Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 12159–12164, October 1996
Biochemistry

A single G-to-C change causes human centromere TGGAA repeats to fold back into hairpins

(human centromere/DNA folding/single-residue loop/G∘A mismatch/GNA motifs)

LEIMING ZHU*, SHAN-HO CHOU†‡, AND BRIAN R. REID*†§

*Chemistry Department, †Biochemistry Department, and ‡Immunology Department, University of Washington, Seattle WA 98195

Communicated by I. Tinoco, Jr., University of California, Berkeley, CA, August 12, 1996 (received for review May 20, 1996)

ABSTRACT **Recently, we established that satellite III (TGGAA)$_n$ tandem repeats, which occur at the centromeres of human chromosomes, pair with themselves to form an unusual "self-complementary" antiparallel duplex containing (GGA)$_2$ motifs in which two unpaired guanines from opposite strands intercalate between sheared G∘A base pairs. In separate studies, we have also established that the GCA triplet does not form bimolecular (GCA)$_2$ motifs but instead promotes the formation of hairpins containing a GCA-turn motif in which the loop contains a single cytidine closed by a sheared G∘A pair. Since TGCAA is the most frequent variant of TGGAA found in satellite III repeats, we reasoned that the potential of this variant to form GCA-turn miniloop fold-back structures might be an important factor in modulating the local structure in natural (TGGAA)$_n$ repeats. We report here the NMR-derived solution structure of the heptadecadeoxynucleotide (G)TGGAATGCAATGGAA(C) in which a central TGCAA pentamer is flanked by two TGGAA pentamers. This 17-mer forms a rather unusual and very stable hairpin structure containing eight base pairs in the stem, only four of which are Watson–Crick pairs, and a loop consisting of a single cytidine residue. The stem contains a (GGA)$_2$ motif with intercalative 14G|4G stacking between two sheared G∘A base pairs; the loop end of the stem consists of a sheared 8G∘10A closing pair with the cytosine base of the 9C loop stacked on 8G. The remarkable stability of this unusual hairpin structure (T_m = 63°C) suggests that it probably plays an important role in modulating the folding of satellite III (TGGAA)$_n$ repeats at the centromere.**

The classical satellite III of human DNA, which has been located in the centromere region of chromosomes (1), is composed of (TGGAA)$_n$ repeats with the consensus sequence 5′- CAACCCGA$^{A}_{G}$(TGGAA)$_n$ (2). From the cooperative UV-melting transition Grady *et al.* (1) found that, in the absence of the complementary TTCCA strand, (TGGAA)$_6$ alone can somehow pair with itself to form a stable duplex with a T_m of 65°C. We recently solved the solution structure of the antiparallel "self-complementary" (TGGAATGGAA)$_2$ duplex containing a tandem TGGAA repeat; this structure contains two (GGA)$_2$ motifs in which the central guanines from opposite strands are unpaired and intercalatively stacked on each other between flanking G∘A pairs (3, 4). The flanking G∘A pairs are in the sheared, or side-by-side, configuration (5–7) and the guanines of the (GGA)$_2$ motif form a four-guanine interstrand stack.

Although the (TGGAA)$_n$ repeat is highly conserved in satellite III, perfect repeats rarely continue for more than 12–15 repeats (60–75 bp) with several variants that differ from the consensus pentamer by only a single base. The most frequent among these variants is TGCAA—i.e., the central GGA triplet [which forms the self-paired (GGA)$_2$ motif] is changed to GCA. Interestingly, we recently found that the triplet GCA is a very strong hairpin promoter (8, 9). Among the four GNA triplets in a d(NAATGNAATG) sequence context, GGA predominantly forms bimolecular duplexes containing a (GGA)$_2$ motif, as described previously (3, 4), while GCA forms a unimolecular hairpin (8, 9) and GAA and GTA form a mixture of mostly hairpin in equilibrium with a minor (GNA)$_2$ duplex form (9). If (TGGAA)$_n$ repeats self-pair to form antiparallel (GGA)$_2$ motifs *in vivo*, we reasoned that occasional TGCAA pentamers interspersed among (TGGAA)$_n$ repeats might promote folding back of the chain on itself by forming GCA-turn loops, and thus modulate centromere folding. In the present paper, we have used NMR, distance geometry, restrained molecular dynamics, and back-calculation refinement to study the solution structure of the DNA heptadecamer (G)TGGAATGCAATGGAA(C), which contains three centromeric pentamer repeats, the middle one of which is changed from TGGAA to TGCAA. In contrast to (G)(TGGAA)$_3$(C), which forms an antiparallel duplex, this heptadecanucleotide forms an extremely stable hairpin (T_m = 63°C) despite several unusual structural features, including a single nucleotide loop and an 8-bp stem containing four non-Watson–Crick pairs.

