Nucleic Acids Research, 1995, Vol. 23, No. 21 **4303–4311**

DNA CTG triplet repeats involved in dynamic mutations of neurologically related gene sequences form stable duplexes

G. Kenneth Smith[1], Ji Jie, George E. Fox[1] and Xiaolian Gao*

Department of Chemistry and [1]Department of Biochemical and Biophysical Sciences, University of Houston, Houston, TX 77204-5641, USA

Received August 16, 1995; Accepted September 22, 1995

ABSTRACT

DNA triplet repeats, 5′-d$(CTG)_n$ and 5′-d$(CAG)_n$, are present in genes which have been implicated in several neurodegenerative disorders. To investigate possible stable structures formed by these repeating sequences, we have examined d$(CTG)_n$, d$(CAG)_n$ and d(CTG)•d$(CAG)_n$ (n = 2 and 3) using NMR and UV optical spectroscopy. These studies reveal that single stranded $(CTG)_n$ (n > 2) forms stable, antiparallel helical duplexes, while the single stranded $(CAG)_n$ requires at least three repeating units to form a duplex. NMR and UV melting experiments show that the T_m increases in the order of $[(CAG)_3]_2$ < $[(CTG)_3]_2$ << $(CAG)_3$•$(CTG)_3$. The $(CTG)_3$ duplex is stable and exhibits similar NMR spectra in solutions containing 0.1–4 M NaCl and at a pH range from 4.6 to 8.8. The $(CTG)_3$ duplex, which contains multiple-T•T mismatches, displays many NMR spectral characteristics similar to those of B-form DNA. However, unique NOE and ^1H-^{31}P coupling patterns associated with the repetitive T•T mismatches in the CTG repeats are discerned. These results, in conjunction with recent *in vitro* studies suggest that longer CTG repeats may form hairpin structures, which can potentially cause interruption in replication, leading to dynamic expansion or deletion of triplet repeats.

INTRODUCTION

The discovery of the association between trinucleotide repeat expansions and several hereditary neurological disorders has far-reaching implications for disease diagnosis and treatment (1,2). These neurodegenerative diseases often exhibit an increase in repeat unit length, a phenomenon known as dynamic mutation, and a related increase in severity and/or decrease in the age of onset of the disease with successive generations. There has been an intensive effort focused on ascertaining the functional role that nucleic acid repeat sequences may play in the mutational mechanisms by which certain inherited human disorders are transmitted. The DNA trinucleotide repeat 5′-$(CAG)_n$ is found in the open reading frame of gene sequences, which are associated with the onset of at least six neurodegenerative disorders (3–8). Myotonic dystrophy (9) shows a correlation to increased copies of CTG repeats, but these are located in a 3′ untranslated region of the MD gene, which has sequence homology to a cyclic-AMP dependent protein kinase (10). Another class of triplet expansion diseases are characterized by CCG/CGG triplet repeats, which are associated with hypermethylation of CG islands adjacent to the repeating sequence (11,12). This induces formation of fragile sites, such as FRAXA and FRA16A, found in X syndrome (11) and Jacobsen syndrome (12) related genes, respectively.

Trinucleotide repeat expansion is a well-recognized, but poorly characterized, mutational mechanism implicated in human genetic diseases. A fundamental question is whether these triplet repeats would associate in an unconventional structure, which may serve as a basic structural motif in formation of stable or transient *in vivo* molecular assemblies. The discernment of such structural motifs could provide a clue to the mechanisms that lead to the manifestation of the degenerative disease states. It has been recognized that certain repeating sequences may form stable structures which, while lacking perfect Watson–Crick complementarity, may be essential to cellular function (13–15). NMR studies have recently demonstrated that a heptamer strand d(CGACGAC) containing the triplet repeat CGA or GAC forms a parallel-stranded (denoted Π) DNA duplex at low pH (16,17). The CGA trinucleotide has been shown to be able to promote the formation of Π DNA duplexes in a variety of sequence contexts. Furthermore, long $(CTG)_n$ repeats (n = 26–250) have been shown to have a high probability of forming nucleosome assemblies compared to the flanking random sequences in plasmid DNA (18). The creation of such strong nucleosome positioning signals in regions containing a high number of CTG repeats may interfere with a cascade of biological processes, leading to dynamic expansion/deletion of triplet repeats. These results demonstrate the need for understanding the structure–function relationships of certain DNA repeats.

In light of the crucial functional role of the triplet repeats, we are interested in using NMR spectroscopy to investigate whether short stretches of DNA triplet repeats can form stable secondary structures. Our selection of the triplet sequences in this initial study originated from the following considerations. Statistically, all 60 unique triplets formed by the four natural nucleotides (excluding four homo-trinucleotides) can be expressed in 20 oligonucleotide strands or 10 complementary duplexes. Among these sequences, there are six unique triplets which fall into the

* To whom correspondence should be addressed

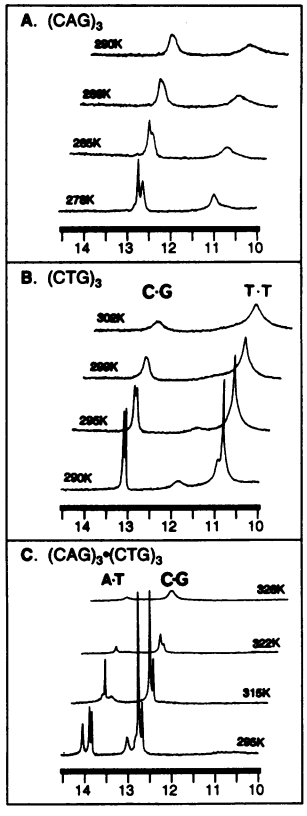


Figure 1. Representative melting profiles of the imino proton region of the nonamer sequences. (**A**) $(CAG)_3$, (**B**) $(CTG)_3$ and (**C**) the complementary $(CAG)_3$•$(CTG)_3$ duplex. NMR samples have a concentration of ~ 0.6–1 mM in single strand DNA. Spectra were recorded in H_2O in the presence of 0.1 M NaCl, 10 mM sodium phosphate and 0.1 mM EDTA at pH 6.3.

category of either CNG or GNC (N = A, C, G and T). Thus far, CTG/CAG and CCG/CGG repeats have been found to be associated with triplet repeat expansion diseases. We have initiated studies of the solution conformation of the six CNG and GNC sequences and found very different spectral characteristics from these triplet repeats (19). A new duplex motif, the *e*-motif, has been discovered for the $d(CCG)_2$ sequence (20). This report focuses on the NMR and UV studies of $d(CTG)_n$, $d(CAG)_n$ and $d(CTG)$•$d(CAG)_n$ (n = 2 and 3, d for deoxy is omitted in the following text). In particular, the relative stability of the $(CAG)_n$ and $(CTG)_n$ duplexes, a complete analysis of the two dimensional (2-D) NMR spectra of the stable $(CTG)_3$ duplex and the structural features of the T•T mismatch in CTG repeats are presented.

MATERIALS AND METHODS

Sample preparation

All DNA strands were synthesized on DNA synthesizers (Biosearch 8600 and Cruachem PS100) using phosphoramidite chemistry. The oligonucleotides were purified using C18 reverse-phase HPLC with 4,4′-dimethoxyltrityl on and then off, followed by size exclusion and Na^+ exchange chromatography. After purification, the oligonucleotides were dissolved in 0.1 M NaCl, 10 mM sodium phosphate, and 0.1 mM EDTA D_2O or 90% H_2O–10% D_2O solutions. The pH values were accurate to ± 0.1 U and those reported in D_2O were uncorrected pH meter readings.

