its three-way junctions (bottom) with 2-aminopurine as a function of temperature using a λ_{ex} of 307 nm and a λ_{em} of 370 nm. These intensities are corrected for the inherent temperature-dependent changes of 2-aminopurine using the single-stranded DNA reference [SS-CAG (Figure 1C)].

This trend advances a hairpin structure (Figure 2 and Figure S5 of the Supporting Information). First, substitutions in the second, fifth, eleventh, and fourteenth repeats that flank the center exhibit intensities that are more similar to the double-stranded reference. This similarity suggests that base stacking is a significant

driving force for folding, consistent with characteristic B-form signature in circular dichroism spectra (Figure S4 of the Supporting Information). Higher intensities for these emissive variants relative to the double-stranded reference suggest that purine-purine mismatches increase the level of solvent exposure (*27*). Second, substitutions in the seventh and eighth repeats show that intensities increase toward the center of the primary sequence and approach the reference intensity of single-stranded DNA. Together, these observations are consistent with symmetric folding to produce a duplex stem connected by a central loop.

All 2-aminopurine oligonucleotides attain a common denatured state at sufficiently high temperatures, and temperature-dependent fluorescence provides energetic signatures of the stem and loop. An inherent temperature effect on fluorescence quantum yield is removed by also collecting emission from a single-stranded DNA reference with limited base stacking (Figure 1C) (*18*). Two types of environments are revealed in fluorescence thermograms of $(CAG)_{15}$: a relatively unstructured loop evidenced by modest intensity changes for 8-AP_{HP} and a base-stacked stem as indicated by enhanced fluorescence for 2-AP_{HP}, 5-AP_{HP}, 11-AP_{HP}, and 14-AP_{HP}. Furthermore, a transition between these motifs is evident from intermediate enhancement of 7-AP_{HP} (Figure 2A). Thermodynamic analysis used limiting intensities to derive the fractional conversion from folded to unfolded forms, and resulting enthalpy and entropy changes show no distinction between the hyperfluorescent oligonucleotides with an average ΔH of 40.6 ± 1.9 kcal/mol and a ΔS of 123 ± 5 cal mol^{-1} K^{-1} at 58.1 ± 0.5 °C (Table 2). This similarity indicates cooperative denaturation, as also corroborated by similar globally based thermodynamic parameters determined from absorbance measurements (Table 2). Derived van't Hoff parameters agree with previously reported values, although they differ from model-free parameters determined by calorimetry (*25*, *28*). Cooperative unfolding deduced from our experiments suggests that other factors such as heat capacity changes due to solvation effects may be important in the unfolding process (*28*).

To extrinsically probe the hairpin structure of $(CAG)_{15}$, we quenched 2-aminopurine fluorescence with acrylamide (Table 2). Solvent accessibility dictates the degree of quenching, and the single- and double-stranded references again distinguish the structural motifs in $(CAG)_{15}$. Substitutions in the second and fourteenth repeats are more protected and quenched least efficiently because they behave most like the double-stranded reference with a quenching constant of 4.2 ± 1.7 M^{-1}. Converging toward the center, quenching efficiency increases and culminates for 8-AP_{HP}, which behaves most similarly to the single-stranded reference with a quenching constant of 9.6 ± 1.0 M^{-1}. This behavior suggests a gradual opening of the duplex stem in the vicinity of the loop, as also observed for $(CAG)_8$ (*19*).