MATERIALS AND METHODS

DNA samples were synthesized at the 3-μmol scale on an Applied Biosystems 380B DNA synthesizer by standard phosphoramidite methods and were purified by well-established procedures (10). The samples were dissolved in 0.40 ml of 25 mM sodium phosphate buffer (pH 6.8) containing 100 mM NaCl and 0.4 mM EDTA.

A double quantum-filtered correlated spectroscopy (DQF-COSY) and five nuclear Overhauser enhancement (NOE) spectroscopy (NOESY) experiments (mixing times of 60, 120, 180, 240, and 600 ms) were carried out in ^2H_2O at 25°C on a Bruker AM500 NMR spectrometer; 2048 complex points in *t*2 and 400 complex points in *t*1 were collected. The spectral width in both dimensions was 4386 Hz. For each *t*1 incrementation, 16 scans were averaged.

A ^{31}P–^{1}H correlation spectrum [detected in the inverse mode (11)] was also acquired on the AM500 spectrometer at 25°C. For this, 2048 complex points in the ^{1}H (*t*2) dimension and 100 complex points in the ^{31}P (*t*1) dimension were collected. For each *t*1 incrementation, 128 scans were averaged. The spectral widths are 833 Hz for the ^{31}P dimension and 4386 Hz for the ^{1}H dimension. The acquired data were transferred to an IRIS 4D workstation and processed by the software program FELIX (Biosym Technologies, San Diego).

The solution structure of GTGGAATGCAATGGAAC was determined by the combined use of distance geometry (DG) and restrained molecular dynamics (RMD) methods. The initial distance bounds were obtained from the initial buildup

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: COSY, correlated spectroscopy; DQF-COSY, double quantum-filtered COSY; NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; DG, distance geometry.

§To whom reprint requests should be addressed.

of NOE intensities. From these initial distance bounds, together with very generous bounds for dihedral angle constraints derived from qualitative analysis of the ^{31}P–^{1}H correlation spectrum, an initial structure was obtained by DG using the program DGII (Biosym Technologies). This initial structure was refined by restrained molecular dynamics/mechanics using the program DISCOVER (Biosym Technologies). After back-calculating the NOESY spectra of the interim structure using the simulation program BIRDER (12) and comparing them with experimental data, a modified bounds file was created. This process was repeated until the simulated NOESY spectra matched the experimental spectra with a satisfactory *R* factor.

RESULTS AND DISCUSSION

Spectral Analysis

Using the information from through-space NOE connectivities and through-bond *J*-coupling connectivities, all the nonexchangeable protons (except some H5′/5″ protons) of GTGGAATGCAATGGAAC could be unambiguously assigned, and their chemical shifts are listed in Table 1.

The expanded aromatic-H1′/H3′ (*A*) and aromatic-H2′/H4′ (*B*) regions of the NOESY spectrum at 600 ms are shown in Fig. 1. The sequence contains two GGA triplets and one GCA triplet, and several striking characteristics that distinguish the GCA triplet from the GGA triplets stand out. In the GGA triplets, the anomeric H1′ protons of 4G and 14G are markedly shifted upfield to ≈4.6 ppm and the preceding 3G and 13G H1′ resonances are also shifted upfield, but only to ≈5.0 ppm; in contrast, in the GCA triplet the H1′ resonances of 8G and 9C are not shifted upfield, occurring between 5.4 and 5.6 ppm.

The strong H8-H3′ NOEs for 4G and 14G (marked by the arrows a and b in Fig. 1*A*) indicate C3′-*endo* sugar conformations for these two residues (confirmed by *J*-coupling patterns). The weaker H6–H3′ NOE for 9C (peak c in Fig. 1*A*) combined with normal *J*-coupling patterns establishes that 9C, like the other residues, has a C2′-*endo* sugar conformation. Interestingly, in our earlier systematic studies of GNA triplets in the sequence context d(NAATGNAATG), where N is G, A, T, or C (9), it was found that the H1′ chemical shift and the sugar conformation of the N residue are important and diagnostic criteria that can be used to distinguish between the (GNA) hairpin loop motif and the bimolecular duplex $(GNA)_2$ motif. In the $(GNA)_2$ duplex pairing motif, the H1′ resonance of the N residue is shifted upfield and its sugar is in the unusual C3′-*endo* conformation, while in the (GNA)-turn loop motif, the H1′ resonance of the N residue is not shifted upfield and its sugar is in the normal C2′-*endo* conformation (9). In the present case the C3′-*endo* sugar conformation for residues 4G and 14G is further confirmed by DQF-COSY experiments (see below). Despite these differences there are, however, some common NOE features between these turn-loop and antiparallel pairing motifs. For example, the base protons of residues 4G, 14G, and 9C all show very strong NOEs to the preceding residue H1′ proton (Fig. 1*A*). As will be described below, additional common features can be found in the base-H2′/H2″/H4′ NOESY region, reflecting other structural features common to both the $(GGA)_2$ duplex pairing motif and the (GCA) loop motif.

