**Published online July 21, 2005**

*Nucleic Acids Research, 2005, Vol. 33, No. 13* ***4065–4077***
*doi:10.1093/nar/gki716*

# Length-dependent energetics of $(CTG)_n$ and $(CAG)_n$ trinucleotide repeats

**Samir Amrane, Barbara Saccà, Martin Mills[1], Madhu Chauhan[1], Horst H. Klump[1] and Jean-Louis Mergny***

Laboratoire de Biophysique, Muséum National d'Histoire Naturelle USM 503, INSERM UR 565, CNRS UMR 8646, 43 rue Cuvier, 75231 Paris cedex 05, France and [1]Department of Molecular and Cell Biology, University of Cape Town, P.B. Rondebosh 7701, Republic of South Africa

Received February 28, 2005; Revised and Accepted June 30, 2005



## ABSTRACT

**Trinucleotide repeats are involved in a number of debilitating diseases such as myotonic dystrophy. Twelve to seventy-five base-long $(CTG)_n$ oligodeoxynucleotides were analysed using a combination of biophysical [UV-absorbance, circular dichroism and differential scanning calorimetry (DSC)] and biochemical methods (non-denaturing gel electrophoresis and enzymatic footprinting). All oligomers formed stable intramolecular structures under near physiological conditions with a melting temperature that was only weakly dependent on oligomer length. Thermodynamic analysis of the denaturation process by UV-melting and calorimetric experiments revealed an unprecedented length-dependent discrepancy between the enthalpy values deduced from model-dependent (UV-melting) and model-independent (calorimetry) experiments. Evidence for non-zero molar heat capacity changes was also derived from the analysis of the Arrhenius plots and DSC profiles. Such behaviour is analysed in the framework of an intramolecular 'branched-hairpin' model, in which long CTG oligomers do not fold into a simple long hairpin–stem intramolecular structure, but allow the formation of several independent folding units of unequal stability. We demonstrate that, for sequences ranging from 12 to 25 CTG repeats, an intramolecular structure with two loops is formed which we will call 'bis-hairpin'. Similar results were also found for CAG oligomers, suggesting that this observation may be extended to various trinucleotide repeats-containing sequences.**

## INTRODUCTION

Recent molecular genetic studies have revealed a correlation between spontaneous expansion of several DNA trinucleotide repeats and a variety of debilitating human diseases [for reviews see (1–4)]. This class of diseases was first characterized in Fragile-X syndrome (5) and later in myotonic dystrophy and other disorders. Myotonic dystrophy type 1 (DM1) is caused by the expansion of a (CTG)–(CAG) repeat in the *DMPK* gene (6). To date, at least nine distinct loci show instability with the same (CAG)–(CTG) repeat. These diseases increase in severity with earlier onset in successive generations and have no cure. Although the pathological states show different characteristics, one common feature among them is that the affected repetitive DNA unit has expanded beyond the number of repeats found in the healthy population. Therefore, the expansion of triplet repeats represents a novel mutational mechanism. More recently, other disorders have been associated with the expansion of non-trinucleotide motifs, such as the CCTG tetranucleotide in myotonic dystrophy type 2 (7).

DNA secondary structures may be considered as a common and causative factor for triplet expansion (8–10) but the molecular mechanisms causing the instability are unknown and remain a subject of intensive study. Even if no therapeutic approach is currently available to prevent or revert repeat expansion, *in vitro* studies suggest that repeat deletion could be induced by various chemotherapeutic agents (11,12); thus, opening a new field of study aimed at the design of trinucleotide repeat-specific ligands. Preliminary results suggest that selective recognition of trinucleotide repeat structures is possible (S. Amrane, unpublished data), but the rational design of such ligands should be facilitated by knowledge of the structure and energetics of their nucleic acid target.

Repetitive CNG sequences are susceptible to the formation of duplexes by self-folding, forming two Watson–Crick G–C pairs and one mismatch pair (13–16). Long hairpins have long

*To whom correspondence should be addressed. Tel: +33 1 40 79 36 89; Fax: +33 1 40 79 37 05; Email: faucon@mnhn.fr
Present address:
Barbara Saccà, Laboratoire de Stabilité des Génomes, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris cedex 15, France

lifetimes and inhibit duplex reannealing (17,18). A controversy remains on how different are these structures from classical B-DNA. Besides forming 'slipped duplexes' (19), CGG repeats have been reported to form quadruplexes (20–23). However, a recent study demonstrated that CGG repeats are reluctant to form tetraplexes under physiological conditions and this structure is unlikely to be involved in the disease (24): these sequences preferentially fold into antiparallel homoduplexes or hairpins in a length-dependent manner.

Concerning the structure of the bimolecular $(CTG–CAG)_n$ duplex, several results revealed a 'polyhairpin concept'. Studies on $(CAG–CTG)_{30,50}$ showed that these duplexes formed alternative DNA duplex structures named SDNA and SiDNA (9,19,25–27). This conclusion was reached using atomic force microscopy, electron microscopy, native gel electrophoresis, Mung Bean (MB) and T7 endonuclease cleavage assays. SDNA is a well-described bimolecular polyhairpin structure composed of multiple short $(CTG)_n$ and $(CAG)_n$ ($n = 1–10$) slipped-out structures with a hairpin-like single-stranded character. SDNA is formed in the absence of replication when the two complementary strands have the same length. Instead, SiDNA is formed during replication between two complementary strands with a different number of repetitions and results in repeat expansion or deletion. SiDNA is composed of one major slipped-out structure containing 20–30 repeats of the longest $(CTG)_n$ or $(CAG)_n$ strand.

To date, only the short hairpin structures (containing <10 repeats) involved in the formation of SDNA have been clearly established (15,28) revealing an intramolecular stem with several repetitions of a T·T or A·A mismatch sandwiched between 2 G–C base pairs. In contrast, the nature of the structures observed for longer repeats involved in the formation of SiDNA is still unclear. Even if several studies have previously demonstrated that long trinucleotide repeats adopt more compact structures than the short ones, suggesting the hypothesis of a multi-folded structure (16,29), but little is known about their exact conformation. The elucidation of this point is the purpose of our study; i.e. the analysis of the conformational properties of individual $(CTG)_n$ or $(CAG)_n$ strands, which may constitute the single slipped-out structures of particular DNA regions.

Initial thermodynamic studies showed the stabilities of CAG and CTG hairpins to be nearly identical under physiological salt concentrations *in vitro* (17). In contrast, Völker *et al*. (30) showed that, within a conformationally confined system, $(CAG)_6$ and $(CTG)_6$ form stable, ordered structures with the former triplet less stable than the latter, as also recently confirmed by another group (31). In this study, we analysed the folding of $(CTG)_n$ and $(CAG)_n$ individual sequences by using biophysical [UV-absorbance, circular dichroism (CD) and microcalorimetry] and biochemical techniques (native gel electrophoresis and enzymatic footprinting). In agreement with the previous results, we found that the folding of these sequences is intramolecular rather than bimolecular and that these structures are stable in physiological conditions. The study of the thermodynamic data obtained by thermal denaturation and microcalorimetry led us to propose an intramolecular 'bis-hairpin' like model for $(CTG)_n$ and $(CAG)_n$ individual strands ($n = 12–25$), with longer strands possibly giving rise to the formation of multibranched hairpins. These structures are actually distinct from the multiple hairpins found in SDNA and SiDNA, in which several short hairpins protrude from a DNA duplex at different positions. Our model suggests that each slipped-out single-stranded region found in SDNA and SiDNA can actually fold into structures more complex than previously thought, as soon as 10–12 repeats are present in a single protruding loop. Similar results were also found for $(CAG)_n$ repeats.

## MATERIALS AND METHODS

### Oligodeoxynucleotides

Oligodeoxynucleotide probes were synthesized by Eurogentec (Belgium) on the 0.2 or 1 μmol scale. As all oligomers studied here correspond to DNA, the 'd-' prefix was omitted from most sequences. Purity was checked by gel electrophoresis. All concentrations were expressed in strand molarity using a nearest-neighbour approximation for the absorption coefficients of the unfolded species (32).

### Thermal difference spectra for CTG and CAG trinucleotide repeats

The thermal difference spectrum (TDS) of a nucleic acid is obtained by simply recording the UV-absorbance spectra of the unfolded and the folded states at temperatures, respectively, above and below its melting temperature ($T_m$). The difference between these two spectra is defined as the TDS. The TDS has a specific shape that is unique for most structures (33,34) (J.L. Mergny *et al*., manuscript in preparation); thus, providing a simple, inexpensive and rapid method to gain structural insight into nucleic acid structures, both DNA and RNA, ranging from short oligomers to polynucleotides. Spectra were recorded between 220 and 335 nm with a Kontron Uvikon 940 UV/Vis spectrophotometer using quartz cuvettes with an optical pathlength of 0.2 or 1 cm. The differential spectrum of a CTG repeat structure gives a maximum differential absorbance at an unusually high wavelength (~277 nm) (34).

### UV-melting experiments

The thermal stability of the different trinucleotide repeat structures was estimated by heating/cooling experiments, recording the UV-absorbance at several wavelengths as a function of temperature using a Kontron Uvikon 940 spectrophotometer thermostated with an external ThermoNeslab RTE111 or ThermoHaake Phoenix C25P1 waterbath. The temperature of the bath was typically increased or decreased at a rate of 0.2°C/min, using 0.2 or 1 cm pathlength quartz cuvettes. All experiments were carried out in 10 mM sodium cacodylate buffer (pH 7.0) containing 30–500 mM KCl. Taking into consideration the TDSs and the high strand concentrations used for some UV-melting experiments, we chose to record the denaturation process at 290 nm (using cuvettes of 0.2 cm optical pathlength for the highest concentrations) in order to obtain an absorbance between 0.1 and 1.5. On the contrary, for the thermodynamic analysis (see below), lower strand concentrations were used and the denaturation process was followed at 275 nm where the signal is maximal. However, the observed melting temperatures were in excellent agreement (usually within 0.5°C, data not shown) with the ones determined at other wavelengths (e.g. 260 nm). All melting profiles were

perfectly reversible at the chosen temperature gradient, demonstrating that these curves correspond to true equilibrium curves (35).