$(CAG)_{15}$ Three-Way Junction. By establishing the inherent folding propensity of $(CAG)_{15}$ alone, we evaluate its degree of perturbation within the three-way junction. Using identical substitutions made in the isolated form of $(CAG)_{15}$, fluorescence studies distinguish two environments in the integrated form (Figures 1D and 2). Stem formation is demonstrated by intensities from 2-AP_J/14-AP_J and 5-AP_J/11-AP_J that are most similar to that of a double-stranded DNA reference. Intrastrand folding is further substantiated by similar fluorescence intensities from these distantly placed modifications and by their enhanced fluorescence that accompanies denaturation (Figure 2 and Figure S6 of the Supporting Information). Loop formation in the repeat tract is supported by two measurements on 8-AP_J: its emission intensity changes little with temperature and is most similar to that of the single-stranded DNA reference. Transition between these two motifs is corroborated by an intermediate fluorescence intensity and a weak, monophasic response to temperature that is exhibited by 7-AP_J. Importantly, these same intensities and intensity changes are exhibited by the analogous variants of the $(CAG)_{15}$ oligonucleotide alone, indicating that the inherent structure of this repeated sequence is retained when it is incorporated in the duplex. In addition to probes in the repeat region of the three-way junction, the structure and stability in the supporting duplex arms were evaluated using substitutions in the β, α, and $\alpha 3'$ repeats. All three modifications have emissions similar to that of the double-stranded DNA reference. Thus, base stacking is a dominant interaction in the supporting arms, and this effect extends into the repeat region, as indicated by comparably low intensities from 1-AP_J. These intensities are slightly higher than that of the double-stranded reference and increase in the vicinity of the repeat tract, which indicates that the repeat tract influences base pairing and stacking (Figure 2).

Segregation into distinct domains is also supported through stability measurements. Absorbance studies at 260 nm exhibit two resolved transitions during thermal denaturation (Figure S3 and Table S1 of the Supporting Information). The high-temperature transition at 71.9 ± 0.6 °C has a ΔH of 235 ± 3 kcal/mol and a ΔS of 712 ± 8 cal mol^{-1} K^{-1}, which is similar to the ΔH of 252 ± 12 kcal/mol and a ΔS of 694 ± 30 cal mol^{-1} K^{-1} measured at 75.4 ± 1.1 °C for the component duplex without $(CAG)_{15}$ (Figure 1E and Table S1 of the Supporting Information). Their differences may reflect compromised base stacking and pairing at the junction. In relation to $(CAG)_{15}$ alone, the low-temperature transition at 65.6 ± 0.7 °C is more thermodynamically restrained with a ΔH of 110 ± 9 kcal/mol and a ΔS of 326 ± 24 cal mol^{-1} K^{-1}, and this difference may be due to anchoring and overall stabilization by the duplex arms. To further clarify these structural motifs and the transitional regions between them, thermodynamic analysis was conducted using fluorescence (Figure 3 and Table 3). As in the absorbance measurements, the majority of fluorescence thermograms exhibit biphasic profiles (Figure S2 of the Supporting Information). Within these profiles, melting temperatures, enthalpy changes, and entropy changes for the two transitions are similar across the different modifications, indicating cooperative relaxation within the domains. The breadth of the lower-temperature transition may be related to local premelting of the mismatched pairs prior to overall denaturation of the hairpin region (*19*). The structural basis of these two transitions is established via two peripheral modifications with monophasic profiles (Figure 3). First, β-AP_J within the duplex arms is distant from the junction, and its single transition exhibits thermodynamic properties that mimic those of the high-temperature transition. This structural and thermodynamic correlation suggests that the high-temperature transition reflects denaturation of the duplex arms. Duplex integrity is altered by the repeat sequence, as shown by a higher fluorescence intensity and a lower thermodynamic stability for β-AP_J in relation to its duplex analogue without $(CAG)_{15}$ (Figures 1D,E and 3 and Table 3). The second peripheral reference is provided by 7-AP_J, which is close to the apex of the hairpin and thus also distant from the junction. Its monophasic transition reflects the thermodynamic properties associated with the low-temperature transitions in the other variants; thus, the low-temperature transition is assigned to denaturation of the repeat stem. This assignment is validated by

