Influence of Hairpins on Template Reannealing at Trinucleotide Repeat Duplexes: A Model for Slipped DNA†

A. Marquis Gacy‡ and Cynthia T. McMurray*,‡,§

Departments of Biochemistry and Molecular Biology and of Pharmacology, Mayo Foundation and Mayo Graduate School, Rochester, Minnesota 55905

Received January 21, 1998; Revised Manuscript Received March 27, 1998

ABSTRACT: Hairpin stabilization of polymerase slippage has been proposed as part of the mechanism for large-scale expansion of CG-rich (CNG, where N = A, T, G, or C) trinucleotide repeats. However, hairpin formation does not entirely account for why long repeats but not short repeats or palindromes expand. Using ultraviolet spectroscopic methods, we examine the thermodynamic and kinetic properties of repeating trinucleotides to evaluate their behavior at a slippage site. We find that CNG trinucleotide repeats associated with expansion form stable hairpins whether they are short (with as few as 10 repeats) or long. However, long repeating stretches exist as single strands up to 2 orders of magnitude longer than sequences with either short repeats or random DNA. Thus, long hairpins have long lifetimes even in the presence of their complementary strands and inhibit duplex reannealing at a slippage site. The kinetic properties explain why expansion occurs with high frequency at long repeats but not at short repeats or palindromes.

The large-scale expansion of contiguous repeating trinucleotide units in genomic DNA has been linked to 10 genetic neurological diseases, including Huntington's disease (HD) and fragile X (for review see refs *1*−*5*). Unlike hereditary forms of colon cancer, trinucleotide instability in neurological disease is a DNA-directed mutation that initiates from improper DNA structure at the repeat region (*6, 7*). Secondary structures inhibit DNA replication (*7*) within the repeat regions, facilitating polymerase slippage and/or sister chromatid exchange (*7*). GAA repeats in Friedreich's ataxia form a YRY triple helix containing non-Watson−Crick pairs (*7*). However, all other trinucleotide diseases arise from instability at CNG repeats (where N is either A, C, T, or G) that form intramolecular hairpins (*1*−*5*).

Trinucleotide repeat expansion in disease occurs at a limited number of sequences and requires a long repeat stretch. We (*6, 7*) and others (*8, 9*) have reported that the sequence characteristics of disease for the CNG-type repeats may be explained by DNA hairpin formation (*6*−*9*). CNG sequences that expand are those that form hairpin structures *in vitro* (*6*). *In vivo*, expansion of repeating CNG trinucleotides appears to require a hairpin of threshold length before a significant probability for expansion exists (*1, 6*). Hairpins form more readily in haplotypes predisposed to disease in HD (*10*) and sequence interruptions (*11*) associated with stable alleles destabilize hairpins at repeating regions (*6*). Thus, secondary structure explains the sequence selectivity of CNG expansion.

However, the requirement for a long repeat length is less clear. Why are only long trinucleotide repeats particularly prone to expansion? Polymerase slippage is part of many mechanistic models for instability (*1, 12*−*17*). Slippage is more probable if the repeat stretch is long (*1, 18*). However, the probability of slippage between 29 and 39 repeats, at the threshold for CAG instability in disease, is only modestly increased (*1*) while the frequency of instability in HD (for example) increases dramatically in this range (*1, 18*−*21*). Hairpins can stabilize a slippage event (*1*). It is possible that long repeat lengths are required to form hairpins of sufficient stability. However, if instability occurs by a single, large hairpin-stabilized event, why is expansion not typically associated with palindromic regions that also can form hairpins? Alternatively, expansion may be the sum of many small slippage events, too small to support hairpin formation at palindromes. However, if short hairpins form at several slips along a long repeat, then instability should occur at repeat lengths well below the CNG threshold value (29−35 repeats) (*1*−*5*). Both short stretches of repeats and palindromic sequences can theoretically form hairpins but neither typically expands. Therefore, the ability to form DNA secondary structure is necessary but not sufficient to explain the characteristics of expansion in human disease. The repeating duplex DNA itself may have unique properties that may contribute to expansion. Alternatively, hairpins formed from long trinucleotide repeats may interact differently with their environment (including its complementary strand) or have unique structural properties that direct expansion.

In this report, we focus on what property of long CNG repeats renders them better templates for instability relative to random sequence DNA or short repeats. If slippage

† This work was supported by the Mayo Foundation, the Eagle's Cancer Fund, the National Science Foundation Grant IBN 9222848 and the NIH Grant DK 43694-01A2 to C.T.M.

* To whom correspondence should be addressed: E-mail mcmurray.cynthia@mayo.edu; Fax 507-284-9111; Telephone 507-284-1597.

‡ Department of Biochemistry and Molecular Biology.

§ Department of Pharmacology.

S0006-2960(98)00157-3 CCC: $15.00
Published on Web 06/11/1998

Table 1: Oligonucleotides Used in Thermal Melting and Kinetics Experiments

Name	Sequence (5′ → 3′)
CAG25	CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG
CTG25	CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG
CAG25ran	ACG GCA ACG CAG CAG GAC AGC AGC CGA GAC CGA CAG AGC CAG ACG GCA ACG GAC ACG GCA GAC CGA GAC GCA AGC
CTG25ran	GCT TGC GTC TCG GTC TGC CGT GTC CGT TGC CGT CTG GCT CTG TCG GTC TCG GCT GCT GTC CTG CTG CGT TGC CGT
CGG25	CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG
CCG25	CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG
GAC25	GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC
GTC25	GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC
CA37	CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA
TG37	TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG TG
AAG25	AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG
CTT25	CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT
CAG10	CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG
CTG10	CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG
CAG10ran	ACG AGC AGC GCA ACG AGC ACG ACG GAC GCA
CTG10ran	TGC GTC CGT CGT GCT CGT TGC GCT GCT CGT

occurs, then the slipped strand competes for binding to itself (hairpin formation) or with its heteroduplex partner for duplex reannealing. Since CNG repeats form hairpins, we evaluated the properties of CNG repeating heteroduplexes and how hairpin formation influences template reannealing. We find that hairpin structure in long repeats inhibits the rate of heteroduplex formation up to 2 orders of magnitude compared to sequences that are incapable of forming hairpins. In contrast, short repeat sequences not associated with expansion can form stable hairpins, yet hairpin structure does not alter the kinetics of duplex formation relative to random, unstructured sequences. The data suggest a model in which the steady-state lifetime rather than the ability to form hairpins is the major driving force for expansion. Additionally, the model provides a reason long repeats but not short repeats or palindromes expand.

EXPERIMENTAL PROCEDURES

Oligonucleotide Selection, Synthesis, and Preparation. Sixteen single-stranded oligonucleotides were synthesized by the Mayo Foundation oligonucleotide synthesis core facility. Each repeating oligonucleotide is designated by the sequence and the copy number. For example, CAG25 is a 75 bp oligonucleotide of 25 CAG repeats and CA37 is a 74 bp oligonucleotide of 37 CA repeats. The 12 repeating oligonucleotides are CAG25, CTG25, CGG25, CCG25, GAC25, GTC25, CA37, TG37, AAG25, CTT25, CAG10, and CTG10. The random CAG/CTG sequences (Table 1) were designed by randomizing the position of the bases within each triplet. This process generates oligonucleotide sequences with identical base composition and similar base distribution but with no specific secondary structure. No useful randomizations exist for repeating CCG or CGG due the large degree of CG base pairing redundancy.