### Thermodynamic analysis

For all parameters listed below, the assumed direction of the thermal process is the single-strand-to-hairpin transition. The thermal reversibility and the known molecularity of the process allow one to calculate the value of the equilibrium constant ($K_a$) assuming a simple two-state transition model. One must first convert absorbance measurements into folded fraction by manually selecting two baselines corresponding to the completely folded and unfolded form. An uncertainty may therefore arise because of the subjectivity in baseline determination. Starting from the classical Gibbs enthalpy equation,

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$$

one can write, for a reversible reaction [in which $\Delta G^\circ = -RT\ln(K_a)$], the following van't Hoff equation:

$$\ln(K_a) = -\frac{\Delta H^\circ_{VH}}{R}(1/T) + \frac{\Delta S^\circ_{VH}}{R},$$

where $T$ is the temperature in Kelvin, while $\Delta H^\circ_{VH}$ and $\Delta S^\circ_{VH}$ are, respectively, the standard enthalpy and the entropy change of the reaction. In other words, provided that $\Delta H^\circ_{VH}$ and $\Delta S^\circ_{VH}$ are temperature-independent (see below), the so-called van't Hoff representation or Arrhenius plot [$\ln(K_a)$ versus $1/T$] should give a straight line, with a slope of $-\Delta H^\circ_{VH}/R$ and a $y$-axis intercept of $\Delta S^\circ_{VH}/R$. The $\Delta H^\circ_{VH}$ of this reversible reaction is called the van't Hoff enthalpy and is defined by

$$\Delta H^\circ_{VH} = -\frac{R\mathrm{d}\ln(K_a)}{\mathrm{d}(T^{-1})}.$$

In the case of temperature-dependent enthalpies and entropies, one will obtain a significant deviation from linearity. Nevertheless, the above equation is still valid, but the slope at each point may be different, leading to temperature-dependent $\Delta H^\circ_{VH}$ and $\Delta S^\circ_{VH}$ values. van't Hoff enthalpies and entropies are said to be 'model-dependent': they rely on a two-state equilibrium hypothesis. They are, therefore, less robust than the 'model-independent' thermodynamic values provided by calorimetry (see below).

### Determination of $\Delta C^\circ_p$

The linear fit of an Arrhenius plot assumes that $\Delta H^\circ$ is temperature-independent, which in turn means that $\Delta C^\circ_p = 0$. $\Delta C^\circ_p$ is the heat capacity change at constant pressure which is occurring during the thermal process (considered in the single-strand-to-hairpin direction). $\Delta H^\circ_{VH}$ and $\Delta S^\circ_{VH}$ are linked to $\Delta C^\circ_p$ by the following relations:

$$\frac{\mathrm{d}(\Delta H^\circ_{VH})}{\mathrm{d}(T)} = \Delta C^\circ_p.$$

$$\frac{\mathrm{d}(\Delta S^\circ_{VH})}{\mathrm{d}[\ln(T)]} = \Delta C^\circ_p.$$

The $\Delta C^\circ_p = 0$ hypothesis (widely assumed for nucleic acids) is now challenged. Several experiments have demonstrated that the $\Delta H^\circ_{VH}$ of a duplex (36–40) or a triplex (41) may significantly depend on the temperature. In the case of UV-melting curves it is often difficult to provide an estimate of $\Delta C^\circ_p$ (35) because of baseline assumption problems. A small experimental error in $\ln(K_a)$ may obscure the temperature dependence of $\Delta H^\circ_{VH}$ (42). Nevertheless, the Arrhenius curves analysed here were so deviant from linearity that reliable non-zero $\Delta C^\circ_p$ values could be extracted by fitting the plot of $\ln(K_a)$ versus $1/T$ with Kaleidagraph 3.5, according to the following equation:

$$\ln(K_a) = \frac{-\Delta H^{T_m}_{VH} + \Delta C^\circ_p T_m}{RT} - \frac{\Delta C^\circ_p}{R}\ln\left(\frac{1}{T}\right) + \frac{[\Delta S^{T_m}_{VH} - \Delta C^\circ_p \ln(T_m)]}{R} - \frac{\Delta C^\circ_p}{R},$$

where $T_m$ is the melting temperature (in Kelvin) of the single-strand-to-hairpin renaturation process, while $\Delta H^{T_m}_{VH}$ and $\Delta S^{T_m}_{VH}$ are, respectively, the enthalpy and entropy change of the thermal transition at the melting temperature. In our case, they can be reasonably considered equivalent to the $\Delta H^\circ_{VH}$ and $\Delta S^\circ_{VH}$, respectively. The heat capacity change obtained by non-linear fitting of the Arrhenius plot will be referred to in the text as $\Delta C^\circ_{pVH}$.

Other methods may be used to measure the heat capacity change such as differential scanning calorimetry (DSC), which uses the difference between the pre- and the post-transition baselines (37), or even better, isothermal titration calorimetry (ITC). However, the DSC method is not always straightforward and requires an optimized experimental setting (40) whereas ITC is not simply applicable to intramolecular reactions.

### Differential scanning calorimetry

Microcalorimetry experiments were performed on the $(CTG)_n$ oligomers using a Nano DSC-II microcalorimeter (CSC) driven by a DSC-run software. The oligonucleotides were dissolved at concentrations ranging from 40 to 200 μM in 10 mM sodium cacodylate buffer at pH 7 containing 30 mM KCl. Buffer and oligo solutions were carefully degassed prior to their utilization and their thermal profiles were analysed in the 0–95°C temperature range at a scan rate of 1°C/min. An initial calibration of the instrument was performed by filling both cells with the buffer solution and accumulating several scans until the balance was reached. Then, the DSC profile of the oligo was obtained by loading the solution of the oligonucleotide into the sample cell during the last cooling scan of the calibration experiment, leaving the buffer solution into the reference cell (load-on-the-fly method). With this method it was possible to execute a preliminary setting up of the instrument and to perform a real experiment in the same set of scans, without compromising the shape and the calorimetric values of the sample curve. A minimum of six scans was collected for each experiment. Substraction of the baseline and calculation of the thermodynamic parameters were carried out using the Cp-Calc software (Applied Thermodynamics). We observed that for all sequences tested here, the $T_m$ values found by

DSC and UV-melting analysis were in good agreement. The systematic small (≈2°C) difference in favour of the DSC $T_m$ is the result of the calorimetric definition of the melting temperature, which corresponds to the temperature of maximum heat release/uptake rather than to the temperature of half association/dissociation. The real $T_m$ (half association/dissociation temperature) is therefore ∼2°C below the DSC $T_m$, referred to in the text as '$T_{max}^{cal}$'. The calorimetric enthalpy ($\Delta H_{cal}^{\circ}$) and entropy ($\Delta S_{cal}^{\circ}$) for the transition process were determined in a model-independent way from the DSC curve. Comparison of the $\Delta H_{cal}^{\circ}$ with the van't Hoff value obtained by the UV-thermal curve ($\Delta H_{VH}^{\circ}$) allowed us to confirm or infirm the correctness of the previously assumed two-state model used to describe the entire thermal process ($x = \Delta H_{VH}^{\circ}/\Delta H_{cal}^{\circ} = 1$). Additional fitting of the experimental DSC curve with a more general model equation (provided by the Cp-Calc software) and its resolution by deconvolution analysis allowed us to obtain the value of $\Delta C_{pCal}^{\circ}$ for the whole process, as well as the thermal profile of the 'daughter' subcurves corresponding to the intermediate transitions. Deconvolution of the DSC curves represents one of the major advantages of the calorimetric analysis over the spectroscopic one. In fact, while by UV analysis small differences in melting temperature between independent subunits may be masked by an even smaller conformational change in the overall structure (giving rise to an apparent two-state transition profile of the UV-thermal curve), deconvolution of the calorimetric curve into its components can allow for unravelling intermediate states.

### Nuclease susceptibility assays

Prior to structural probing reactions, 5′ end-labelled oligonucleotides were subjected to denaturation by heating the samples at 90°C for 3 min and then slowly cooled to room temperature. $(CTG)_n$ sequences (with $n = 15$, 16, 20 or 25) were analysed by enzymatic probing in the presence of increasing concentrations of S1 nuclease (0.1–4 U/μl) and MB nuclease (1–10 U/μl). For both nucleases (Promega), limited digestions of 5′ end-labelled oligonucleotides were performed at 30°C for 5 min in a buffer composed of 10 mM Tris–HCl (pH 7.2), 10 mM $MgCl_2$ and 30 mM KCl. All the reactions were stopped by adding an equal volume of 100% formamide solution and heating for 3 min at 90°C. The nuclease digestion products were subjected to electrophoretic runs (70 W, 1500 V, 90 min) on denaturing polyacrylamide gels (15%) containing 7 M urea/TBE 1× and revealed on a phosphorimager screen (Typhoon; Molecular Dynamics). The cleavage sites for the two nucleases were determined by comparison of the enzymatic digestion products with size markers corresponding to $(CAG)_n$ treated with formic acid, $(CTG)_n$ treated with DMS and seven non-treated size markers: $(CTG)_{12}$, $(CTG)_{10}$, $(CTG)_8$, $(CTG)_7$, $(CTG)_6$, $(CTG)_4$ and $(CTG)_2$. This determination was not straightforward because of the differences in the cleavage sites between chemical reactions (HCOOH or DMS) and enzymatic reactions (S1 and MB nucleases). On the one hand, the chemical reactions produce oligonucleotides with a 3′-phosphate end, whereas on the other hand, the enzymatic reactions produce sequences with a 3′-OH end. Therefore, an untreated size marker and its corresponding one obtained upon chemical cleavage of the parent oligo may not be equally accelerated by the electric field, although they have the same sequence [e.g. the $(CTG)_7$ untreated size marker and its corresponding one obtained upon treatment of $(CTG)_{20}$ with formic acid].

## RESULTS

### Confirmation of the folded form and intramolecular folding of CTG repeats

Several groups have previously demonstrated that repeats adopt an intramolecular duplex structure [for a recent example see (31)]. For this reason, we will very briefly describe the experiments that were carried out to confirm this model in our system. As shown in Supplementary Figures S1A and S1B, DNA composed of pure $(CTG)_n$ repeats ($n = 4$–10) exhibit fast, strand concentration-independent, mobility on polyacrylamide gels (29,43–45) in good agreement with an intramolecular structure. This increased mobility is almost completely lost under denaturing conditions (Supplementary Figures S1C and S1D). CD spectra of CTG oligomers are similar to the CD spectral signature of B-DNA with a negative peak ∼255 nm and a positive peak ∼285 nm (Supplementary Figure S2) (31). The concentration dependence of the melting temperature, $T_m$, was measured to clarify the type of structure formed (Figure 1). Plots of $T_m$ versus $C_0$ (where $C_0$ is the total strand concentration) for a $(CTG)_n$ series of oligomers, where $n = 8$, 15 or 25, produced near-zero slopes (Figure 1B), indicating that unimolecular denaturations as also reported previously (31,46).