In Fig. 1*B*, it is clear that the NOEs between H2′/H2″ protons and the base H8/H6 proton of the following residue, normally observed in B-DNA, are either missing or very weak for residues 4G, 14G and 9C (indicated by × symbols). Furthermore, the 5AH2–14GH4′/5′/5″ and 15AH2–4GH4′/5′/5″ "cross-strand" NOEs are obvious in the bottom left corner of Fig. 1*B*. Much weaker 10AH2–9CH4′/H5′ NOEs can also be observed. Perhaps the most interesting feature in Fig. 1*B* is that the 2- to 3-ppm region, which usually contains only deoxyribose 2′/2″ protons (13), now contains the H4′ resonances of 4G, 9C, and 14G which are extraordinarily upfield-shifted by over 1000 Hz to ≈2 ppm (top of Fig. 1*B*); the H5′/5″ resonances of 4G, 9C, and 14G are also signficantly upfield-shifted to the 3- to 3.5-ppm region (bottom of Fig. 1*B*). The assignments of the unusually shifted 4G, 14G, and 9C H4′ protons were confirmed and corroborated by H1′–H2′/2″–H3′–H4′ *J*-coupling connectivities and by $(n-1)$H3′–$(n)^{31}$P–(n)H4′ *J*-coupling connectivities (see below). The NOE data of Fig. 1 *A* and *B* indicate that qualitatively the stem of this hairpin contains a structural motif very similar to that of the $(GGA)_2$ motifs in the antiparallel satellite III $(GTGGAATGGAAC)_2$ duplex (3, 4)—i.e., 4G is intercalated between 14G and the sheared 3G∘15A pair, and conversely 14G is intercalated between 4G and the sheared 13G∘5A pair. On the other hand, the 8G-9C-10A loop appears to be qualitatively very similar to the single-residue GCA hairpin loop that we reported recently (8). A schematic representation of this hairpin is presented in Fig. 1*C*, in which Watson–Crick base pairs are represented by solid bars, sheared G∘A pairs are represented by striped bars, and unpaired intercalated guanosines are represented by arrows.

Fig. 2 shows the expanded NOESY and DQF-COSY spectra for the H1′/3′ to H2′/2″/4′ region. The horizontal lines

Table 1. Chemical shifts of nonexchangeable protons and ^{31}P of the stem-harpin d(GTGGAATGCAATGGAAC) at 25°C

	Shift, ppm							
Nucleotide	H6/H8	H5/H2/M5	H1′	H2′	H2″	H3′	H4′	^{31}P
1G	7.80		5.78	2.29	2.50	4.59	4.00	
2T	6.90	0.99	5.96	1.44	2.08	4.60	4.09	−4.32
3G	7.77		5.10	2.43	2.23	4.70	4.27	−4.98
4G	7.23		4.66	2.29	1.38	4.44	2.11	−3.90
5A	7.70	7.82	5.99	2.20	2.68	4.77	4.11	−3.55
6A	8.20	7.61	5.60	2.20	2.57	4.77	4.18	−4.54
7T	6.76	1.03	5.79	1.52	2.19	4.63	4.08	−4.36
8G	7.90		5.43	2.46	2.19	4.65	4.19	−5.02
9C	7.05	5.04	5.60	1.39	1.92	4.20	1.77	−4.69
10A	7.76	7.87	6.06	2.60	2.77	4.64	4.15	−4.56
11A	8.12	7.55	5.59	2.14	2.49	4.67	4.19	−4.91
12T	6.74	0.86	5.83	1.40	2.08	4.59	4.09	−4.37
13G	7.76		4.99	2.39	2.15	4.65	4.17	−5.01
14G	7.17		4.60	2.18	1.50	4.47	2.29	−3.70
15A	7.75	7.90	6.01	2.14	2.61	4.75	4.15	−3.44
16A	8.16	7.67	5.67	2.37	2.61	4.81	4.20	−4.54
17C	7.13	5.10	5.93	1.86	1.96	4.25	3.86	−4.13

