

Unexpected Formation of Parallel Duplex in GAA and TTC Trinucleotide Repeats of Friedreich's Ataxia

Eric M. LeProust^{1†}, Christopher E. Pearso^{2,3†}, Richard R. Sinden² and Xiaolian Gao^{1*}

¹Department of Chemistry
University of Houston, 4800
Calhoun St., Houston
TX 77204-5641, USA

²Center for Genome Research
Institute of Biosciences and
Technology, Texas A&M
University, Texas Medical
Center, 2121 West Holcombe
Houston, TX 77030-3303, USA

³Department of Genetics
Hospital for Sick Children, 555
University Avenue, Toronto
Ontario, Canada M5G 1X8

The onset and progress of Friedreich's ataxia (FRDA) is associated with the genetic instability of the (GAA)·(TTC) trinucleotide repeats located within the *frataxin* gene. The instability of these repeats may involve the formation of an alternative DNA structure. Poly-purine (R)/poly-pyrimidine (Y) sequences typically form triplex DNA structures which may contribute to genetic instability. Conventional wisdom suggested that triplex structures formed by these poly-purine (R)/poly-pyrimidine (Y) sequences may contribute to their genetic instability. Here, we report the characterization of the single-stranded GAA and TTC sequences and their mixtures using NMR, UV-melting, and gel electrophoresis, as well as chemical and enzymatic probing methods. We show that the FRDA GAA/TTC repeats are capable of forming various alternative structures. The most intriguing is the observation of a parallel (GAA)·(TTC) duplex in equilibrium with the antiparallel Watson-Crick (GAA)·(TTC) duplex. We also show that the GAA strands form self-assembled structures, whereas the TTC strands are essentially unstructured. Finally, we demonstrate that the FRDA repeats form only the YRY triplex (but not the RRY triplex) at neutral pH and the complete formation of the YRY triplex requires the ratio of GAA to TTC strand larger than 1:2. The structural features presented here and in other studies distinguish the FRDA (GAA)·(TTC) repeats from the fragile X (CGG)·(CCG), myotonic dystrophy (CTG)·(CAG) and the Huntington (CAG)·(CTG) repeats.

© 2000 Academic Press

Keywords: Friedreich's ataxia; trinucleotide repeat; parallel-stranded DNA; triplex; NMR

*Corresponding authors

Introduction

DNA has long been thought of as a genetic reservoir of information, being passive and stable through numerous cellular processes. However, this picture is changing. Since 1991, more than ten human neurological/muscular diseases have been associated with chain length expansions of trinucleotide repeats (CGG)_n·(CCG)_n, (CAG)_n·(CTG)_n, and (GAA)_n·(TTC)_n (reviewed by Pearson & Sinden, 1998a). Unstable disease-associated trinucleotide repeats (TRs) are found in various regions of the associated genes, such as 3' and 5'-untranslated regions, exons and introns. The length of these repeats is closely related to the onset and progression of clinical symptoms. Usually, the process of TR mutation runs through several generations, from stable to pre-mutation and ultimately full expansion. In recent years there have been intense efforts in understanding the molecular mechanism(s) of triplet repeat instability.

Friedreich's ataxia (FRDA) patients have an unstable repeat in the first intron of the X25 gene, which contains seven exons and encodes different protein isoforms (Campuzano *et al.*, 1996). The most common isoform is a protein called frataxin whose absence in affected patients can be caused by an expansion of the GAA tract. Among the TR expansion disorders, FRDA has very unique properties on the molecular level. Foremost, it is associated with (GAA)_n·(TTC)_n, which is presently the

†These two authors contributed equally to this work.
Abbreviations used: TR, trinucleotide repeat; FRDA, Friedreich's ataxia; ps, parallel-stranded; aps, antiparallel-stranded; NOESY, nuclear Overhauser enhancement spectroscopy; FM, fast migrating, DMS, dimethylsulfate.

E-mail addresses of the corresponding authors:
cepearson@genet.sickkids.on.ca; gao@kitten.chem.uh.edu

only non GC-rich triplet to be associated with expansion. Furthermore, it is the only disease-associated repeat located within an intron. Lastly, unlike all other diseases associated with TR expansions, FRDA is an autosomal recessive disease, requiring expansions (or mutations) at both alleles (Montermini *et al.*, 1997). The TR distribution of (GAA)·(TTC) lengths is mosaic (Cossee *et al.*, 1997), with the normal-size alleles containing 7-34 repeats (most frequent $n = 9$), the intermediate unstable alleles (pre-mutation) contain more than 34 repeats, and the expanded alleles (full mutation) contain from 200 to more than 1000 repeats. In addition, long GAA tracts interrupted by (GAG-GAA) $_n$ repeats have been shown to be genetically stable (Cossee *et al.*, 1997) with partial or complete suppression of disease expression.

The molecular mechanism of trinucleotide repeat instability is poorly understood. One hypothesis is that the trinucleotide repeats form a stable non-B-DNA structure(s), which can lead to aberrant DNA replication, repair, or recombination (Sinden & Wells, 1992; Richards & Sutherland, 1994; Samadashwily *et al.*, 1997; Jakupciak & Wells, 1999; reviewed by Pearson & Sinden, 1998b). Indeed, recent solution structure studies of single-stranded and cloned TRs suggest the formation of hairpins and duplexes (Smith *et al.*, 1995; Zheng *et al.*, 1996; Mariappan *et al.*, 1998), hairpins (Nadel *et al.*, 1995; Yu *et al.*, 1995a; Lian *et al.*, 1996; Pertruska *et al.*, 1996; Darlow & Leach, 1998; Suen *et al.*, 1999), triplexes (Ohshima *et al.*, 1996; Bidichandani *et al.*, 1998; Mariappan *et al.*, 1999), quadruplexes (Kettani *et al.*, 1995; Usdin & Woodford, 1995; Usdin, 1998), slipped structures (Pearson & Sinden, 1996; Pearson *et al.*, 1997, 1998a,b) and other unusual structures (Gao *et al.*, 1995; Mitchell *et al.*, 1995; Sakamoto *et al.*, 1999).

There has been accumulating evidence that the homo-purine/homo-pyrimidine (GAA)·(TTC) repeats of FRDA can form secondary structures such as intra or inter-allelic triplex of the RRY or YRY types (Hanvey *et al.*, 1988a,b; Ohshima *et al.*, 1996; Bidichandani *et al.*, 1998; Mariappan *et al.*, 1999; Sakamoto *et al.*, 1999). The formation of these triple-stranded structures may explain the observed *in vitro* replication pausing (Ohshima *et al.*, 1998) and stabilization of the inter-plasmid aggregate DNA (Sakamoto *et al.*, 1999). Presumably, these alternative DNA structures can also inhibit transcription of the *frataxin* gene, resulting in decreases in RNA transcripts (Campuzano *et al.*, 1996; Ohshima *et al.*, 1998) and/or in abnormal RNA processing (Bidichandani *et al.*, 1998).

To understand the mechanism of FRDA-associated (GAA)·(TTC) repeat instability it is important to understand the DNA structures formed by this repeat. In this report we analyzed structure formation by individual strands and complementary mixtures of (GAA) $_n$ and (TTC) $_n$ oligonucleotides. Under physiological conditions, through the use of NMR, UV-melting, and gel electrophoresis, as well as chemical and enzymatic probing methods, we

demonstrate that in addition to the expected anti-parallel duplex and YRY triplex, these sequences can form a parallel-stranded duplex.

Results

Sequence design

In this study we used two 13-mer sequences; (GAA) $_4$ G and C(TTC) $_4$ (abbreviated (GAA) and (TTC), respectively). These were studied both alone as individual strands and as complementary strand mixtures. The sequences have pseudosymmetry, and thus present no bias towards sequence alignment in either orientation in duplexes or triplexes. For instance, an antiparallel-stranded duplex (aps-duplex) (Figure 1(a)) involving A·T and G·C base-pairs and a parallel-stranded duplex (ps-duplex) (Figure 1(b)) of the same base-pairs may have an equal chance of forming without shifting base alignments. Similarly, in either the RRY triplex (Figure 1(c)) or the YRY triplex (Figure 1(d)), the third strand (R or Y) may be in an aps or a ps orientation relative to the R strand of the duplex (Sinden, 1994; Soyfer & Potaman, 1995). These two forms would also have an equal chance to compete. In all, the 13-mer sequences would potentially allow the formation of six possible structures in a linear form (Figure 1). Additional structures may be possible if hairpin formation is involved.

NMR and UV experiments

Single-stranded GAA and TTC

The formation of possible structures was monitored by ^1H one-dimensional (1D) NMR spectroscopy. The individual strands were first studied at pH 6.4 in NMR buffer (see Materials and Methods). Temperatures were varied between 0 and 25°C. At all temperatures the one-dimensional spectra of the C(TTC) $_4$ strand contained only a broad hump at ~10.8 ppm (Figure 2(a)) and no resonances downfield from this point. In contrast, the (GAA) $_4$ G strand exhibited a clear temperature transition: at room temperature the spectra revealed broad resonances at 10.5 ppm (data not shown) while at low temperature (below 10°C) ^1H NMR signals centered around 9.5 and 11 ppm appeared (Figure 2(b)). The position and chemical shift dispersion showed that these are hydrogen bonded (H-bonded) protons from G NH (imino H) and G/A NH $_2$ (amino H) resonances (Maskos *et al.*, 1993). The temperature transition of the (GAA) $_4$ G strand suggested the formation of an intra-strand structure. To further investigate this the spectra of the shorter (GAA) $_2$ and (GAA) $_3$ strands were also investigated. At all temperatures the spectra for the (GAA) $_2$ and (GAA) $_3$ were similar to the spectra for the (GAA) $_4$ G at room temperature, showing broad resonances at 10.5 ppm and no resonances at further downfield spectral region (data not shown). This indicates a length-dependence of

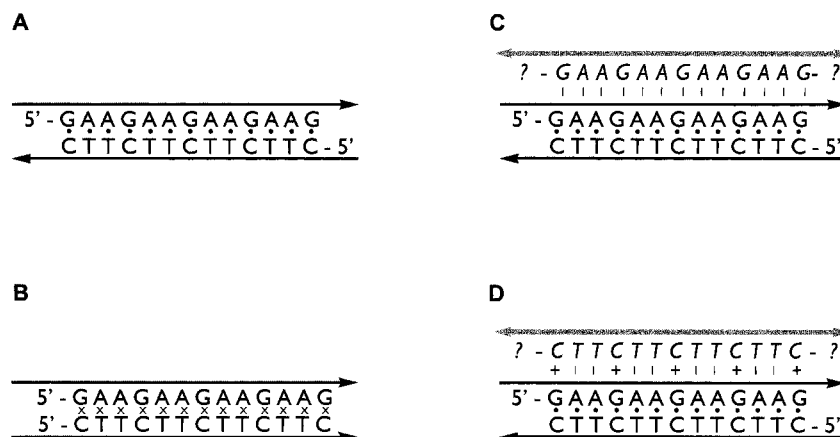


Figure 1. Possible secondary structures formed by inter-strand interactions of (GAA)₄G and C(TTC)₄. (a) Anti-parallel-stranded duplex, (b) parallel-stranded duplex, (c) RRY triplex, and (d) YRY triplex. The thin black arrows indicate the duplex strand orientation from 5' to 3'. The bold double-headed arrows indicate that the third strand orientation is undetermined. The ·, ×, |, and +, respectively represent standard Watson-Crick base-pairs, reverse Watson-Crick base-pairs, Hoogsten base-pairs and Hoogsten base-pairs involving protonation of cytosine residues.

intra-strand GAA structure formation. UV melting of the isolated (GAA)₄G and C(TTC)₄ sequences were recorded from 0-90 °C. The 13-mer TTC measured nearly null hyperchromicity, while the 13-mer GAA exhibited hyperchromicity of larger

than 10% but no clear transition point (data not shown). These results further support the formation of an intra-strand structure by the GAA and not by the TCC strand.

Titration experiments

The (GAA)₄G strand was mixed with increasing amounts of its complementary strand C(TTC)₄, resulting in various final ratios of (GAA):(TTC) (unless otherwise indicated, the ratio refers to that of (GAA):(TTC) strands) as illustrated on the top of Figure 3. The formation of possible structures in the mixtures was monitored by 1D ¹H NMR at 10 °C and pH 6.4. Upon addition of the TTC strand, the NMR spectra showed simultaneous appearance of the base-paired NH (12-14 ppm) and disappearance of the NH resonances of the free ssGAA at 10 ppm. The final spectrum of the titration at the 1:1 ratio of (GAA):(TTC) is shown in Figure 3(a). Upon addition of excess TTC (a 1:2 stoichiometric ratio of (GAA):(TTC), Figure 3(b)), the duplex resonances were unchanged while an additional set of resonances were observed at 9-10 and 12-15 ppm (Figure 3(b)). These signals are characteristic of NH in protonated C (C⁺, downfield from 14 ppm) and C⁺ NH₂ (8-10 ppm) (Radhakrishnan *et al.*, 1991). Interestingly there was also the appearance of a broad hump at 10.8 ppm, which corresponded to the non-H-bonded NH observed in ssTTC (Figure 2(a)). In order to drive the hybridization to completion, the (GAA):(TTC) ratio was increased to 1:4 (Figure 3(c)). At this ratio, the resonances associated with the 1:1 ratio disappeared and only the resonances associated with the 1:2 ratio and ssTTC were observed.