UV experiments

UV thermal melting studies were conducted on a Cary 3E (Varian Associates, Inc.) spectrophotometer equipped with a peltier 12 cell holder capable of temperature variation from 268 to 368 K. Each oligonucleotide sequence was dissolved in 2 ml of buffer solution containing 0.1 M NaCl, 10 mM sodium phosphate and 0.1 mM EDTA adjusted to pH 6.8. The sample concentrations are 16–18 μM for all n = 3 repeats, 34 μM for $(CTG)_2$ and 9 μM for $(CTG)_8$. Temperature was ramped from low to high and back to low at a rate of 0.2 K per min. A_{260} was monitored. Temperature independent thermodynamic parameters (21) were derived from a two state analysis of the melting profiles using the Meltwin program (a kind gift from D. H. Turner at University of Rochester) (22) and/or a similar approach using the Excel software. T_ms determined by NMR or UV have an error estimation of ± 0.5–1.5 K. The values of thermal energies were within ± 15%.

NMR experiments

All NMR experiments were conducted on a Bruker AMX 600 MHz spectrometer. D_2O or 10% D_2O/90% H_2O were used as solvents for observation of non-exchangeable and exchangeable protons, respectively. Proton chemical shifts were referenced to the HOD resonance (4.70 p.p.m. at 298 K, temperature correction factor –0.0109 p.p.m./K). ^{31}P chemical shifts were referenced relative to an external trimethyl phosphate in an aqueous solution containing 0.1 M NaCl (pH 6.5). NMR data were processed using the UXNMR program (Bruker Instruments, Inc.) and the Felix 2.3 program (Biosym Technologies, Inc.).

One dimensional NMR. Sample concentration of various sequences for one dimensional (1-D) NMR studies ranged between ~0.6 and 1.0 mM in single strands. 1-D melting, pH titration and salt concentration-dependent spectra were recorded in H_2O buffer using the Jump-Return (J-R) pulse sequence for HOD suppression, with maximum excitation frequency centered at 12.8 p.p.m. (23). 1-D spectra of the CAG sequences were recorded in a pH range of 3.3–8.5, while those for the CTG sequences were recorded at pH 4.6–8.8. At pH <5.5, single stranded $(CAG)_2$ and $(CAG)_3$ exhibited very broad multiple signals in the spectral region of 10–13 p.p.m.. This condition prevented further studies. Imino line-broadening as a function of temperature was measured at near neutral pH using a minimum of 10 min equilibration time between temperature changes. NMR melting curves were obtained by plotting the half linewidth of marker imino proton resonances as a function of temperature. In the $(CTG)_3$ duplex, the imino proton resonances of the most internal G•C pair and T•T mismatch were used as markers. To provide adequate comparison, the imino protons of the most internal G•C and A•T base pairs in the $(CTG)_3$•$(CAG)_3$ duplex were monitored. 1-D NMR spectra were recorded at various temperatures and at pH 6.8 and 8.5 with salt concentrations varying from 0.1 to 4 M.

2-D NMR experiments of $[(CTG)_3]_2$. 2-D NMR spectra of $(CTG)_3$ were collected at a sample concentration of 2.0 mM (single strand). Typical experimental parameters were as those reported previously (24). 2-D NOESY spectra of exchangeable protons were acquired at 273 and 280 K (1.5 s relaxation delay, 100 and 170 ms mixing times) and those of non-exchangeable protons (70, 120 and 250 ms mixing times) were acquired at 273 and 283 K to resolve ambiguities in resonance and cross peak assignments, such as those arising from HOD and H3′ overlaps.

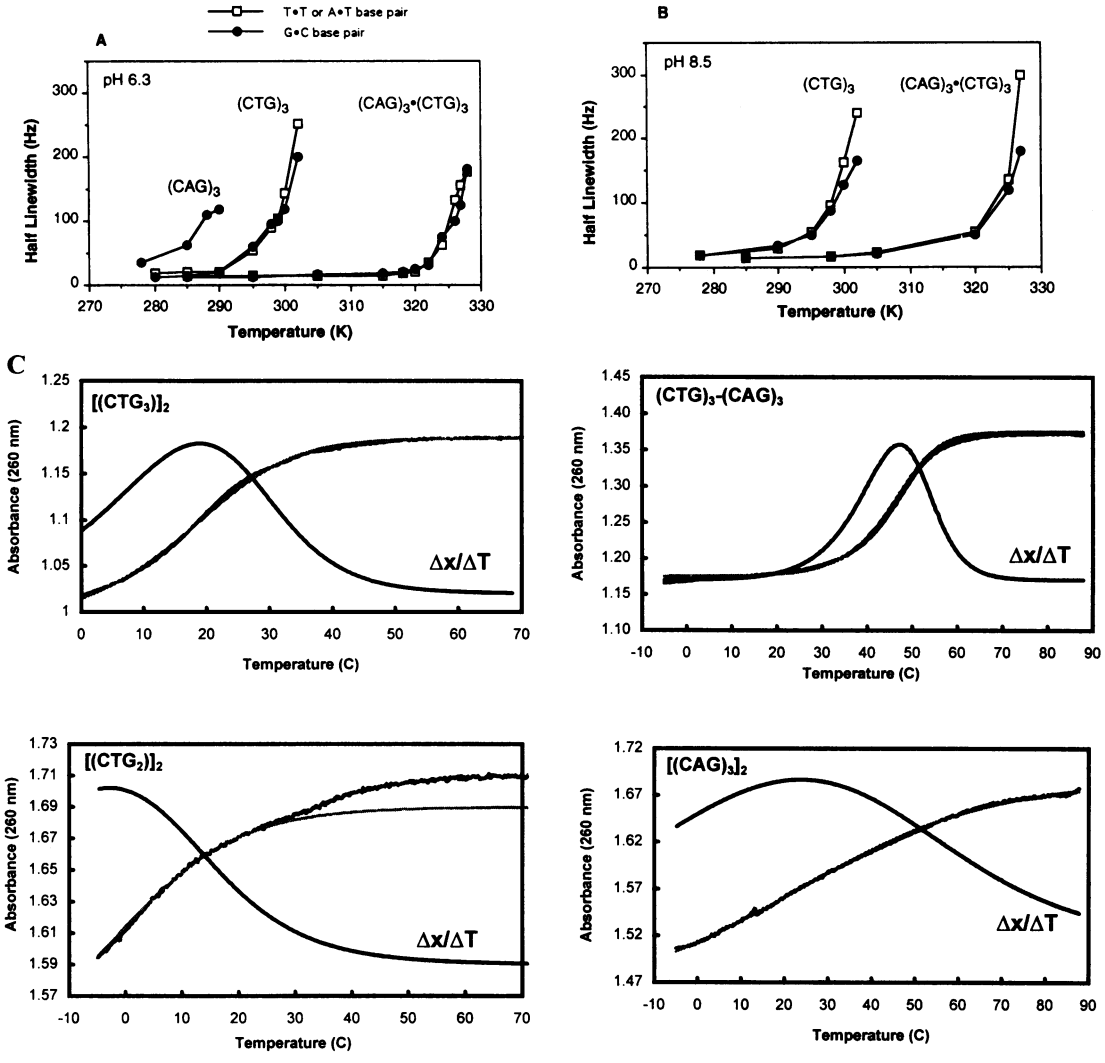


Figure 2. NMR and UV melting curves of CAG and CTG repeat sequences. (**A**) Plots of NMR linewidth (Hz) at half peak height of the internal G or T imino proton resonances versus temperature for the $(CAG)_3$, $(CTG)_3$ and $(CTG)_3$•$(CAG)_3$ duplexes. Data were recorded at pH 6.3. (**B**) Similar NMR melting curves as in (A). Data were recorded at pH 8.5. Experimental conditions are described in Figure 1. Imino proton resonances of G•C base pairs are represented by solid circles and those of either A•T base pair or T•T mismatches are represented by open squares. (**C**) UV melting curves of the four CAG and CTG repeating sequences. The overlay plots show both experimentally measured (light line) and calculated curves (dark line) (see Materials and Methods) and the scaled $\Delta x/\Delta T$ (x = fraction of single strand) versus T curves.