Purification and resuspension of these oligonucleotides has been previously described (*22*). Additionally, the oligonucleotide of the desired size was purified from incomplete oligonucleotide syntheses by denaturing polyacrylamide gel electrophoresis. Approximately 60 nmol of each synthesis was loaded on a 10 M formamide, 7 M urea, 89 mM Tris−borate, 8% acrylamide/bisacrylamide (37.5:1) gel. Samples were visualized by UV shadowing on a TLC plate, and the band corresponding to the proper length oligonucleotide was cut from the gel. The gel slice was crushed, covered with 3 mL of Nanopure water, and heated to 37 °C overnight. The solution was strained through silanized glass wool and passed over a Bio-Rad 10 DG desalting column. Samples were lyophilized and resuspended in 50 μL of Nanopure water. As needed, oligonucleotides were end-labeled by T4 polynucleotide kinase with ^{32}P from [γ-^{32}P]ATP, by techniques described previously (*23*).

For the UV spectroscopy measurements, each oligonucleotide was diluted to the desired concentration (4.6×10^{-7} or 2.3×10^{-6} M) in 10 mM Pipes, 100 mM NaCl, and 0.1 mM EDTA, pH 7.0 (Pipes-10 buffer) to a final volume of 1.0 mL. For the nondenaturing gel electrophoresis experiment, 1 μL of each end-labeled oligonucleotide was diluted in 10 μL of 25 mM Pipes, 0.1 mM EDTA, and 10% glycerol, pH 7.0. For the denaturing gel electrophoresis experiment, 1 μL of each end-labeled oligonucleotide was diluted in 10 μL of 7 M urea, 25 mM Pipes, 0.1 mM EDTA, and 10% glycerol, pH 7.0. For the gel electrophoresis kinetics experiment, CAG25 and CTG25 were diluted into 1.5 mL of 25 mM Pipes and 0.1 mM EDTA, pH 7.0, to a final concentration of 9.2×10^{-7} M. End-labeled CAG25 (1 μL at approximately 2 μM) was then added to the unlabeled CAG25. For all experiments, duplex formation was performed by equimolar quantitative mixing of complementary oligonucleotides (see Kinetics Data Collection and Analysis) (*22*).

Theoretical Predictions of Single-Stranded DNA Secondary Structure. The Wisconsin Sequence Analysis version of the RNA folding program FoldRNA (*24*) was modified to simulate single-stranded DNA folding. The file fold.energy contains the parameters for determining the free energy (ΔG at 37 °C) of a single-stranded RNA molecule. In addition to nearest-neighbor base pairing and stacking energies, the file contains free energy contributions from hairpin loops, bulge loops, and interior loops to determine the overall free energy for the molecule. To adapt the file to DNA, the RNA nearest-neighbor values were replaced with the reported DNA nearest-neighbor energies (*25*). The

fold.energy file contains energies for GT pairs, interior loops, and bulge loops in RNA. These values were not modified. The energy of DNA hairpin loops has been well characterized (*26*−*28*), and reported values for DNA hairpin loop energies (*28*) replace the RNA values. The folded sequences were then visualized using the Squiggles graphic display routine in the Wisconsin Sequence Analysis package.

Polyacrylamide Gel Electrophoresis. Denaturing and nondenaturing polyacrylamide gels were typically made according to standard methods with modifications in the buffers (*23*). Oligonucleotides were separated on a 30 cm × 14 cm × 1 mm polyacrylamide gel. Nondenaturing 12% polyacrylamide/0.6% bisacrylamide (19:1) gel were typically buffered in 89 mM Tris−borate and 0.1 mM EDTA at pH 7.0. For kinetics experiments, the hairpin to duplex transitions in Pipes-10 (10 mM Pipes, 100 mM NaCl, and 0.1 mM EDTA, pH 7.0) were analyzed on gels in which the sodium ion concentration was varied between 45 and 150 mM. We found that separation of the molecules occurred more rapidly than duplex reannealing in this range and samples could be accurately analyzed on gels at any ionic strength in this range. The denaturing 8% acrylamide/0.2% bisacrylamide (37.5:1) gel, containing 10 M formamide and 7 M urea, was buffered in 89 mM Tris−borate and 0.1 mM EDTA at pH 7.0. No dye was added to the samples; instead, bromophenol blue markers were loaded into separate wells to monitor mobility. The gel was run under a constant voltage of 350 V until the desired separation was achieved. End-labeled oligonucleotides were visualized by exposure to film or by phosphorimaging.

Thermal Melting Data Collection and Analysis. Single- and double-stranded oligonucleotide samples were diluted to a concentration of 4.6×10^{-7} M in Pipes-10. Absorbance data collection was performed using a Cary 3 UV−vis spectrophotometer as described previously (*22*). The melting transition temperature [T_m (°C)] for each double-stranded sequence was determined by the maximum of a Gaussian distribution fit to the first derivative of the absorbance versus temperature curve as described previously (*22*).

The enthalpy, entropy, and free energy for hairpin to strand transitions were determined by a modified version of previously described methods (*27*, *29*). The following equation represents the absorbance versus temperature data from the spectrophotometer as a function of $\Delta H°_{vH}$ and $\Delta S°$:

$$\text{abs} = (\text{abs}_F - \text{abs}_I)\frac{1}{1 + e^{[(-\Delta H°_{vH}/RT)+(\Delta S°/R)]}} + \text{abs}_I \quad (1)$$

where abs is the observed absorbance, abs_f is the absorbance of the fully melted strands, abs_I is the absorbance of the fully formed hairpins, $\Delta H°_{vH}$ is the van't Hoff enthalpy, $\Delta S°$ is the entropy, T is the absolute temperature, and R is the gas constant.

The enthalpy, entropy, and free energy for duplex to strand transitions were determined by

$$\text{abs} = (\text{abs}_F - \text{abs}_I)\frac{e^{[(-\Delta H°_{vH}/RT)+(\Delta S°/R)]}}{2[C_T]} \times \left(\sqrt{1 + \frac{4[C_T]}{e^{[(-\Delta H°_{vH}/RT)+(\Delta S°/R)]}}} - 1\right) + \text{abs}_I \quad (2)$$

where the parameters are the same as in eq 1, with the addition of the duplex concentration $[C_T]$.