To probe whether the 2:1 (GAA):(TTC) triplex could form, a 4:1 (GAA):(TTC) sample was examined at 10 °C. The spectra were otherwise identical

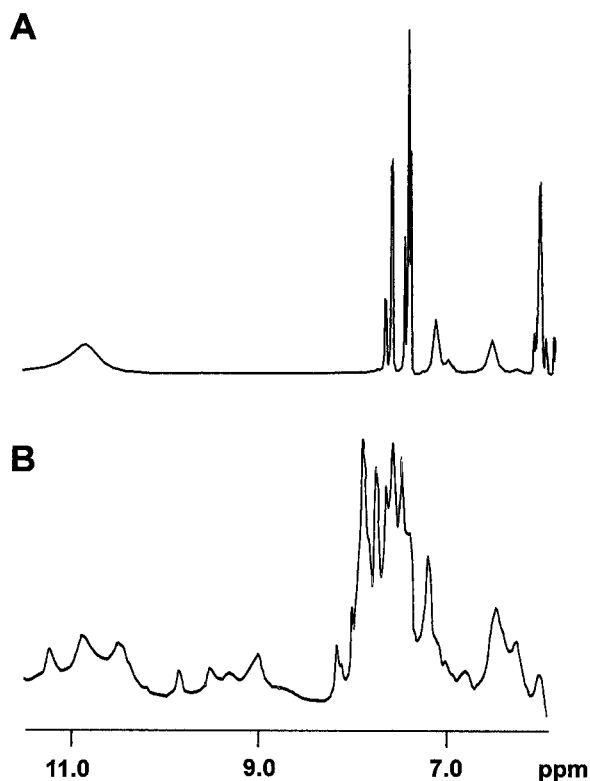


Figure 2. 1D NMR spectra of individual C(TTC)₄ and (GAA)₄G strands in H₂O at 0 °C. The sugar proton region (6-9 ppm), and amino proton regions (9-11.5 ppm) are shown for (a) C(TTC)₄, (b) (GAA)₄G. Salt concentrations are 90 mM KCl, 10 mM NaH₂PO₄, 0.1 mM EDTA (pH 6.4).

NMR Titration

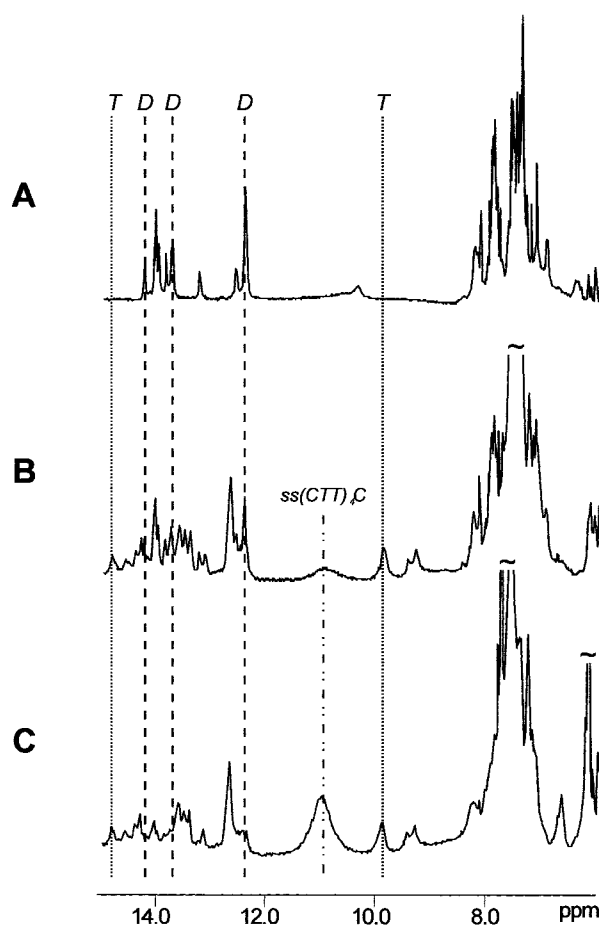


Figure 3. 1D NMR spectra of different (GAA)₄G and C(TTC)₄ strand ratios in H₂O at 10°C. The sugar proton region (6-9 ppm), and imino and amino proton regions (9-15 ppm) are shown for (a) 1:1 (GAA):(TTC), (b) 1:2 (GAA):(TTC) and (c) 1:4 (GAA):(TTC). Vertical lines labeled D and T, respectively represent typical duplex and triplex resonances. The line labeled ssC(TTC)₄ indicates the amino resonance of single-stranded C(TTC)₄ not involved in inter-strand interactions. Salt concentrations are 90 mM KCl, 10 mM NaH₂PO₄, 0.1 mM EDTA (pH 6.4).

with that of 1:1 (GAA):(TTC) (Figure 3(a)), except for additional signals consistent with those observed for ssGAA. The spectra were not affected by pH (pH 5 to 7). The effect of Mg²⁺ on the 4:1 (GAA):(TTC) sample was examined by increasing MgCl₂ concentration from 0 to 2.5 mM at pH 5.9. The NMR spectra of the Mg²⁺-containing samples showed broadened resonances but their chemical shifts were essentially unaffected by added Mg²⁺.

Characterization of the 1:1 and the 1:2 (GAA):(TTC) samples

Both the 1:1 and 1:2 (GAA):(TTC) samples were characterized using NOESY spectra. While detailed structure elucidation was beyond the scope of this study, the purpose of these experiments was to identify whether these samples displayed spectral

patterns typical of canonical duplexes or YRY triplexes.

The NOESY spectrum of the 1:1 sample revealed intra and inter-residue connectivities as expected in a canonical antiparallel-stranded duplex. This is illustrated in an expanded spectral region involving interactions from base (H6/H8) to sugar H1' protons (Figure 4). Although extensive overlaps exist in this sample containing repeat sequences, careful analysis by standard methods (Wüthrich, 1986) allowed derivation of important information. The assignment for the GAA strand began with the terminal G₁ and G₁₃ residues, which were well resolved, allowing the assignments of G₁-A₂-A₃-G₄ and A₁₂-G₁₃ connectivities. Since there was a clear repeating pattern in the spectrum corresponding to the triplet repeats in the sequence (consider the three types of resonances: G, A (3' to G) and A

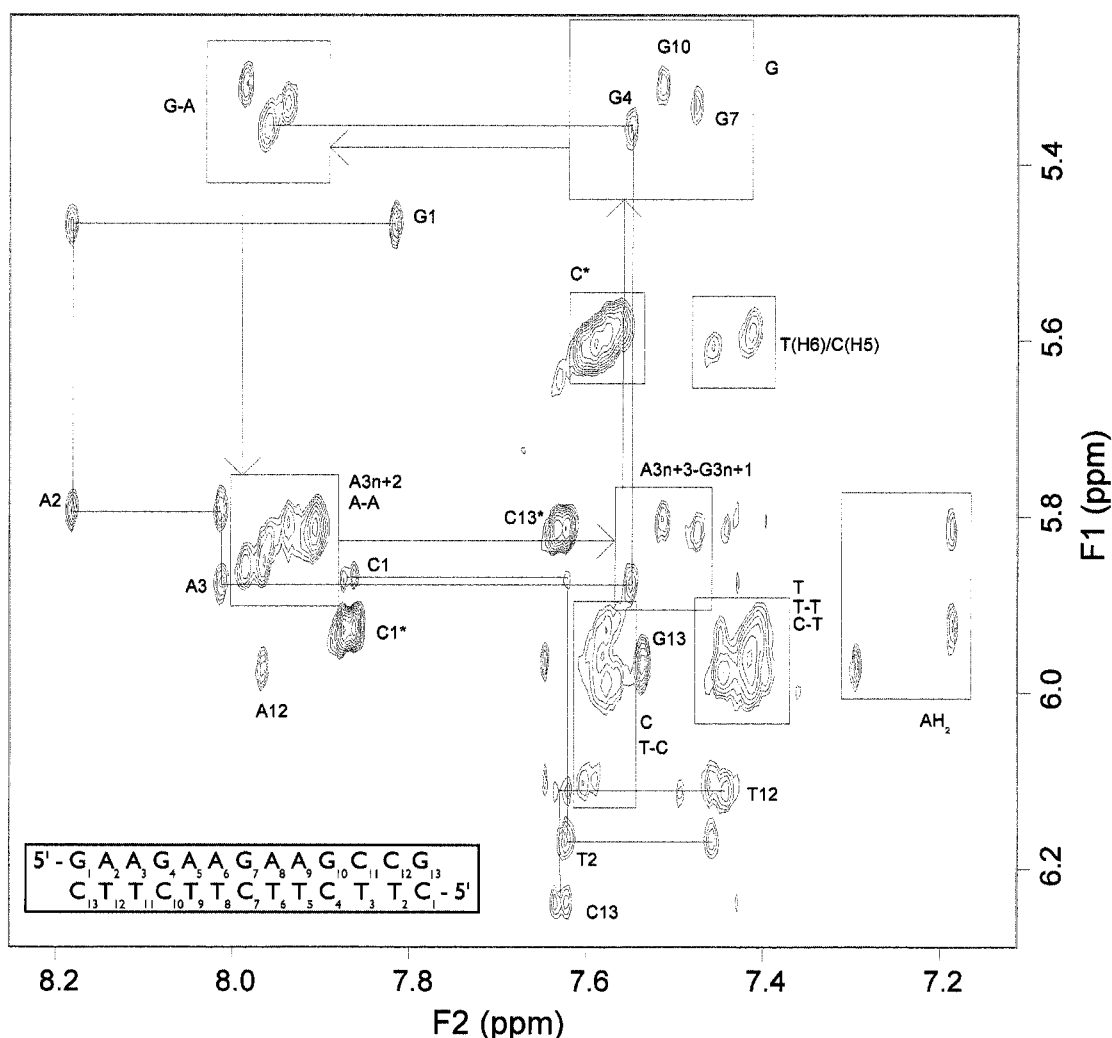


Figure 4. 2D NOESY spectra of 1:1 (GAA)₄G:C(TTC)₄ in ²H₂O recorded at 20 °C with a 250 ms mixing time. An expanded region is shown representing the interactions between H1' sugar protons (F1 dimension) and H6/H8 base protons (F2 dimension). Assigned peaks are indicated by their residue number and, overlapped and/or equivalent cross-peaks are boxed and labeled. Arrows indicate sequential assignments, from 5' to 3' for the GAA strand. C* indicate C H5/C H6 interactions.

(5' to G) and three sequential steps: G-A, A-A, and A-G), these assignments provided leads for the assignments of other resonances. The intra-residue G interactions (G_{3n+1} H8-G_{3n+1} H1') ($n = 1, 2$ or 3) and half the A intra-residue interactions were identified; the remaining resonances of H8 and H1' protons were then assigned according to their types (indicated by boxes in Figure 4). The G H8 resonances were clustered around 7.5 ppm and A H8 were clustered around 8.0 and 7.9 ppm, respectively for the A 3' to G and the A 5' to G, respectively. Following these qualitative assignments, the sequential connectivities at the G-A (G_{3n+1} H1'/A_{3n+2} H8) and A-G (A_{3n+3} H1'/G_{3n+1} H8) were found (Figure 4). The sequential NOE connectivities at the A-A step (A_{3n+2} H1'/A_{3n+2} H8) were only partly distinguishable, since these resonances were overlapped with the A intra-residue NOEs. The NOE cross-peaks of the

C(TTC)₄ strand were analyzed in a way similar to those of the (GAA)₄G strand and the NOEs involving the four terminal residues (C₁, T₂, T₁₂ and C₁₃) were assigned. Although extensive overlaps of the NOE cross-peaks existed for the TTC strand (Figure 4), spectral pattern recognition based on the assignments of the terminal residues provided qualitative information for the intra and inter-residue NOE connectivities (Figure 4).

The 1:2 sample NOESY recorded in ²H₂O contained a set of NOE cross-peaks identical with those detected in the 1:1 sample. A few additional weak cross-peaks were detected but their assignments were not possible due to the uncertainty in assigning the majority of resonances. However, the most informative results were derived from NOESY recorded in H₂O. The spectrum showed cross-peaks connecting NH₂⁺ at 9.5 ppm to sugar protons at 7.5 ppm (data not shown). These are

inter-strand proton-proton contacts characteristic of ps-YRY triplexes (Radhakrishnan *et al.*, 1991).