### Thermal absorbance difference spectra

Another method, we recently proposed, to study the conformation of a nucleic acid is to record the 'thermal absorbance difference spectrum' between its high and low temperature UV-absorbance spectrum (33,34). The normalized TDS has a shape which is specific for each nucleic acid conformation studied so far, ranging from duplexes to quadruplexes: as shown in Figure 2A, the shapes of TDS for $(CTG)_n$ ($n = 8$–20) were all very close (red curves) suggesting that these oligomers adopt a similar folded conformation. Furthermore, these spectra with two maxima ∼275 and 235 nm are highly reminiscent of pure GC-rich B-DNA TDS [blue curve and (34)]. A similar analysis was performed for CAG repeats (Figure 2B); although the TDS were not superimposable (compare panels A and B), these spectra were also reminiscent of GC-rich B-DNA TDS, but completely different from the ones of other structures such as quadruplexes (33). It is interesting to note that the CTG spectra are very close to the 100% GC spectra (47) (Figure 2A, in blue; these duplexes correspond to sequences where only GC base pairs are formed), whereas CAG repeats resemble mixed duplexes (Figure 2B, in black; these duplexes involve 67% GC/33% TA base pairs).

### Analysis of the melting profiles

Careful analysis of the UV-melting profiles revealed several observations. First, the melting temperature of $(CTG)_n$ and $(CAG)_n$ was weakly dependent on oligonucleotide length (example provided in Supplementary Figure S3; see also Figure 1B). For the CTG oligomers tested here, the $T_m$s varied between 54 ($n = 6$) and 58.7°C ($n = 25$) (Table 1), while for $(CAG)_n$ the $T_m$s varied between 51.6 ($n = 6$) and 54°C ($n = 25$) (Supplementary Table S1).

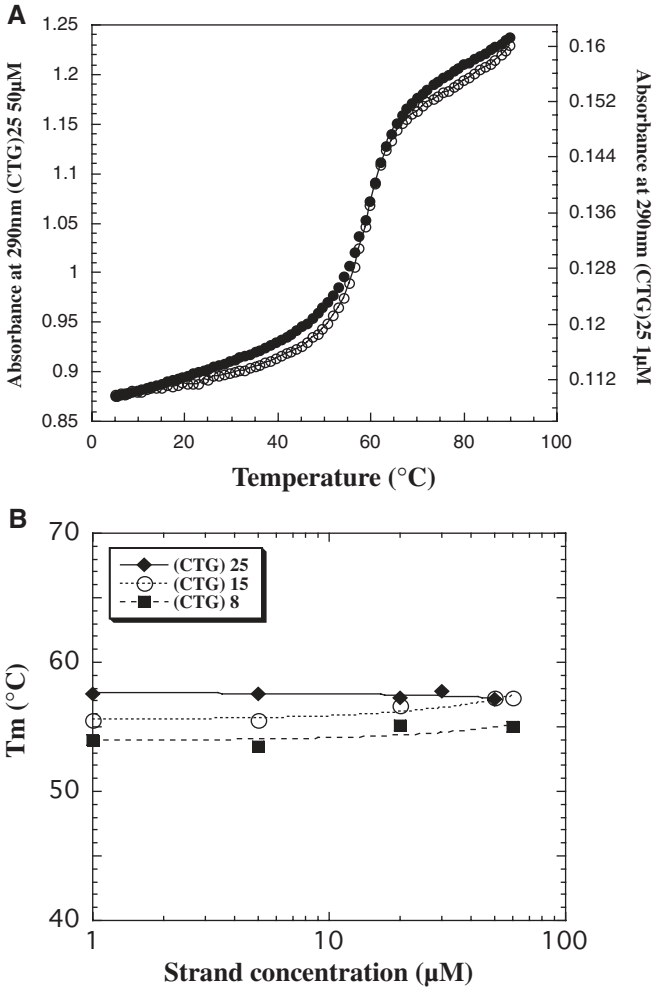


**Figure 1.** UV-melting curves. (**A**) Example of UV-absorbance denaturation profile (recorded at 290 nm) at two different concentrations (50 µM, left *y*-axis and 1 µM, right *y*-axis) of the $(CTG)_{25}$ sequence in a 10 mM sodium cacodylate (pH 7.0) buffer containing 30 mM KCl. (**B**) Concentration dependence of the $T_m$ for three different oligonucleotides: $(CTG)_8$, squares; $(CTG)_{15}$, circles; and $(CTG)_{25}$, diamonds.

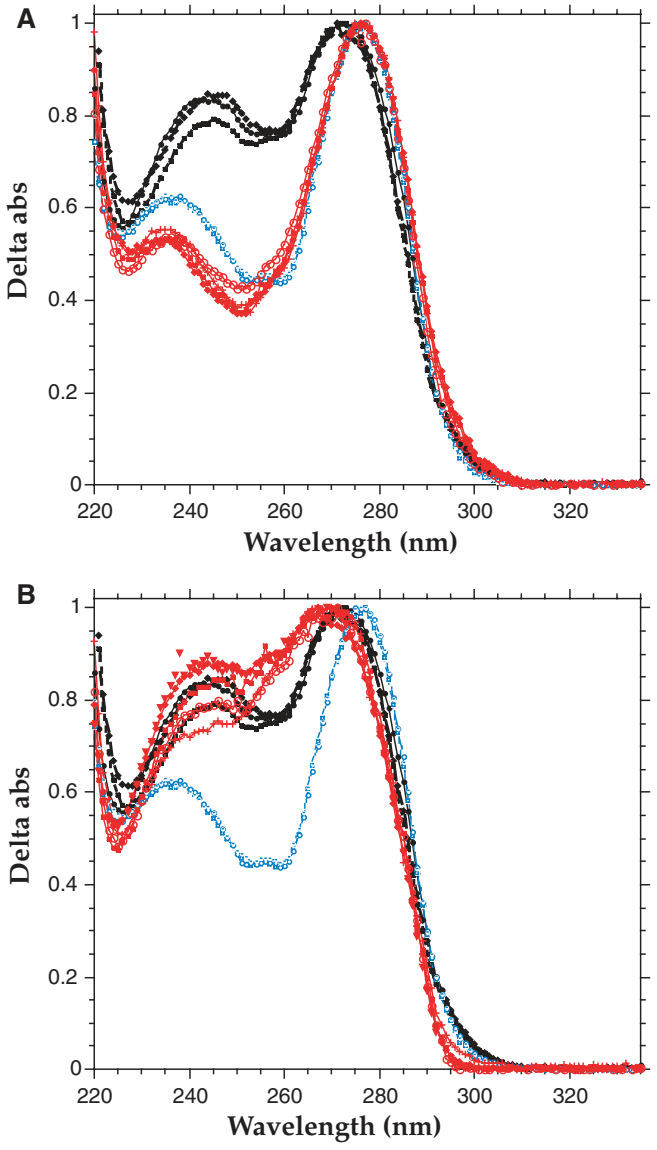


**Figure 2.** Thermal differential absorbance data. Normalized absorbance thermal difference spectra (TDS) of $(CTG)_{8–20}$ (**A**) and $(CAG)_{8–20}$ (**B**) (TDS were obtained by high temperature minus low temperature absorbance spectra in a 10 mM sodium cacodylate buffer at pH 7 containing 30 mM KCl). The trinucleotide spectra are shown in red. Three representative TDS of pure GC duplexes (in blue) or 67% GC duplexes (in black) are also presented for comparison with the trinucleotide sequences.

Second, the Arrhenius representation of the melting curves also revealed a more complex behaviour (35). Determination of the folded fraction for the two samples analysed in Supplementary Figure S3 led to the Arrhenius plots are shown in Figure 3. A significant and reproducible deviation from linearity is seen in both cases. As described previously (Materials and Methods), it is possible to fit these curves to obtain the $\Delta C^\circ_{pVH}$ values (Table 1). Assuming a reaction in the single-strand-to-hairpin direction, these values were found to be always negative, in agreement with the studies on different DNA structures but with surprisingly high absolute values for long sequences [e.g. $(CTG)_{20}$, $(CTG)_{25}$ and $(CAG)_{25}$]. Additionally, even if some discrepancies may be found, the $\Delta C^\circ_p$ values deduced from van't Hoff plots ($\Delta C^\circ_{pVH}$) were generally in good agreement with those deduced from calorimetry ($\Delta C^\circ_{pCal}$; Tables 1 and S1).

Third, the thermal stability of these trinucleotide repeats was analysed as a function of KCl concentration. As expected, the $T_m$ values for the $(CTG)_8$ oligomer were found to be indeed dependent on KCl concentration (Supplementary Figure S4): a 10-fold increase in potassium concentration led to a 7°C increase in $T_m$ (from 50°C at 10 mM KCl to 57°C at 100 mM KCl). This ionic strength dependence is somewhat lower than previously reported (16) and lower than expected for regular double-stranded hairpins.

## Calorimetric analysis of $(CTG)_n$

The calorimetric values of the enthalpy ($\Delta H^\circ_{cal}$) and entropy ($\Delta S^\circ_{cal}$) of renaturation for the CTG oligomers are reported in Table 1 and an example of a DSC denaturation run of $(CTG)_{25}$ is provided in Figure 4A. As in the UV-thermal analysis,

**Table 1.** Thermodynamic parameters for $(CTG)_n$ repeats

| $n$[a] | $T_m^{VH}$ (°C)[b] | $T_{max}^{cal}$ (°C)[c] | $\Delta H^\circ_{VH}$ (kcal/mol)[d] | $\Delta H^\circ_{cal}$ (kcal/mol)[e] | $x = \Delta H^\circ_{VH}/\Delta H^\circ_{cal}$ | $\Delta S^\circ_{VH}$ (cal/mol/K)[d] | $\Delta S^\circ_{cal}$ (cal/mol/K)[e] | $\Delta C^\circ_{pVH}$ (cal/mol/K)[f] | $\Delta C^\circ_{pCal}$ (cal/mol/K)[g] |
|---|---|---|---|---|---|---|---|---|---|
| 6 | 54 | 55.1 | −33.0 | −22.6 | 1.28 | −101 | −69 | −450 | nd |
| 8 | 56.8 | 58.1 | −44.0 | −40.5 | 1.09 | −133 | −122 | −1030 | −380 |
| 10 | 57 | 60.1 | −51.0 | −45.9 | 1.11 | −155 | −138 | −710 | −720 |
| 11 | 57.1 | 59.4 | −53.9 | −51.0 | 1.06 | −165 | −153 | −1360 | −1040 |
| 12 | 57.8 | 60.0 | −56.6 | −62.8 | 0.90 | −171 | −188 | −1410 | −1210 |
| 13 | 57.7 | 60.1 | −52.7 | −62.6 | 0.77 | −159 | −188 | −1700 | nd |
| 14 | 57.6 | 60.8 | −52.6 | −78.7 | 0.67 | −159 | −235 | −1510 | nd |
| 15 | 57.7 | 59.6 | −51.9 | −82.1 | 0.63 | −156 | −293 | −1500 | nd |
| 16 | 57.7 | 61.4 | −51.5 | −95.2 | 0.54 | −156 | −296 | −1500 | nd |
| 20 | 57 | 59.3 | −56.5 | −125.7 | 0.45 | −171 | −385 | −2060 | −1260 |
| 25 | 58.7 | 61.3 | −61.7 | −147.2 | 0.42 | −186 | −440 | −2120 | −1880 |

For all parameters listed, the assumed direction is the single-strand-to-hairpin transition. nd, not determined.
[a]Number of (CTG) repeats.
[b]Melting temperature deduced from the UV-melting curve.
[c]'Melting' temperature deduced from the DSC profile.
[d]Thermodynamic parameters deduced from the UV-melting curves, using a non-linear fit of the Arrhenius plots, where the $\Delta H^\circ_{VH}$ is close to the $\Delta H^\circ$ determined at the $T_m \Delta H^{T_m}_{VH}$. (Error bars for $\Delta H^\circ$ values are shown in Figure 4C; highest relative error of 6.4%.)
[e]Thermodynamic parameters deduced from the DSC profiles. (Error bars for $\Delta H^\circ$ values are shown in Figure 4C; highest relative error of 5.1%.)
[f]Heat capacity change deduced from non-linear fitting of the Arrhenius plots.
[g]Heat capacity change deduced from general model fitting of the DSC profiles.