Fitting the equations to actual absorbance data required the subtraction of both an upper and lower baseline. We chose to have the curve-fitting procedure subtract the baseline simultaneously with finding the best thermodynamic equation fit to the data. In the curve-fitting procedure, upper and lower baseline slopes are graphically determined, supplied as initial values and then allowed to vary, along with the parameters in eq 1 or 2. Curve-fitting with concurrent baseline subtraction yields values that are virtually identical to values obtained through normal baseline subtraction. This method has the advantage that no assumptions are made about the temperature at which the baseline shifts from the lower to the upper baseline, and it is faster and yields more accurate values (lower individual errors, R values, and χ^2 values) than the individual steps of baseline subtraction followed by curve-fitting.

Kinetics Data Collection and Analysis. The kinetics of CAG25/CTG25 duplex formation analyzed by nondenaturing polyacrylamide gel electrophoresis was performed in the following manner. Unlabeled CTG25 and labeled CAG25 were allowed to equilibrate to 37 °C for half an hour. For each time point, 50 μL of CAG25 and CTG25 were mixed in a microfuge tube and held at a constant temperature of 37 °C until gel loading. The longest time point was mixed first and subsequent samples were timed and loaded onto the gel such that all samples corresponding to the indicated times could be separated simultaneously on the same gel. For kinetic gel experiments, the hairpin to duplex transitions in Pipes-10 (10 mM Pipes, 100 mM NaCl, and 0.1 mM EDTA, pH 7.0) were analyzed on gels in which the sodium ion concentration was varied between 45 and 150 mM. We found that separation of the molecules occurred more rapidly than duplex reannealing in this range and samples could be accurately analyzed on gels at any ionic strength in this range. We typically analyzed samples on standard TBE gels (45 mM sodium) as described above (see Polyacrylamide Gel Electrophoresis section). For the ultraviolet spectroscopic experiments, each 1 mL (unlabeled) hairpin sample was placed in a separate chamber of a Hellma 238-QS quartz mixing cell and equilibrated to the appropriate temperature. After temperature equilibration, the cuvette was inverted to ensure complete mixing of the two samples. Mixing diluted the samples another 2-fold as each 1 mL sample was diluted to a final volume of 2 mL. Absorbance data was collected as a function of time. The rate constant was determined by a modified form of a second-order equation fit to the data, derived in detail elsewhere (*22*, *30*). The perfectly paired heteroduplex has a lower extinction coefficient at 260 nm and 37 °C compared to the sum of extinction coefficients of both single-stranded sequences. Thus, duplex formation is observed as a gradual decrease in the absorbance as a function of time. Details of the instrumentation and analysis are previously described (*22*).

For the pseudo-first-order experiments, one sample was diluted to 4.6×10^{-7} M while the other sample was diluted to 2.3×10^{-7} M, 5 times the concentration of the first sample. The equation used for the pseudo-first-order condition experiments is the general second-order equation for reactants of different concentration, derived in detail elsewhere (*22*, *30*).

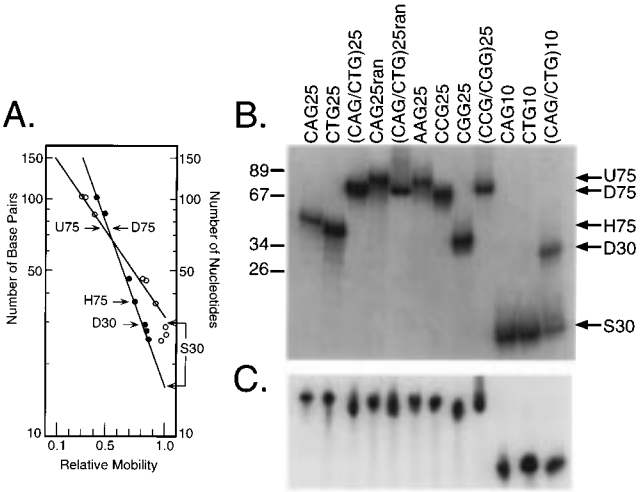

FIGURE 1: Structure of oligonucleotides, on the basis of electrophoretic mobility. (A) Single-stranded oligonucleotide (○) and double-stranded oligonucleotide (●) electrophoretic mobility as a function of length from Maniatis et al. (*35*). Predicted mobilities of oligonucleotides in panel B: U75, an unstructured 75 base oligonucleotide; D75, a 75 base pair oligonucleotide duplex; H75, a 75 base oligonucleotide capable of folding back into a hairpin structure; D30, a 30 base pair oligonucleotide duplex; S30, a 30 base oligonucleotide, structured or unstructured. (B) Nondenaturing polyacrylamide gel electrophoresis of single- and double-stranded oligonucleotides with the observed mobilities as indicated in panel A. (C) Denaturing (10 M formamide, 7 M urea) polyacrylamide gel electrophoresis of the oligonucleotides in panel B, indicating the mobility of sequences in the absence of specific secondary structure for both 75 base and 30 base sequences.

RESULTS

Duplexes of Trinucleotide Repeat Hairpins Do Not Display Unique Properties. To determine if a heteroduplex comprising a long stretch of trinucleotide repeats exhibits unusual properties that differ from those of random sequences, we examined the electrophoretic mobility and melting profiles among a set of oligonucleotides. Included in the set are CAG25/CTG25 and CCG25/CGG25 duplexes, which contain 25 contiguous repeats and are associated with expansion and disease. CAG25ran/CTG25ran is a control sequence for base composition (Table 1). Randomizing the position of the bases within each triplet generates oligonucleotide sequences with identical base composition and similar base distribution but with no specific secondary structure. The GAC25 and GTC25 oligonucleotides have an identical base composition to CAG25 and CTG25 but in the reverse order. The GAC single strands have the theoretical capacity to form hairpins (*6*) but are only found in short stretches in the human genome and have not been associated with expansion (*6*). Instability in AAG/CTT repeats occurs in Friedreich's ataxia (*7*, *31*), but these repeats cannot form hairpins (*6*, *7*). CA/TG repeats are dinucleotide repeats that have not been implicated in large-scale expansion but have been associated with small-scale ($\pm 1-4$ repeats) instability (*1*) in some colon cancers (*32*−*34*).

The gel electrophoresis experiments demonstrate that CAG25/CTG25 and CCG25/CGG25 have single-band electrophoretic mobilities identical to the mobility of the randomized duplex, migrating between the 67 and 89 bp markers (Figure 1B), in agreement with predicted mobilities (Figure 1A) (*35*). Control single strands (S, U) and hairpins (H) migrate according to their molecular mass and conformation (Figure 1B,C). Identical mobilities of both repeating and random DNA indicate two important results. First, a heteroduplex formed from complementary strands is the dominant structure of a trinucleotide repeat in the presence of its complementary sequence. Second, the narrow, precise comigration with the random sequence further indicates that the repeat heteroduplex is perfectly paired, without a significant amount of "slipped" duplex (i.e., single strands forming a duplex offset by one or more repeats). Thus, the ability of the individual strands to form hairpins does not significantly influence heteroduplex structure. Thermal melting experiments confirm the structural similarity of the heteroduplex sequences (Table 2). We find that the melting curves of each heteroduplex are cooperative transitions with similar thermodynamic parameters (Table 2). The T_m (°C) values of each heteroduplex roughly correlate with GC content regardless of either the repeating nature of a sequence or the ability to form hairpins. The CA37/TG37 and AAG25/CTT25 sequences, with 50% or less GC content, display low melting temperatures, while the 100% GC sequence, CGG25/CCG25, displays the highest melting temperature. Specifically, no significant differences were seen among the three duplexes (CAG25/CTG25, CAG25ran/CTG25ran, and GAC25/GTC25) with identical base composition. We conclude that the trinucleotide repeat duplexes associated with expansion are not significantly unique in either structure or stability.