^{31}P NMR spectra of the 1:1, the 1:2 and the 1:4 (GAA):(TTC) samples were recorded (data not shown). These data showed narrowly dispersed resonances within 0.7 ppm, suggesting no major distortions in phosphate backbones in the structures present under the experimental conditions. This indicates that there were no gross backbone alterations (as would be expected for hairpin or hairpin-like structures).

UV spectroscopy

The same samples used for the NMR experiments were also used to study the UV melting behavior of the DNA structures. As noted above, the profile of the free ss(TTC) is remarkably flat ($R = 0.00$) indicating a random coil structure. This is in contrast to the free ss(GAA) profile, which has a hyperchromicity of 0.15, indicating the formation of a defined structure. Figure 5 displays the UV melting profiles of the 1:1, 1:2 and 1:4 (GAA):(TTC) samples measured at pH 6.4. Since isolated C(TTC)₄ hardly showed hyperchromicity, its con-

tribution in the 1:2 and 1:4 samples was ignored. The 1:1 sample presented a single transition ($t_m = 37^\circ\text{C}$, $\Delta G = -8.0$ kcal/mol), whereas the 1:4 (GAA):(TTC) ratio sample displayed two successive transitions ($t_{m,1} = 24^\circ\text{C}$, $\Delta G_1 = -5.0$ kcal/mol and $t_{m,2} = 41^\circ\text{C}$, $\Delta G_2 = -8.0$ kcal/mol). The two transitions of the 1:4 sample were characteristic of a triplex structure (Howard *et al.*, 1964; Blake & Fresco, 1966; Riley *et al.*, 1966; Morgan & Wells, 1968; reviewed by Felsenfeld & Miles, 1968), in agreement with the NMR data of this sample. The melting curve of the 1:2 sample appeared to be intermediate between the 1:1 and 1:4 samples with the first transition less pronounced. The hyperchromicity increased from the 1:1 ($R = 0.155$) to the 1:4 (GAA)·(TTC) ratio ($R = 0.325$).

Gel electrophoresis experiments

pH, strand ratio and Mg^{2+} studies

Polyacrylamide gel electrophoresis was performed on the same samples in parallel with the above NMR and UV experiments. Figure 6 shows the native gel profile of different (GAA):(TTC)

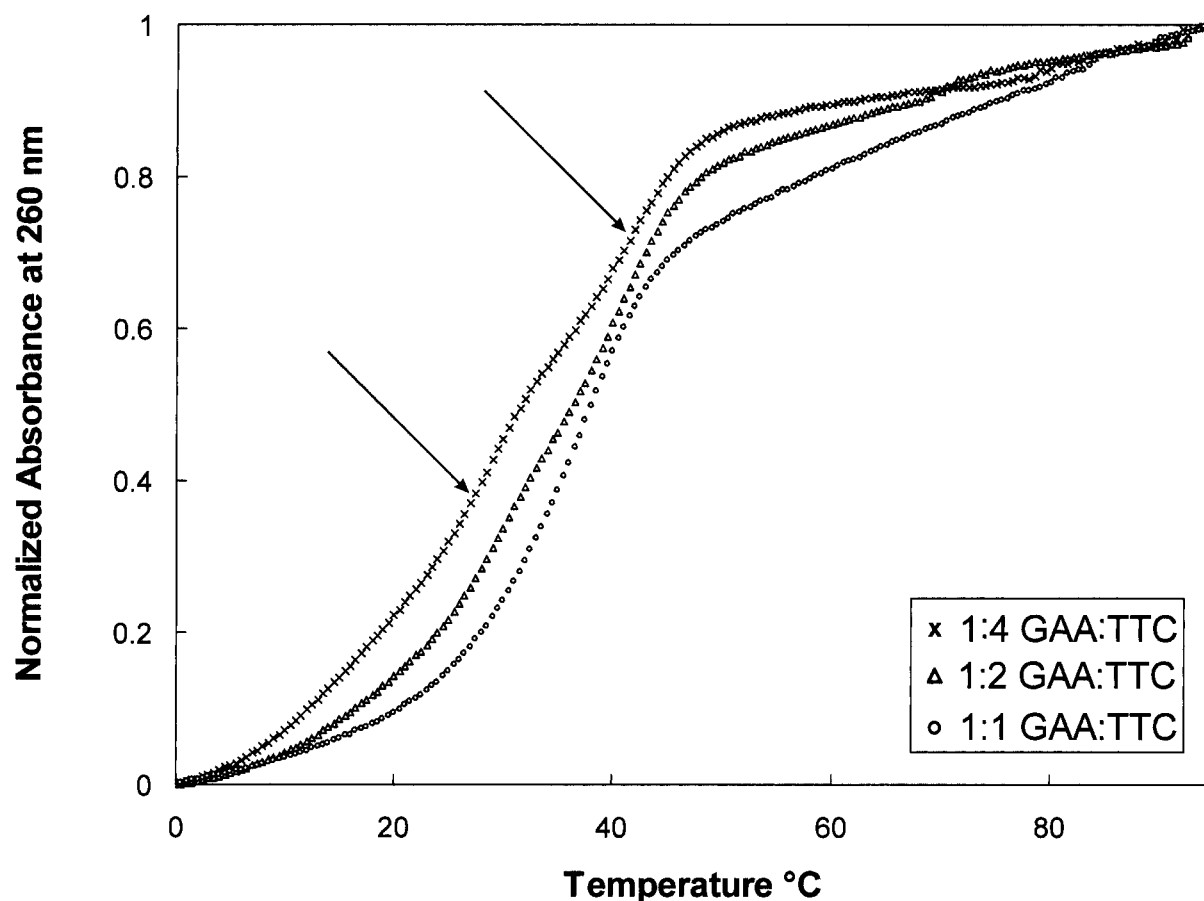


Figure 5. UV melting experiment at 260 nm. The melting curve, from 0 to 95°C with a $0.5^\circ\text{C}/\text{minute}$ gradient, is represented for the three samples of Figure 2, pH 6.4. The hyperchromicity is respectively 0.24, 0.26 and 0.24 for 1:1, 1:2 and 1:4 (GAA):(TTC) strand ratio. The hyperchromicity of the melting profiles was about 24%. t_m (melting temperature) values were derived according to published methods (Zheng *et al.*, 1996).

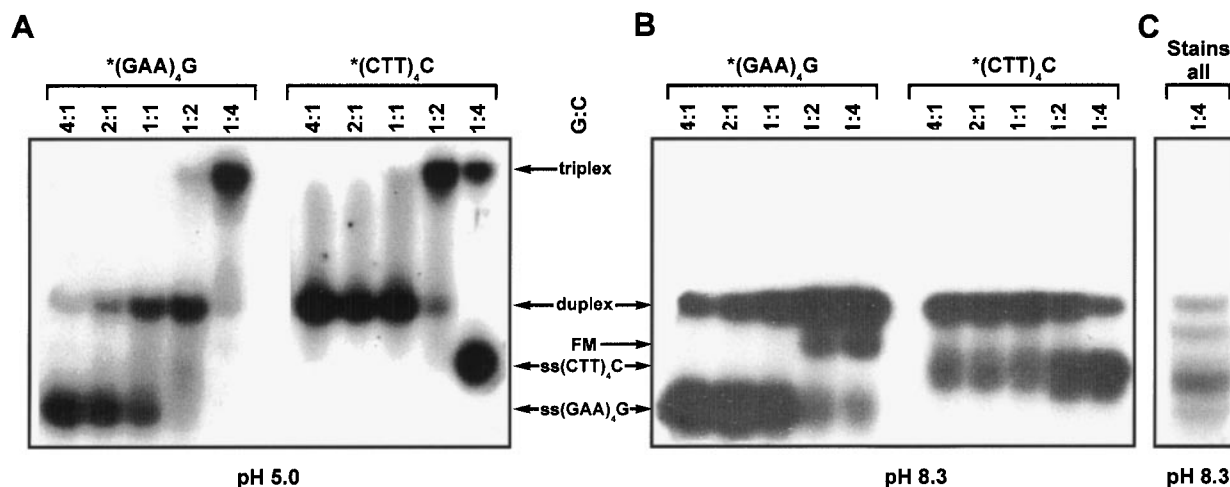


Figure 6. Native 20% polyacrylamide gels of different (GAA):(TTC) strand ratios at constant DNA concentration, hybridized at 20°C and electrophoresed at 4°C. The hybridization and electrophoresis running buffers were: (a) pH 5.0 in TAE; (b) and (c) pH 8.3 in TBE. (a) and (b) Autoradiograms where the * indicates the ^{32}P -labeled strand, and (c) was monitored by Stains-All (Sigma). The DNA concentration is 2 μM in (a) and (b) and 300 μM in (c).

ratios (4:1, 2:1, 1:1, 1:2, and 1:4). At pH 5.0 and 8.3, ssGAA migrated faster than ssTTC (Figure 6(a) and (b), compare 4:1 with 1:4 (GAA):(TTC)). At pH 5.0, increasing the relative proportion of TTC from 4:1 to 1:4 resulted in a decreased amount of the duplex (Figure 6(a)). A slower migrating band appeared only at pH 5.0 and at ratios of 1:2 and 1:4 (Figure 6(a)); this ratio, pH-dependence, and slow electrophoretic migration is characteristic of triplex DNA structures (Soyfer & Potaman, 1995). Triplex formation was complete at a 1:4 (GAA):(TTC) ratio in which free ss(TTC) was still present (Figure 6(a), compare the 1:4 ratios, where the $^{*}(\text{GAA})_4\text{G}$ or the $^{*}\text{C}(\text{TTC})_4$ strands are radiolabeled, as indicated by the *). Triplex formation was never observed at pH 8.3 (Figure 6(b)).

At a pH of 8.3 a fast migrating (FM) product was observed when (GAA):(TTC) ratios were larger than 1:1 and when the ^{32}P was on the GAA strand (Figure 6(b)). Increasing amounts of the FM species were observed with increasing amounts of the TTC strand, with relative amounts of the FM:duplex at 20:80 for the 1:2 and 30:70 for the 1:4 strand ratios. This FM species could not be detected autoradiographically when the ^{32}P was on the TTC strand (Figure 6(b)). The FM species could be detected by Stains-All (Sigma), along with both free single strands and the Watson-Crick duplex (Figure 6(c)). The formation of the FM product absolutely required the presence of both (GAA) and (TTC) strands and was independent of DNA concentration (40 nM to 400 μM) (data not shown). However, the FM species did not appear to contain the (TTC) strand (Figure 6(b), compare the 1:2 and 1:4 ratios where the $^{*}(\text{GAA})_4\text{G}$ or the $^{*}\text{C}(\text{TTC})_4$ strands are radiolabeled). While its formation was observed only in mixtures of the two complementary strands, the FM species seemed to contain only the (GAA) strand, as it was observed by auto-

radiography only when the (GAA) was radiolabeled (Figure 6(b)).

Magnesium ions can affect DNA conformation and alternative structure formation (Sinden, 1994; Soyfer & Potaman, 1995). The addition of 20 mM MgCl_2 during the hybridization and during electrophoresis did not enhance the formation of triplex or of the FM species (data not shown). However, the inclusion of MgCl_2 did alter the relative migration of the ss(GAA) and ss(TTC) such that they co-migrated (data not shown). Also, the FM band disappeared and the duplex formation was enhanced, indicating an effect of Mg^{2+} on the equilibria of structure formation. Spermine can affect DNA triple-stranded structure formation (Sinden, 1994; Soyfer & Potaman, 1995). The addition of spermine (300 μM) during the hybridization did not alter the observed electrophoretic pattern of any of the products (data not shown).