NOESY spectra were used to trace through space connectivities and to obtain chemical shifts of proton resonances. DQF-COSY and COSY-35 data were obtained for scalar coupling analysis. The H1′-H2′ and H1′-H2″ coupling constants were measured from the antiphase multiplets of the COSY-35 spectrum. The coupling information of H2′-H3′, H2″-H3′ and H3′-H4′ were obtained from qualitative analyses of the COSY-35 and DQF-COSY data sets. The proton detected ^1H-^{31}P COSY spectrum was obtained at 283 K. ^1H-^{31}P coupling patterns were analyzed by 1-D simulation of multiplet couplings in the proton dimension based on homo- and hetero-nuclear Karplus equations (25).

Structure elucidation

More than 300 two-fold symmetrical distance/vol and torsion angle restraints were derived from cross peak analyses of NOESY in D_2O (120 and 250 ms mixing time) or in H_2O (100 ms mixing time) and various COSY spectra of the $(CTG)_3$ duplex. Distances were primarily derived from the 120 ms NOESY based on two spin approximation. Fifty percent of the equilibrium distance was added to or subtracted from the equilibrium value to give upper and lower bounds. The maximum and minimum distances were 1.8 and 5.5 Å, respectively. No restraints were applied to the base pair alignment of the T•T mismatches. Standard A- and B-form $(CTG)_3$ duplexes (root-mean-square-deviation, RMSD 5.267 Å) were generated using the Quanta program (Molecular Simulations, Inc., MSI) and used as starting structures. Six structures were generated from each of the two initial starting models by molecular dynamics simulations using random initial velocity assignments. The resulting 12 initial structures have an RMSD value of 2.412 Å, which reflects the diversity among the starting models. A simulated annealing protocol of molecular dynamics (26) was performed using the XPLOR program (MSI). After each generation of calculations, the standard and improper geometry and energies of the molecule were examined. The calculated distances of <5 Å were compared with the distance restraints used in the calculations as well as the complete set of original experimental

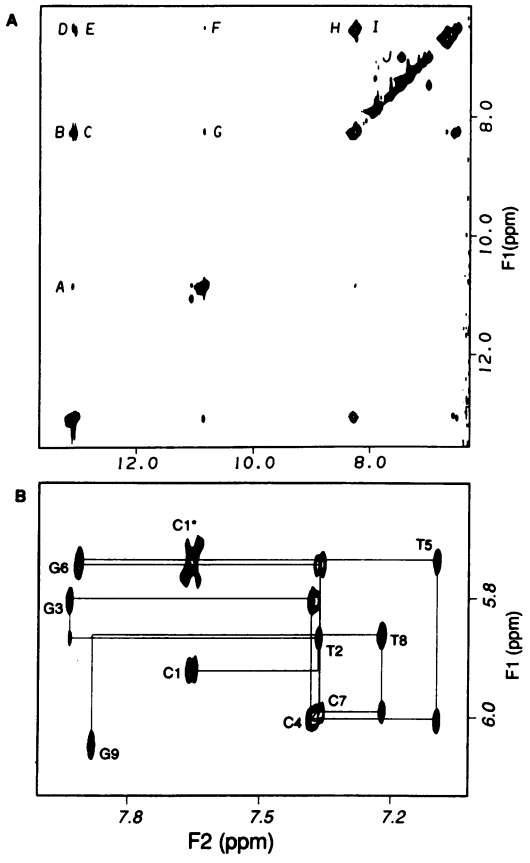


Figure 3. Expanded NOESY spectra of the $(CTG)_3$ duplex (4.4 mM). (**A**) NOESY (170 ms mixing time) recorded in H_2O, pH 6.3, 273 K. NOE assignments of the imino and amino protons are indicated by the labeled cross peaks. Each NOE represents two pairs of symmetrical interproton contacts since the symmetrical residues on the two strands are indistinguishable. *A*, T5(HN)–G6(HN); *B*, G6(HN)–C104(NH_b); *C*, G3(HN)–C107(NH_b); *D*, G6(HN)–C104(NH_{nb}); *E*, G3(HN)–C107(NH_{nb}); *F*, T5(HN)–C104(NH_{nb}); *G*, T5(HN)–C4(NH_b); *H*, *I* and *J*, geminal amino protons of C residues. (**B**). NOESY (120 ms mixing time) recorded in D_2O, pH 6.3, 283 K. Intraresidue and sequential connectivities of H1′ proton resonances (5.7–6.1 p.p.m.) to base proton resonances (7.0 - 8.0 p.p.m.) are shown by solid lines with intraresidue base to H1′ NOEs labeled with residue numbers.

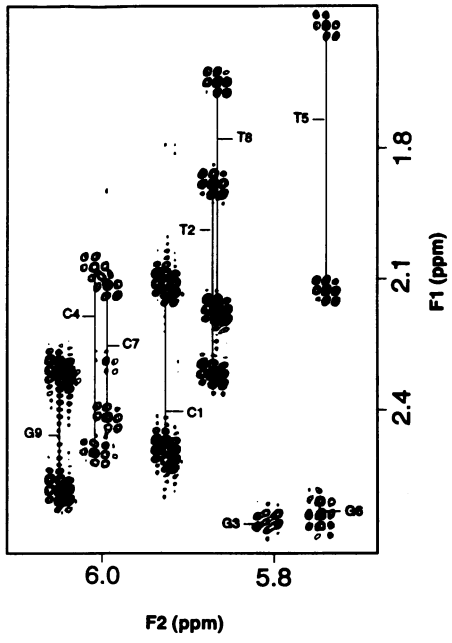


Figure 4. COSY-35 spectrum of the $(CTG)_3$ duplex. The plotted spectral region establishes through bond connectivities between H1′ (5.7–6.1 p.p.m.) and H2′, H2″ (1.4–2.7 p.p.m.) resonances. Coupling constants of H1′-H2′ and H1′-H2″ were derived from the frequency separation of the antiphase components in the spectrum.

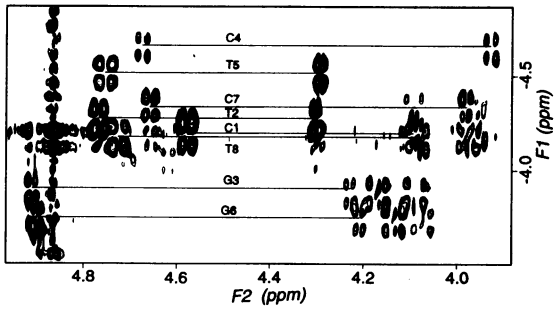


Figure 5. ^1H–^{31}P correlated spectrum for the $(CTG)_3$ duplex. ^{31}P assignments are labelled in the spectrum and the ^{31}P scalar couplings to sugar H3 and H4 protons are shown in solid lines.