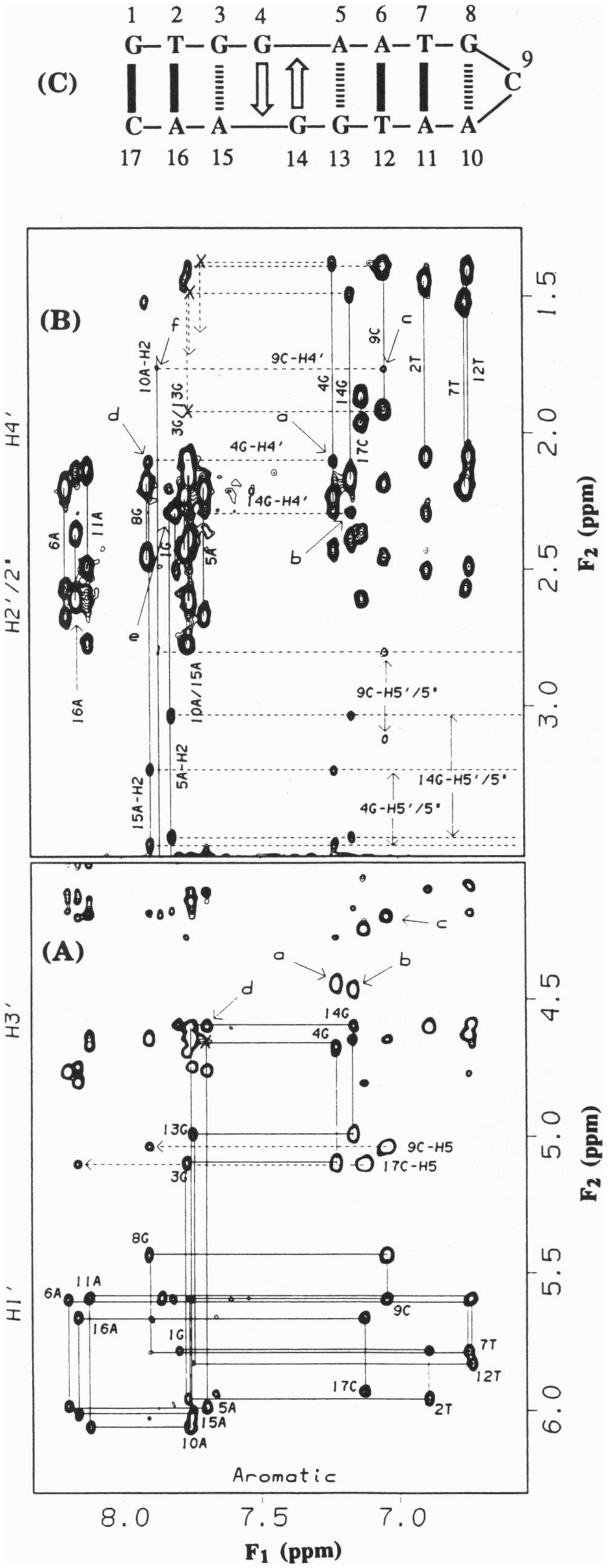

FIG. 1. (*A*) The base–H1′/3′ region of the NOESY spectrum of GTGGAATGCAATGGAAC at 600-ms mixing time. The H3′–H8/H6 connectivity is not traced for the sake of clarity, but the *n*H8/H6 to *n*H3′ NOE cross-peaks for 4G, 14G, and 9C are indicated by the arrows a, b, and c, respectively. The peak labeled d indicates the "cross-strand" 5AH8–14GH1′ NOE (the "cross-strand" 15AH8–4GH1′ cross-peak overlaps the 10AH3′–10AH8 cross-peak). (*B*) The H8/H6 to H2′/2″/4′ region of the same NOESY spectrum. Arrows a, b, and c indicate the *n*H8/H6 to *n*H4′ interactions for 4G, 14G, and 9C. Other labeled peaks are d, 4GH4′ to 15AH2; and e, 14GH4′ to 5AH2. (*C*) Schematic picture of the hairpin structure formed by the 17-mer d(GTGGAATGCAATGGAAC).

connect the H1′–H2′ cross-peak to the H1′–H2″ cross-peak of the same residue; the H3′ and H4′ protons of residues 4G, 14G, and 9C are specifically indicated. The strong four-lobed H3′–H4′ DQF-COSY cross-peaks for 4G and 14G can be easily distinguished from the prevalent H1′–H2′/2″ and H3′–H2′/2″ cross-peaks in this spectral region by their compact cross-peak pattern, which is due to the absence of strong geminal H2′/H2″ coupling. The strong H3′–H4′ (and the weak H1′–H2′) *J*-coupling cross-peaks for 4G and 14G confirm that these two residues are in the C3′-*endo* domain—as was concluded from the NOE data of Fig. 1*A*. For residue 9C, on the other hand, the very weak H3′–H4′ COSY cross-peak, the strong H1′–H2′ COSY cross-peak, and the lack of a detectable H2″–H3′ COSY cross-peak (indicated by ×) all indicate that the 9C sugar is in the normal C2′-*endo* domain. In the above discussion, it is important to note that, as in the $(GGA)_2$ motif described previously (3, 4), certain H2′ and H2″ proton pairs have reversed their usual chemical shifts, with the H2″ chemical shifts of 3G and 4G, as well as 13G and 14G, now more upfield than their geminal H2′. Furthermore, as in our previously described GCA motif loop structure (8), the G8 H2″ chemical shift (but not that of 9C), is also more upfield than its H2′. It is worth noting that there is never any ambiguity in distinguishing the H2′ from the H2″ resonance since, regardless of the sugar conformation, the H2″ proton always has a stronger NOE to H1′ than does the H2′.