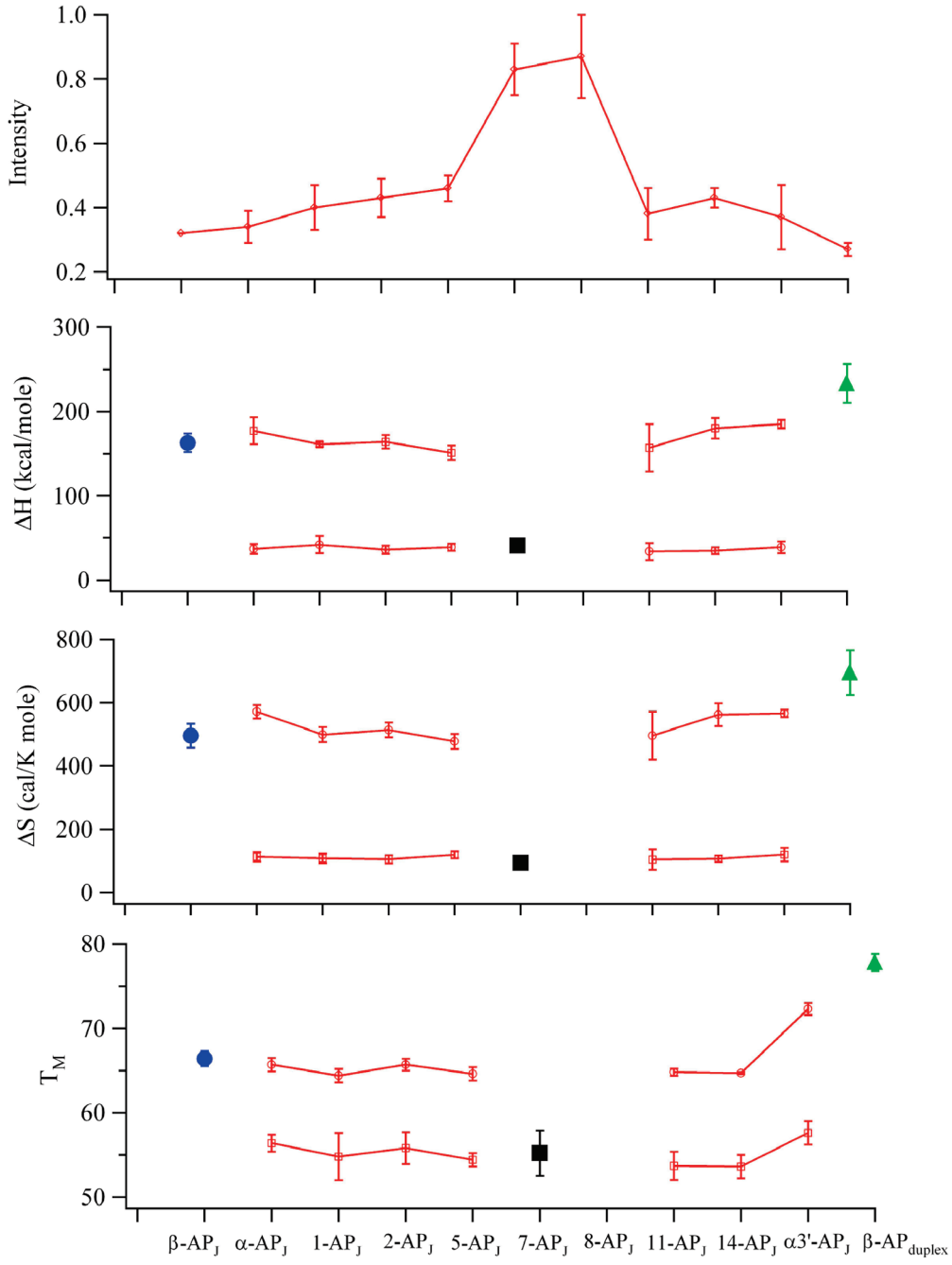

FIGURE 3: Summary of fluorescence intensities and thermodynamic results for the $(CAG)_{15}$ three-way junction. In the top panel, emission intensities increase in the center of the primary sequence, indicative of stem−loop formation. The bottom three panels shows enthalpy (ΔH) and entropy (ΔS) changes and melting temperatures (T_m), respectively, measured throughout the three-way junction. Substitutions in the α, first, second, fifth, eleventh, fourteenth, and α3′ repeats (empty red squares) exhibit similar thermodynamic changes. Modifications in the β (blue circles) and seventh (black squares) repeats exhibit monophasic transitions that mimic the high- and low-temperature transitions, respectively, in the other fluorescent variants. Parameters associated with double-stranded DNA without $(CAG)_{15}$ are indicated by green triangles. Error bars represent standard deviations derived from at least three measurements.

comparable thermodynamic properties for the isolated $(CAG)_{15}$ hairpin. For substitutions in the second, fifth, eleventh, and fourteenth repeats, the three-way junction has an average ΔH of 36.0 ± 2.1 kcal/mol and an average ΔS of 109 ± 7 cal mol^{-1} K^{-1} at 56 ± 2 °C, which are comparable to the analogous averages for $(CAG)_{15}$ alone ($\Delta H = 41 \pm 2$ kcal/mol and $\Delta S = 123 \pm 6$ cal mol^{-1} K^{-1} at 57.9 ± 0.4 °C). Further reflecting the properties of the hairpin, thermodynamic integrity is almost lost by the peripheral eighth repeat, as indicated by its weak and structureless response to temperature changes. These similarities between the integrated and isolated forms of $(CAG)_{15}$ provide strong evidence that the inherent folding propensity of this longer repeat tract is retained in the three-way junction.