NOE cross peaks. This allowed extraction of new distances and removal of ambiguities in the previous input restraints. The calculations were repeated at least three times until NMR restraints were satisfied and a good convergence of structures was reached. The pairwise RMSD of the final 12 structures (six from the A-form and six from the B-form) is 0.92 Å. The final structures satisfy all dihedral angle restraints and no NOE violations were found within a limit of 0.3 Å. A full account of the relaxation matrix structure refinement of the $(CTG)_3$ duplex under hydration conditions followed by molecular dynamics simulations calculations will be reported separately (27).

RESULTS AND DISCUSSION

We report temperature, pH and salt (NaCl) dependence studies of $(CTG)_n$ and $(CAG)_n$ sequences and 2-D NMR structural characterization of the $(CTG)_3$ duplex. The following discussions present several points unique to the CTG and CAG repeats: (i) sequence dependent formation of a duplex structure by the CTG and CAG triplet repeat sequences and their relative stabilities; (ii) formation of the stable CTG repeat duplexes which contain multiple non-W-C base pairs; (iii) interesting spectral features of the T•T mismatch in CTG repeat sequences and their comparison with those reported in the literature (28–31).

Temperature transition of the CTG and CAG triplet repeats suggests formation of stable secondary structures

^1H imino proton spectra provide evidence for the presence of a $(CTG)_2$ duplex and mostly random coiled $(CAG)_2$. The $(CAG)_2$ strand shows a major broad imino resonance at ~11 p.p.m., 273 K, which is indicative of predominantly random structures. In contrast, under similar conditions, the spectrum of the $(CTG)_2$ strand shows three major, moderately sharp resonances at 10.9, 11.1 and 13.2 p.p.m. These resonances belong to W-C base paired G imino protons (13.2 p.p.m.) and T imino protons (10.9 and 11.1 p.p.m.) [assignments based on ^1H chemical shifts of the $(CTG)_3$ duplex, Table 1].

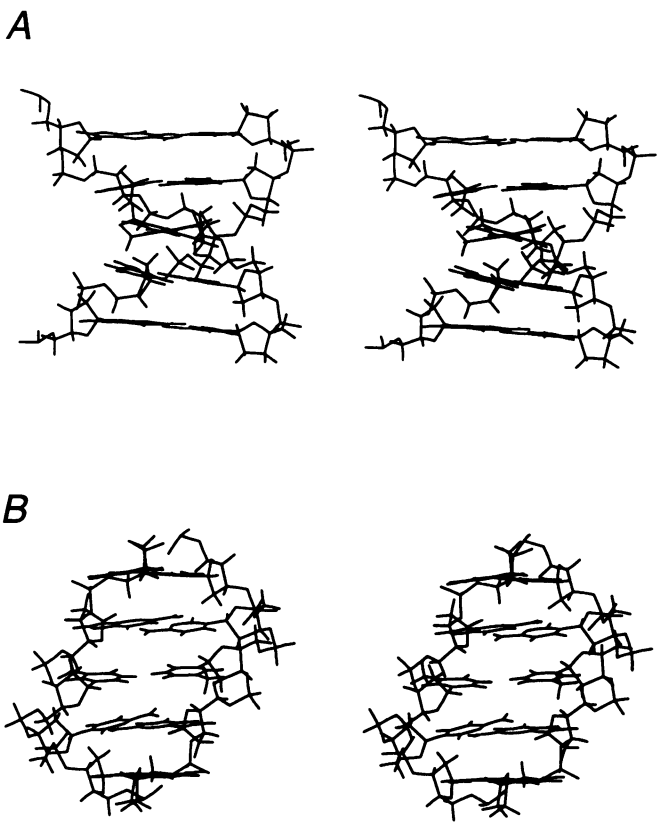


Figure 6. Stereo-drawings (cross-eye) of the central 5'-GCTGC•GCTGC pentanucleotide in the $(CTG)_3$ duplex. The structural drawing is representative of the 12 calculated structures derived from NMR restrained molecular dynamics simulations. The global helical feature and the alignment of the T•T mismatch in the duplex are demonstrated. (**A**) A side view of the pentamer sequence illustrating the relative orientation of the T base moieties in the center of the helix and the alignment of the T•T and G•C base pairs. (**B**) A view into the minor groove illustrating the in-plane alignment of the T•T mismatch. Two hydrogen bonds between the imino protons (HN) and the carbonyl oxygen atoms (O2 and O4) are identified (H-N–O distances <1.78 ± 0.25 Å and H-N–O angle ~166 ± 5°). The cross strand C1′-C1′ separation at the T•T mismatch site is 8.6 ± 0.3 Å, while the C1′-C1′ separation for a A•T base pair in the B-form DNA is 10.7 Å.

^1H imino proton spectra reveal that $(CAG)_3$, $(CTG)_3$ and $(CTG)_3$•$(CAG)_3$ form duplexes. Single stranded $(CTG)_3$ and $(CAG)_3$ exhibit resonances in the 12.6 and 13.1 p.p.m. region (Fig. 1A and B), indicating formation of W-C G•C base pairs characteristic of antiparallel duplexes (32). The two sharp signals at 11 p.p.m. in the spectrum of the $(CTG)_3$ sequence (Fig. 1B) have been assigned to T(HN) protons by a complete NMR spectral analysis as discussed later.

NMR and UV T_m comparisons demonstrate that the relative stability of the mismatch-containing duplexes is sequence dependent. Figure 1 shows representative 1-D NMR spectra plotted as a function of temperature for the imino protons of the three duplexes, $[(CAG)_3]_2$, $[(CTG)_3]_2$ and $(CTG)_3$•$(CAG)_3$. The corresponding plots of linewidth as a function of temperature demonstrate NMR melting profiles for the three duplexes (Fig. 2A and B). In comparison, Figure 2C also shows the UV equilibrium melting curves of $[(CTG)_2]_2$, $[(CAG)_3]_2$, $[(CTG)_3]_2$ and $(CTG)_3$•$(CAG)_3$. The UV measurements show that the melting of $[(CTG)_2]_2$, $[(CTG)_3]_2$ and the complementary duplex is closer to a two state transition than that of $[(CAG)_3]_2$, which displays a broad melting transition, probably due to premelting of the A•T base pairs detected by NMR. Table 2 summarizes T_m and free energy derived from these data. There is a good agreement between NMR and UV melting results. The UV T_m of $(CTG)_8$ included in Table 2 provides support for formation of stable duplexes by CTG repeats. Analysis of data shown in Table 2 and in Figures 1 and 2 suggests that duplex stability increases in the order of $[(CAG)_3]_2 < [(CTG)_3]_2 << (CTG)_3$•$(CAG)_3$. This order of relative stability of the CAG and CTG repeating sequences is also observed by electrophoretic measurements of the longer $(CAG)_{15}$ and $(CTG)_{15}$ sequences (33). Thus, the presence of A•A mismatches in the sequence context of $(CXG)_n$ has a more profound destabilization effect compared to that of the T•T mismatch. Monitoring ^1H linewidth as a function of temperature reveals that the linewidth and intensity of the most internal T•T imino resonances are comparable to those of the most internal G•C imino protons at all temperatures (Fig. 1B). In contrast, the imino proton resonances (14 p.p.m.) of the A•T base pair in the complementary $(CAG)_3$•$(CTG)_3$ duplex disappear much faster than those of the G•C base pair (Fig. 1C), suggesting partial premelting at the A•T base pairs.