Table 2: Melting Temperatures of Double-Stranded Repeating Oligonucleotides[a]

	T_m (°C)	ΔH (kcal/mol)	ΔS [cal/(deg·mol)][c]
CAG25/CTG25	79.1 ± 1.0	187 ± 1.2	501 ± 20.5
CAG25ran/CTG25ran	85.2 ± 1.6	251 ± 2.5	670 ± 17.6
CGG25/CCG25	89.9 ± 1.1	392 ± 3.8	1051 ± 35
GAC25/GTC25	79.0 ± 1.2	274 ± 3.6	748 ± 27.8
CA37/TG37	68.2 ± 0.6	155 ± 2.1	423 ± 7.8
AAG25/CTT25	67.8 ± 0.7	280 ± 1.2	791 ± 17.4
CAG10/CTG10	74.6 ± 2.0	*b*	*b*
CAG10ran/CTG10ran	79.6 ± 1.0	*b*	*b*

[a] Summary of the thermodynamics of double-stranded trinucleotide and dinucleotide sequences. Reactions were performed at a concentration of 0.7 μM in 10 mM Pipes, 100 mM NaCl, and 0.1 mM EDTA (Pipes-10) buffer. Each reported value is the mean of four experiments. [b] The data could not be fit to eq 2.

Trinucleotide Repeat Hairpins Display a Wide Range of Energies of Formation. As previously reported (*6*, *8*), however, the individual strands of CNG repeating trinucleotides associated with expansion form intramolecular hairpin structures with a repeating motif of two GC pairs and a single base-base mismatch (Figure 2A). The six sequences predicted to form hairpins (CAG25, CTG25, CCG25, CGG25, GAC25, and GTC25) display thermal melting profiles similar to that of CAG25, shown in Figure 2C. This melting profile reflects hairpin structure since CAG25ran, with identical base composition but no theoretical capacity for hairpin structure (Figure 2B), displays no discernible T_m (°C) (Figure 2D; Table 3). The CAG25ran melting profile is a noncooperative absorbance increase with temperature due to only isolated base pairings and modest base stacking. CAG25, CTG25, CCG25, and GAC25 all have melting temperatures near 50 °C while CGG25 has a melting temperature of 75 °C and GTC25 has a melting temperature of 43 °C (Table 3). Gel electrophoresis confirms hairpin formation of these oligo-

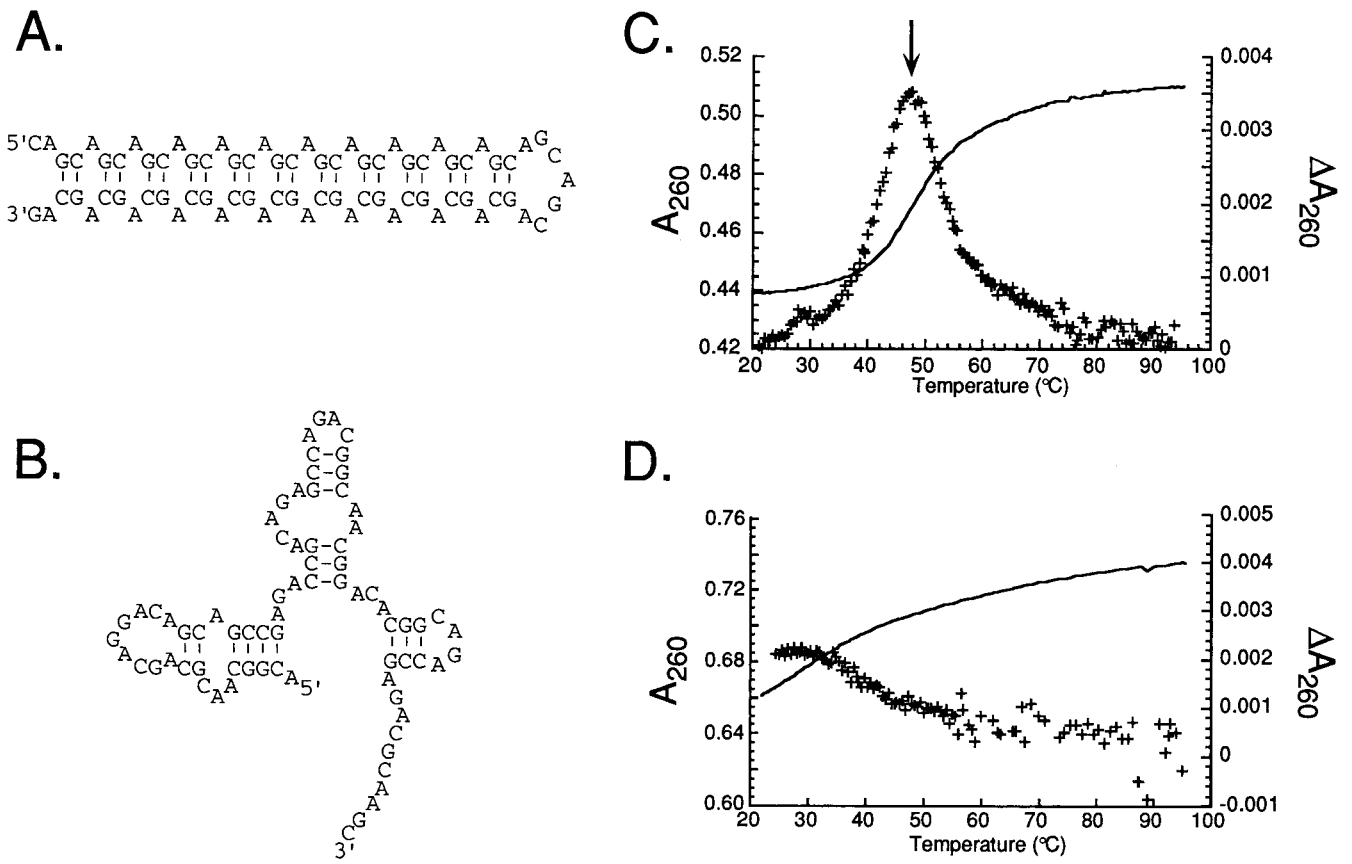

FIGURE 2: Predicted structures and observed thermal melting profiles of oligonucleotides of 25 repeating CAGs. For both (A) CAG25 and (B) CAG25ran, free energy minimization of the single-stranded structure is graphically represented, on the basis of base connections made by foldRNA from the GCG package with the energies modified for DNA. The thermal melting profiles of (C) CAG25 and (D) CAG25ran in Pipes-10 at 1.4 μM. The solid line represents the absorbance at 260 nm (A_{260}) as a function of temperature and the plus sign indicates the first derivative of the absorbance at 260 nm as a function of temperature (ΔA_{260}).