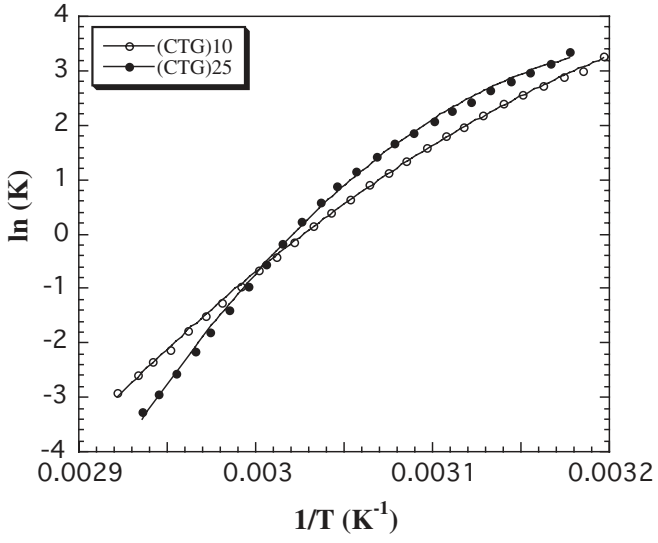


**Figure 3.** Arrhenius plots [$(\ln(K_a)$ versus $1/T)$] for $(CTG)_n$ sequences. These plots were deduced from the UV-absorbance denaturation profiles recorded at 275 nm of two different oligonucleotides: 5 μM $(CTG)_{10}$ (open circles) and 5 μM $(CTG)_{25}$ (closed circles).

the $T_{max}^{cal}$ determined by DSC was almost sequence-length independent (Figure 4B). Next, the van't Hoff enthalpy determined by the analysis of the UV-thermal transition curve ($\Delta H^\circ_{VH}$) was compared to the calorimetric value obtained by DSC ($\Delta H^\circ_{cal}$) (Figure 4C and D). For relatively short sequences ($n = 4$–12) the ratio $x = \Delta H^\circ_{VH}/\Delta H^\circ_{cal}$ was close to 1 (1.08 ± 0.13; Table 1). The similarity of the two values validated the hypothesis of the two-state model used in the van't Hoff analysis to describe the thermal transition of the intramolecular trinucleotide repeat. However, in the case of longer sequences, the $\Delta H^\circ_{VH}/\Delta H^\circ_{cal}$ ratio dramatically dropped to values ~0.4 (Figure 4D). The $\Delta H^\circ_{VH}$ failed to become more negative with increased repeat number, whereas $\Delta H^\circ_{cal}$ continued to increase (in absolute terms) explaining why the agreement was lost (Figure 4C and Table 1). The difference was highly significant as the errors bars (Figure 4C) on $\Delta H^\circ_{VH}$ and $\Delta H^\circ_{cal}$ were 6% or lower.

### Calorimetric analysis of $(CAG)_n$

The thermal behaviour observed for the CTG repeats was compared with that obtained from CAG repeats of identical length under the same experimental conditions. An example of a DSC denaturation run for a $(CAG)_n$ oligo is provided in Figure 5A. Again, we observed that, for all sequences tested, the $T_{max}^{cal}$ and the $T_m$ values found by DSC and UV-melting analysis were in good agreement and almost independent on sequence-length (Figure 5B). As shown in Figure 5C and D, the values of $\Delta H^\circ_{VH}$ and $\Delta H^\circ_{cal}$ for the $(CAG)_n$ thermal denaturations were compared and plotted against the number of triplet repeats, leading to the same conclusions drawn for the $(CTG)_n$ analogs (see also Supplementary Table S1).

### Enzymatic probing of long CTG repeats

Nuclease sensitivity studies were performed with S1 and MB nuclease that are two single-strand specific endonucleases. The bases contained in the single-stranded regions of the structures adopted by long CTG repeats are preferentially cleaved with respect to the ones contained in the double-stranded regions. The cleavage sites for the two nucleases were determined by comparison with size markers (Materials and Methods). The enhanced MB digestion of $(CTG)_{15}$ occurred at three main regions (Figure 6A): (i) two sites at the 3′ end, G13 and G14; (ii) five sequence-centred sites around the $(CTG)_7$ (T4) marker, i.e. C7, T7, G7, C8 and T8 resulting from a TGCT-loop; and (iii) two sites around the 5′ end region identified as G1 and C2 representing a CTG-loop. S1 digestion of $(CTG)_{15}$ presents only two sites around the same 5′ end region cleaved by the MB nuclease: G1 and C2 (Figure 6B). Similarly, two loops were characterized for $(CAG)_{15}$: one central loop composed of two repeats

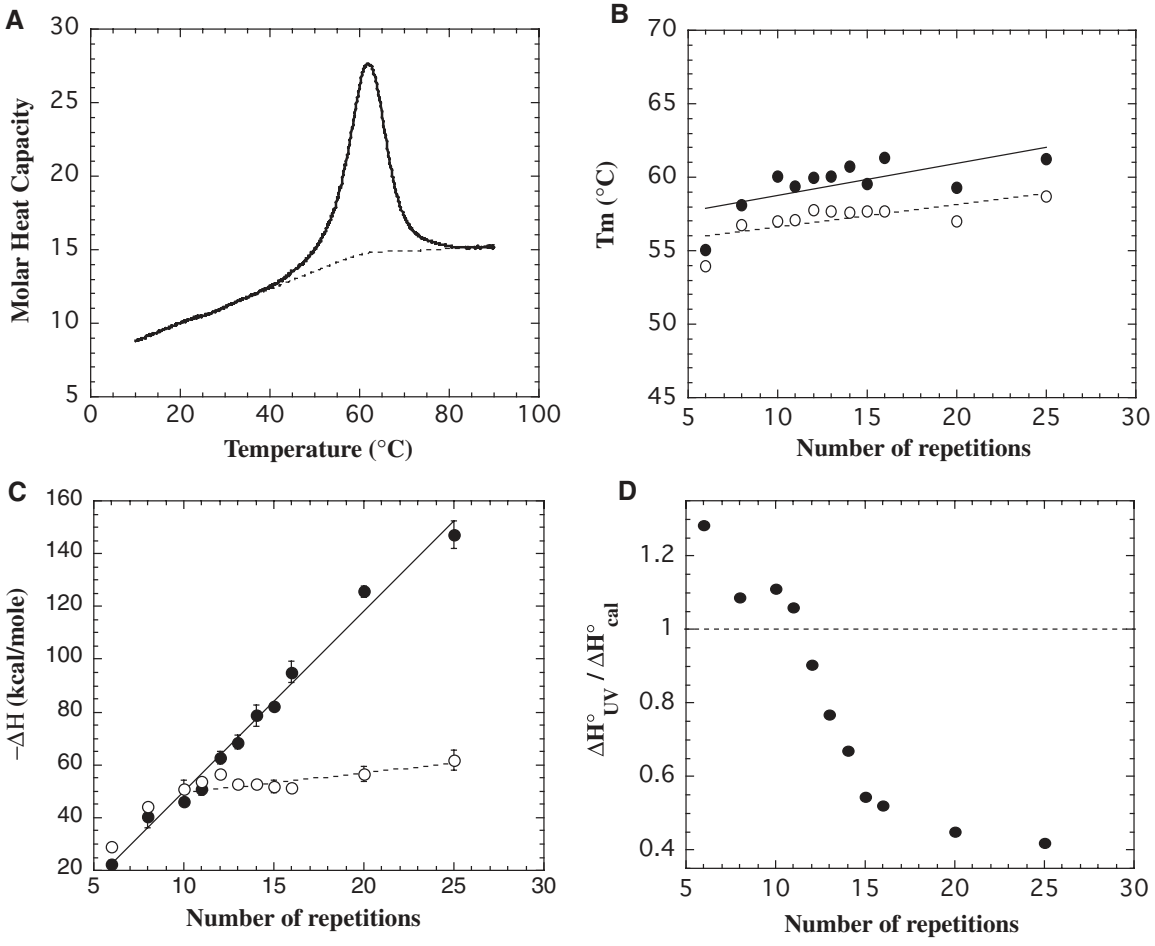


**Figure 4.** Enthalpy determination for $(CTG)_n$. (**A**) Typical differential scanning calorimetric profile for $(CTG)_{25}$ denaturation. The solid line is the thermogram and the dashed line is the determined baseline. Integration of the area between the thermogram and the baseline yields to the model-independent calorimetric $\Delta H^\circ_{cal}$ of the transition (in this case, in the hairpin-to-single-strand direction; hence, a positive $\Delta H^\circ_{cal}$; the opposite reaction—formation of the hairpin—is associated with a negative $\Delta H^\circ_{cal}$). Thermograms were obtained in a 10 mM sodium cacodylate buffer at pH 7 containing 30 mM KCl. (**B**) Comparison between the $T^{cal}_{max}$ and the $T_m$ found by DSC (closed circles) and by UV-melting profiles (open circles), respectively, as a function of sequence-length (number of CTG repeats). Linear fits (which are poor, $R = 0.65$, solid line and 0.69, dashed line for DSC and UV data, respectively) are shown. (**C**) Comparison between the calorimetric model-independent $\Delta H^\circ_{cal}$ (closed circles) and the $\Delta H^\circ_{VH}$ deduced from a van't Hoff analysis of the UV-melting profiles (open circles) as a function of sequence-length (number of CTG repeats). The renaturation process is considered in the single-strand-to-hairpin direction; hence, negative values of $\Delta H^\circ$. Errors bars on $\Delta H^\circ_{VH}$ and $\Delta H^\circ_{cal}$ were 6% or lower. (**D**) Ratio between the $\Delta H^\circ_{VH}$ and the $\Delta H^\circ_{cal}$ as a function of sequence-length (number of repeats).

(CAGCAG) and one CAG loop at the 5′ end (Supplementary Figure S5).