Table 1. Chemical shifts of protons and ^{31}P of the $(CTG)_3$ duplex[a]

	H8	H6	H5	H1′	H2′	H2″	H3′	H4′	HN	NH_{nb}	NH_b	^{31}P
C1/101		7.66	5.74	5.92	2.11	2.49	4.57	4.03		7.00	7.48	–4.20
T2/102		7.36	1.65	5.87	1.88	2.30	4.76	4.09	11.07			–4.24
G3/103	7.94			5.81	2.67	2.62	4.91	4.30	13.05			–4.02
C4/104		7.37	5.30	6.01	2.07	2.50	4.66	4.23		6.57	8.27	–4.47
T5/105		7.09	1.56	5.74	1.52	2.13	4.75	3.93	10.85			–4.38
G6/106	7.91			5.75	2.66	2.60	4.90	4.29	13.09			–3.93
C7/107		7.37	5.29	5.99	2.11	2.42	4.65	4.21		6.51	8.25	–4.28
T8/108		7.22	1.63	5.86	1.65	2.17	4.72	3.98	11.07			–4.17
G9/109	7.88			6.05	2.58	2.31	4.60	4.07				

[a]Chemical shift assignments (p.p.m.) of non-exchangeable proton and ^{31}P resonances were measured at 283 K. Exchangeable proton chemical shifts were measured at 273 K. Fine multiple splitting was detected for some resonances in both the H_2O and D_2O spectra. However, the differences were too small to give distinguishable chemical shifts.

Table 2. NMR and UV summary of the $(CTG)_n$ and $(CAG)_n$ sequences

Sequence	$T_{m,nmr}$ (K)[a]	$T_{m,UV}$ (K)[b]	ΔG(kCal/mol)	pH effect[c]
$(CAG)_2$	~273	n.a.	n.a.	n.a.
$(CTG)_2$	>273	268	–2.90	n.a.
$(CAG)_3$	288	296–300[d]	–5.98	n.a.
$(CTG)_3$	300	290	–4.85	@300 K, Intensity T(HN):G(HN) 1:2
$(CAG)_3$•$(CTG)_3$	325	324	–9.50	@325 K, Intensity T(HN):G(HN) 1:3
$(CTG)_8$[e]	n.a.	332	–10.29	n.a.

[a]$T_{m,nmr}$ (K) was derived from line broadening of imino proton resonances as a function of temperature (Figs 1 and 2). $T_{m,nmr}$ (K) is the point where the first derivative of the curve is approaching maximum. All NMR samples contain 0.1 M NaCl, 10 mM phosphate.
[b]UV samples contain 0.1 M NaCl, 10 mM phosphate, pH 6.8. n.a. means information not available.
[c]pH effect was examined to probe the behaviour of T•T mismatch and its comparison with the A•T base pair.
[d]$(CAG)_3$ exhibits broad transitions, T_m thus derived is likely to be higher than the actual T_m.
[e]T_m is concentration independent with 12× range, indicating formation of a hairpin structure.

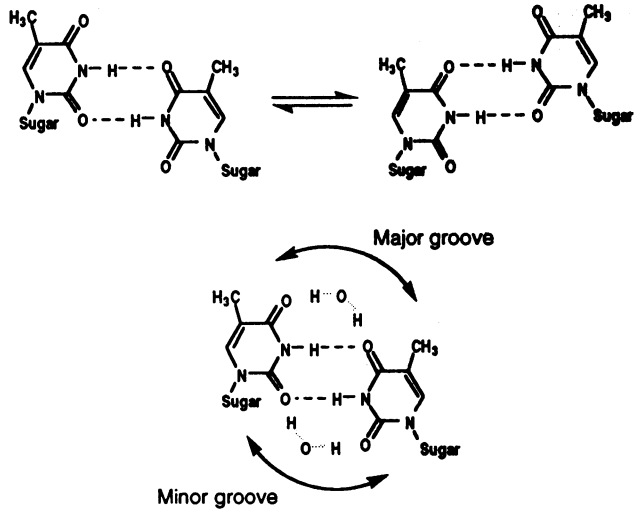
Basic pH ^{1}H spectra show stable T•T imino protons

The imino proton spectra of the $(CTG)_3$ and $(CAG)_3$•$(CTG)_3$ duplexes were recorded at pH 8.5 (NaOH as the base) to probe line broadening due to base catalyzed exchange of imino protons with solvent water. Under these conditions, OH^- catalyzes solvent H-transfer to monomer T at a rate 10^5–10^6 [s^{-1}] (= collision rate constant × [OH^-]). As expected, under basic pH conditions, the terminal imino proton resonances are rather broad. The interesting feature is that T•T imino proton resonances at ~11 p.p.m. maintain, at high pH, comparable linewidth to that of the G•C imino protons (13 p.p.m.). These experimental observations contradict conventional thinking of T•T mismatch being an 'open' base pair (29–31). In those cases, T•T imino proton resonances would be too broad to be observed under basic conditions, while the W-C base paired G imino proton resonances would give sharp signals (29,31). The unusual behavior of the T•T imino proton resonances may be attributed to two causes: the formation of a tightly hydrogen bonded T•T mismatch or the restricted access of these imino protons by the solvent water molecules. The first cause is unlikely, since the $(CTG)_3$ duplex is much less stable than the complementary duplex (Table 2). Thus, the internal T imino protons of the T•T mismatch in CTG repeats may be protected from OH^- catalyzed H-exchange. The lower limit of the exchange life time (τ_{ex}) of imino proton resonances is estimated to be 0.13 ms ($1/\tau_{ex} = \Delta\upsilon \times \pi / 2^{1/2}$) at temperatures <290 K, based on the 3.5 kHz chemical shift separation between HOD and the sharp T•T imino proton resonances. It is possible that water molecules are involved in the T•T pairing in a way similar to that observed in the crystal G•T wobble base pair (34). The putative water in the major and/or the minor grooves (Scheme 2) may be stabilized by weak hydrogen bonds between water protons and acceptor groups, such as the T carbonyl groups. However, other hydrogen bonding schemes, such as those involving the nearest neighbor C or G residues, cannot be ruled out. Molecular dynamics simulations using hydration conditions and further NMR experiments are underway to examine the presence of water in helical grooves (35–37) and to investigate the molecular origin of the stable T•T mismatch protons in the $(CTG)_3$ duplex.

5'

C1	T2	G3	C4	T5	G6	C7	T8	G9
•	x	•	•	x	•	•	x	•
G109	T108	C107	G106	T105	C104	G103	T102	C101

Scheme 1. • and x represent Watson–Crick and mismatch base pairs, respectively.



Scheme 2. T•T mismatch alignment and plausible hydration pattern.