Acrylamide quenching constants also support a repeat hairpin that emanates from the duplex arms (Table 3). Modifications in the seventh and eighth positions have the highest quenching constants that are similar to that of the single-stranded reference, thus consistent with greater base exposure in a loop. Quenching constants of 1-AP$_J$, 2-AP$_J$, 5-AP$_J$, 11-AP$_J$, 14-AP$_J$, α-AP$_J$, β-AP$_J$, and α3′-AP$_J$ are similar to that of the fully sequestered base in the double-stranded reference. These features were also exhibited by $(CAG)_{15}$ alone.

Table 3: Thermodynamic Data for Thermal Denaturation and Acrylamide Quenching Constants for $(CAG)_{15}$-Based Three-Way Junctions[a]

position[b]	T_{m1} (°C)	ΔH_1 (kcal/mol)	ΔS_1 (cal mol^{-1} K^{-1})	T_{m2} (°C)	ΔH_2 (kcal/mol)	ΔS_2 (cal mol^{-1} K^{-1})	K_q (M^{-1})[d]
β-AP_J				66.4 ± 0.9	163 ± 11	496 ± 38	4.6 ± 0.3
α-AP_J	56.4 ± 1.0	37 ± 6	113 ± 15	65.7 ± 0.8	177 ± 16	572 ± 21	3.9 ± 0.2
1-AP_J	54.8 ± 2.8	42 ± 10	108 ± 15	64.4 ± 0.8	161 ± 3	499 ± 24	5.3 ± 0.3
2-AP_J	55.8 ± 1.9	36 ± 5	105 ± 13	65.7 ± 0.7	164 ± 8	514 ± 24	4.1 ± 0.2
5-AP_J	54.4 ± 0.8	39 ± 4	119 ± 11	64.6 ± 0.8	151 ± 2	478 ± 4	5.6 ± 0.2
7-AP_J	55.2 ± 2.7	41 ± 4	94 ± 9				8.7 ± 0.2
8-AP_J	ND[c]	ND[c]	ND[c]	ND[c]	ND[c]	ND[c]	8.3 ± 0.2
11-AP_J	53.7 ± 1.7	34 ± 10	104 ± 32	64.8 ± 0.4	157 ± 28	496 ± 76	4.6 ± 0.2
14-AP_J	53.6 ± 1.4	35 ± 4	107 ± 1	64.7 ± 0.1	180 ± 12	562 ± 36	4.0 ± 0.2
α3′-AP_J	57.6 ± 1.4	39 ± 7	120 ± 21	72.3 ± 0.7	185 ± 4	565 ± 10	4.6 ± 0.2
β-AP_D				77.8 ± 1.0	233 ± 23	695 ± 70	