The enhanced digestion of $(CTG)_{16}$ occurred at three main regions (Figure 7A): (i) two sites at the 3′ end: G14 and G15; (ii) two sequence-centred sites around the $(CTG)_8$ (T3) marker: T8 and G8 resulting from a CTG-loop; and (iii) two sites around the 5′ end region identified as G1 and C2, indicating a CTG-loop. These two CTG-loops are conserved in $(CTG)_{20}$ (Figure 6A and B) with the following digestion sites: (i) two sequence-centred sites around the $(CTG)_{10}$ (T2) marker: T10 and G10 and (ii) two sites around the 5′ end region identified as G1 and C2.

Finally, $(CTG)_{25}$ presents the same kind of loops as $(CTG)_{15}$ (Supplementary Figure S6): (i) four sequence-centred sites around the $(CTG)_{12}$ (T1) marker: T12, G12, C13 and T13 resulting from a 4 nt loop; and (ii) two sites around the 5′ end region identified as G1 and C2, representing a CTG-loop. Thus, our data show that all these sequences present one centred loop of 3 or 4 nt and one loop at the 5′ end containing 3 nt (Figures 6C and 7B). In theory, this terminal loop could also correspond to DNA end fraying, but in this case we would have obtained four cleavage sites: C1, T1, G1 and C2 at the 5′ end region instead of the only two sites G1 and C2. Furthermore, two NMR studies on short CTG repeats sequences have demonstrated that the T–T mismatches, all over the stem of the hairpin, are very well stacked between the two adjacent CpG base pairs and are bound to each other by two hydrogen bonds (35). The DNA end fraying hypothesis is therefore unlikely, as will be confirmed in the following part.

### $(CTG)_{15}$ variant sequences

In order to confirm these results and to ensure that the proximal cuts of the 5′ end nucleotides do not represent DNA end fraying, the C and G bases of the single-strand regions of $(CTG)_{15}$ previously defined by footprinting experiments

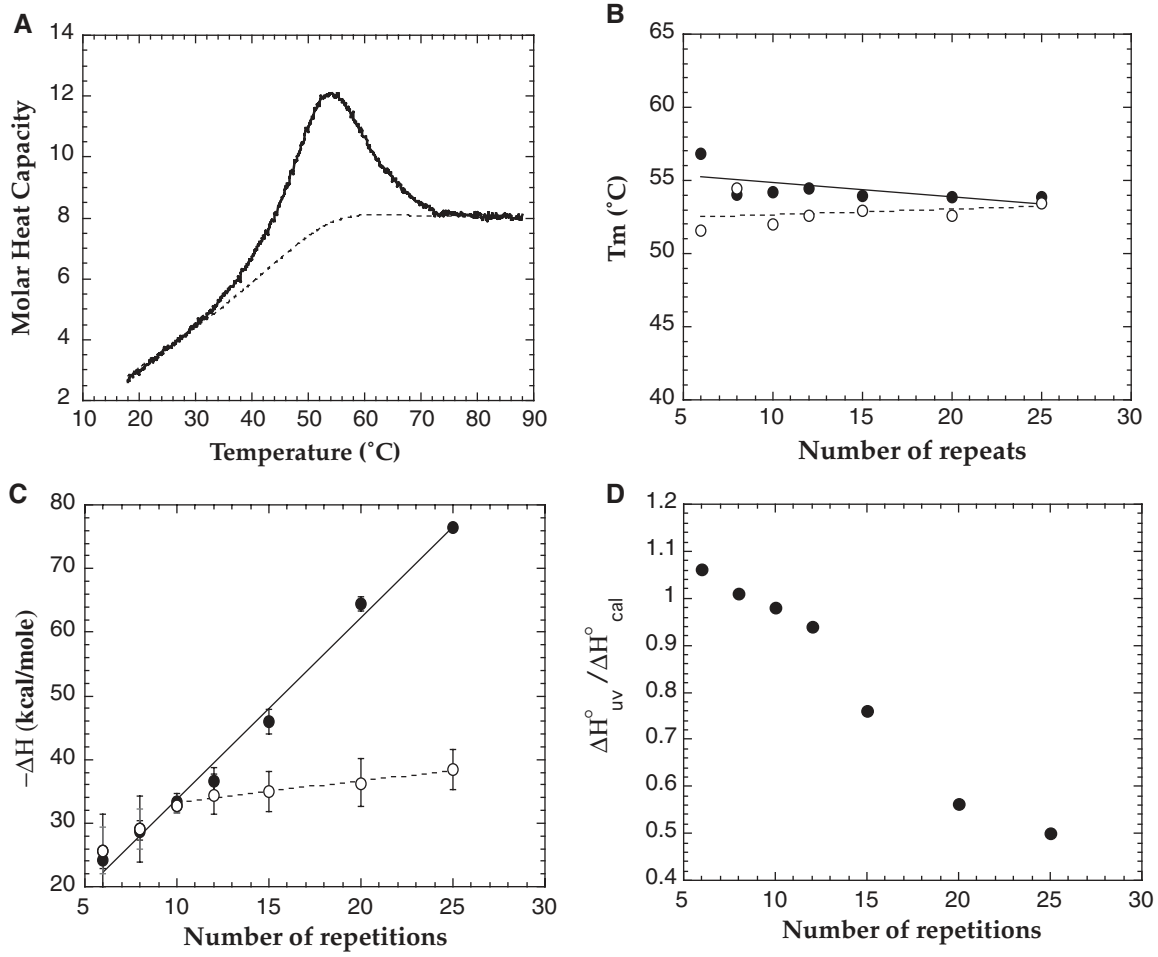


**Figure 5.** Enthalpy determination for $(CAG)_n$. (**A**) Typical differential scanning calorimetric profile for $(CAG)_{25}$ denaturation. Identical legend as in Figure 4. Thermograms were obtained in a 10 mM sodium cacodylate buffer at pH 7containing 30 mM KCl. (**B**) Comparison between the $T^{cal}_{max}$ and the $T_m$ found by DSC (closed circles) and by UV-melting profiles (open circles), respectively, as a function of sequence-length (number of CAG repeats). (**C**) Comparison between the calorimetric model-independent $\Delta H^\circ_{cal}$ (closed circles) and the $\Delta H^\circ_{VH}$ deduced from a van't Hoff analysis of the UV-melting profiles (open circles) as a function of sequence-length (number of CAG repeats). The renaturation process is considered in the single-strand-to-hairpin direction, hence negative values of $\Delta H^\circ$. (**D**) Ratio between the $\Delta H^\circ_{VH}$ and the $\Delta H^\circ_{cal}$ as a function of sequence-length (number of repeats).

were replaced by T bases. The $\Delta T^{cal}_{max}$ and renaturation enthalpy of the modified oligonucleotide sequences were then measured by DSC and compared to the ones of the unmodified analog (Table 2). Replacement of C2 and G2 (as well as of G7 and C8) by two T bases did not influence the renaturation enthalpy. In fact, despite a little decrease in the $\Delta T^{cal}_{max}$ (−2.5°C) of the modified sequences, their enthalpy remained around −80 kcal/mol as for the unmodified $(CTG)_{15}$. This suggests that these four bases, i.e. C2, G2, G7 and C8, do not belong to the stem region of the hairpin but to the unpaired regions of the structure. The small difference in $\Delta T^{cal}_{max}$ could be the result of little variations in the ionic conditions between the modified and the unmodified samples or errors in data acquisition/baseline determination. More importantly, this $\Delta T^{cal}_{max}$ is defined as the maximum of the DSC curve (Materials and Methods) and does not necessarily correspond to the real $T_m$ (half association/dissociation temperature): it is therefore not the best parameter to assess the stability of a structure. At this purpose, the model-independent $\Delta H^\circ_{cal}$ of the renaturation process is a more suitable parameter, directly related to the number of base pairs involved in the molecule. Thus, the constancy of this value for T-replacement of C2, G2, G7 and C8, undoubtedly indicates that these four bases do not belong to the stem region of the hairpin but to unpaired regions of the structure. In contrast, we obtained a dramatic decrease, in absolute value, of the renaturation enthalpy (from −80 till −54.8, −51.8 and −51.4 kcal/mol) by replacing C3, C7 and G8 by a T base, thus showing that these three bases belong instead to the stem region of the structure.

## DISCUSSION

The goal of this study was to analyse the assumption of a quasi 'normal' hairpin duplex for trinucleotide repeats, namely the CTG and CAG repeat motifs. Little is known about the exact conformation of long trinucleotide repeats. Several studies have previously demonstrated that long trinucleotide repeats adopt more compact structures than the short ones, suggesting the hypothesis of a multi-folded structure rather