Fig. 3 shows the ^{31}P–^{1}H correlation spectrum obtained in the inverse-detection mode. The horizontal lines connect the $(n)^{31}$P–$(n-1)$H3′ cross-peak to the $(n)^{31}$P–(n)H4′ cross-peak. Fig. 3 corroborates and verifies the assignments for the H3′ and H4′ resonances discussed above. The unusual H4′ chemical shifts (≈2 ppm) and H5′ chemical shifts (≈3 ppm) for 4G, 14G, and 9C are indicated by arrows. Thus the anomalously shifted H4′ resonances of 4G and 14G in the stem $(GGA)_2$ motif, and of 9C in the loop-GCA-turn motif have been established three ways—by through-space NOEs, by through-bond ^{1}H–^{1}H couplings, and by ^{31}P–^{1}H *J*-couplings. Unusual H4′ chemical shifts have also been observed in Z-DNA (14), in a two-residue hairpin loop (15), and in a one-residue GAA hairpin loop (16).

The ^{31}P–^{1}H *J*-couplings shown in Fig. 3 also provide important information on the backbone conformation. Except for residue 10A, all the intranucleotide four-bond n^{31}P–nH4′ *J*-coupling cross-peaks are detectable, indicating that the P-O5′–C5′–C4′–H4′ part of the backbone linker is in a planar "W" form conformation (17, 18)—i.e., the torsion angle β is in the *trans* domain and γ is in the g^{+} domain for all residues except 10A. The 10A ^{31}P–H4′ *J*-coupling cross-peak was not detected, and the intermediate-strength NOEs from 10AH2′ to 10AH5″ and from 10AH2″ to 10AH5′ (data not shown) indicate that the γ torsion angle for 10A is in the *trans* domain instead of the g^{+} domain (the nH2′–nH5″/H5′ NOEs for all other residues were either not detected or very weak). Since no $(n-1)$H2′–$(n)^{31}$P *J*-coupling peaks were detected, the $(n-1)$H2′–$(n-1)$C2′–$(n-1)$C3′–$(n-1)$O3′–(n)P linkage does *not* possess the planar "W" conformation, indicating that none of the ε torsion angles (at least for the 15 C2′-*endo* sugars where the C2′–C3′ torsion angle is known) are in the g^{+} domain (19). Last, since none of the ^{31}P resonances are shifted downfield below -3.4 ppm, it appears that none of the α or ζ torsions are in the *trans* conformation (7, 20).

Structure Determination

Using distance bounds derived from NOE build-up rates and generous backbone dihedral angle bounds derived from the

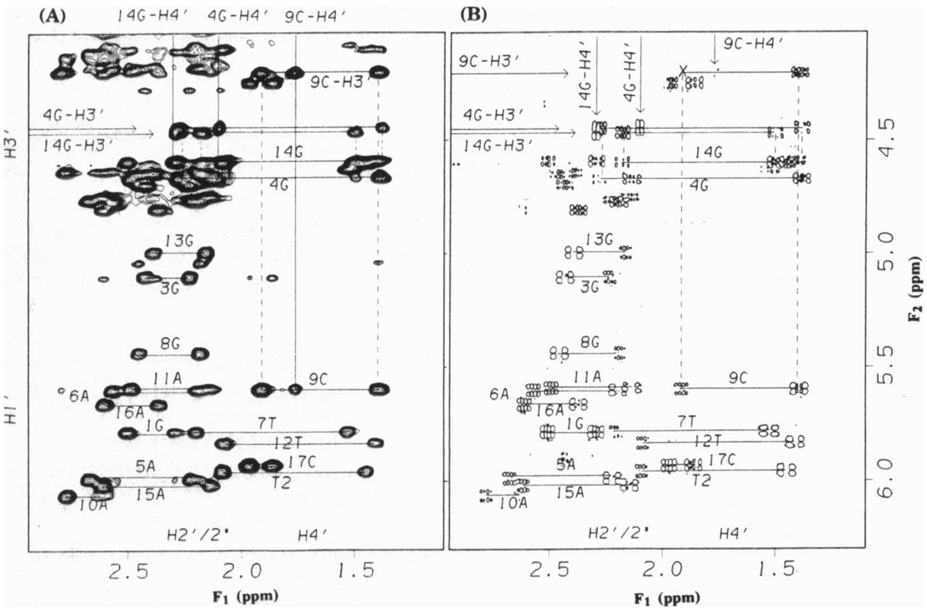

FIG. 2. The expanded NOESY (*A*) and DQF-COSY (*B*) spectra of GTGGAATGCAATGGAAC in the H1'/3'–H2'/2"/4' region. Note that, unlike in normal B-form DNA, the H2" resonances of 3G, 4G, 13G, 14G, and 8G (but not 9C) are upfield of their H2' chemical shifts.