[a]Parameters measured using fluorescence at a λ_{ex} of 307 nm and a λ_{em} of 370 nm. Standard deviations are derived from a minimum of three measurements. Melting temperatures (T_m), enthalpy changes (ΔH), and entropy changes (ΔS) were derived from fluorescence changes. Quenching constants (K_q) were derived from acrylamide quenching of 2-aminopurine fluorescence. Subscripts 1 and 2 correspond to the low- and high-temperature transitions, respectively. [b]See Figure 1D for the location of modifications in $(CAG)_{15}$. [c]Values not determined because of a poorly defined denaturation profile. [d]Quenching constants for the double-stranded (DS-CAG) and single-stranded (SS-CAG) standards are 4.2 ± 1.7 and 9.6 ± 1.0 M^{-1}, respectively.

DISCUSSION

These studies demonstrate the inherent propensity of longer CAG repeats to fold, both as isolated oligonucleotides and within three-way junctions. From a structural standpoint, fluorescence intensity and acrylamide quenching studies with six 2-aminopurine variants support hairpin formation in both contexts. Solvent-sequestered, stemlike properties are exhibited by substitutions in the second and fourteenth and in the fifth and eleventh repeats, which have comparably low intensities and quenching constants in relation to those of double-stranded DNA analogues. Adjoining loops encompassing the seventh and eighth repeats are discerned by their high level of solvent exposure when compared with that of the single-stranded DNA reference. Hairpin folding in both forms of $(CAG)_{15}$ is further substantiated by stability studies. Most importantly, similar thermodynamic properties for the isolated hairpin and the repeat tract in the three-way junction support the formation of a repeat hairpin domain within the three-way junction. The global structure of the three-way junction is inferred by mapping solvent exposure through the junction region. Base stacking and pairing is maintained throughout the junction, and prior studies of DNA three-way junctions suggest that the three arms should be fully extended (*29*). Thus, for the $(CAG)_{15}$-based three-way junction, two arms are canonical duplexes while the third is the repeat hairpin (Figure 1C).

These results broaden our understanding of repeat folding in slipped intermediates that are postulated to develop during DNA replication. CAG tracts incorporated into duplex DNA adopt associated hairpin and unassociated loop forms (*13*). In support of unstructured loops, the repeat region is preferentially digested by single-strand specific enzymes (*13*). In vivo studies using overexpressed proteins and in vitro studies that measure DNA structure via electron microscopy support a preferential affinity of single-strand binding proteins for CAG repeats (*13*, *30*). Spectroscopy-based studies using 2-aminopurine showed that $(CAG)_8$ within a three-way junction is uniformly open and solvent-exposed, and the results suggest that the duplex component dominates both structure and stability within the composite three-way junction (*18*). Collectively, these studies suggest that CAG repeats inherently favor open loops, possibly induced by purine−purine mismatches that are expected to reduce hairpin stability (*7*). However, such structures should be considered in a larger context, as the longer $(CAG)_{15}$ retains its native hairpin structure when incorporated in a three-way junction. Cooperative interactions that drive intrastrand association may be the origin of this stark structural difference that depends on the number of repeats. For isolated hairpins, stability increases with the number of repeats because of the more favorable enthalpic interactions from base stacking and pairing (*25*, *28*). Our results suggest that cooperativity has an impact on secondary structure in repeat tracts of three-way junctions. Assessing the significance of cooperativity on folding could be accomplished using interruptions such as CAA in place of CAG, which are known to influence in vivo genetic instability (*31*, *32*). Because CAG repeats form one of the least stable folded hairpins, cooperative effects are expected to be more significant for other repeated sequences.