FM component analysis by ^{32}P -relabeling

The fact that the formation of the FM species required the presence of both the (TTC) and the (GAA) strand but was visible only when the (GAA) strand was radiolabeled presented somewhat of a conundrum: Why was the FM species not observed autoradiographically when the (TTC) strand was radiolabeled? We performed the same experiment as that of Figure 6(b) with a 1:4 ratio of radiolabeled (^{32}P) $^{*}(\text{GAA})_4\text{G}$ strand to a non-radioactively phosphorylated (^{31}P) $\text{C}(\text{TTC})_4$ strand (Figure 7(a)). The presence of a 5' phosphate on the $\text{C}(\text{TTC})_4$ strand inhibited the formation of the FM species. This result agrees with the inability to detect the FM species when the $\text{C}(\text{TTC})_4$ strand is radiolabeled with a 5' phosphate (compare Figure 7(a) with Figure 6(b)). To determine the strand composition of the novel FM species,

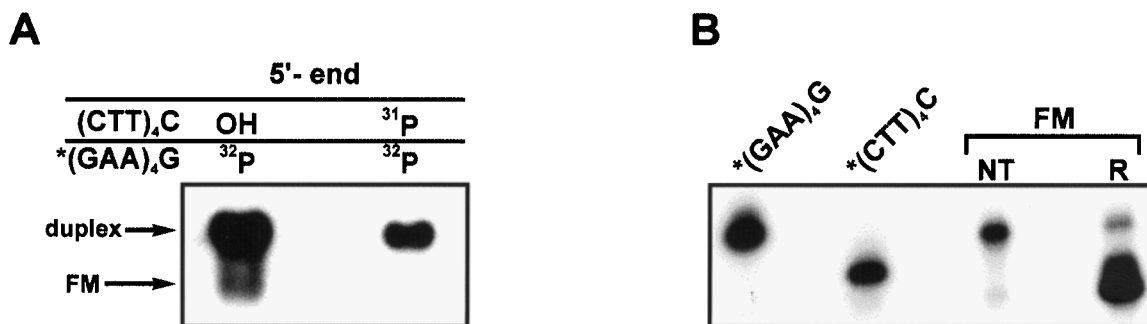
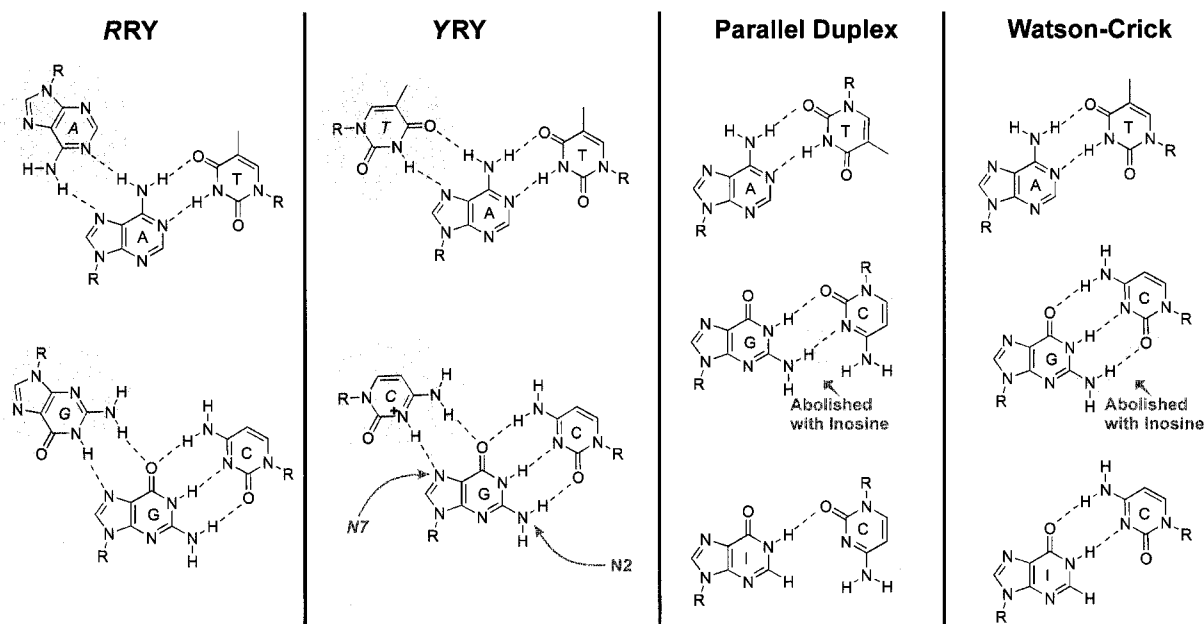


Figure 7. Role of 5' phosphate in the (TTC) strand and strand composition on the FM species. (a) Hybridization of [³²P]phosphate-labeled (GAA)₄G strand with C(TTC)₄ strand kinased at the 5' end with non-radioactive [³¹P]phosphate. The * indicates the ³²P-radiolabeled strand. Strands were hybridized at a 1:4 ratio and analyzed on a native 20% polyacrylamide gel as in Figure 6(b). (b) 8 M urea, 20% denaturing polyacrylamide gel analysis of gel-purified FM subjected to: NT, no treatment, and R, a second round of labeling with [³²P]ATP by T4 polynucleotide kinase. The individual labeled single-stranded (GAA)₄G and C(TTC)₄ are included as reference.

preparative quantities of FM were made and gel-purified. An aliquot of the gel-purified FM species was heat-denatured and subjected to a second round of radiolabeling on the 5' end. These samples were analyzed on a sequencing gel to compare the gel-purified FM species before and after the secondary labeling (Figure 7(b)). Taking advantage of the differential migration of the single strands on a sequencing gel, Figure 7(b) shows that the relabeled FM is a mixture of both complementary single strands. Therefore, the FM species is composed of both the (GAA) and the (TTC) strands but is limited to only (TTC) strands without a 5'-phosphate. That the FM species can be formed with the (GAA) strand regardless of its 5' status is evidenced by its detection both by autoradiography [³²P]phosphate) (Figure 6(b)) and by Stains-All (no phosphate) (Figure 6(c)).

Chemical probing of the base-pairing of the FM species

G to I (inosine) substitutions in the GAA strands (G7/I7 in one duplex) (Figure 8) and G4/I4 and G7/I7 in a second duplex (not shown) were performed to provide evidence of the possible parallel duplex nature of FM. Inosine, which lacks the N2 amino group, is an analog of G. As a result, parallel duplex formation is largely destabilized due to the instability of I·C parallel base-pairs (Scheme 1). By contrast, the antiparallel I·C Watson-Crick base-pair is only slightly destabilized since two hydrogen bonds are still present. Therefore, I substitution is expected to destabilize the formation of the FM band more than the antiparallel duplex. It was found experimentally that a single I substitution of the central G7 residue was sufficient



Scheme 1. Base-pairs discussed in the text. The third strand bases in the YRY and RRY triplexes are bound to the purine strand of the duplex by Hoogsteen base-pairing. Broken lines indicate H bonds; R, sugar moiety.

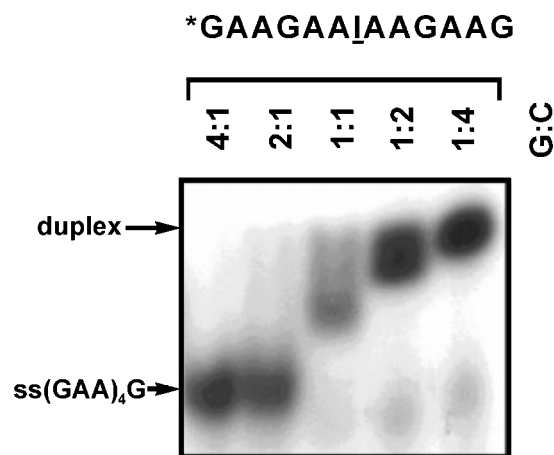


Figure 8. Chemical probing, effect of G N2 positions probed by inosine substitution. Native 20% polyacrylamide gels of different $^*(\text{GAAGAAIAAGAAG})$:(TTC) strand ratios. Conditions are similar to Figure 6(b), where $^*(\text{GAA})$ strand is ^{32}P -labeled: constant 2 μM DNA concentration, hybridized at 20 °C and electrophoresed at 4 °C in TBE (pH 8.3).

to abolish the FM formation (Figure 8), thereby supporting the notion that the FM species was in fact a parallel duplex. (Substitution of both G4 and G7 with I had an identical effect.) At 1:4 G*:C ratio, only the antiparallel duplex band is present when the GAA strand contains I substitution(s), while both the antiparallel duplex and the FM band are present when the GAA strand contains no substitutions (Figure 6(b)). Although a lesser effect, the destabilization of the antiparallel duplex by the I substitution in the GAA strand is clearly detectable in these gel electrophoresis experiments. As the ratio of the two single-stranded DNAs varied from G*:C (or G:C*) 1:4 to 4:1, the duplex to single strand transition was observable from the gradual disappearance of the duplex band and appearance of an average band migrating between the duplex and the labeled single strand. This reflects an increased melting rate of the duplex in relation to gel migration time. In the experiments using unmodified GAA strand, the duplex and single strands were observed as isolated bands, indicating the melting of the duplex was very slow during gel migration.

Chemical modification of N7 positions of G residues with dimethylsulfate (DMS) was performed to determine if Hoogsteen base-pairing contributed to the DNA structures. DNA structures involving Hoogsteen base-pairs are destabilized by DMS methylation of the G N7 position. Parallel-stranded duplex DNA is not expected to be affected by DMS methylation, since there are interactions only between Watson-Crick sites (Scheme 1). The classical approach to chemical modification consists of treating the already formed structure with DMS in a protection-type experiment. However, our

approach was to first methylate to completion all G N7 positions of the (GAA) single strand, and then study the products of its hybridization with the (TTC) single strand. As expected, at pH 5.0 the methylation abolished triplex formation by preventing Hoogsteen base-pair formation (Figure 9(a)). However, methylation did not inhibit FM formation at either pH 5.0 (Figure 9(a)) or pH 8.3 (Figure 9(b)). In fact, methylation favored the FM structure relative to the antiparallel form: following DMS treatment, the relative amounts of the FM parallel species relative to the antiparallel species was 50:50 and 70:30 at pH 5.0 and pH 8.3, respectively (compare this to the unmethylated DNAs having relative amounts of 0:100 and 30:70 at pH 5.0 and pH 8.3, respectively) (Figure 9(a) and (b)). Thus, in the FM structure, there are no hydrogen bonds involving G N7 positions, consistent with a parallel-stranded duplex structure.

A characteristic difference between antiparallel and parallel-stranded DNAs is their differential ability to bind the drug BBI-258 and their similar ability to bind ethidium bromide (Ramsing & Jovin, 1988; Jovin *et al.*, 1990). These were used as staining agents in order to elucidate the structural characteristics of the FM species. We were unable to detect binding of either BBI-258 or of ethidium bromide to any of our (GAA)·(TTC) species (data not shown). The lack of binding by BBI-258 may be due to its requiring a stretch of four consecutive

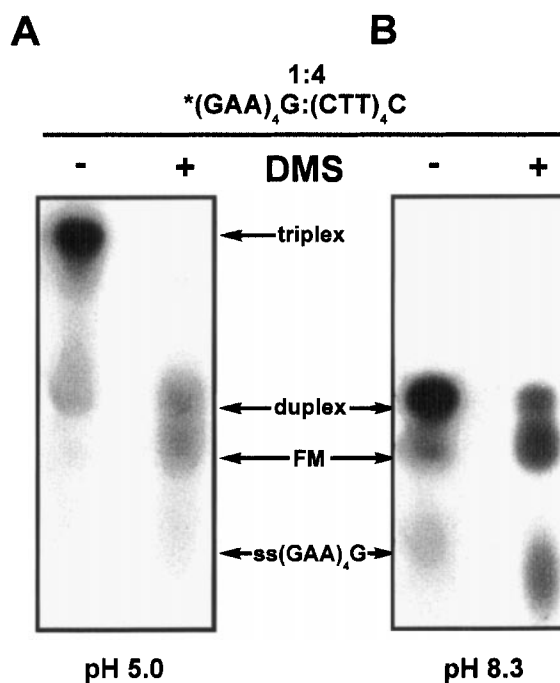


Figure 9. Chemical probing, effect of G N7 positions. Native 20% polyacrylamide gels of 1:4 (GAA):(TTC) with (+) or without (-) DMS treatment: (a) pH 5.0 in TAE; (b) pH 8.3 in TBE. Prior to hybridization with $(\text{TTC})_4$ the $(\text{GAA})_4\text{G}$ strand was treated with DMS to methylate N7 positions of G residues.

A·T residues (Wang & Teng, 1990), which is absent in our sequences. The inability to detect binding of ethidium bromide by (GAA)·(TTC) may be specific to these sequences, and may not necessarily be due to reduced binding but to abolished fluorescence (Latimer & Lee, 1991).

Enzymatic probing of the structure of FM

Parallel-stranded DNAs exhibit substrate properties which differ from regular B-DNA antiparallel duplexes (Rippe & Jovin, 1989). Parallel-stranded DNAs are insensitive to *Escherichia coli* exonuclease III (*Exo* III) due to the absence of a distinct minor or major groove. We determined the *Exo* III sensitivity of the various structural species in a 1:4 (GAA):(TTC) mixture at pH 5.0 (Figure 10(a)) and pH 8.3 (Figure 10(b)). At pH 5.0 the slow migrating triplex was completely digested by the addition of low levels of *Exo* III, whereas the duplex was resistant (Figure 10(a), see 1× and 10× concentration). As expected, higher *Exo* III concentrations also digested the duplex form (Figure 10(a), see 100× and 1000× concentration). For samples hybridized at pH 8.3 (Figure 10(b)), the FM species was strikingly resistant to *Exo* III digestion, even at concentrations which completely digested the antiparallel duplex form (Figure 10(b), compare relative sensitivities of the duplex and FM species at 10× and 100× *Exo* III concentrations). The resistance to *Exo* III digestion of the FM species is consistent with a parallel duplex DNA (Rippe & Jovin, 1989).