^{31}P–^{1}H correlation spectrum, DG was used to generate initial structures by using the program DGII (Biosym Technologies). Individual sugar conformations were constrained by intrasugar ^{1}H–^{1}H distances together with H1'/H3'–aromatic proton distances and loose bounds from vicinal proton coupling. The glycosidic χ torsion angles were constrained by H8/H6–H2'/H2"/H1'/H3' distances and H5–H1'/H2'/H2" distances. Base–base stacking was constrained by (n)H6/H8–(n + 1)H6/H8 and (n)H6/H8–(n + 1)H5 constraints. Generous distance bounds (1.7–2.5 Å) were used for hydrogen bonds. The backbone conformation was constrained by qualitative analysis of the ^{31}P–^{1}H correlation spectrum, as described above. Vicinal proton–proton *J*-coupling information was used in only a very qualitative way, and only to supplement the NOE data. We have shown that the use of methine–methylene *J*-coupling data to determine precise sugar conformations is questionable for biopolymers in the slow-tumbling regime (21–23) due to dipolar modulation of the measured ^{1}H–^{1}H splitting.

The initial DG structures were further refined by molecular dynamics/mechanics and back-calculation refinement using the simulation program BIRDER (12). The back-calculated NOESY spectra of the refined final structures match the experimental NOESY data very well, with *R* factors of ≈0.25, and all the final structures were highly converged. Fig. 4 *Upper* shows 10 superimposed structures that were independently embedded from, and annealed against, the final bounds file. Fig. 4 *Lower* shows the final structures obtained by further refinement of the above structures by restrained molecular dynamics and energy minimization (MD/EM) using the AMBER force field. The pairwise rms deviations within these families of structures are ≈1.5 ± 0.5 Å before MD/EM refinement and ≈0.5 ± 0.3 Å after MD/EM refinement.

Structural Features

The heptadecamer GTGGAATGCAATGGAAC forms an interesting hairpin with several unusual structural features. Fig. 5 shows stereo views of one of the final structures in three orientations. In these stereo views the Watson–Crick A·T pairs are blue, all guanosine residues are yellow, the adenines of sheared G∘A pairs are magenta, and the cytidine of the single-residue loop is white (the terminal residues 1G and 17C are not shown). As shown in the Fig. 5*A* view into the major groove of the (GGA)$_2$ segment, the bases 4G and 14G do not form a base pair; instead they stack on each other intercalatively. The cross-strand G|G stack in turn stacks on the G residues of the flanking sheared G∘A pairs above and below, leading to a continuous four-guanine stack. The deoxyribose

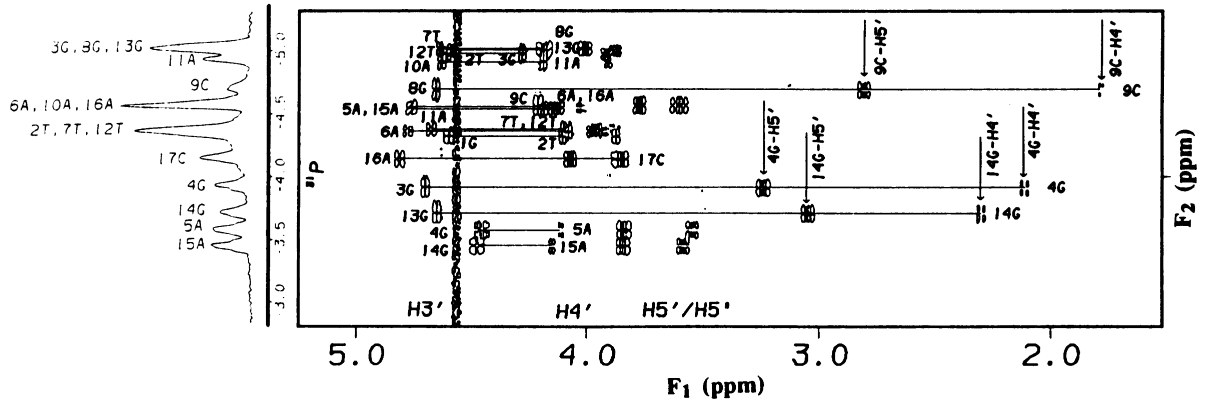

FIG. 3. The ^{31}P–^{1}H correlation spectrum of GTGGAATGCAATGGAAC detected in the inverse mode. The (n − 1)H3'-(n)^{31}P and (n)H4'-(n)^{31}P cross-peaks are connected by horizontal lines. The H4' and H5' resonances of 4G, 14G, and 9C are specifically indicated by arrows.