[NaCl] effect on the melting of $[(CTG)_3]_2$ and $(CAG)_3$•$(CTG)_3$ is insignificant

High salt conditions are known to induce B to Z conformational transitions of $(CG)_n$ sequences (38,39). 1-D proton spectra for the observation of imino proton resonances were recorded under various salt conditions ([NaCl] = 0.1–4.0 M), at pH 6.8 and 8.4 and a temperature range from 273 to 305 K, for both $(CTG)_3$ and

$(CAG)_3$•$(CTG)_3$ duplexes. There are only subtle changes in the spectra under these conditions and only under extreme conditions, such as high salt concentration (> 1.0 M), high temperature (close to T_m) and high pH (8.4), are there moderate universal line-broadening of the imino proton resonances with a concomitant 3–4 K reduction in T_m for both $(CTG)_3$ and $(CAG)_3$•$(CTG)_3$ duplexes. Thus, T•T mismatches in the $(CTG)_3$ duplex are not particularly sensitive to sodium ion concentration. There is no structural transition in the $(CTG)_3$ duplex under various salt concentrations.

2-D Spectral characterization of the $(CTG)_3$ duplex

NOE connectivity analysis reveals helical alignment and a stack-in conformation for T•T mismatch pairs. A complete accounting of the proton chemical shifts (except for those of H5′ and H5″) of the $(CTG)_3$ duplex is given in Table 1. The contour plot (6.5–14.0 p.p.m.) of the NOESY spectrum (100 ms mixing time) of the $(CTG)_3$ duplex in H_2O buffer is shown in Figure 3A. Cross peaks B, C, D and E are assigned to inter-proton interactions between G imino protons and C amino protons of the same base pair, indicating that the G and C residues flanking the T•T mismatch sites are W-C base paired. Several NOEs related to the T imino proton resonances are identified. These interactions between the T imino and the adjacent G imino protons (peak A, Fig. 3A) and between the T imino and the adjacent C amino protons (peaks G and F, Fig. 3A) provide evidence for a stacked duplex at the T•T mismatch site. Furthermore, it is noted that in canonical B- or A-form duplexes, such as CTG•CAG, distances between T imino protons of W-C A•T pairs and the adjacent C amino protons are longer than 4 Å. Usually no corresponding NOEs are observed in NOESY spectra recorded in H_2O. Thus, NOE cross peaks G and F (Fig. 3A) observed at the T•T mismatch site of the $(CTG)_3$ duplex indicate a different stacking pattern at the T•T mismatch site.

The observed NOEs related to T imino protons are consistent with proposed T•T mismatch pairing (31) (Scheme 2), which is stabilized by two hydrogen bonds: one between the imino H3 and the carbonyl O4, and the other between the second H3 and the carbonyl O2 of the opposite strands with glycosidic bonds in an anti-configuration. Experimentally, only one imino proton resonance was observed for each of the T•T mismatches. Thus, the T imino protons in either orientation (Scheme 2) are indistinguishable, either because of the symmetry of the two strands, or the fast equilibrium between the two H-bonding states. It is also possible that the chemical shifts of the two imino proton resonances in different hydrogen bonding states are coincidentally degenerate.

NOEs of non-exchangeable protons show alternating patterns of non-B-structural features. The non-exchangeable proton resonances of the $(CTG)_3$ duplex are well-resolved, as shown by the expanded NOESY plot recorded at 283 K and pH 6.5 (Fig. 3B). The chemical shift assignments were obtained by conventional methods (40) and are given in Table 1. The proton chemical shifts of the two strands are virtually indistinguishable, thus, for simplicity only one of the two symmetrical residues will be mentioned in the following discussion. The chemical shifts of T(H2′) resonances, especially that of the T5 residue, are upfield shifted with T5(H2′) resonating at 1.52 p.p.m. [in the corresponding W-C duplex containing A•T base pair, T5(H2′) is at 1.91 p.p.m.], while T5(H4′) resonance appears upfield compared to its own H5′,5″ protons (assigned through ^{1}H-^{31}P COSY). Another unusual feature is that the H2′ and H2″ resonances of the G residues exhibit an inverted chemical shift pattern with H2′ being downfield shifted from the H2″ resonances. These observations are likely to be associated with the structural features of a CTG repeat duplex that differs from a canonical B-form DNA duplex. The inverted chemical shifts of H2′ and H2″ of G residues are also observed in the third strand of a DNA triplex (41).

The observed sequential NOEs between base and sugar protons confirm a right-handed helical structure of the $(CTG)_3$ duplex. The NOESY (120 ms mixing time) spectral region in Figure 3B shows the connectivities between base protons (pyrimidine H6 and purine H8) at 7.0–8.1 p.p.m. (F2) and sugar H1′ protons at 5.6–6.1 p.p.m. (F1). Each base proton exhibits NOEs to its own and to the 5′ flanking sugar H1′ protons (40), and there is no interruption at the mismatch sites. An interesting feature in this spectral region is that the intensities of intraresidue NOEs are equal to or greater than those of inter-residue NOEs for T and G residues (G6 residue exhibits overlapped intra- and inter-residue NOEs, Fig. 3B), opposite to the pattern observed in canonical B-DNA duplexes. This suggests that the helical alignment of the 5′-CT and 5′-TG steps differs from that of the 5′-GC steps in the $(CTG)_3$ duplex. The NOEs in the base to H2′,2″ spectral region show a similar pattern, which deviates from what is expected for a typical B-form helix at the 5′-CT and 5′-TG steps.

^{1}H-^{1}H scalar coupling analysis finds C1′-exo and O4′- endo sugar pucker throughout the helix. The analysis of the COSY-35 and DQF-COSY data is summarized in Table 3. An expanded COSY-35 contour plot which establishes through bond connectivities between sugar H1′ (5.6–6.2 p.p.m.) and H2′, H2″ (1.4–2.7 p.p.m.) resonances is plotted in Figure 4. The coupling constants of H1′ to H2′ and H2″ were measured from the frequency separation of antiphase components of the COSY-35 cross peaks along the F2 dimension (Table 3). Additionally, the examination of the fine spectral patterns discernible from the COSY-35 and DQF-COSY spectra yielded qualitative information for H2′-H3′, H2″-H3′ and H3′-H4′ couplings (Table 3). The H2′ and H2″ of the G residues are less well-resolved (Table 1), but the absence of the H3′-H4′ coupling cross peaks for these residues indicates that the sugars are likely to be in the C2′-endo region (42). The pyrimidine residues show comparable H1′-H2′ and H1′-H2″ couplings, as well as moderate H2′-H3′ and H3′-H4′ couplings. These data define a preferred sugar conformation in the range from C1′-exo to O4′-endo, similar to the conformation usually observed for pyrimidine residues in canonical B-DNA duplexes. These results suggest that the accommodation of the three T•T mismatches in CTG repeats does not require significant change in the sugar conformation.

^{1}H-^{31}P spectral analysis shows a generally unperturbed backbone and an alternating pattern of backbone torsion angles. The ^{1}H-^{31}P COSY spectrum shows well-resolved cross peaks (Fig. 5), establishing connectivities from H3′, H4′ and H5′,5″ to backbone phosphates. The chemical shifts of ^{31}P are given in Table 1 ([^{31}P$_i$] is the phosphate linking O3′ of residues *i* with O5′ of residue *I* + *1*). Although the chemical shift dispersion of ^{31}P resonances is only ~0.5 p.p.m., their cross peaks are well-resolved, revealing a certain pattern. The ^{31}P resonances of G3-C4 and G6-C7 (5′-GC type) appear at the most upfield region, while those of C4-T5 and T5-G6 (T5 related type) are the most downfield shifted (Fig. 5). The ^{31}P resonances of the terminal two

residues are located in the center of the spectrum shown in Figure 5. In according with the differences in the chemical shifts, the cross peaks of the three types of residues (the 5′-GC type, the T5 type and the terminal type) show distinctly different patterns. The 5′-GC ^{31}P resonances show relatively weak coupling to H4′ and moderate couplings to both H5′ and H5″ (4.1 p.p.m.), while the T5 related ^{31}P resonances exhibit strong H4′-P and extremely weak H5′,5″-P couplings (too weak to be seen in Fig. 5), respectively. It is interesting that the alternating pattern observed in ^{1}H-^{31}P couplings is consistent with that of sequential NOEs (*vide supra*).