Table 3: Thermal Stability of Repeating Oligonucleotides[a]

	T_m (°C)	$\Delta H°_{vH}$ (kcal/mol)	$\Delta S°$ [cal/(deg·mol)]	$G°$ at 37 °C (kcal/mol)
CAG25	50.1 ± 0.8	−60.8	−179	−5.3
CTG25	51.4 ± 0.9	−100	−308	−7.8
CGG25	75.1 ± 1.2	−120	−347	−12.4
CCG25	48.9 ± 0.7	−88.1	−274	−3.2
GAC25	54.9 ± 1.7	−103	−313	−6.0
GTC25	43.0 ± 3.6	−124	−393	−2.2
CAG10	50.1 ± 0.8	−28.9	−88	−2.2
CTG10	51.6 ± 0.7	−65.6	−170	−4.3

[a] Summary of the thermodynamic properties of single-stranded trinucleotide and dinucleotide sequences. All parameters are determined from thermal melting experiments. Each reaction was performed at a concentration of 1.4 μM in Pipes-10 buffer. Each reported value is the mean of four experiments. The data for CAG25ran, CTG25ran, CA37, TG37, AAG25, CTT25, CAG10ran, and CTG10ran did not form hairpins and could not be fit to eq 1.

nucleotides used in the melting experiments (Figure 1B,C). Since each repeat differs only in the central mismatch pair (all are CNG repeats), the differences in the thermodynamic parameters (Table 3) are due to differences in base stacking and hydrogen bonding associated with a T-T, A-A, or G-G mismatched base pair (*36*). We observed up to a 2-fold variation in the enthalpy or entropy associated with different CNG hairpins, with CGG25 the highest and CAG25 the lowest. In contrast to the heteroduplexes, significant differences in structure occur among repeating hairpins.

The Rate of Duplex Formation Is Reduced by Hairpin Structures. Expansion is thought to involve slipped hairpin formation on either template or daughter strand during replication. As hairpin formation occurs in the presence of the complementary strand, we evaluated the effect of hairpin structure on the rate of reannealing at trinucleotide repeat segments. Pure populations of each complementary hairpin (CAG25 and CTG25, for example) were equilibrated at 37 °C in a mixing cell, mixed, and monitored for duplex formation. Mixing results in an immediate decrease in absorbance that progresses over approximately 1 h. A plot of absorbance versus time for the formation of CAG25/CTG25 at 37 °C is shown in Figure 3B. To ensure that the absorbance decrease was due to duplex formation and to ensure the integrity of our starting material, we monitored the conformational state of the sample at several time points of the reaction using gel electrophoresis (Figure 3A). The reactions were analyzed on 12% polyacrylamide gels and the relative fraction of hairpin was measured by densitometry. The densitometry trace of individual lanes from the gel electrophoresis analysis of the hairpin sample are shown before and at selected time points after mixing (Figure 3B). We found that the relative fraction of hairpin and duplex corresponded well with the observed decrease in absorbance (Figure 3B). Before the start of the reaction, all of the oligonucleotide is in the hairpin configuration (Figure 3A, lane H). At approximately 25−30 min, we observed half of the total absorbance decrease and the amounts of hairpin

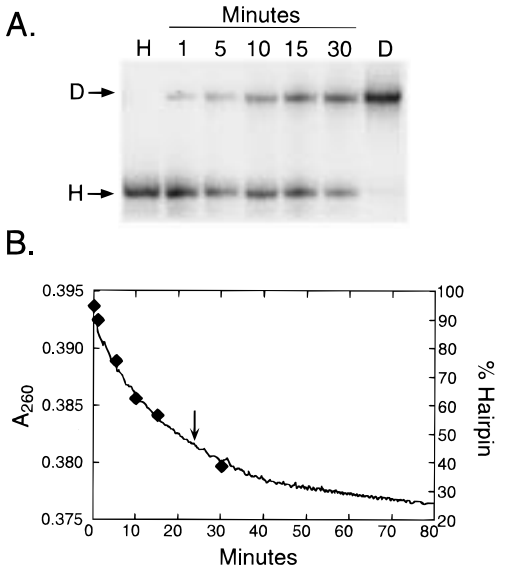

FIGURE 3: Time course of the hairpin to duplex transition. (A) Polyacrylamide gel electrophoresis of a CAG25/CTG25 duplex at indicated times after equimolar mixing. H, hairpin structure in the absence of complementary strand; D, duplex 24 h after equimolar mixing. (B) Formation of duplex measured by the decrease in absorbance at 260 nm (A_{260}) as a function of time (minutes). The arrow indicates the half-time determined from a second-order fit to the absorbance curve. The symbols (◆) indicate the calculated hairpin percentage as determined from the gel electrophoresis results in panel A.

Table 4: Kinetics of Duplex Formation from Complementary Oligonucleotides at 37 °C

	k^a (M^{-1} s^{-1})	R^b
CAG25/CTG25	8500	0.99
CAG25ran/CTG25ran	104 800	0.99
CGG25/CCG25	3000	0.99
GAC25/GTC25	5500	0.99
CA37/TG37	300 000	0.99
AAG25/CTT25	220 000	0.98
CAG10/CTG10	40 000	0.98
CAG10ran/CTG10ran	45 000	0.98
palindromic Enk enhancer[c]	7300	0.99

[a] Rate constant of duplex formation. Complementary oligonucleotides (1 mL, 0.7 μM) in Pipes-10 buffer are mixed to form the duplex. [b] Correlation coefficient of the best-fit curve to the data. [c] Data taken from ref *22*. The 23 bp enhancer of the human proenkephalin gene has a 74% GC content.

and duplex measured by densitometry are roughly equal, and after 24 h (Figure 3A, lane D) essentially all of the oligonucleotide is in the duplex form. The gel electrophoresis experiment also demonstrates the two-state nature of heteroduplex formation. Only two species, the hairpin and the duplex, are seen at any time throughout the course of the reaction, indicating that duplex formation occurs without the detectable presence of any intermediate structure.

Using the equation previously described for second-order reaction kinetics (*22*), we calculated the rates constants of heteroduplex formation for the repeating sequences as determined by ultraviolet spectroscopy (Table 4). Duplexes formed from repeating stretches that form hairpins exhibit similar rates of formation. These rates range from 1 to 2 orders of magnitude slower than the rates of duplex formation from unstructured sequences (CA37/TG37, CAG25ran, and AAG25/CTT25), indicating that hairpin structure indeed inhibits duplex formation. This inhibitory effect is due to the presence of structure and not repeating sequence, as CA37/TG37 and AAG25/CTT25 display kinetics approximately equal to the kinetics of the randomized sequence.