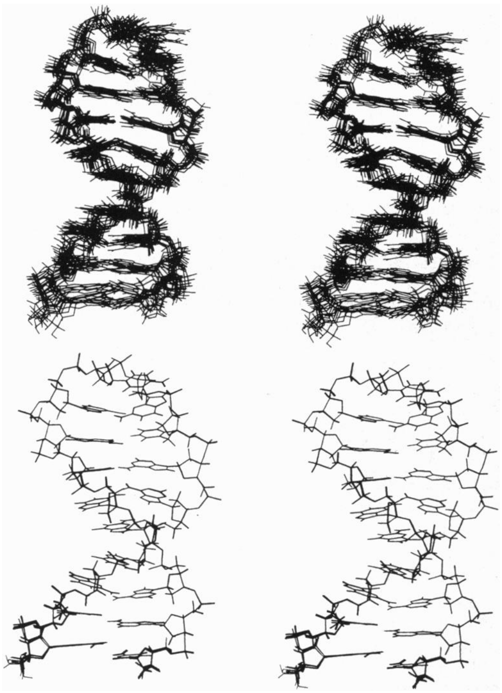

FIG. 4. Wide-eye stereo views of 10 superimposed structures produced by embedding from and refining against the distance bounds (*Upper*) and then further refined by molecular dynamics (*Lower*).

sugars of the unpaired 4G and 14G residues fulfill an interesting potential structure-stabilizing function by stacking on the purine rings of 15A and 5A, with the O4′ lone pair interacting with the adenine heterocycles—as clearly shown in Fig. 5*B*, where the O4′ atoms of 4G, 14G, and 9C are red. Stabilization of such lone pair–base interactions by $n \rightarrow \pi^*$ hyperconjugation has been discussed recently (24). Thus the stem of this hairpin contains the same $(GGA)_2$ motif found in the bimolecular self-paired $(TGGAA)_2$ pentamer repeats (3, 4).

In the single cytidine loop, the base of 9C stacks on the 8G base of the closing sheared G∘A pair, while the sugar O4′ atom of 9C stacks on the following 10A base in a manner similar to the 4GO4′/15A and 14GO4′/5A "interstrand" interactions—see Fig. 5 *B* and *C*. The loop structure in this hairpin is thus virtually identical to our previous GCA loop structure determined in a hairpin containing a simpler stem sequence devoid of $(GGA)_2$ motifs (8). Although the "double-stranded" $(GGA)_2$ motif in the stem and the (GCA)-turn motif in the loop have quite different geometries and quite different backbone traces, they share sugar O4′–base stacking as a common structural feature. Fig. 5*C* also reveals that the amino protons of the two unpaired guanines 4G and 14G (small yellow spheres) are very likely to be involved in hydrogen bonding to the opposite-strand phosphate oxygens of 15A and 5A (small magenta spheres), respectively. Even though there is no direct experimental evidence for this hydrogen bond, which arose spontaneously during the embedding and refining process, its proposed stabilization effect nicely explains our observation that the $(GGA)_2$ motif is more stable in a duplex than the analogous $(GAA)_2$ motif—the latter forming a hairpin–duplex equilibrium containing predominantly hairpin in most sequence contexts (9).