Discussion

Using NMR, UV and gel electrophoresis, we have examined both (GAA)_n and (TTC)_n strands

individually or combined at (GAA):(TTC) ratios of 1:1-1:10 at pH 5.0-8.3. These studies have identified the following structural characteristics of GAA and CTT repeats: (a) the isolated (GAA)_n is capable of self-association to form meta-stable structure(s) at a chain length longer than $n = 3$. ssGAA is most likely not a hairpin; (b) C(TTC)₄ does not exhibit low-energy, self-associated forms; (c) mixtures of 1:1 (GAA):(TTC) form a stable aps-duplex at all pHs and temperatures used in this study; (d) mixtures of 1:2 (GAA):(TTC) form a YRY triplex, which has a transition point at pH ~ 7.0 and is more stable at lower pH; (e) the RRY triplex, containing 2:1 (GAA):(TTC) is too unstable to form; (f) limited screening of the effect of Mg²⁺ found no evidence that the ion has a special effect on any structure formation (except for a small effect on the electrophoretic migration of ssGAA); and (g) most interestingly, gel electrophoresis, and chemical and enzymatic probing identified a parallel-stranded duplex formed between one GAA and one TTC strand, which does not contain a 5'-PO₄. This ps-duplex migrates as a fast migrating (FM) species on a polyacrylamide gel.

Single-stranded oligonucleotides

The 13-mer C(TTC)₄ sequence exhibited no UV hyperchromicity, no H-bonding by NMR and faster gel migration than a 12-mer duplex marker. These results clearly establish that the ssTTC sequence does not form any stable or meta-stable structures. Although antiparallel T·T and C·C base-pairing are possible mismatches (Saenger, 1984) and could allow ssTTC to align to form an antiparallel duplex, there is no evidence in this

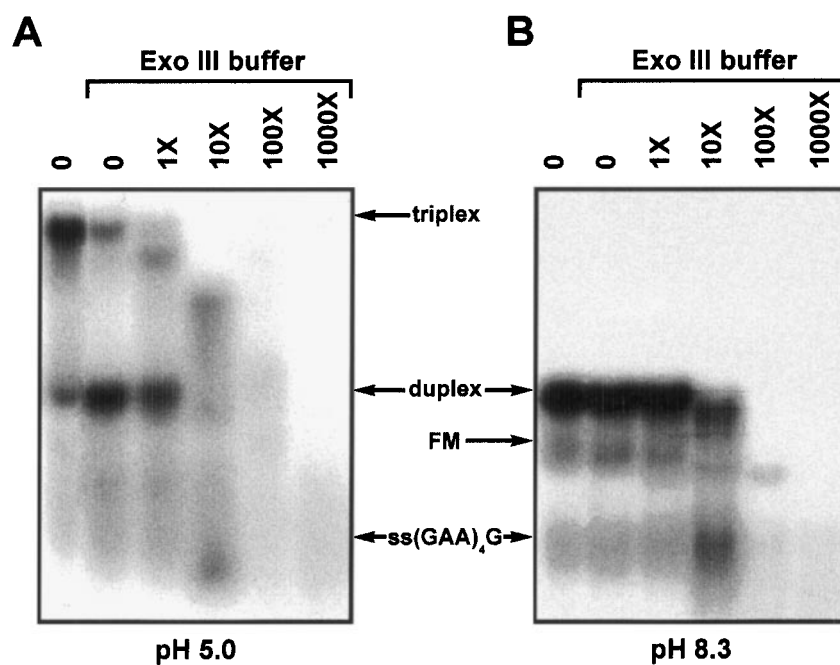


Figure 10. Exonuclease III experiment. Native 20% polyacrylamide gels of 1:4 (GAA):(TTC) after 25 minutes reaction with different *Exo* III concentration at: (a) pH 5.0 in TAE; (b) pH 8.3 in TBE. Where indicated, *Exo* III buffer is added to a final concentration of 66 mM Tris (pH 8.0) and 5 mM MgCl₂. The 1× concentration of enzyme corresponds to 0.2 unit/nmol of DNA. In every lane, only the (GAA)₄G strand is labeled.

study to support the formation of this structure by the ssTTC strands.

In contrast, the 13-mer (GAA)₄G exhibited different UV, NMR and electrophoretic features. The UV melting of (GAA)₄G was broad and displayed a hyperchromicity of >10%, indicating extensive base stacking in the sequence. 1D NMR experiments (at 10°C) showed, in particular, imino and amino proton resonances at 11 and 9–10 ppm, respectively (resonance assignment by 2D NOESY in H₂O, data not shown). The chemical shift dispersion and their downfield resonance frequency indicated that these functional groups are involved in H-bonding. The native gel mobility of (GAA)₄G is faster than the C(TTC)₄, although the mass of the (GAA)₄G is larger and its mobility is expected to be slightly slower or the same as the C(TTC)₄. Under denaturing conditions, ssGAA indeed migrates slower than ssTTC. Thus, the fast gel mobility of (GAA)₄G under non-denaturing conditions may be associated with the formation of somewhat more compact structure by GAA repeats. This assessment was further supported by the Mg²⁺ sensitivity of (GAA)₄G on gel electrophoresis, at pH 8.3. The gel mobility of (GAA)₄G was also in line with what we have demonstrated previously for the (CGG)_n, (CCG)_n, (CAG)_n, and (CTG)_n TRs: the structured homo-stranded multiplexes tend to migrate faster than the non-structured single-stranded sequences (Zheng *et al.*, 1996).

Shorter (GAA)_n repeats ($n = 2$ and 3) were also examined by NMR. Unlike the 13-mer (GAA)₄G, the 1D NMR spectra of these sequences contained only a broad exchangeable resonance, clearly H-bonds were not formed. The difference in the NMR spectral appearance of (GAA)_n sequences of various chain lengths suggest two points: (a) the (GAA)_n of four to six repeats are unlikely to form hairpins; and (b) there is a chain-length dependence on the biophysical stability of self-structured GAA. These properties of the (GAA)_n translate into a higher energy barrier for the (GAA)_n strand to self-fold than that for other disease-related TRs such as (CGG)_n, (CCG)_n, (CAG)_n, and (CTG)_n.

Other triplet repeat sequences (including (CGG)_n, (CCG)_n, (CAG)_n, and (CTG)_n repeats) form both inter-molecular (same strand) homoduplexes and intra-molecular hairpins (reviewed by Gao *et al.*, 1998; Pearson & Sinden, 1998b). Our previous studies demonstrated that these repeats of similar lengths (i.e. $n = 2$ and 3) form antiparallel same-strand homo-duplexes whose high-resolution structures have been elucidated (Gao *et al.*, 1998). On native polyacrylamide gels the migration of inter-molecular duplexes is characteristically slower than that of their corresponding Watson-Crick duplexes (Chen *et al.*, 1995; Petruska *et al.*, 1996; Mariappan *et al.*, 1998), whereas the migration of the intra-molecular hairpins is faster than that of their complementary base-paired duplexes (Mitchell *et al.*, 1995; Mitas *et al.*, 1995a,b; Yu *et al.*, 1995a,b; Zheng *et al.*, 1996; Nadel *et al.*,

1995; Petruska *et al.*, 1996). Using isolated GAA or TTC oligonucleotides we were unable to detect either inter-molecular (same-strand homoduplexes) or intra-molecular hairpins. This contrasts with the finding by Mitas and colleagues (Suen *et al.*, 1999), who reported the formation of hairpins by single-stranded GAA oligonucleotides at 5°C but not at >25°C. Our inability to detect hairpin formation by the GAA strand is likely due to differing experimental conditions. However, in hybridization mixtures of the two complementary strands we demonstrated here that (GAA)₄G and C(TTC)₄ triplet repeat sequences are able to form various structures including the Watson-Crick antiparallel duplex, triplex DNA and parallel-stranded DNA.

Triple helix formation

Two types of triplex, YRY and RRY, could potentially form between the GAA and TTC trinucleotide repeats (Figure 1 and Scheme 1). The characterization of these structures using NMR, UV and gel electrophoresis has been well established (Radhakrishnan & Patel, 1994; Soyfer & Potaman, 1995). Triplexes exhibit much slower electrophoretic mobility than their corresponding duplexes due to their larger mass. In temperature-dependent UV melting, triplexes are likely to give a two-state transition corresponding to the dissociation of the third strand and then the dissociation of the duplex. The NMR spectra of triplexes contain many unique features, but the most distinct and easily recognizable is NOEs linking exchangeable NH and NH₂ with sugar protons. These NOEs are not present in canonical duplexes. The RRY and YRY triplexes are distinguishable, since the YRY triplex is pH-dependent due to the requirement of C N3 protonation and the amino and the N⁺H resonances of the protonated C (C⁺) appear at 9–10 and >14 ppm, respectively. These resonances are unusual for the canonical antiparallel Watson-Crick duplexes.

The orientation of the triplex formed between (GAA)₄G and C(TTC)₄ may be either parallel or antiparallel; it has been reported that the latter orientation is preferred when numerous GA steps are present (Sun *et al.*, 1991). Several studies using a single oligonucleotide composed of mixtures of both GAA and TTC repeats (Gacy *et al.*, 1998; Mariappan *et al.*, 1999) reported triplex formation only in the parallel orientation. However, triplex formation by these oligonucleotides was forced by the juxtaposition of the two repeats in a single chain. Furthermore, the parallel alignment of this triplex was predetermined by the relative orientation of the sequences within the chain (Gacy *et al.*, 1998). It has previously been shown that (GAA)₉·(TTC)₉ inserts in recombinant plasmids were able to form a parallel YRY triplex (Hanvey *et al.*, 1988b). The high degree of overlap in our NMR spectra prevented the determination of the orientation of the third strand as described elsewhere (Feigon *et al.*, 1995).

In our experiments, *Exo* III showed increased nuclease activity on the triplex compared to the duplex. Triplex formation was reported to transiently protect a part of a linear purine-rich plasmid from *Exo* III digestion (Takabatake *et al.*, 1992). However, in other triplexes the third strand involved in only Hoogsteen base-pairing within a triplex can present sensitivity towards *Exo* III (Roy, 1993, 1994) or other enzymes (Krasilnikov *et al.*, 1997). The increased sensitivity of oligonucleotide (GAA)·(TTC) triplexes may be due to the short size of the repeat.

Our efforts to form the 2:1 (GAA):(TTC) RRY triplex were unsuccessful. Our conditions included variation of pH from 5.0 to 8.3, variation of the Mg^{2+} concentration up to 20 mM and inclusion of the well-recognized triplex-stabilizer spermine (Soyfer & Potaman, 1995). Similarly, Bidichandani *et al.* (1998) did not observe any inter-molecular RRY triplex using a cloned (GAA)₁₁·(TTC)₁₁ repeat. The inability to form the RRY triplex may be due to the low GC-content of (GAA)·(TTC) causing this system to be deficient in the main triplex-stabilizing effects characteristic of GGC base-pairing (Xodo *et al.*, 1993).

Complementary strand parallel duplex

Our experimental results suggest that the FM species we observed is a parallel-stranded duplex formed between GAA and the TTC (Figure 1). As first described by Jovin and colleagues in 1988 (van de Sande *et al.*, 1988; reviewed by Rippe & Jovin, 1992) parallel-stranded DNA is a duplex where the polarity of the two base-paired strands is the same (Figure 1). This is opposite to the normal Watson-Crick duplex where the polarity is antiparallel. In parallel-stranded DNA the base-pairs are in a reverse Watson-Crick fashion (Scheme 1), such that one nucleotide (the base and the ribose sugar) has essentially flipped 180° with respect to the complementary nucleotide. The three hydrogen bonds present in G·C Watson-Crick base-pairs are reduced to two in G·C parallel base-pairs. Therefore, I substitution of G residue(s) allows chemical probing of the hydrogen bonding status of G N2 in the FM species. Compared to G·C base-pairs, reverse I·C base-pairs are strongly destabilized and parallel duplex formation is reduced correspondingly. However, the absence of N2 amino in I only moderately destabilizes I·C Watson-Crick base-pairs and antiparallel duplex formation. Furthermore, in neither Watson-Crick nor in reverse Watson-Crick base-pairing schemes is the N7 position of the G residue used for hydrogen bonding. Hoogsteen base-pairs use the C6-N7 face of the purine for hydrogen bonding with the N3-C4 face of the pyrimidine (Scheme 1), thus explaining their susceptibility to N7 modification agents such as DMS. Neither Hoogsteen nor reverse Hoogsteen base-pairing schemes, both of which use G N7 positions, is likely to participate in the formation of the FM species.