Table 3. Coupling constant analysis of the sugar protons of the $(CTG)_3$ duplex[a]

	H1′-H2′	H1′-H2″	H2′-H3′	H2″-H3′	H3′-H4′
C1	6.0	6.0	s–(w+)	x (w+)	s (m)
T2	7.2	7.3	m (w)	w (x)	m (m–)
G3	7.2	7.0	w (w–)	x (x)	x (w–)
C4	7.4	6.5	m (x)	w–(w–)	m+(m–)
T5	7.1	7.1	m (w)	w (x)	s (m)
G6	9.1	7.0	x (w)	x (x)	x (w–)
C7	8.0	6.1	m (x)	w–(w–)	m+(m–)
T8	7.1	7.0	m+(w–)	m–(w–)	s (m)
G9	7.1	7.3	s (w)	s (w+)	s (m)

[a]Quantitative coupling constants (± 0.5 Hz) were obtained from the measurements of the frequency separations between the antiphase multiplets in the COSY-35 spectrum. Coupling cross peak intensities are expressed in s = strong; m = medium; w = weak; x = not seen or overlapped. Information in the parentheses was derived from the DQF-COSY spectrum.

^{31}P chemical shifts are sensitive to bond geometry measured by backbone dihedral angle parameters ζ and α (C3′-**O3′-P-O5′** and **O3′-P-O5′**-C5′, bond angle in bold face) and β and ε (**P-O5′-C5′**-C4′ and C4′-**C3′-O3′-P**) (43–45). For a B-type duplex, the ε angle is approximately –150°, which translates to a 30° H3′-C3′-O3′-P torsion angle and a 10 Hz coupling constant. Our experimental results are consistent with this ε angle orientation except for the C4(H3′)-P. The reduced cross peak intensity (Fig. 5) indicates a reduction in the coupling, resulting from an increase in H3′-P torsion angle as the ε angle approaches 180°. The B-form β angle is approximately –170°, which would place H5′ and H5″ at 70 and 50°, respectively, with respect to the phosphate and results in coupling constants of ~9 Hz according to the heteronuclear Karplus equations (25). Experimentally, both couplings are detectable as composite cross peaks of the couplings of H5′,5″-P, H5′,5″-H4′ and H5′-H5″ proton pairs. This is the case for the ^{31}P resonances at the lower half of the spectrum (Fig. 5), however, the cross peaks linking 5′ phosphates of T5, G6 and T8 residues (cross peaks labeled C4, T5 and C7) with H5′ and H5″ protons can only be found at very low spectral level. Two alternatives may be responsible for this observation. Since the predicted coupling approaches a minimal value of ~2 Hz, as the H-C-O-P torsion angle reaches 90 or 270° angles, increasing the β angle from –170° to a value close to –150° would result in a 90° H5′(5″)-O5′-C5′-P torsion angle, causing significant intensity reduction. Alternatively, one of the H5's could be oriented 90° with respect to the phosphate while the other geminal proton was –30° with respect to the phosphate, corresponding to a 10 Hz coupling. It may be that the β angle fluctuates in the range of –150 to 150°, resulting in an averaged apparent coupling constant of small value. This seems to be reasonable, since the T residues are likely in fast exchange between the two base pairing states (Scheme 2).

Structure refinement of the $(CTG)_3$ duplex reveals hydrogen bonded T•T mismatches in a B-type helix and a narrower groove width at the mismatch site

The structure of the $(CTG)_3$ duplex has been elucidated using molecular dynamics simulations based on distances and dihedral angles derived from NMR spectral analysis. The stereo-drawing of the central pentanucleotide 5′-GCTGC•GCTGC in a representative structure of the $(CTG)_3$ duplex is shown in Figure 6. These residues are well-converged, while terminal residues exhibit relatively large deviation in the 12 calculated structures, consistent with end-fraying of these residues. Each of the three T•T mismatches may be present in at least two pairing forms (Scheme 2), which, in combination, results in four possible forms for the duplex of identical $(CTG)_3$ strands. Therefore, it is important to keep in mind that the structure drawn in Figure 6 is a snap shot of the conformational equilibrium of the $(CTG)_3$ duplex and should resemble an averaged form of the conformation assembly.

The T•T mismatch of $(CTG)_3$ is well-accommodated in the helical duplex with both central T bases stacked in the helix. Although no hydrogen bonding restraints for the T•T mismatch were applied during the calculation, the two bases on the opposite strand are aligned in such a way that two hydrogen bonds between the imino protons and the carbonyl oxygen atoms are identified (H-N–O distances <1.78 ± 0.25 Å and H-N–O angle ~166 ± 5°). The two T residues of opposite strands are not equivalent, because the base moiety of one of the two T residues is pushed towards the major groove, adopting an orientation supported by the NOEs between the T imino proton and the C amino protons. The formation of the T•T mismatch base pair is accompanied by a reduced groove width. The cross strand C1′-C1′ separation at the T•T mismatch site is 8.6 ± 0.3 Å compared to 10.7 Å of a B-form W-C A•T base pair in the same sequence context. The narrowing of the groove width suggests direct hydrogen bonds between the T residues of opposite strands, as opposed to indirect hydrogen bonds involving a bridging water found in the C•U mismatch (46).

SUMMARY

UV melting, NMR spectral analysis and restrained molecular dynamics structure refinement reveal the formation of the stable, antiparallel duplexes of CTG repeats, which are stabilized by W-C G•C base pairs and H-bonded T•T mismatches in a right handed helix. In comparison with CTG repeats, CAG repeats form less stable duplexes at near neutral pH. The distinctive features of the T•T mismatches observed in this study are that the T imino proton resonances are not much more broadened more than those of the C•G pair at elevated temperature or basic pH (8.5). This suggests that a hairpin structure may readily form in longer CTG repeat sequences. This is supported by recent electrophoretic results (33) and primer extension studies (47), which indicate that CTG repeats cause DNA polymerase to pause along such tracts during replication in *E.coli*. The observed expansion and contraction of the CTG repeats may be attributable to a slippage mechanism, possibly due to the formation of stable

hairpin structures and differential stabilities between $(CTG)_n \bullet (CTG)_n$ and $(CAG)_n \bullet (CAG)_n$.

ACKNOWLEDGEMENTS

This research is supported in part by grants from the Welch Foundation and Triplex Pharmaceutical Co. to X. G., NASA graduate Student Researchers Fellowship Award NGF51085 to G. K. S and NASA grant NAGW-2108 to G. E. F. The 600 MHz NMR spectrometer at the University of Houston is funded in part by the W. M. Keck Foundation. The computation was supported in part by the W.M. Keck Center for Computational Biology. The authors thank Minxue Zheng for UV measurements, Dr Wells and his research group (Texas A&M University) for many stimulating discussions. X. G. thanks Dr Patel (Sloan-Kettering Cancer Center) and Dr Wang (Columbia University) for discussions of their preliminary results, in which similar stable T•T mismatches in $G(CTG)_{2,5}G$ sequences were observed in NMR spectra recorded in H_2O.