Table 5: Individual Effect of Complementary Oligonucleotides on the Rate of Duplex Formation at 37 °C

	k^a (M^{-1} s^{-1})	R^b
CAG25−CTG25	11 200	0.99
5×CAG25−CTG25	12 700	0.99
CAG25−5×CTG25	10 900	0.98
CGG25−CCG25	2700	0.92
5×CGG25−CCG25	2200	0.92
CGG25−5×CCG25	2600	0.88

[a] Rate constant of duplex formation. Complementary oligonucleotides (1 mL) at 0.3 μM or 1.5 μM (5×) in Pipes-10 buffer are mixed to form the duplex. [b] Correlation coefficient of the best-fit curve to the data.

Complementary Hairpins Contribute Equally to the Rate of Duplex Formation. Since complementary hairpin structures display different degrees of stability and base stacking (Table 3), we next asked whether hairpin structure on one specific strand is rate-limiting for heteroduplex formation. To address this question, we measured the rate of heteroduplex formation under pseudo-first-order conditions. We compared the rates of oligonucleotides containing trinucleotide repeats associated with expansion with those of randomized sequences. Reactions are prepared as before, except that either one of the oligonucleotides is in 5-fold excess. The rate of duplex formation is determined by fitting the data to a second-order reaction equation, but with reactants at different concentrations and adjusted for extinction coefficients (*22*). For either CAG/CTG or for CGG/CCG, we observed no large differences between either pseudo-first-order rate constant compared to the equimolar rate constant (Table 5). These results indicate that hairpins forming on either strand have a similar inhibitory effect on reannealing. Since the hairpins are stable, duplex formation is a simple second-order process and nucleation is the rate-limiting step of duplex formation.

Short Trinucleotide Repeats Form Hairpins but Do Not Inhibit Duplex Formation. Trinucleotide sequences of less than 10 repeats have not been implicated in large scale expansion (*1*−*5*). We then wished to consider if the inability of short repeats to expand was due to hairpin stability. To address this question, we generated two sets of 30 bp oligonucleotides. CAG10 and its complement CTG10 represent a short stretch of trinucleotide repeats. CAG10ran and its complement CTG10ran (Table 1) represent a randomized sequence, similar to CAG25ran and CTG25ran. We find that CAG10 and CAG10/CTG10 sequences have the predicted electrophoretic mobilities for single- and double-stranded sequences, respectively (Figure 1B,C). However, at only 30 bases, differences between the mobility of unstructured single strands and hairpin sequences are not well resolved (Figure 1A). Both CAG10 and CTG10 have cooperative melting transitions (Table 3) indicative of structure formation, but the randomized sequences do not. While CAG10 can form a hairpin of equal stability to CAG25, the rates of duplex formation for the repeat sequence and the randomized sequence are virtually identical. In

contrast to the long repeats, short, repetitive hairpins rapidly melt to form duplexes in the presence of their complements (Table 4).

DISCUSSION

Both genetic and biochemical data indicate that trinucleotide repeat instability is a replication/repair error that is dependent on improper DNA secondary structure formation at the repeating region of the affected gene (*1*−*5*, *7*, *13*, *19*, *37*−*42*). While the DNA itself is a central part of the mechanism, it was not clear why a critical length of repeat DNA is necessary for DNA instability and how this requirement is related to secondary structure formation.

In this paper, we have focused primarily on CNG repeats that form hairpins. Since most models of DNA instability include DNA slippage, the properties of both the duplex and the individual strands at repeating segments are relevant to a model of expansion and may contribute to the observed length dependence. However, we show that heteroduplexes of trinucleotides associated with expansion do not display properties that differentiate them either from repeating sequences with no capacity for structure or from random sequences. Our results indicate that instability and its length dependence must be related in some way to the ability of trinucleotides to form structure. Of the CNG repeats associated with expansion, CAG25, CTG25, and CGG25 form stable hairpins. We observed that CCG25 displays an electrophoretic mobility unlike that of either the duplex or hairpin sequences. However, evidence for both a unique double-stranded CCG structure (the e-motif) (*43*) and a distorted helix (*44*) have been reported for this repeat, either of which may explain the aberrant mobility.

Our results indicate several important consequences of hairpin formation that may shed light on the length dependence of expansion and a mechanism. First, hairpin formation does not distinguish long "expansion-capable" repeats from short, stable repeats. We show here that both short and long CNG repeats form hairpins of equal stability although only long repeats tend to expand. Specifically, hairpins formed from CAG10 and CTG10 have stability roughly equal to CAG25 and CTG25. However, CAG10/CTG10 hairpin rapidly reanneals with its complementary strand at a rate that is indistinguishable from randomized CAG10ran/CTG10ran or random DNA. In contrast, CAG25/CTG25 forms a duplex up to 2 orders of magnitude more slowly than the randomized CAG25ran/CTG25ran. Our results suggest that hairpin stability is necessary but not sufficient for expansion. Rather, in evaluating the influence of hairpin formation at a slip, we find that a long repeating segment competes better for self-pairing than for heteroduplex formation with its complementary strand. Self-pairing inhibits template reannealing, extending the lifetime of the hairpin. Slow reannealing is also observed for "slipped-strand" structures that form when the genomic DNA of myotonic dystrophy patients is thermally melted (*45*, *46*). Although these "slipped-strand" structures contain longer repeats and are more complex in sequence than our simple system, the observed structures resemble fold-back structures (*45*) and are stable for long periods (*46*). We find that short repeats that form stable hairpins reanneal to their complementary strands at a rate that is indistinguishable from random DNA or from repeats that form no structure. The data suggest that the kinetics of template reannealing distinguishes short repeats that are stably propagated from long repeats that are not. Hairpins of long lifetime do not reanneal to their complementary strands fast enough to prevent replication/repair errors.

Second, complementary CAG and CTG repeats form hairpins of similar stability and inhibit reannealing to the repeating heteroduplex to an equal extent. It is known that orientation of DNA is a factor in the degree of instability (*15*, *17*). For example, in both bacteria and yeast, CTG in the lagging strand is more unstable and tends to delete relative to CAG in the lagging strand (*15*, *17*). It has been suggested that the greater hairpin stability of CTG repeats may be responsible for this difference (*15*, *17*). However, our data indicate that CTG and CAG hairpins not only are similar in their stability but also have an equal ability to inhibit reannealing to the repeating heteroduplex if hairpins form. The lack of differential effects imposed by hairpins after they form points to a model in which orientation has a differential influence on the rate of formation of CTG relative to CAG hairpins. Recent reports indicate that secondary structure is more likely to form in the lagging strand (*47*, *48*). If this applies at repeats, then it is possible that differential protein−protein interactions (such as binding of single-stranded binding proteins) impose as yet unknown constraints on the formation of CAG relative to CTG hairpins. It is now critical to determine whether instability is limited to the lagging strand or is associated with both orientations.