CONCLUSION

At the genetic level, structures adopted by abnormally long repeated DNA sequences are implicated in a large class of inherited neurological diseases, and slipped intermediates that form during DNA replication provide one avenue for sequence expansion. Folding that facilitates expansion depends on base sequence and DNA context, and our studies demonstrate that sequence length also dictates intrastrand association within the repeat tract. We suggest that base stacking promotes base pairing between distant repeats and cooperative interactions drive such folding in longer repeat tracts. Length-dependent folding is expected to influence other repeat tracts, thereby having an impact on processes that produce and recognize abnormally long repeat tracts.

SUPPORTING INFORMATION AVAILABLE

Supplemental figures describing structural characterization of the three-way junction, deconvolution of fluorescence and absorbance thermograms, circular dichroism spectra of $(CAG)_{15}$, and composite fluorescence spectra of $(CAG)_{15}$ and of the three-way junction variants and a table summarizing thermodynamic information related to the three-way junction. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Bacolla, A., and Wells, R. D. (2009) Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol. Carcinog. 48*, 273–285.
2. Wang, G., and Vasquez, K. M. (2006) Non-B DNA structure-induced genetic instability. *Mutat. Res. 598*, 103–119.
3. Mirkin, S. M. (2007) Expandable DNA repeats and human disease. *Nature 447*, 932–940.