Several lines of experimental data suggest that the FM species we observed at pH 8.3 is a parallel duplex. First, using gel electrophoresis we demonstrated that the FM structure is composed of both the (GAA) and (TTC) strands. Second, the faster electrophoretic mobility of the FM species relative to the antiparallel duplex was also consistent with a parallel duplex (Rippe & Jovin, 1992). Third, the FM requires hydrogen bonding of the G N2 position as indicated by G to I substitution. Fourth, the methylation of all G N7 positions by DMS inhibited triplex but not FM formation, confirming that the Hoogsteen sites are not involved in the latter. (Interestingly, the FM species was observed at pH 5.0 only when DMS-inhibited triplex formation was inhibited by methylation of all G N7 positions, indicating a possible shift in equilibrium.) Finally, the FM species was resistant to *Exo* III, a characteristic of parallel duplex (Rippe & Jovin, 1992).

Interestingly, the formation of the FM species required the presence of both the (TTC) and the (GAA) strands but could be detected by autoradiography only when the (GAA) strand was radiolabeled. Our experiments indicated that the FM species is composed of both strands but that its formation was specific to only the (TTC) strands containing 5'-OH (i.e. those that were not radiolabeled) and excluded (TTC) strands with a 5'-PO₄. Such an end-effect has previously been reported (Yoon *et al.*, 1993) for triplex formation, such that a 5' end phosphate on the strand involved in both Hoogsteen and Watson-Crick base-pairing completely inhibited the binding of the third strand relative to the same strand with a 5'-OH. A similar effect of the 5' phosphorylation status of the (TTC) strand may affect parallel duplex formation.

Our NMR analysis of (GAA):(TTC) mixtures did not reveal the formation of a structure typical of parallel-stranded duplex. This may be due to two reasons: the relatively unstable nature of the ps-duplex compared to that of the aps-duplex and/or stability of the ps-structure is chain length-dependent. Electrophoresis gel matrix and buffer conditions may help to stabilize the ps-duplex but these conditions cannot be duplicated in NMR experiments. We are presently extending our studies to longer repeats in which the ps-structure may gain sufficient stability to be studied at high resolution by NMR. Several sequences have been reported to form parallel-stranded duplexes, these ps-motifs include C-rich (C⁺·C), AT-rich (ps A·T), dinucleotide CG (ps C⁺·C and G·G), (AG)·(AG) (ps G·G, A⁺·A⁺) and (GA)·(GT) dinucleotide repeats (ps G·G and A·T), GAC trinucleotide repeats (ps G·G, A·A, and C⁺·C) (Zheng, 1997), or GATC tetranucleotides (ps G·G, A·A, T·T and C⁺·C) (Cruse *et al.*, 1983; Rippe & Jovin, 1992; Dolinnaya *et al.*, 1997; Germann *et al.*, 1995; Gao *et al.*, 1998; Robinson & Wang, 1993; Robinson *et al.*, 1992). These diverse sequence compositions of ps-duplexes manifest different properties. Some ps-duplexes have been reported to be stable at low to neutral pH conditions, partly due to the

required protonation of C and A residues. Several ps-structures have been elucidated to provide atomic details of ps-base-pair formation and unique features of ps-duplexes (Cruse *et al.*, 1983; Germann *et al.*, 1995; Zheng, 1997; Robinson & Wang, 1993; Robinson *et al.*, 1992; Wang & Patel, 1994). Our results and those reported in the literature (*vide supra*) suggest that ps-duplexes are stable structural motifs and their formation may be more common than conventionally thought. The (GAA)·(TTC) ps-duplex adds a new structure to the alternative DNA structures formed by trinucleotide repeats.

Structural diversity and dynamic equilibrium

Table 1 summarizes the dependence of structure formation upon changes in pH and strand ratio. Given the right conditions, an equilibrium may exist, for example, among the ps and aps-duplexes, ssTTC and the YRY triplex (neutral pH). At pH 8.3 in electrophoretic experiments, the aps-duplex is surprisingly in equilibrium with both ssGAA and ssTTC. At lower pH, in NMR and gel electrophoresis experiments the 1:2 (GAA):(TTC) YRY triplex formed is in equilibrium with the aps-duplex and ssTTC. The self-structured ssGAA imposes another equilibrium to the system, competing with the formation of other alternative structures. Clearly, TRs, such as (CGG)_n, (CCG)_n, (CAG)_n, and (CTG)_n are characterized by diverse structural forms in dynamic equilibrium. In (CGG)·(CCG) and (CAG)·(CTG) TRs, single-stranded sequences may

self-associate to form hairpins or bi-stranded duplexes, or a new duplex motif, the e-motif (for CCG TRs), containing mismatches in every third position along the sequence (Gao *et al.*, 1998). The fluxional mismatches in these homo-stranded duplexes create a broad spectrum of sequence-dependent properties which do not exist in canonical complementary-stranded duplexes. The longer complementary duplexes of these trinucleotide repeats exhibited anomalous gel migrating bands due to slippage of single strands in the duplex (Pearson & Sinden, 1998a). It is, therefore, interesting to find that (GAA)·(TTC) share the common feature of being structurally diverse and dynamic with other disease-related TRs.

While the determination of the thermodynamic parameters for the formation of each structure was beyond the scope of this study, the relative stability of the various structures deserves comment. First, as expected, the aps-duplex, when G N7 is not methylated, is more favored than the ps-duplex, since it is present in all conditions and all strand ratios. G N7 methylation results in a change of the relative order of stability so that the ps-duplex is in excess compared to the aps-duplex. Second, the aps-duplex favorably competes with the YRY triplex even at a 1:2 (GAA):(TTC) stand ratio at pH 6.4 and above. Finally, the ps-duplex and the YRY triplex form at different pH conditions with high pH favoring the ps-duplex and lower pH favoring the triplex. However, when the triplex formation is prohibited at pH 5.0 by G N7 methylation, ps-stranded DNA is observed. This

Table 1. Summary of simultaneous structure formation at different (GAA):(TTC) ratios and pH

(GAA):(TTC)	pH 5.0-6.4	pH 8.3
1:0	Self-structured (GAA) ^{b,c,d}	Self-structured (GAA) ^{b,c}
4:1	Antiparallel Self-structured (GAA) ^{a,c}	Antiparallel Self-structured (GAA) rc(TTC) ^c
2:1	Antiparallel Self-structured (GAA) ^c	Antiparallel Self-structured (GAA) rc(TTC) ^c
1:1	Antiparallel ^{a,c}	Antiparallel Self-structured (GAA) rc(TTC) ^c
1:2	Antiparallel rc(TTC) YRY triplex ^{a,b,c}	Parallel Antiparallel rc(TTC) ^c
1:4	Parallel ^e rc(TTC) YRY triplex ^{a,b,c}	Parallel Antiparallel rc(TTC) ^c
0:1	rc(TTC) ^{a,b,c}	rc(TTC) ^c

(GAA) refers to (GAA)₄G and (TTC) refers to C(TTC)₄. The structures are abbreviated as follows: rc(TTC), random coil (TTC); antiparallel, antiparallel duplex; and parallel, parallel duplex. The structures were observed by ^a = NMR, ^b = UV, ^c = PAGE, ^d = NMR only at 0°C, and ^e = only when the YRY triplex formation was inhibited by G N7 methylation.

suggests that the ps-duplex could form under the conditions in which the YRY triplex forms as long as the formation of the triplex could be suppressed, as achieved by G N7 methylation. However, the possibility that methylation improved the stability of the ps-duplex cannot be ruled out. The analyses of the equilibrium among the various structured forms demonstrates that their presence is highly sensitive to a fine energy balance that is greatly influenced by related conditions or environmental factors of these molecular systems.

Biological implications

(GAA)·(TTC) is a unique family of triplet repeats that is not GC-rich and yet undergoes dynamic expansion in the X25 gene and leads to hereditary neurological/neuromuscular diseases. Although various mechanisms of the genetic mutation are under debate, a common theme is that the triplet is able to form stable non-B-DNA structures. Thus, triplet repeat instability specific to (CGG)·(CCG), (CAG)·(CTG) and (GAA)·(TTC) is likely to depend on the formation of an unusual DNA structure during DNA replication, repair or recombination. In the case of replication, prior to secondary structure formation in either the leading or the lagging strand, replication pausing occurs (Samadashwily *et al.*, 1997), which may cause polymerase dissociation, strand slippage and secondary structure formation, followed by resynthesis of templated repeats. There is increasing evidence that longer repeat tracts generate more stable structures with greater probability for expansion (Pearson & Sinden, 1998a,b). Our investigation demonstrates that several forms of stable structures may be associated with the GAA and TTC repeats. These include the self-structured ssGAA, the aps and the ps-duplexes of ssGAA and ssTTC strands, and the YRY triplex containing one ssGAA and two ssTTC strands. We have demonstrated that ssGAA may not easily fold into hairpins like other disease-related TR sequences and that the various structures that do form co-exist in dynamic equilibria, depending on environmental conditions. These are unique features of GAA and TTC repeats, the expression of which should affect their genetic stability and their presentation to other interacting molecules, such as proteins. The solution properties of the GAA and TTC oligonucleotides revealed by this and other studies form a basis for further characterization and examination of longer TRs and provide insights into the biological activities of these mutable sequences.

Our design of GAA and TTC TR oligonucleotides has permitted revelation of a parallel-stranded structure formed from complementary strands, whose formation is sensitive to the 5'-end chemistry (i.e. OH or PO₄) of the TTC strand. This effect could be argued to be due to the altered "fraying" at the ends of the DNA. It may not be simply a coincidence that the 5'-ends of the Okazaki fragments, produced during semiconservative

DNA replication, are transiently dephosphorylated (Pohjanpelto & Holttä, 1996). Perhaps this transient state could facilitate alternative structure formation during DNA replication or Okazaki fragment processing. Interestingly, the activity of FEN-1, a protein required for Okazaki processing, is sensitive to the 5'-end chemistry of the Okazaki fragment (Wu *et al.*, 1996). It was suggested that a 5'-end phosphate at a nick facilitates DNA fraying and favors the loading of FEN-1 to cleave exonucleolytically (as opposed to cleaving endonucleolytically). *Saccharomyces cerevisiae* that are mutant in FEN-1 (rad27) allow for large expansions of (CTG)·(CAG) repeats (Freudenreich *et al.*, 1998).

In conclusion, (GAA)·(TTC) triplet repeat sequences have the versatility to form various alternative structures including the YRY triplex and parallel-stranded DNA formed from complementary strands. This information will prove to be of interest towards offering a complete molecular basis for instability of (GAA)·(TTC) and other TRs.

Materials and Methods

Oligonucleotide synthesis and purification

Oligodeoxyribonucleotides 5'-(GAA)₂-3', 5'-(GAA)₃-3', 5'-(GAA)₄G-3', and 5'-C(TTC)₄-3' were synthesized according to standard phosphoramidite procedures implemented on an Expedite 8909 synthesizer (PerSeptive Biosystem). Cleavage, deprotection and HPLC purification of the 5' DMT-oligonucleotides were performed as previously described (Zheng *et al.*, 1996). Oligodeoxyribonucleotides 5'-(GAAGAAIAAGAAG)-3' (G7/I7 strand) and 5'-(GAAIAAIAAGAAG)-3' (G4/I4 and G7/I7 strand) were purchased already purified from Cyber-Syn. For UV and NMR experiments, the samples were further purified after DMT-deprotection by preparative HPLC purification (Zheng *et al.*, 1996). Salt-free oligonucleotides were obtained after desalting with size-exclusion and ion-exchange columns, and samples were dissolved in a solution containing 90 mM KCl, 10 mM Na₂HPO₄/NaH₂PO₄, 0.1 mM EDTA (called NMR buffer) in either 90%:10% H₂O:²H₂O (called H₂O solvent) or 100% ²H₂O. pH was adjusted using diluted H/²HCl or NaOH/²H. The reported pH in NMR experiments was based on calibrated correlation of ³¹P chemical shift (10 mM sodium phosphate at pH 7.0, referenced to trimethyl phosphate at 0.0 ppm, 25°C) versus pH meter reading. MgCl₂ was used for Mg²⁺ experiments. For all other experiments, the HPLC purified oligonucleotides were further resolved on 8 M urea/20% polyacrylamide denaturing preparative gels. The desired bands were visualized using UV-shadowing at 254 nm, excised and eluted overnight in 1 M NaOAc, 2 mM EDTA at 37°C. These samples were de-salted by EtOH-precipitation (75%) at -70°C then resuspended in 10 mM Tris-HCl (pH 7.5) to a concentration of 120 μM. The extinction coefficients at 260 nm (ε₂₆₀) of (GAA)₄G and C(TTC)₄ were estimated by the nearest-neighbor method (Cantor *et al.*, 1970) to be 1.52 × 10⁵ and 1.01 × 10⁵ l/(mol·cm), respectively. Where indicated in the text, stoichiometric strand ratios reported were determined by P₁-nuclease digestion (one unit of P₁-nuclease in 50 mM NaOAc, 0.1 mM ZnCl₂ (pH 5.2) for two hours

at 55 °C) yielding monophosphate nucleotides. Further digestion by phosphatase (two units of phosphatase, pH 9.0 for 1.5 hours at 37 °C) gave deoxyribonucleotides which were resolved using analytical HPLC (data not shown). The ratio of the peak areas of dG:dC at 278 nm determined (GAA):(TTC) strand ratio. The digestion and HPLC of the single-stranded sequences with a known GC composition, such as d(CGG)₅ and d(CCG)₅, provided reference.