Finally, our kinetics results provides an explanation as to why palindromes that form stable hairpins do not typically expand. During most of the life cycle of a cell, the DNA is perfectly paired to its complementary strand. However, strand unpairing during replication and transcription provides a window of opportunity for hairpin formation. At a slippage event (Figure 4, panel IA), hairpin formation on the leading strand or the replicated lagging strand must compete with its complement for duplex reformation (Figure 4, panel IC). On the lagging strand, hairpin formation on the template is in competition with the binding of single-stranded binding proteins (SSB), whose role is to prevent secondary structure (Figure 4, panel IB) (*1*, *13*). If hairpin formation is to compete effectively with duplex reannealing or with protein binding, hairpins must form quickly and must be slow to melt back to an unstructured strand (Figure 4, panels IB and IC). On the basis of the data presented here, only long repeats capable of hairpin formation fulfill both these criteria. Both long and short repeats may readily form hairpin structures because any pairing of repeats throughout the repetitive region may initiate hairpin formation (Figure 4, panels IIA and IIB). However, only the long repeats are slow to melt resulting in a hairpin of long lifetime. Our kinetic data clearly show that long but not short hairpins inhibit duplex formation. Since the hairpin itself is the competing structure, hairpin lifetime distinguishes long from short repeats and likely determines the probability of expansion. The hairpin structures extend the lifetime of the single-stranded state up to 2 orders of magnitude relative to strands composed of random sequences, repetitive sequences lacking stable secondary structure, or short repeats capable of forming hairpins. Once formed, the GC-rich hairpins

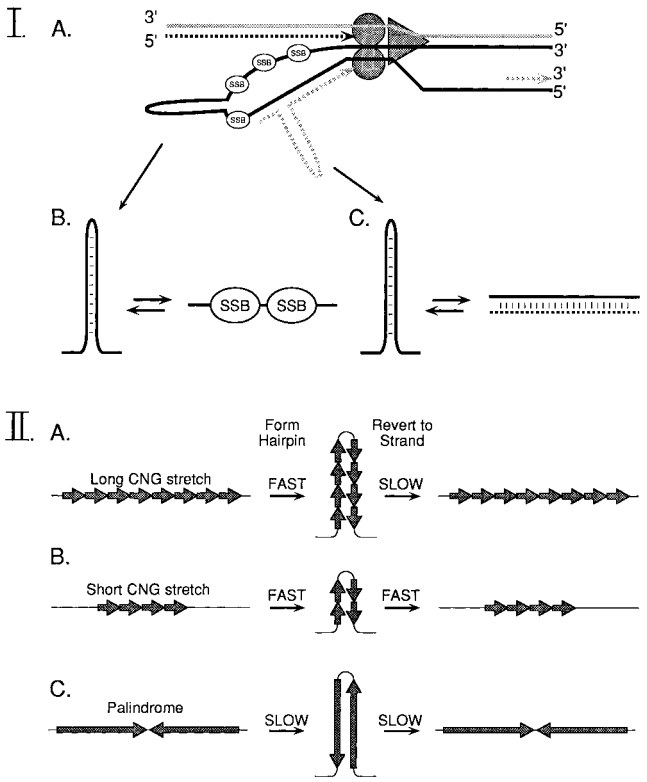

FIGURE 4: Only long repeats are likely to form hairpins during replication. (Panel I) Competition for hairpin formation during (A) replication by either (B) single-stranded binding proteins if the hairpin forms on the lagging strand or (C) the complementary strand if the hairpin forms on a replicated leading strand or replicated lagging strand. (Panel II) Likelihood of structure formation during replication. (A) A long stretch of repeats can quickly form a hairpin, due to the degeneracy of pairing possibilities. Once formed, it will only slowly return to the single-stranded state. (B) A short stretch of repeats can quickly form a hairpin but also quickly reverts to the single-stranded state. (C) A palindromic region will only slowly form a hairpin due to the exact pairing required for formation. Once formed, however, it would be stable and slow to return to the single-stranded state.

sequences are slow to unpair. Thus, polymerase can rebind and continue replication before hairpin dissociation occurs, trapping the secondary structure until repair.

Like the long repeat hairpin, a palindromic DNA sequence is very stable once the hairpin has formed (Figure 4, panel IIC). However, during replication, a palindromic hairpin rarely forms since only one pairing configuration will initiate hairpin formation. An entire half of the palindrome must be unpaired before hairpin formation is possible. We have directly examined the kinetics of a GC-rich palindrome within the proenkephalin enhancer (*22*, *49*). The proenkephalin enhancer is known to control transcriptional activity by switching into a hairpin conformation (*23*, *49*−*52*). No duplication or expansion of this region is observed in human populations (*53*). For the enhancer palindrome, duplex formation is 50-fold slower than for random sequences (*22*) and is similar to the rate of CAG25 or CTG25 hairpins (Table 4). Although the enkephalin hairpin is slow to dissociate, the size of the slip required to initiate its formation is 23 bp, larger than the typical size of a slippage event not stabilized by secondary structure (*14*). Thus, the rate of hairpin formation at a nonrepetitive, palindromic slip is slow (*22*), and duplex reannealing or binding of SSB is likely to occur before hairpin formation. However, insertion and deletion mutations are known to occur at palindromic regions (*54*). On the basis of our results, single insertion and deletion events are likely to represent "low-frequency" instability mediated by an occasional hairpin formation event. Because hairpin formation at palindromic regions is slow, the resulting deletion and insertion events will occur at much lower frequency than is observed for DNA instability at repetitive DNA. Again, kinetic parameters can explain why expansions are not observed at palindromes, which are abundant in the human genome.

In summary, multiple pairing possibilities allow hairpin formation to occur quickly at long repeats. Once formed, the GC-rich hairpins are slow to melt. Kinetic properties distinguish longer repeats from short repeats. Sequences that are too short cannot expand even if they form stable hairpins because the hairpins do not have sufficient lifetime in solution. Longer CNG hairpins form quickly and dissociate slowly compared to short repeats or random sequence DNA. CNG hairpins have a sufficient lifetime that polymerase can rebind and continue replication before hairpins reanneal to their complementary strands. Thus, trapped secondary structure can serve as substrates for repair, giving rise to expansion and contraction of a repeat.