4. Castel, A. L., Cleary, J. D., and Pearson, C. E. (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol. 11*, 165–170.
5. Kovtun, I. V., and McMurray, C. T. (2008) Features of trinucleotide repeat instability in vivo. *Cell Res. 18*, 198–213.
6. Gatchel, J. R., and Zoghbi, H. Y. (2005) Diseases of Unstable Repeat Expansion: Mechanisms and Common Principles. *Nat. Rev. Genet. 6*, 743–755.
7. Mitas, M. (1997) Trinucleotide repeats associated with human disease. *Nucleic Acids Res. 25*, 2245–2254.
8. Mirkin, S. M. (2006) DNA structures, repeat expansions and human hereditary disorders. *Curr. Opin. Struct. Biol. 16*, 351–358.
9. Kovtun, I. V., and McMurray, C. T. (2001) Trinucleotide expansion in haploid germ cells by gap repair. *Nat. Genet. 27*, 407–411.
10. Panigrahi, G. B., Slean, M. M., Simard, J. P., Gileadi, O., and Pearson, C. E. (2010) Isolated short CTG/CAG DNA slip-outs are repaired efficiently by hMutSβ, but clustered slip-outs are poorly repaired. *Proc. Natl. Acad. Sci. U.S.A. 107*, 12593–12598.
11. Sinden, R. R., Potaman, V. N., Oussatcheva, E. A., Pearson, C. E., Lyubchenko, Y. L., and Shlyakhtenko, L. S. (2002) Triplet repeat DNA structures and human genetic disease: Dynamic mutations from dynamic DNA. *J. Biosci. 27*, 53–65.
12. 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.
13. Pearson, C. E., Tam, M., Wang, Y.-H., Montgomery, S. E., Dar, A. C., Cleary, J. D., and Nichol, K. (2002) Slipped-strand DNAs formed by long (CAG)-(CTG) repeats: Slipped-out repeats and slip-out junctions. *Nucleic Acids Res. 30*, 4534–4547.
14. Pearson, C. E., Edamura, K. N., and Cleary, J. D. (2005) Repeat instability: Mechanisms of dynamic mutations. *Nat. Rev. Genet. 6*, 729–742.
15. Sinkeldam, R. W., Greco, N. J., and Tor, Y. (2010) Fluorescent Analogs of Biomolecular Building Blocks: Design, Properties, and Applications. *Chem. Rev. 110*, 2579–2619.
16. Rist, M. J., and Marino, J. P. (2002) Fluorescent Nucleotide Base Analogs as Probes of Nucleic Acid Structure, Dynamics and Interactions. *Curr. Org. Chem. 6*, 775.
17. Ward, D. C., Reich, E., and Stryer, L. (1969) Fluorescence studies of nucleotides and polynucleotides. I. Formycin, 2-aminopurine riboside, 2,6-diaminopurine riboside, and their derivatives. *J. Biol. Chem. 244*, 1228–1237.
18. Ballin, J. D., Bharill, S., Fialcowitz-White, E. J., Gryczynski, I., Gryczynski, Z., and Wilson, G. M. (2007) Site-Specific Variations in RNA Folding Thermodynamics Visualized by 2-Aminopurine Fluorescence. *Biochemistry 46*, 13948–13960.
19. Degtyareva, N. N., Reddish, M. J., Sengupta, B., and Petty, J. T. (2009) Structural Studies of a Trinucleotide Repeat Sequence Using 2-Aminopurine. *Biochemistry 48*, 2340–2346.
20. Degtyareva, N. N., Barber, C. A., Sengupta, B., and Petty, J. T. (2010) Context dependence of trinucleotide repeat structures. *Biochemistry 49*, 3024–3030.
21. Fox, J. J., Wempen, I., Hampton, A., and Doerr, I. L. (1958) Thiation of Nucleosides. I. Synthesis of 2-Amino-6-mercapto-9-β-D-ribofuranosylpurine ("Thioguanosine") and Related Purine Nucleosides. *J. Am. Chem. Soc. 80*, 1669–1675.
22. Yen, W. S., and Blake, R. D. (1980) Analysis of high-resolution melting (thermal dispersion) of DNA. Methods. *Biopolymers 19*, 681–700.
23. Mergny, J.-L., and Lacroix, L. (2003) Analysis of Thermal Melting Curves. *Oligonucleotides 13*, 515–537.
24. Lakowicz, J. R. (1983) Principles of Fluorescence Spectroscopy, Plenum Press, New York.
25. Paiva, A. M., and Sheardy, R. D. (2004) Influence of sequence context and length on the structure and stability of triplet repeat DNA oligomers. *Biochemistry 43*, 14218–14227.
26. Lycksell, P. O., Graslund, A., Claesens, F., McLaughlin, L. W., Larsson, U., and Rigler, R. (1987) Base pair opening dynamics of a 2-aminopurine substituted EcoRI restriction sequence and its unsubstituted counterpart in oligonucleotides. *Nucleic Acids Res. 15*, 9011–9025.
27. 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.
28. Amrane, S., Sacca, B., Mills, M., Chauhan, M., Klump, H. H., and Mergny, J.-L. (2005) Length-dependent energetics of $(CTG)_n$ and $(CAG)_n$ trinucleotide repeats. *Nucleic Acids Res. 33*, 4065–4077.
29. Lilley, D. M. J. (2008) Analysis of branched nucleic acid structure using comparative gel electrophoresis. *Q. Rev. Biophys. 41*, 1–39.
30. Andreoni, F., Darmon, E., Poon, W. C. K., and Leach, D. R. F. (2010) Overexpression of the single-stranded DNA-binding protein (SSB) stabilises CAG-CTG triplet repeats in an orientation dependent manner. *FEBS Lett. 584*, 153–158.
31. Jarem, D. A., Huckaby, L. V., and Delaney, S. (2010) AGG Interruptions in $(CGG)_n$ DNA Repeat Tracts Modulate the Structure and Thermodynamics of Non-B Conformations in Vitro. *Biochemistry 49*, 6826–6837.
32. Sobczak, K., and Krzyzosiak, W. J. (2005) CAG Repeats Containing CAA Interruptions Form Branched Hairpin Structures in Spinocerebellar Ataxia Type 2 Transcripts. *J. Biol. Chem. 280*, 3898–3910.