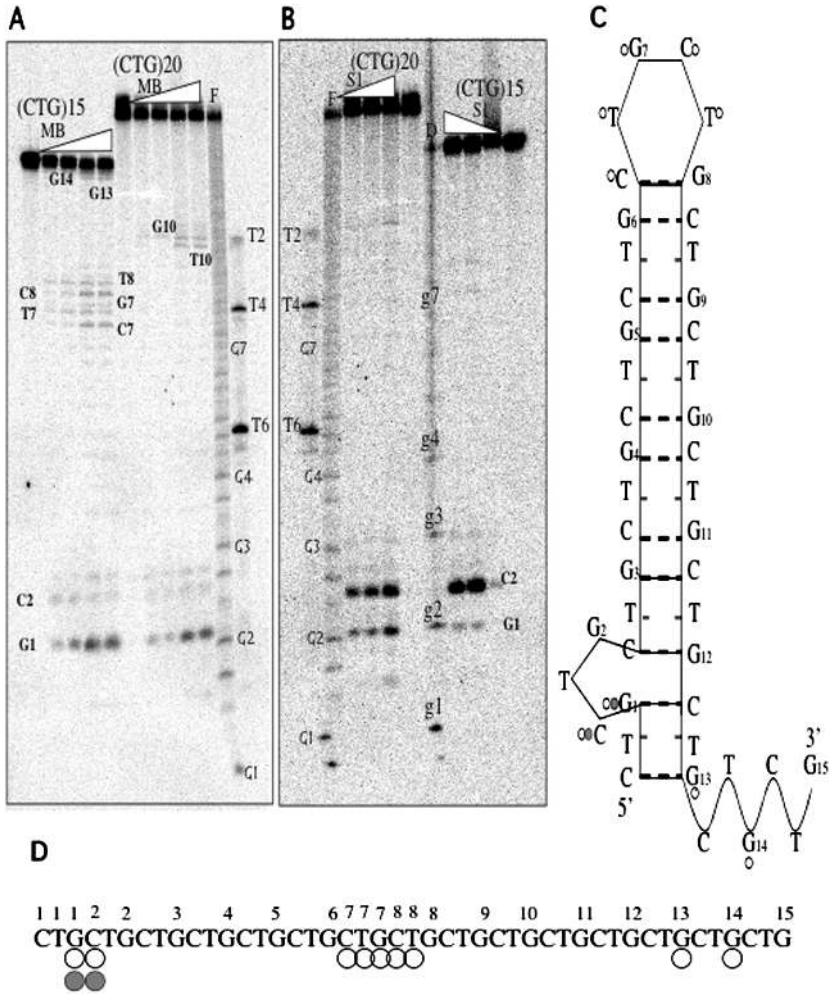


**Figure 6.** Enzymatic probing of $(CTG)_{15}$ and $(CTG)_{20}$. Two sequences, $(CTG)_{15}$ and $(CTG)_{20}$, were 5′ radiolabelled, incubated at 30°C for 5 min in a buffer composed of 10 mM Tris–HCl, pH 7.2, 10 mM $MgCl_2$ and 30 mM KCl with (**A**) increasing concentrations of MB nuclease 1–3.5–7–10 U/µl. (**B**) Increasing concentrations of S1 nuclease (NS1) 0.2–0.4–0.8 U/µl then loaded on a denaturing 15% polyacrylamide gel. Length markers were also loaded in order to identify the residues cleaved by the nucleases. Lane F is a guanine- and adenine-specific ladder corresponding to $(CAG)_{20}$ treated with formic acid (where g*i* indicates the G residue from the *i*-th CTG repeat); lane D is a guanine-specific ladder corresponding to $(CTG)_{15}$ treated with DMS (where g*i* indicates the G residue from the *i*-th CTG repeat) and T1, T2, T3, T4, T5, T6 and T7 correspond to non-treated size markers, respectively, $(CTG)_{12}$ $(CTG)_{10}$, $(CTG)_8$, $(CTG)_7$, $(CTG)_6$, $(CTG)_4$ and $(CTG)_2$. (**C**) The schematic diagram shows the $(CTG)_{15}$ folding pattern which agrees with the nuclease susceptibility assays. Rather than a simple hairpin–loop model, thermodynamic analysis as well as enzymatic studies suggest a more complicated 'bis-hairpin' model. (**D**) Summary of the cleavage pattern of $(CTG)_{15}$. Nucleotides cleaved by MB nuclease (open circles) or S1 nuclease (closed circles) are shown.

than a single-hairpin (16,29). Few NMR or crystallographic studies have actually analysed 'pure' CNG repeats, as they are usually embedded into different sequence motifs [e.g. GCGGTTTGCGG in (48) and TGGCGGC in (49); for a review see (50)]. Disease-related CNG repeats exhibit a propensity for folding at chain lengths as short as 12 residues (14). CTG-containing sequences have been studied with a variety of other techniques, including PAGE, $KMnO_4$ modification, P1 nuclease digestion, UV-absorbance and molecular dynamic simulations [reviewed in (51)]. To date, the hairpin structure of sequences shorter than 10 repeats has been clearly established (15,28), but the nature of the compact structures observed for longer repeats is still unclear.

In this paper, CD, electrophoresis, UV and DSC data revealed sharp structural transitions, in agreement with the formation of a rather simple canonical B-DNA hairpin, with a stem length growing with the repeat number. However, independent results indicate that the energetics and/or structure of these intramolecular trinucleotide repeats were significantly altered when compared with the canonical B-DNA, especially for (relatively) long $(CTG)_n$ sequences. Qualitatively similar results were obtained for $(CAG)_n$ repeats, suggesting that this behaviour may be a general phenomenon for $(CNG)_n$ oligonucleotides.

First, the thermal stability of this structure (its $T_m$) was very weakly dependent on the oligonucleotide length (Figures 4B and 5B and Supplementary Figure S3). The weak sequence-length dependence of the $T_m$ of these repeats was previously observed for $(CTG)_{10}$ and $(CTG)_{30}$ (16); $(CTG)_{10}$ and $(CTG)_{25}$ (17); $(CTG)_{10}$ and $(CTG)_{15}$ (31) as well, but has never been systematically studied [with a notable exception for $(CUG)_n$ RNAs; $n = 5$–69 (52)]. In our experience, the thermal stability of DNA hairpins increases rapidly with the stem length, in complete opposition to what is observed here. This behaviour

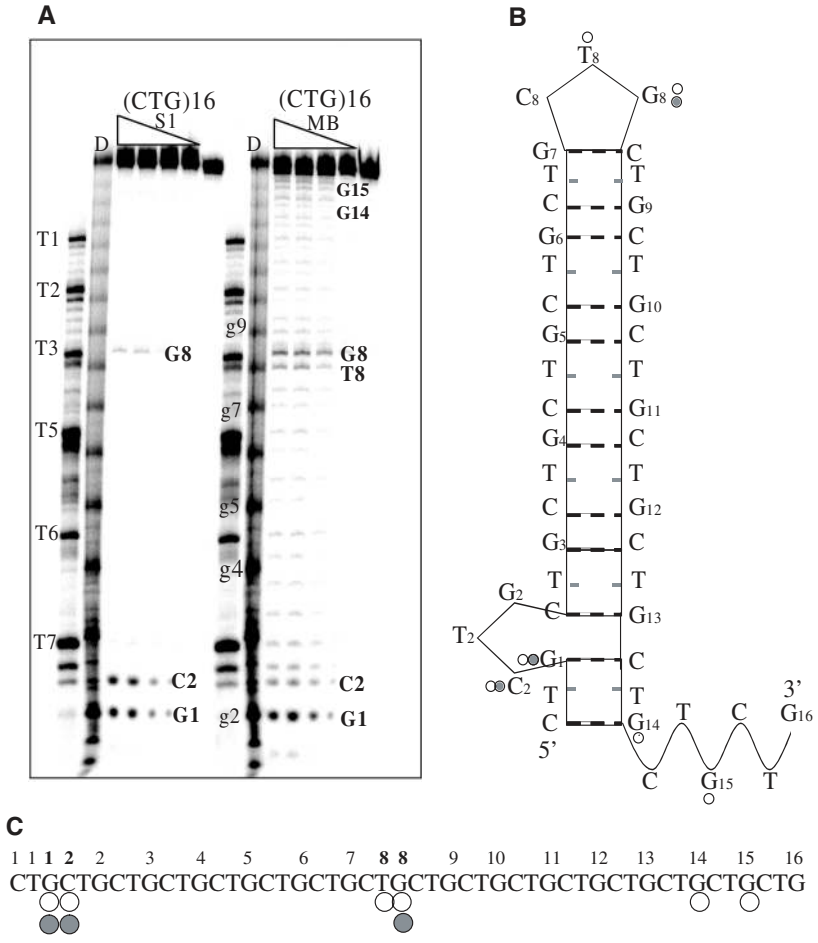


**Figure 7.** Enzymatic probing of $(CTG)_{16}$ **(A)** $(CTG)_{16}$ was 5′ radiolabelled, incubated at 30°C for 5 min in a buffer composed of 10 mM Tris–HCl, pH 7.2, 10 mM $MgCl_2$ and 30 mM KCl with increasing concentrations of MB nuclease (1–3.5–7–10 U/µl) and increasing concentrations of S1 nuclease (0.2–0.4–0.8 U/µl). Samples were loaded on a denaturing 15% polyacrylamide gel and analysed as in Figure 6. **(B)** The schematic diagram shows the $(CTG)_{16}$ folding pattern which agrees with the nuclease susceptibility assays. **(C)** Summary of the cleavage pattern of $(CTG)_{16}$. Nucleotides cleaved by MB nuclease (open circles) or S1 nuclease (closed circles) are shown.

**Table 2.** Thermodynamic parameters of $(CTG)_{15}$ variant sequences

| Sequence | $T_{cal}^{max}$ (°C)[a] | $\Delta H^{\circ}_{cal}$(kcal/mol)[a] |
|---|---|---|
| $(CTG)_{15}$ | 60.9 | −81 |
| $(CTG)\mathbf{T}T\mathbf{T}CTG(CTG)_{12}$ | 58.4 | −78 |
| $(CTG)CT\mathbf{TT}TG(CTG)_{12}$ | 57.5 | −54.8 |
| $(CTG)CTG\mathbf{T}TG(CTG)_{12}$ | 58.7 | −51.8 |
| $(CTG)_6CT\mathbf{TT}TG(CTG)_7$ | 58.4 | −82.2 |
| $(CTG)_6\mathbf{T}TGCT\mathbf{T}(CTG)_7$ | 51.1 | −51.4 |
| $(CTG)_6\mathbf{T}TGCTG\ (CTG)_7$ | 56 | −56.1 |

[a]Values deduced from the DSC profiles.

is therefore highly irregular for a simple hairpin system. Second, the analysis of the melting curves did not lead to simple Arrhenius plots, as significant and reproducible curvatures were observed. We actually had to manually select an unlikely lower baseline in order to 'linearize' the Arrhenius representations (Supplementary Figure S7). At least three non-exclusive reasons may be proposed to explain this non-linear behaviour: (i) a non-two-state transition (partially melted structures are significantly populated), (ii) two (or more) intermingled transitions occur or (iii) a single two-state transition occurs with a highly negative $\Delta C^{\circ}_{pVH}$ (i.e. a temperature-dependent $\Delta H^{\circ}$). This last possibility which was, until recently (53,54), overlooked for nucleic acids transitions, should be seriously considered here. The comparison of the model-dependent and model-independent $\Delta H^{\circ}$ values provides interesting clues.

## Length-dependent discrepancy between van't Hoff and calorimetry enthalpy

Several groups have reported important discrepancies between the van't Hoff enthalpy ($\Delta H^{\circ}_{VH}$) and the calorimetry enthalpy ($\Delta H^{\circ}_{cal}$) for various nucleic acids structures (55), although not for trinucleotides. For $(CTG)_n$ and $(CAG)_n$ repeats, we observed that the melting temperatures found by DSC and van't Hoff analysis of the UV-melting curves were always in reasonable agreement; however, the model-dependent and the model-independent enthalpies were not always identical (Figures 4C and 5C and Tables 1 and S1). For relatively short $(CTG)_n$ or $(CAG)_n$ sequences ($n = 4$–12) the $\Delta H^{\circ}_{VH}/\Delta H^{\circ}_{cal}$ ratio was close to 1, validating the hypothesis of the two-state model used in the van't Hoff analysis. In the

case of longer sequences, the $\Delta H^{\circ}_{VH}/\Delta H^{\circ}_{cal}$ ratio dramatically dropped (Figures 4D and 5D) to values ~0.4 (for CTG repeats) or 0.5 (for CAG repeats). This length-dependent difference between the two enthalpy values suggests the hypothesis of a non-two-state transition model for longer sequences. That is, above a critical threshold number of trinucleotide repetitions ($n$ = 10–12), a change in the structural organization of the oligonucleotides occurs, going from a simple hairpin–stem model to a structure of higher complexity, composed of different independent units. Therefore,

(i) Below 12 repeats, no significant difference between the two $\Delta H^{\circ}$ values was observed, leading to their similar increase with sequence-length. For these relatively short sequences the two-state transition model was confirmed by deconvolution of the DSC profiles (Figure 8A) into a single transition, indicating their folding into a simple hairpin structure.