REFERENCES

1. McMurray, C. T. (1995) *Chromosoma 104*, 2−13.
2. Gusella, J. F., and MacDonald, M. E. (1996) *Annu. Rev. Med. 47*, 201−209.
3. Bates, G., and Lehrach H. (1994) *Bioessays 16*, 277−284.
4. Warren, S. T. (1996) *Science 271*, 1374−1375.
5. Timchenko, L. T., and Caskey, C. T. (1996) *Faseb J.*, 1589−1597.
6. Gacy, A. M., Goellner, G. M., Juranic, N., Macura, S., and McMurray, C. T. (1995) *Cell 81*, 533−540.
7. Gacy A. M., Goellner, G. M., Spiro, C., Chen, X., Gupta, G., Bradbury, E. M., Dyer, R. B., Mikesell, M. J., Yao, J. Z., Johnson, A. J., Richter, A., Melancon, S. B., and McMurray, C. T. (1998) *Mol. Cell 1*, 583−593.
8. Mitas, M., Yu, A., Dill, J., Kamp, T. J., Chambers, E. J., and Haworth, I. S. (1995) *Nucleic Acids Res. 23*, 1050−1059.
9. Chen, X., Mariappan, S. V. S., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoun, A., Smith, S. S., Bradbury, E. M., and Gupta, G. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 5199−5203.
10. Goldberg, Y. P., et al. (1995) *Hum. Mol. Genet. 4*, 1911−1918.
11. Snow, K., Tester, D. J., Kruckeberg, K. E., Schaid, D. J., and Thibodeau, S. N. (1994) *Hum. Mol. Genet. 3*, 1543−1551.
12. Kunkel, T. A., (1993) *Nature 365*, 207−208.
13. Gordenin, D. A., Kunkel, T. A., and Resnick, M. A. (1997) *Nat. Genet. 16*, 116−118.
14. Sia, E. A., Jinks-Robertson, S., and Petes, T. D. (1997) *Mutat. Res. 383*, 61−70.
15. Kang, S., Jaworski, A., Ohshima, K., and Wells, R. D. (1995) *Nat. Genet. 10*, 213−218.
16. Strand, M., Prolla, T. A., Liskay, R. M., and Petes, T. D. (1993) *Nature 365*, 274−276.
17. Freudenreich, C. H., Stavenhagen, J. B., and Zakian, V. A. (1997) *Mol. Cell. Biol. 17*, 2090−2098.
18. Sia, E. A., Kokoska, R. A., Dominska, M., Greenwell, P., and Petes, T. D. (1997) *Mol. Cell. Biol. 17*, 2851−2858.
19. Goellner, G. M., Tester, D., Thibodeau, S., Almqvist, E., Goldberg, Y. P., Hayden, M. R., and McMurray, C. T. (1997) *Am. J. Hum. Genet. 60*, 879−890.
20. Brinkmann, R. R., Mezei, M. M., Theilmann, J., Almqvist, E., and Hayden, M. R. (1997) *Am. J. Hum. Genet. 60*, 1202−1210.
21. Goldberg, Y. P., Telenius, H., and Hayden, M. R. (1994) *Curr. Opin. Neurol. 7*, 325−332.

22. Gacy, A. M., and McMurray, C. T. (1994) *Biochemistry 33*, 11951−11959.
23. Spiro, C., Richards, J. P., Chandrasekaran, S., Brennan, R. G., and McMurray, C. T. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 4606−4610.
24. Zucker, M., and Steigler, P. (1981) *Nucleic Acids Res. 9*, 133−148.
25. Breslauer, K. J., Frank, R., Blöcker, H., and Marky, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746−3750.
26. Baxter, S. M., Greizerstein, M. B., Kushlan, D. M., and Ashley, G. W. (1993) *Biochemistry 32*, 8702−8711.
27. Xodo, L. E., Manzini, G., Quadrifoglio, F., van der Marel, G., and van Boom, J. (1991) *Nucleic Acids Res. 19*, 1505−1511.
28. Rentzeperis, D., Alessi, K., and Marky, L. A. (1993) *Nucleic Acids Res. 9*, 133−148.
29. Marky, L. A., and Breslauer, K. J. (1987) *Biopolymers 26*, 1601−1620.
30. Jencks, W. P. (1975) in *Catalysis in Chemistry and Enzymology*, pp 605−612, Dover, New York.
31. Campuzano, V., et al. (1996) *Science 271*, 1423−1426.
32. Fearon, E. R., Cho, K. R., Nigro, J. M., Kern, S. E., Simons, J. W., Ruppert, J. M., Hamilton, S. R., Preisinger, A. C., Thomas, G., Kinzler, K. W., and Vogelstein, B. (1990) *Science 247*, 49−56.
33. Aaltonen, L. A., Peltomäki, P., Leach, F. S., Sistonen, P., Pylkkänen, L., Mecklin, J.-P., Järvinen, H., Powell, S. M., Jen, J., Hamilton, S. R., Petersen, G. M., Kinzler, K. W., Bogelstein, B., and de la Chapelle, A. (1993) *Science 260*, 812−816.
34. Thibodeau, S. N., Bren, G., and Shaid, D. (1993) *Science 260*, 816−819.
35. Maniatis, T., Jeffrey, A., and van de Sande, H. (1975) *Biochemistry 14*, 3787−3794.
36. Abou-ela, F., Koh, D., Tinoco, I., Jr., and Martin, F. H. (1985) *Nucleic Acids Res. 13*, 4811−4824.
37. Kuhl, D. P. A., and Caskey, C. T. (1993) *Curr. Opin. Genet. Dev. 3*, 404−407.
38. Mangiarini, L., Sathasivam, K., Mahal, A., Mott, R., Seller, M., and Bates, G. P. (1997) *Nat. Genet. 15*, 197−200.
39. Monckton, D. G., Coolbaugh, M. I., Ashizawa, K. T., Siciliano, M. J., and Caskey, C. T. (1997) *Nat. Genet. 15*, 193−196.
40. Richards, R. I., and Sutherland, G. R. (1994) *Nat. Genet. 6*, 114−116.
41. Wells, R. D. (1996) *J. Biol. Chem. 271* (6), 2875−2878.
42. Hancock, J. M. (1996) *BioEssays 18* (5), 421−425.
43. Zheng, M., Huang, X., Smith, G. K., Yang, X., and Gao, X. (1996) *J. Mol. Biol. 264*, 323−336.
44. Yu, A., Barren, M. D., Romero, R. M., Christy, M., Gold, B., Dai, J. L., Gray, D. M., Haworth, I. S., and Mitas, M. (1997) *Biochemistry 36* (12), 3687−3699.
45. Pearson, C. E., Wang, Y., Griffith, J. D., and Sinden, R. R. (1998) *Nucleic Acids Res. 26*, 816−823.
46. Chastain, P. D., Eichler, E. E., Kang, S., Nelson, D. L., Levens, S. D., and Sinden, R. R. (1995) *Biochemistry 34*, 16125−16131.
47. Veaute, X., and Fuchs, R. P. P. (1994) *Science 266*, 598−600.
48. Trinh, T. Q., and Sinden, R. R. (1994) *Nature 352*, 544−547.
49. McMurray, C. T., Juranic, N., Chandrasekaran, S., Macura, S., Li, Y., Jones, R. L., and Wilson, W. D. (1994) *Biochemistry 33*, 11960−11970.
50. Spiro, C., Wu, X., Bazett-Jones, D. J., and McMurray, C. T. (1995) *J. Biol. Chem. 270* (46), 27702−27710.
51. Spiro, C., and McMurray, C. T. (1997) *J. Biol. Chem. 272*, 33145−33152.
52. McMurray, C. T., Wilson, W. D., and Douglass, J. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 666−670.
53. Mikesell, M., Sobell, J., Sommer, S., and McMurray, C. T. (1996) *Am. J. Med. Genet. 67*, 459−467.
54. Cooper, D. N., and Krawczak, M. (1991) *Hum. Genet. 87*, 409−415.

BI980157S