NMR experiments

NMR experiments were performed on a Bruker AMX-II 600 MHz spectrometer using NMR buffer in either H₂O or ²H₂O solvents to observe exchangeable and non-exchangeable protons, respectively. Unless stated otherwise, 1D and 2D ¹H and ³¹P NMR data acquisition, water signal suppression and data processing were performed as previously described (Zheng *et al.*, 1996). 1D ¹H spectra were recorded for titration of the complementary sequences, and for monitoring of temperature-dependent spectral changes. 2D NOESY spectra of the 1:1 (GAA):(TTC) and 1:4 (GAA):(TTC) in H₂O and ²H₂O (120 and 250 ms mixing times) were recorded at 5 and 20 °C for through-space connectivities. COSY-35, DQF-COSY and TOCSY (80 ms mixing time) were also recorded in ²H₂O at 10 °C for through bond connectivities.

UV experiments

UV experiments were carried out on a Varian Cary 3E spectrophotometer equipped with a peltier temperature-controlled 12-cells holder. UV samples were prepared by diluting the corresponding NMR sample using NMR buffer to reach a 260 nm absorption between 0.5 and 1 (concentrations ~3–10 µM depending on sequences). Thermal melting curves were recorded at 260 nm from 0 to 95 °C and back to 0 °C with a gradient of 0.5 deg. C/minute and thermodynamic parameters were extracted as previously described (Zheng *et al.*, 1996).

Gel electrophoresis experiments

Denaturing gel electrophoresis experiments were performed using 39 cm × 33 cm and 0.4 mm thick gel prepared with 8 M urea and 20 % polyacrylamide (38:2 (w/w) acrylamide:bis-acrylamide). 1× TBE (89 mM Tris, 90 mM borate, 2 mM EDTA, pH 8.3) was used as the running buffer. Each gel was pre-run for 30 minutes and electrophoresed at 2000 V. DNA samples were dried under vacuum, resuspended in formamide (5 µl), denatured at 95 °C for three minutes, transferred onto ice, and loaded immediately.

Native gel electrophoresis experiments were performed on 9 cm × 13 cm and 1.5 mm thick gels prepared with 20 % polyacrylamide (29:1 acrylamide:bis). Either 1× TBE (pH 8.3) or TAE (15 mM Tris, 33 mM NaOAc, 2 mM EDTA, pH 5.0) was used as running buffer. Each gel was pre-run for 30 minutes and electrophoresed at 4 °C at 104 V for ~22 hours. For experiments including MgCl₂, 20 mM MgCl₂ was added and 2.5 l/hour recirculation applied between the upper and lower tanks. Unless stated otherwise, unlabeled and labeled single-stranded samples were diluted to 5 µM and 25 nM (~200,000 cpm), respectively. For each lane, 3 µl of the appropriate 5'-end labeled strand (8 × 10⁻⁵ nmol) was mixed with a total of 4 µl (2 × 10⁻² nmol) of unlabeled

strands (different (GAA):(TTC) ratio at constant DNA concentration). The mixture was diluted with 1 µl of 10× NMR buffer and distilled water to a total volume of 10 µl. After 30 minutes of hybridization at room temperature, 5 µl of 40 % (w/w) sucrose was added and the samples loaded. Running markers (bromophenol blue and xylene cyanol) were not directly added to the samples but were loaded in separate lanes. Each gel also included a radiolabeled 12-mer duplex ((CG)₂T₄(GC)₂·(GC)₂A₄(CG)₂) and a radiolabeled 15-mer duplex ((CGG)₅·(CCG)₅) as size markers. Following electrophoresis, gels were wrapped in cellophane and exposed to X-ray film (Kodak). Where indicated gels were stained by Stains-All (Sigma), as described (Zeiger *et al.*, 1974) and photographed using an Eagle-Eye camera.

Enzyme reactions

³²P-labeling

5'-End labeling was performed typically using 400 ng of oligonucleotide using [γ-³²P]ATP (NEN), T4 polynucleotide kinase (USB) and conditions recommended by the manufacturer. Excess ATP was removed by NAP-5 columns (Pharmacia Biotech) and oligonucleotides were purified by ethanol-precipitation.

Enzymatic digestions experiments

Samples were prepared as described in the Gel electrophoresis experiments section and the mixed ssGAA and ssTTC (pH 5.0 or 8.0 as indicated in the text) were allowed to hybridize for 30 minutes. Appropriate amounts of *Exo* III (New England Biolabs, MA) in a buffer solution (final concentration 66 mM Tris (pH 8.0), 5 mM MgCl₂) was added to oligonucleotides (final volume of 10 µl). The enzyme concentrations used were 0×, 1×, 10×, 100× and 1000×, where 1× corresponds to 0.02 unit of enzyme per nmol of DNA. After 25 minutes of the enzymatic reaction at room temperature, the reaction was stopped by adding EDTA (5 µl, 45 mM) in aqueous sucrose solution (40 % (w/w)) and stored on ice until loading onto a gel.

Methylation reactions

The methylation of G N7 in ssGAA was carried out using DMS and ³²P-labeled ssGAA. The oligonucleotide (48 µmol of ³²P-labeled (GAA)₄G, 3 × 10⁶ cpm) was dissolved in a solution (50 µl) containing sodium cacodylate (50 mM, pH 7.0), EDTA (1 mM) and incubated with 4 µl of DMS at room temperature for 30 minutes. This step was repeated once with an additional 4 µl of DMS. The reaction was stopped by addition of a solution containing NaAcOH (25 µl, 1.5 M, pH 7.0) and β-mercaptoethanol (1 M). The methylated oligonucleotide was ethanol-precipitated twice. The concentration of the sample was adjusted with 10 mM Tris-HCl (pH 7.5) to 5 mM (as measured by the radioactivity) and used immediately after the reaction as described in Results. The complete methylation of all G N7 positions was confirmed by treating an aliquot of the DMS-treated (GAA)₄G sample with piperidine under standard conditions (Maxam & Gilbert, 1980). Denaturing polyacrylamide gel analysis of this sample showed the characteristic bands (data not shown).

Intercalator binding assay

DNA samples of different (GAA):(TTC) ratios (0.2 nmol) were resolved on native gel electrophoresis as described above and stained essentially as described (Rippe & Jovin, 1992). Briefly, the gel was first stained in a 15 µg/ml BBI-258 (Sigma, MO) solution for 15 minutes, destained 30 minutes, then stained in a 1 µg/ml ethidium bromide (Sigma, MO) solution for 15 minutes. Bands were visualized using a UV transilluminator and photographed using the appropriate filter.

Acknowledgments

The 600 MHz NMR spectrometer at the University of Houston is funded by the W. M. Keck Foundation. Acknowledgment is made to the NIH (R01GM54652) (X.G.), the Robert A. Welch Foundation (E-1270) (X.G.), the Muscular Dystrophy Association (C.E.P.), and the Medical Research Council of Canada (C.E.P.) for financial support, and to the Institute of Molecular Design and the W. M. Keck Center for Computational Biology at the University of Houston for computer resource support. C.E.P. is an MRC Scholar. The authors thank Dr Vladimir Potaman for helpful discussions, Dr Mingxue Zhen for providing some of the 1D NMR spectra, and Mr Hungjoo Lee for providing some of the 1D NMR spectra and contributing to Figure 8.

References

- Bidichandani, S. I., Ashizawa, T. & Patel, P. I. (1998). The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. *Am. J. Hum. Genet.* **62**, 111-121.
- Blake, R. D. & Fresco, J. R. (1966). Polynucleotides. VII. Spectrophotometric study of the kinetics of formation of the two-stranded helical complex resulting from the interaction of polyadenylate and polyribouridylylate. *J. Mol. Biol.* **19**, 145-160.
- Campuzano, V., Montermini, L., Molto, M. D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F. *et al.* (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science*, **271**, 1423-1427.
- Cantor, C. R., Warshaw, M. M. & Shapiro, H. (1970). Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers*, **9**, 1059-1077.
- Chen, X., Mariappan, S. V., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoun, A., Smith, S. S., Bradbury, E. M. & 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.
- Cossee, M., Schmitt, M., Campuzano, V., Reutenauer, L., Moutou, C., Mandel, J.-L. & Koenig, M. (1997). Evolution of the Friedreich's ataxia trinucleotide repeat expansion: founder effect and premutations. *Proc. Natl Acad. Sci. USA*, **4**, 7452-7457.
- Cruse, W. B., Egert, E., Kennard, O., Sala, G. B., Salisbury, S. A. & Viswamitra, M. A. (1983). Self base-pairing in a complementary deoxydinucleoside monophosphate duplex: crystal and molecular structure of deoxycytidylyl-(3'-5')-deoxyguanosine. *Biochemistry*, **12**, 1833-1839.
- Darlow, J. M. & Leach, D. R. (1998). Evidence for two preferred hairpin folding patterns in d(CGG)·d(CCG) repeat tracts *in vivo*. *J. Mol. Biol.* **275**, 17-23.
- Dolinnaya, N. G., Ulku, A. & Fresco, J. R. (1997). Parallel-stranded linear homoduplexes of d(A⁺-G)_n > 10 manifesting the contrasting ionic strength sensitivities of poly(A⁺. A⁺) and DNA. *Nucl. Acids Res.* **25**, 1100-1107.
- Feigon, J., Koshlap, K. M. & Smith, F. W. (1995). ¹H NMR spectroscopy of DNA triplexes and quadruplexes. *Methods Enzymol.* **261**, 225-255.
- Felsenfeld, G. & Miles, H. T. (1968). The physical and chemical properties of nucleic acids. *Annu. Rev. Biochem.* **36**, 407-447.
- Freudenreich, C. H., Kantrow, S. M. & Zakian, V. A. (1998). Expansion and length-dependent fragility of CTG repeats in yeast. *Science*, **279**, 853-856.
- Gacy, A. M., Goeliner, G. M., Spiro, C., Chen, X., Gupta, G., Bradbury, E. M., Dyer, R. B., Mikesell, M. J., Yao, J. Z., Johnson, A. J., Richter, A., Melancon, S. B. & McMurray, C. T. (1998). GAA instability in Friedreich's ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. *Mol. Cell*, **1**, 583-593.
- Gao, X., Huang, X., Smith, G. K., Zheng, M. & Liu, H. (1995). A new antiparallel duplex motif of DNA CCG repeats that is stabilized by extrahelical bases symmetrically located in the minor groove. *J. Am. Chem. Soc.* **117**, 8883-8884.
- Gao, X., Huang, X., Smith, G. K. & Zheng, M. (1998). Structure and dynamics of single-stranded nucleic acids containing trinucleotide repeats. In *Genetic Instabilities and Hereditary Neurological Diseases* (Wells, R. D. & Warren, S. T., eds), pp. 623-646, Academic Press, San Diego, CA.
- Germann, M. W., Zhou, N., van de Sande, J. H. & Vogel, H. J. (1995). Parallel-stranded duplex DNA: an NMR perspective. *Methods Enzymol.* **261**, 207-225.
- Hanvey, J. C., Klysik, J. & Wells, R. D. (1988a). Influence of DNA sequence on the formation of non-B right-handed helices in oligopurine·oligopyrimidine inserts in plasmids. *J. Biol. Chem.* **263**, 7386-7396.
- Hanvey, J. C., Klysik, J. & Wells, R. D. (1988b). Intramolecular DNA triplexes in supercoiled plasmids. *Proc. Natl Acad. Sci. USA*, **85**, 6292-6296.
- Howard, F. B., Frazier, J., Lipsett, M. N. & Miles, H. T. (1964). Infrared demonstration of two- and three-strand helix formation between poly C and Guanosine mononucleotides and oligonucleotides. *Biochem. Biophys. Res. Commun.* **17**, 93-102.
- Jakupciak, J. P. & Wells, R. D. (1999). Genetic instabilities in (CTG)·(CAG) repeats occur by recombination. *J. Biol. Chem.* **274**, 23468-23479.
- Jovin, T. M., Rippe, K., Ramsing, N. B., Klement, R., Elhorst, W. & Vojtisková, M. (1990). Parallel stranded DNA. *Struct. Methods*, **3**, 155-174.
- Kettani, A., Kumar, R. A. & Patel, D. J. (1995). Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat. *J. Mol. Biol.* **254**, 638-656.
- Krasilnikov, A. S., Panyutin, I. G., Samadashwily, G. M., Cox, R., Lazurkin, Y. S. & Mirkin, S. M. (1997). Mechanisms of triplex-caused polymerization arrest. *Nucl. Acids Res.* **25**, 1339-1346.