(ii) Beyond 12 repeats, instead, an increasing discrepancy between the $\Delta H^{\circ}$ values appeared, with the model-independent $\Delta H^{\circ}_{cal}$ still increasing in a length-dependent manner, whereas the two-state model-dependent $\Delta H^{\circ}_{VH}$ remained quasi-constant. In contrast with a previous report (31), no plateau was observed in $\Delta H^{\circ}_{cal}$ for sequences longer than 15 trinucleotide repeats (Figures 4C and 5C). For long oligomers, the two-state hypothesis was no longer valid, suggesting the formation of a more complex structure with at least two independent units. This observation was confirmed by deconvolution of the DSC profiles into a three-state equilibrium with two intermediate transitions (Figure 8B).

In other words, for short sequences, folding into a single-hairpin structure is relatively simple, leading to a good concordance between $\Delta H^{\circ}_{cal}$ and $\Delta H^{\circ}_{VH}$. For long sequences (at least 12), rather than adopting a regular long stem, the oligonucleotide seems to maintain a relatively short and constant length of the stem region while the remaining part fold into several (at least two) independent folding units, of not necessarily the same length, which melt in an independent fashion. In that case, a multi-branched hairpin should have a relatively apparent constant van't Hoff enthalpy, as seen by UV-melting analysis, but an increased calorimetric enthalpy, as observed by DSC. This different folding trend for longer sequences may be explained in terms of a favourable enthalpy–entropy compensation. In fact, although a long single-hairpin structure should indeed have a more favourable renaturation enthalpy (owing to maximization of the number of base pairs), it has not necessarily the most favourable free energy. On the contrary, branched hairpin structures may offer additional degrees of freedom, possibly lowering the unfavourable entropic term and thus explaining their preferential formation for longer sequences.

## Nuclease assays

The enzymatic assays on long sequences (Figures 6 and 7 and Supplementary Figures S5 and S6), additionally confirmed by the DSC analysis of the $(CTG)_{15}$-modified sequences (Table 2), suggested that a simple hairpin was not necessarily the chosen structure, as other bases close to the 5′ end of the sequence were cleaved by single-strand specific nucleases.

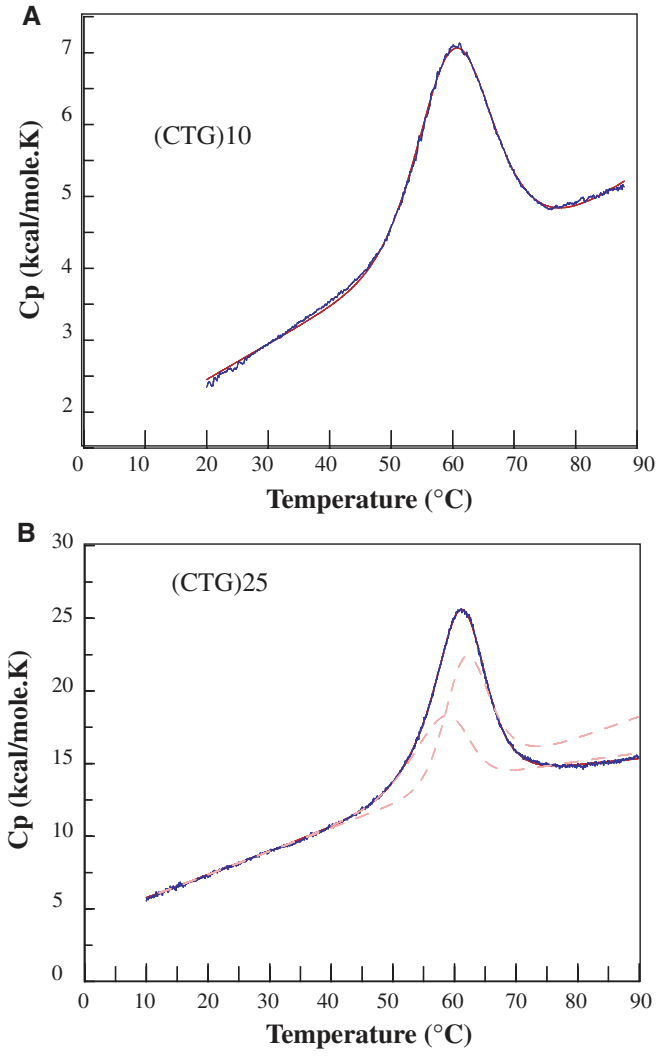


**Figure 8.** Analysis of complex differential scanning calorimetry (DSC) data. The deconvolution of the $(CTG)_{25}$ and $(CTG)_{10}$ DSC profiles were carried out using the deconvolution program of the Cp-Calc software (Applied Thermodynamics). In the two cases, the reconstructed curves (dashed lines) are in good agreement with the experimental ones. (**A**) $(CTG)_{10}$ is deconvoluted with only one transition indicating a single structural domain. Deconvolution data: $\Delta H(\text{total})_{\text{deconv}}$ = 50.8 kcal/mol; $T_m$ = 59.9°C; and experimental data: $\Delta H(\text{total})_{\text{exp}}$ = 47.2 kcal/mol. (**B**) $(CTG)_{25}$ is deconvoluted with two transitions corresponding to the melting of two independent domains of the structure. Deconvolution data: transition 1, $\Delta H$ = 56.4 kcal/mol, $T_m$ = 57.9°C; and transition 2, $\Delta H$ = 79.9 kcal/mol, $T_m$ = 61.7°C. $\Delta H(\text{total})_{\text{deconv}}$ = 136.3 kcal/mol. Experimental data: $\Delta H(\text{total})_{\text{exp}}$ = 142.8 kcal/mol.

However this is in contrast with the previous results obtained using other enzymatic and chemical probes (43,56). According to the biophysical and biochemical data obtained in this study, we propose a possible 'bis-hairpin' folding pattern for $(CTG)_{15}$ and $(CTG)_{16}$, as shown in Figures 6C and 7B.

The different number of cut sites between the $(CTG)_n$ sequences ($n$ = 15, 16, 20 or 25) probably reflect the different number of bases in the loops for the odd and even number of repeats (57). At least two NMR studies (15,28) have demonstrated that $(CTG)_n$ sequences ($n$ = 1–10), with an even number of repeats, fold into a hairpin structure containing a 4 nt TGCT-loop whereas an odd number of repeats leads to a

3 nt CTG-loop hairpin. It is interesting to note that, in our study, this tendency is reversed for longer sequences: $(CTG)_{15}$ and $(CTG)_{25}$ present a centred 4 nt TGCT-loop, while $(CTG)_{16}$ and $(CTG)_{20}$ present a centred 3 nt CTG-loop. Thus, this inversion confirms the presence of a second loop composed of an odd number of CTG repeats at the 5′ end of the sequence.

## CONCLUSIONS

Above a critical threshold number of trinucleotide repetitions ($n = 12$–15), a change in the structural organization of the oligonucleotides occurs going from a simple hairpin–stem model to a structure of higher complexity composed of different independent units such as a bis- (or multi-branched) hairpin-like structure. In contrast with a recent report (31), our calorimetric data suggest that the enthalpic stability of this structure is not compromised as the length of the hairpin overcomes 15 repeats. The possible intramolecular bis-hairpin model could be generalized as an intramolecular 'multi-branched' hairpin model for pathological sequences as long as 3000 repetitions. The 3 or 4 nt bulges disseminated all over the repeated sequence could be involved in the instability process during DNA replication through the formation of slipped-out DNA structures, but may also represent an important feature to discriminate trinucleotide repeats against canonical B-DNA for targeting by small repeat-specific bulge-ligands. This instability has been ascribed to the energy of the structure but in our case it could be explained in terms of number of hairpin loops that introduce an additional steric parameter to the energetic one. These results have encouraged us to carry on further investigations for a better understanding of the thermodynamics of these repetitions, and stimulating additional studies directed to the structural elucidation of long trinucleotide repeat models. We will now pursue similar studies on other DNA and RNA trinucleotides [$(CGG)_n$, $(CCG)_n$ and $(CUG)_n$ (58,59)] to see whether these observations may be extended to other CNG repeats.

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

## ACKNOWLEDGEMENTS

We thank T. Garestier, J. Pylouster, A. De Cian, L. Guittat and L. Lacroix (MNHN, Paris, France) for helpful discussions. S.A. is the recipient of a 'Fondation Jérome Lejeune' PhD fellowship. This work was supported by an ARC grant (no. 3365), an INSERM 'Equipement Mi-Lourd' grant (to J.L.M.) and a French–South African exchange grant (to J.L.M. and M.M.). Funding to pay the Open Access publication charges for this article was provided by INSERM.

*Conflict of interest statement*. None declared.