REFERENCES

1 Caskey, C. T., Pizzuti, A., Fu, Y.-H., Fenwick, Jr., R. G. and Nelson, D. L. (1992) *Science*, **256**, 784–789
2 Willems, P. J. (1994) *Nature Genet.*, **8**, 212–214.
3 Orr, H. T., Chung, M. Y., Banfi, S., Kwiatkowski, T. J., Servakio, A., Beaudet, A. L., McCall, L.A., Duvick, Ranum, L.P. and Zoghbi, H. Y. (1993) *Nature Genet.*, **4**, 221–226.
4 LaSpada, A. E., Wilson, E. M., Lubahn, D. B., Harding, A. E. and Fischbeck, K. H. (1991) *Nature*, **352**, 77–79.
5 The Huntington's Disease Collaborative Research Group (1993) *Cell*, **72**, 971–983.
6 Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hatashi, T., Saito, M., Tomoda, A., Miike, T., Naito, H. and Ikuta, F. (1994) *Nature Genet.*, **6**, 9–13.
7 Burke, J. R., Wingfield, M. S., Lewis, K. E., Roses, A. D., Lee, J. E., Hulette, C., Pericak-Vance, M. A. and Vance, J. M. (1994) *Nature Genet.*, **7**, 521–524.
8 Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S. Kawakami, H, Nakamura, S., Nishimura, M., Akiguchi, I., Kimura, J., Narumiya, S. and Kakizuka, A. (1994) *Nature Genet.*, **8**, 221–228.
9 Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, N., Barcelo, J., O'Hoy, K., Leblond, S., Earle-Macdonald, J., DeJong, P. J., Weiringa, B. and Korneluk, R. G. (1992) *Science*, **5**, 1253–1255.
10 Brook, J. D., McCurrash, A. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J.-P., Hudson, T. *et al.* (1991) *Science*, **252**, 1097–1102.
11 Oberle, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M. F. and Mandel, J. L. (1991) *Science*, **252**, 1097–1102.
12 Jones, C., Penny L., Mattina, T., Yu, S., Baker, E., Voullaire, L., Langdon, W. Y., Sutherland, G. R., Richards. R. I. and Tunnacliffe, A. (1995) *Nature*, **376**, 145–149.
13 Wells, R. D., Larson, J. E. and Grant, R. C. (1970) *J. Mol. Biol.*, **54**, 465–497.
14 Shimizu, M., Hanvey, J. C. and Wells, R. D. (1990) *Biochemistry*, **29**, 4704–4713.
15 Sinden, R. R. and Wells, R. D. (1992) *Curr. Opin. Biotech.*, **3**, 612–622.
16 Robinson, H., van der Marel, G. A., van Boom J. H. and Wang, A. H.-J. (1992) *Biochemistry*, **31**, 1565–1566.
17 Robinson, H., van Boom J. H. and Wang, A. H.-J. (1994) *J. Am. Chem. Soc.*, **116**, 1565–1566.
18 Wang, Y. H., Amirhaeri, S., Kang, S., Wells, R. D. and Griffith, J. D. (1994) *Science*, **265**, 669–670.
19 unpublished UV, gel electrophoresis and NMR results in this laboratory.
20 Gao, X., Huang, X., Smith, G. K., Zheng, M. and Liu. H. (1995) *J. Am. Chem. Soc.*, **117**, 8883–8884.
21 Breslauer. K. J. (1994) *Methods Mol. Biol.*, **26**, 347–372.
22 Petersheim, M. and Turner, D. H. (1983) *Biochemistry*, **22**, 253–263.
23 Plateau, P. and Gueron, M. (1982) *J. Am. Chem. Soc.*, **104**, 7310–7311.
24 Gao, X., Stassinopoulos, A., Rice, J. S. and Goldberg, I. H. (1995) *Biochemistry*, **34**, 40–49.
25 Majumdar A. and Hosur, R. V. (1992) *Prog. NMR Spec.*, **24**, 109–158.
26 Kuszewski, J. K., Nilges, M. and Brünger, A. T. (1992) *J. Biol. NMR* **2**, 33–56.
27 Ji, J., Smith, K. and Gao, X. (1995) manuscript in preparation.
28 Aboul-ela, F., Koh, D. and Tinoco, Jr, I. (1985) *Nucleic Acids Res.*, **13**, 4811–4824.
29 Werntges, H., Steger, G., Riesner, D. and Fritz, H.-J. (1986) *Nucleic Acids Res.*, **14**, 3773–3790.
30 Arnold, F. H., Wolk, S., Cruz, P. and Tinoco, Jr, I. (1987) *Biochemistry*, **26**, 4068–4075.
31 Kouchakdjian, M., Li, B. F. L., Swann, P. F. and Patel, D. J. (1988) *J. Mol. Biol.*, **202**, 139–155.
32 20% polyacrylamide nondenaturing gel electrophoresis performed in this laboratory shows that $(CAG)_3$ and $(CTG)_3$ migrate faster than the corresponding complementary duplex at 4°C.
33 Mitas, M., Yu, A., Dill, J., Kamp, T. J., Chambers, P. E. J. and Haworth, I. S. (1995) *Nucleic Acids Res.*, **23**, 1050–1059.
34 Rabinovich, D., Haran, T., Eisenstin, M. and Shakked, Z. (1988) *J. Mol. Biol.*, **200**, 151–161.
35 Otting, G., Liepinsh, E. and Wuthrich, K. (1991) *Science*, **254**, 974–980.
36 Kubinec, M. G. and Wemmer, D. E. (1992) *J. Am. Chem. Soc.*, **114**, 8739–8740.
37 Mohan, V., Smith, P. E. and Pettitt, M. (1993) *J. Am. Chem. Soc.*, **115**, 9297–9298.
38 Wang, A. H.-J., van Boom, J. H., van der Marel, G., Rich, A., Quilgley, G. J. and Kolpack, F. J. (1979) *Nature*, **282**, 680–686,
39 Reid, D. G., Salisbury, S. A. and Williams, D. H. (1983) *Nucleic Acids Res.*, **11**, 3779–3793.
40 Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids.* John Wiley & Sons Inc., New York, pp. 203–255.
41 Dittrich, K., Gu, J., Tinder, R., Hogan, M. and Gao, X. (1994) *Biochemistry*, **33**, 4411–4120.
42 Chary, K. V. R., Hosur, R. V., Govil, G., Zu-Kun, T. and Miles, H. T. (1987) *Biochemistry*, **26**, 1315–1322.
43 Giessner-Prettre, C., Pullman, B., Cheng, D. M., Iuorno, V. and Ts'O, P. O. P. (1984) *Biopolymers*, **23**, 377–388.
44 Gorenstein, D. G. (1984) *Phosphorus-31 NMR. Principles and Applications.* Academic Press, New York.
45 Lankhorst, P. P., Haasnoot, C. A. G., Erkelens, C. and Altona, C. (1984) *J. Biomol. Struct. Dyn.*, **1**, 1387–1405.
46 Cruse, W. B. T., Saludjian, P., Biala, E., Strazewski, P., Prange, T. and Kennard, O. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 4160–4164.
47 Kang, S., Jaworski, A., Ohshima, K. and Wells, R. D. (1995) *Nature Genet.*, **10**, 213–218.