- Latimer, J. P. & Lee, J. S. (1991). Ethidium bromide does not fluoresce when intercalated adjacent to 7-deazaguanine in duplex DNA. *J. Biol. Chem.* **266**, 13849-13851.
- Lian, C., Robinson, H. & Wang, A. H.-J. (1996). Structure of actinomycin D bound with (GAAGCTTC)₂ and (GATGCTTC)₂ and its binding to the (CAG)_n:(CTG)_n triplet sequence as determined by NMR analysis. *J. Am. Chem. Soc.* **118**, 8791-8801.
- Mariappan, S. V., Chen, X., Catasti, P., Bradbury, E. M. & Gupta, G. (1998). Structural studies on the unstable triplet repeats. In *Genetic Instabilities and Hereditary Neurological Diseases* (Wells, R. D. & Warren, S. T., eds), pp. 647-690, Academic Press, San Diego, CA.
- Mariappan, S. V., Catasti, P., Silks, L. A., III, Bradbury, E. M. & Gupta, G. (1999). The high resolution structure of the triplex formed by the GAA/TTC triplet repeat associated with Friedreich's ataxia. *J. Mol. Biol.* **285**, 2035-2052.
- Maxam, A. M. & Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**, 499-560.
- Mitas, M., Yu, A., Dill, J., Kamp, T. J., Chambers, E. J. & Haworth, I. S. (1995). Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: (CTG)₁₅. *Nucl. Acids Res.* **23**, 1050-1059.
- Mitas, M., Yu, A., Dill, J. & Haworth, I. S. (1995a). The trinucleotide repeat sequence d(CTG)₁₅ forms a heat-stable hairpin containing G^{syn}·C^{anti} base-pairs. *Biochemistry*, **34**, 12803-12811.
- Mitchell, J. E., Newbury, S. F. & 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. *Nucl. Acids Res.* **23**, 1876-1881.
- Maskos, K., Gunn, B. M., LeBlanc, D. A. & Morden, K. M. (1993). NMR study of G·A and A·A pairing in (dGCGAATAAGCG)₂. *Biochemistry*, **32**, 3583-3595.
- Montermini, L., Andermann, E., Labuda, M., Richter, A., Pandolfo, M., Cavalcanti, F., Pianese, L., Iodice, L., Farina, G., Monticelli, A., Turano, M., Filla, A., De Michele, G. & Cocozza, S. (1997). The Friedreich ataxia GAA triplet repeat: premutation and normal alleles. *Hum. Mol. Genet.* **6**, 1261-1266.
- Morgan, A. R. & Wells, R. D. (1968). Specificity of the three-stranded complex formation between double-stranded DNA and single-stranded RNA containing repeating nucleotide sequences. *J. Mol. Biol.* **37**, 63-80.
- Nadel, Y., Weisman-Shomer, P. & Fry, M. (1995). The fragile X syndrome single-strand d(CGG)_n nucleotide repeats readily fold back to form unimolecular hairpin structures. *J. Biol. Chem.* **270**, 28970-28977.
- Ohshima, K., Kang, S., Larson, J. E. & Wells, R. D. (1996). Cloning, characterization, and properties of seven triplet repeat DNA sequences. *J. Biol. Chem.* **271**, 16773-16783.
- Ohshima, K., Montermini, L., Wells, R. D. & Pandolfo, M. (1998). Inhibitory effects of expanded GAA·TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication *in vivo*. *J. Biol. Chem.* **273**, 14588-14595.
- Pearson, C. E. & 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.
- Pearson, C. E. & Sinden, R. R. (1998a). Slipped strand DNA, dynamic mutations, and human disease. In *Genetic Instabilities and Hereditary Neurological Diseases* (Wells, R. D. & Warren, S. T., eds), pp. 585-621, Academic Press, San Diego, CA.
- Pearson, C. E. & Sinden, R. R. (1998b). Trinucleotide repeat DNA structures: dynamic mutations from dynamic DNA. *Curr. Opin. Struct. Biol.* **8**, 321-330.
- Pearson, C. E., Ewel, A., Acharya, S., Fishel, R. A. & Sinden, R. R. (1997). Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. *Hum. Mol. Genet.* **6**, 1117-1123.
- Pearson, C. E., Wang, Y.-H., Griffith, J. D. & Sinden, R. R. (1998a). Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)_n·(CAG)_n repeats from the myotonic dystrophy locus. *Nucl. Acids Res.* **26**, 816-823.
- Pearson, C. E., Eichler, E. E., Lorenzetti, D., Kramer, S. F., Zoghbi, H. Y., Nelson, D. L. & Sinden, R. R. (1998b). Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand (S-DNA) formation. *Biochemistry*, **37**, 2701-2708.
- Pertruska, J., Arnheim, N. & Goodman, M. F. (1996). Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases. *Nucl. Acids Res.* **24**, 1992-1999.
- Pohjanpelto, P. & Holtta, E. (1996). Phosphorylation of Okazaki-like DNA fragments in mammalian cells and role of polyamines in the processing of this DNA. *EMBO J.* **15**, 1193-1200.
- Radhakrishnan, I. & Patel, D. J. (1994). Hydration sites in purine.purine.pyrimidine and pyrimidine.purine.pyrimidine DNA triplexes in aqueous solution. *Structure*, **2**, 395-405.
- Radhakrishnan, I., Gao, X., de los Santos, C., Live, D. & Patel, D. J. (1991). NMR structural studies of intramolecular (Y⁺)_n·(R⁺)_n·(Y⁻)_n DNA triplexes in solution: imino and amino proton and nitrogen markers of G·TA base triple formation. *Biochemistry*, **30**, 9022-9030.
- Ramsing, N. B. & Jovin, T. M. (1988). Parallel-stranded duplex DNA. *Nucl. Acids Res.* **16**, 6659-6676.
- Richards, R. I. & Sutherland, G. R. (1994). Simple repeat DNA is not replicated simply. *Nature Genet.* **6**, 114-116.
- Riley, M., Maling, B. & Chamberlain, M. J. (1966). Physical and chemical characterization of two- and three-stranded adenine-uracil homopolymer complexes. *J. Mol. Biol.* **20**, 359-389.
- Rippe, K. & Jovin, T. M. (1989). Substrate properties of 25-nt parallel-stranded linear DNA duplexes. *Biochemistry*, **28**, 9542-9549.
- Rippe, K. & Jovin, T. M. (1992). Parallel-stranded duplex DNA. *Methods Enzymol.* **211**, 199-220.
- Robinson, H. & Wang, A. H.-J. (1993). 5'-CGA sequence is a strong motif for homo base-paired parallel-stranded DNA duplex as revealed by NMR analysis. *Proc. Natl Acad. Sci. USA*, **90**, 5224-5228.
- Robinson, H., van der Marel, G. A., van Boom, J. H. & Wang, A. H.-J. (1992). Unusual DNA conformation at low pH revealed by NMR: parallel-stranded DNA duplex with homo base-pairs. *Biochemistry*, **31**, 10510-10517.
- Roy, C. (1993). Inhibition of gene transcription by purine rich triplex forming oligodeoxyribonucleotides. *Nucl. Acids Res.* **21**, 2845-2852.

- Roy, C. (1994). Triple-helix formation interferes with the transcription and hinged DNA structure of the interferon-inducible 6-16 gene promoter. *Eur. J. Biochem.* **220**, 493-503.
- Saenger, W. (1984). *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, NY.
- Sakamoto, N., Chastain, P. D., Parniewski, P., Ohshima, K., Pandolfo, M., Griffith, J. D. & Wells, R. D. (1999). Sticky DNA: self-association properties of long (GAA):(TTC) repeats in R·R·Y triples structures from Friedreich's ataxia. *Mol. Cell*, **3**, 465-475.
- Samadashwily, G. M., Raca, G. & Mirkin, S. M. (1997). Trinucleotide repeats affect DNA replication *in vivo*. *Nature Genet.* **17**, 298-304.
- Shea, R. G., Ng, P. & Bischoffberger, N. (1990). Thermal denaturation profiles and gel mobility shift analysis of oligodeoxynucleotide triplexes. *Nucl. Acids Res.* **18**, 4859-4866.
- Sinden, R. R. (1994). *DNA Structure and Function*, Academic Press, San Diego, CA.
- Sinden, R. R. & Wells, R. D. (1992). DNA structure, mutations, and human genetic disease. *Curr. Opin. Biotechnol.* **3**, 612-622.
- Smith, G. K., Jie, J., Fox, G. E. & Gao, X. (1995). DNA CTG triplet repeats involved in dynamic mutations of neurologically related gene sequences form stable duplexes. *Nucl. Acids Res.* **23**, 4303-4311.
- Soyfer, V. N. & Potaman, V. N. (1995). *Triple-helical Nucleic Acids*, Springer-Verlag, New York.
- Suen, I. S., Rhodes, J. N., Christy, M., McEwen, B., Gray, D. M. & Mitas, M. (1999). Structural properties of Friedreich's ataxia d(GAA) repeats. *Biochim. Biophys. Acta*, **1444**, 14-24.
- Sun, J. S., De Bizemont, T., Duval-Valentin, G., Montenay-garestier, T. & Hélène, C. (1991). Extension of the range of recognition sequences for triple helix formation by oligonucleotides containing guanines and thymines. *C.R.H. Acad. Sci. Paris, Série III*, **313**, 585-590.
- Takabatake, T., Asada, K., Uchimura, Y., Ohdata, M. & Kusukawa, N. (1992). The use of purine-rich oligonucleotides in triplex-mediated DNA isolation and generation of unidirectional deletions. *Nucl. Acids Res.* **21**, 5853-5854.
- Usdin, K. (1998). NGG-triplet repeats form similar intrastrand structures: implications for the triplet expansion diseases. *Nucl. Acids Res.* **26**, 4078-4085.
- Usdin, K. & Woodford, K. J. (1995). CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis *in vitro*. *Nucl. Acids Res.* **23**, 4203-4209.
- van de Sande, J. H., Ramsing, N. B., Germann, M. W., Elhorst, W., Kalisch, B. W., Kitzing, E. V., Pon, R. T., Clegg, R. C. & Jovin, T. M. (1988). Parallel stranded DNA. *Science*, **241**, 551-557.
- Wang, A. H.-J. & Teng, M.-K. (1990). Molecular recognition of DNA minor groove binding drugs. In *Crystallographic and Modeling Methods in Molecular Design* (Bugg, C. E. & Ealick, S. E., eds), pp. 123-150, Springer-Verlag, New York.
- Wang, Y. & Patel, J. P. (1994). Solution structure of the d(T-C-G-A) duplex at acidic pH. *J. Mol. Biol.* **242**, 508-526.
- Wu, X., Li, J., Li, X., Hsieh, C. L., Burgers, P. M. & Lieber, M. R. (1996). Processing of branched DNA intermediates by a complex of human FEN-1 and PCNA. *Nucl. Acids Res.* **24**, 2036-2043.
- Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, pp. 236-245, John Wiley & Sons, New York.
- Xodo, L. E., Alunni-Fabbroni, M., Manzini, G. & Quadrifoglio, F. (1993). Sequence-specific DNA-triplex formation at imperfect homopurine-homopyrimidine sequences within a DNA plasmid. *Eur. J. Biochem.* **212**, 395-401.
- Yoon, K., Hobbs, C. A., Walter, A. E. & Turner, D. H. (1993). Effect of a 5'-phosphate on the stability of triple helix. *Nucl. Acids Res.* **21**, 601-606.
- Yu, A., Dill, J., Wirth, S. S., Huang, G., Lee, V. H., Howorth, I. S. & Mitas, M. (1995a). The trinucleotide repeat sequence d(GTC)₁₅ adopts a hairpin conformation. *Nucl. Acids Res.* **23**, 2706-2714.
- Yu, A., Dill, J. & Mitas, M. (1995b). The purine-rich trinucleotide repeat sequences d(CTG)₁₅ and d(GAC)₁₅ form hairpins. *Nucl. Acids Res.* **23**, 4055-4057.
- Zeiger, R. S., Salomon, R., Dingman, C. W. & Peacock, A. C. (1974). Role of base composition in the electrophoresis of heat-treated deoxyribonucleic acid from HeLa and mouse cells in composite polyacrylamide gels. *Biochemistry*, **13**, 3388-3393.
- Zheng, M. (1997). Studies of the structure and stability of DNA trinucleotide repeat sequence by nuclear magnetic resonance spectroscopy. Thesis, University of Houston.
- Zheng, M., Huang, X., Smith, G. K., Yang, X. & Gao, X. (1996). Genetically unstable CXG repeats are structurally dynamic and have a high propensity for folding. An NMR and UV spectroscopic study. *J. Mol. Biol.* **264**, 323-336.

Edited by I. Tinoco

(Received 22 May 2000; received in revised form 19 July 2000; accepted 24 July 2000)