## REFERENCES

1. Timchenko,L.T. and Caskey,C.T. (1999) Triplet repeat disorders: discussion of molecular mechanisms. *Cell. Mol. Life Sci.*, **55**, 1432–1447.
2. Cummings,C.J. and Zoghbi,H.Y. (2000) Trinucleotide repeats: mechanisms and pathophysiology. *Annu. Rev. Genomics Hum. Genet.*, **1**, 281–328.
3. Bowater,R.P. and Wells,R.D. (2001) The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders. In Moldave,K. (ed.), *Progress in Nucleic Acid Research*. Academic Press Inc., San Diego, CA, Vol. 66, pp. 159–202.
4. Everett,C.M. and Wood,N.W. (2004) Trinucleotide repeats and neurodegenerative disease. *Brain*, **127**, 2385–2405.
5. Oberle,I., Rousseau,F., Heitz,D., Kretz,C., Devys,D., Hanauer,A., Boue,J., Bertheas,M. and Mandel,J. (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science*, **252**, 1097–1102.
6. Brook,J.D., McCurrach,M.E. and Harley,H.G. (1992) Molecular basis of myotonic dystrophy expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family number. *Cell*, **68**, 799–808.
7. Liquori,C.L., Ricker,K., Moseley,M.L., Jacobsen,J.F., Kress,W., Naylor,S.L., Day,J.W. and Ranum,L.P.W. (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science*, **293**, 864–867.
8. McMurray,C.T. (1999) DNA secondary structure: a common and causative factor for expansion in human disease. *Proc. Natl Acad. Sci. USA*, **96**, 1823–1825.
9. Sinden,R.R. (1999) Biological implications of the DNA structures associated with disease causing triplet repeats. *Am. J. Hum. Genet*, **64**, 346–353.
10. Cleary,J.D. and Pearson,C.E. (2005) Replication fork dynamics and dynamic mutations: the fork-shift model of repeat instability. *Trends Genet.*, **21**, 272–280.
11. Hashem,V.I. and Sinden,R.R. (2002) Chemotherapeutic induced deletion of expanded triplet repeats. *Mutat Res.*, **508**, 107–119.
12. Hashem,V.I., Pytlos,M.J., Klysik,E.A., Tsuji,K., Khajav,M., Ashizawa,T. and Sinden,R.R. (2004) Chemotherapeutic deletion of CTG repeats in lymphoblast cells from DM1 patients. *Nucleic Acids Res.*, **32**, 6334–6346.
13. 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) Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. *Proc. Natl Acad. Sci. USA*, **92**, 5199–5203.
14. Zheng,M.X., Huang,X.N., Smith,G.K., Yang,X.Y. and Gao,X.L. (1996) Genetically unstable CXG repeats are structurally dynamic and have a high propensity for folding. An NMR and UV spectroscopic study. *J. Mol. Biol.*, **264**, 323–336.
15. Mariappan,S.V.S., Garcia,A.E. and Gupta,G. (1996) Structure and dynamics of the DNA hairpins formed by tandemly repeated CTG triplets associated with myotonic dystrophy. *Nucleic Acids Res.*, **24**, 775–783.
16. Petruska,J., Arnheim,N. and Goodman,M.F. (1996) Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases. *Nucleic Acids Res.*, **24**, 1992–1998.
17. Gacy,A.M. and McMurray,C.T. (1998) Influence of hairpins on template reannealing at trinucleotide repeat duplexes: a model for slipped DNA. *Biochemistry*, **37**, 9426–9434.
18. Paiva,A.M. and Sheardy,R.D. (2005) The influence of sequence context and length on the kinetics of DNA duplex formation from complementary hairpins possessing (CNG) repeats. *J. Am. Chem. Soc.*, **127**, 5581–5585.
19. 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.
20. Fry,M. and Loeb,L.A. (1994) The fragile X syndrome $d(CGG)_n$ nucleotide repeats form a stable tetrahelical structure. *Proc. Natl Acad. Sci. USA*, **91**, 4950–4954.
21. Chen,F.M. (1995) Acid-facilitated supramolecular assembly of G-quadruplexes in $d(CGG)_4$. *J. Biol. Chem.*, **270**, 23090–23096.
22. Darlow,J.M. and Leach,D.R.F. (1998) Secondary structures in d(CGG) and d(CCG) repeat tracts. *J. Mol. Biol.*, **275**, 3–16.
23. Weisman-Shomer,P., Naot,Y. and Fry,M. (2000) Tetrahelical forms of the fragile X syndrome expanded sequence $d(CGG)_n$ are destabilized by two heterogeneous nuclear ribonucleoprotein-related telomeric DNA-binding proteins. *J. Biol. Chem.*, **275**, 2231–2238.
24. Fojtík,P., Kejnovská,I. and Vorlícková,M. (2004) The guanine-rich fragile X chromosome repeats are reluctant to form tetraplexes. *Nucleic Acids Res.*, **32**, 298–306.
25. Pearson,C.E., Wang,Y.H., Griffith,J.D. and Sinden,R.R. (1998) Structural analysis of slipped-strand DNA (SDNA) formed in

$(CTG)_n$·$(CAG)_n$ repeats from the myotonic dystrophy locus. *Nucleic Acids Res*., **26**, 816–823.
26. Tam,M., Montgomery,S.E., Kekis,M., Stollar,B.D., Price,G.B. and Pearson,C.E. (2003) Slipped (CTG)–(CAG) repeats of the myotonic dystrophy locus: surface probing with anti-DNA antibodies. *J. Mol. Biol*., **332**, 585–600.
27. 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.
28. Chi,L.M. and Lam,S.L. (2005) Structural roles of CTG repeats in slippage expansion during DNA replication. *Nucleic Acids Res*., **33**, 1604–1617.
29. Mitchell,J.E., Newbury,S.F. and McClellan,J.A. (1995) Compact structures of $d(CNG)_n$ oligonucleotides in solution and their possible relevance to fragile X and related human genetic diseases. *Nucleic Acids Res*., **23**, 1876–1881.
30. Völker,J., Makube,N., Plum,G.E., Klump,H.H. and Breslauer,K.J. (2002) Conformational energetics of stable and metastable states formed by DNA triplet repeat oligonucleotides: implications for triplet expansion diseases. *Proc. Natl Acad. Sci. USA*, **99**, 14700–14705.
31. 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.
32. Cantor,C.R., Warshaw,M.M. and Shapiro,H. (1970) Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers*, **9**, 1059–1077.
33. Mergny,J.L., Phan,A.T. and Lacroix,L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett*., **435**, 74–78.
34. Alberti,P., Hoarau,M., Guittat,L., Takasugi,M., Arimondo,P.B., Lacroix,L., Mills,M., Teulade-Fichou,M.P., Vigneron,J.P., Lehn,J.M. *et al*. (2002) Triplex vs. quadruplex specific ligands and telomerase inhibition. In Bailly,C., Demeunynck,M. and Wilson,D. (eds), *Small Molecule DNA and RNA Binders: From Synthesis to Nucleic Acid Complexes*. Wiley VCH, Weinheim, pp. 315–336.
35. Mergny,J.L. and Lacroix,L. (2003) Analysis of thermal melting curves. *Oligonucleotides*, **13**, 515–537.
36. Holbrook,J.A., Capp,M.W., Saecker,R.M. and Record,M.T. (1999) Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: interpretation in terms of coupled processes of formation and association of single-stranded helices. *Biochemistry*, **38**, 8409–8422.
37. Chalikian,T.V., Volker,J., Plum,G.E. and Breslauer,K.J. (1999) A more unified picture for the thermodynamics of nucleic acid duplex melting: a characterization by calorimetric and volumetric techniques. *Proc. Natl Acad. Sci. USA*, **96**, 7853–7858.
38. Rouzina,I. and Bloomfield,V.A. (1999) Heat capacity effects on the melting of DNA. 1. General aspects. *Biophys. J*., **77**, 3242–3251.
39. Rouzina,I. and Bloomfield,V.A. (1999) Heat capacity effects on the melting of DNA. 2. Analysis of nearest- neighbor base pair effects. *Biophys. J*., **77**, 3252–3255.
40. Jelesarov,I., Crane-Robinson,C. and Privalov,P.L. (1999) The energetics of HMG box interactions with DNA: thermodynamic description of the target DNA duplexes. *J. Mol. Biol*., **294**, 981–995.
41. Shindo,H., Torigoe,H. and Sarai,A. (1993) Thermodynamic and kinetic studies of DNA triplex formation of an oligohomopyrimidine and a matched duplex by filter binding assay. *Biochemistry*, **32**, 8963–8969.
42. Chaires,J.B. (1997) Possible origin of differences between van't Hoff and calorimetric enthalpy estimates. *Biophys. Chem*., **64**, 15–23.
43. Mitas,M., Yu,A., Dill,J., Kamp,T.J., Chambers,E.J. and Haworth,I.S. (1995) Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: $(CTG)_{15}$. *Nucleic Acids Res*., **23**, 1050–1059.
44. Chastain,P.D.,II, Eichler,E.E., Kang,S., Nelson,D.L., Levene,S.D. and Sinden,R.R. (1995) Anomalous rapid electrophoretic mobility of DNA containing triplet repeats associated with human disease genes. *Biochemistry*, **34**, 16125–16131.
45. Chastain,P.D. and Sinden,R.R. (1998) CTG repeats associated with human genetic disease are inherently flexible. *J. Mol. Biol*., **275**, 405–411.
46. Gacy,A.M., Goellner,G., Juranic,N., Macura,S. and McMurray,C.T. (1995) Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*. *Cell*, **81**, 533–540.
47. Riesner,D. and Roemer,R. (1973) Differential melting techniques and typical melting curves. In Duchesne,J. (ed.), *Physico-Chemical Properties of Nucleic Acids*. Academic Press, NY, Vol. 2, pp. 277–318.
48. Kettani,A., Kumar,R.A. and Patel,D.J. (1995) Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat. *J. Mol. Biol*., **254**, 638–656.
49. Patel,P.K., Bhavesh,N.S. and Hosur,R.V. (2000) Cation-dependent conformational switches in d-TGGCGGC containing two triplet repeats of Fragile X Syndrome: NMR observations. *Biochem. Biophys. Res. Commun*., **278**, 833–838.
50. Patel,D., Bouaziz,S., Kettani,A. and Wang,Y. (1999) Structures of guanine-rich and cytosine-rich quadruplexes formed *in vitro* by telomeric, centromeric, and triplet repeat disease DNA sequence. In Neidle,S. (ed.), *Oxford Handbook of Nucleic Acid Structure*, Oxford University Press, Oxford, pp. 389–454.
51. Mitas,M. (1997) Trinucleotide repeats associated with human disease. *Nucleic Acids Res*., **25**, 2245–2253.
52. Tian,B., White,R.J., Xia,T.B., Welle,S., Turner,D.H., Mathews,M.B. and Thornton,C.A. (2000) Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. *RNA*, **6**, 79–87.
53. Mikulecky,P.J. and Feig,A.L. (2004) Heat capacity changes in RNA folding: application of perturbation theory to hammerhead ribozyme cold denaturation. *Nucleic Acids Res*., **32**, 3967–3976.
54. Tikhomirova,A., Taulier,N. and Chalikian,T.V. (2004) Energetics of nucleic acid stability: the effect of delta $C_p$. *J. Am. Chem. Soc*., **126**, 16387–16394.
55. Haq,I., Chowdhry,B.Z. and Jenkins,T.C. (2001) Calorimetric techniques in the study of high-order DNA–drug interactions. In Chaires,J.B. and Waring,M.J. (eds), *Drug Nucleic Acid Interaction*. Academic Press Inc., San Diego, CA, Vol. 340, pp. 109–149.
56. Yu,A., Dill,J., Wirth,S.S., Huang,G., Lee,V.H., Haworth,I.S. and Mitas,M. (1995) The trinucleotide repeat sequence $d(GTC)_{15}$ adopts a hairpin conformation. *Nucleic Acids Res*., **23**, 2706–2714.
57. Hartenstine,M.J., Goodman,M.F. and Petruska,J. (2000) Base stacking and even/odd behavior of hairpin loops in DNA triplet repeat slippage and expansion with DNA polymerase. *J. Biol. Chem*., **275**, 18382–18390.
58. Napierala,M., Michalowski,D., de Mezer,M. and Krzyzosiak,W.J. (2005) Facile FMR1 mRNA structure regulation by interruptions in CGG repeats. *Nucleic Acids Res*., **33**, 451–463.
59. Sobczak,K. and Krzyzosiak,W.J. (2005) CAG repeats containing CAA interruptions form branched hairpins structures in spinocerebellar ataxia type 2 transcripts. *J. Biol. Chem*., **280**, 3898–3910.