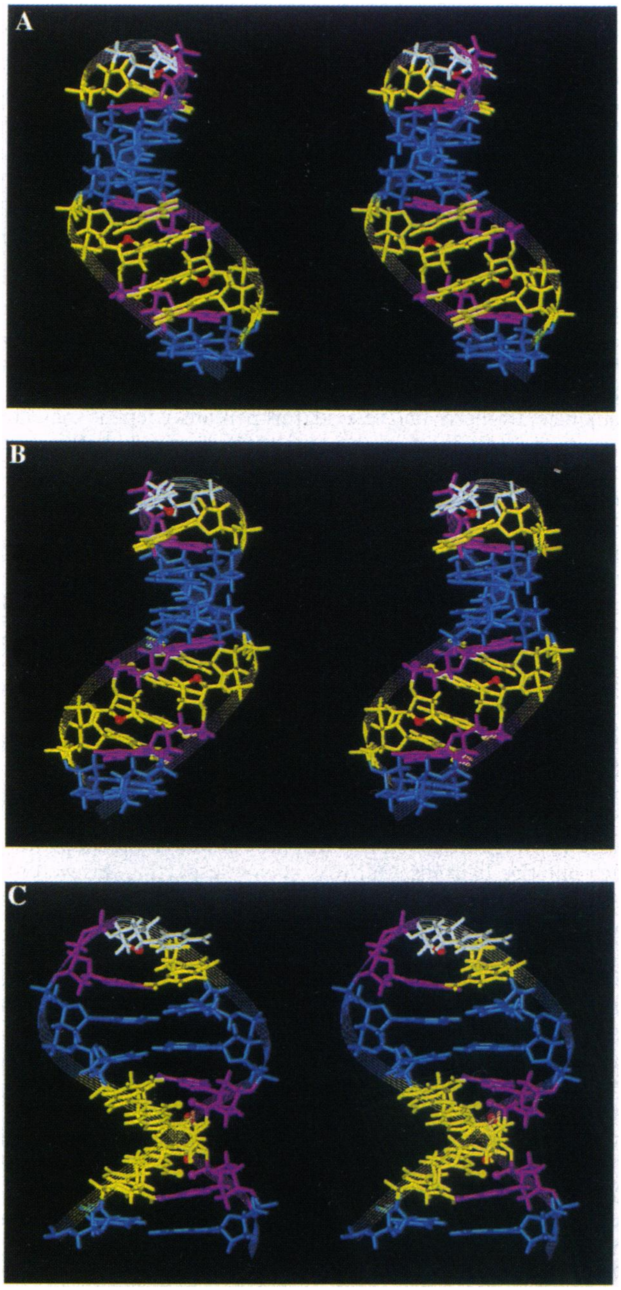

FIG. 5. Wide-eye stereo view of one of the final refined hairpin structures in three orientations. Watson–Crick A·T pairs are blue; deoxyguanosines are yellow; deoxyadenosines in sheared pairs are magenta; and the deoxycytidine residue in the loop is white (the terminal 1G and 17C are not shown). In *A*, the major groove of the stem at the $(GGA)_2$ region (i.e., the [d(3G-4G-5A)∘d(13G-14G-15A)] segment) is toward the reader. In this perspective, the costacking of the four guanine bases is seen most clearly. In *B*, the minor groove at the $(GGA)_2$ section of the stem is toward the reader. The (deoxyribose) O4′–base interactions are clearly shown in this orientation; the O4′ atoms of 4G, 14G, and 9C are red. The geometry of the loop is best seen in *C*; the base of 9C stacks on the base 8G of the closing sheared G∘A pair, while the sugar O4′ atom of 9C stacks on the 10A base. *C* also shows the hydrogen bonding of the amino protons of 4G and 14G (small yellow spheres) to the "cross-strand" phosphate oxygens of 15A and 5A (small magenta spheres), respectively.

In terms of sugar conformations, an interesting structural feature is that the sugars of residues 4G and 14G in the stem $(GGA)_2$ motif are in the C3′-*endo* or N conformation ($P = 25°$ ± 3°), while the 9C sugar in the GCA turn loop is in the normal C2′-*endo* or S conformation—as are all the remaining sugars in this DNA hairpin. The stretching of the GpA step in the $(GGA)_2$ motif (to accommodate the intercalation of the opposite-strand guanine base) is mainly accomplished by the change in δ on going from a C2′-*endo* to a C3′-*endo* conformation in the sugars of the two middle guanosines, with the backbone conformation deviating only slightly from standard B-DNA values, as discussed previously (3, 4). In contrast, the 180° turn made by the GCA loop is accomplished by changing the phosphate backbone at the CpA step from the $\varepsilon(t)$ $\zeta(g^-)$ $\alpha(g^-)$ $\beta(t)$ $\gamma(g^+)$ range found in typical B-DNA to an $\varepsilon(t)$ $\zeta(g^+)$ $\alpha(g^+)$ $\beta(g^-)$ $\gamma(t)$ conformation, with the loop sugar conformations remaining C2′-*endo*. The individual ε, ζ, α, β, and γ values for the 9C-p-10A backbone conformation in the loop are −170°, 124°, 72°, −108°, and −179°, respectively. From crystallographic (25–27) and NMR (19, 28–30) studies of various sized loops, it would appear that the last (3′) loop phosphodiester that links it to the 5′-end of the stem is generally close to the $\zeta(g^+)$ $\alpha(g^+\text{-}t)$ $\beta(t\text{-}g^-)$ conformation.

Biological Implications

Grady *et al.* (1) demonstrated two important properties of the human satellite III $(TGGAA)_n$ tandem repeat; first, it occurs at human centromeres and second, this pentamer repeat somehow pairs with itself to form a purine-rich duplex with the same stability as the corresponding Watson–Crick duplex formed with the complementary pyrimidine strand—suggesting that the $(TGGAA)_n$ strand might possibly function alone *in vivo*. We showed that the bimolecular $[(TGGAA)_n]_2$ duplex was an antiparallel duplex with two self-paired $(GGA)_2$ motifs per turn; each motif contains four costacked guanines whose N7, O^6, N1H, and N^2H are exposed to solvent and available for liganding (3, 4). If, at some point in the cell cycle, the antiparallel $[(TGGAA)_n]_2$ duplex functions *in vivo*, then folding back of the pentamer repeats into hairpins at several loci would become important, since the TGGAA runs occur only on one strand. Since TGCAA is the most frequent variant among TGGAA repeats (2), especially in the centromeric repeats immediately adjacent to and contiguous with human α satellite DNA (31), we reasoned that this variant pentamer might function to "turn the corner" at the loop end of such putative hairpins. Our reasons for suspecting this were the extremely "tight turn" loops (containing only one residue in the loop) formed by the triplet GCA (8). The present structure confirms that these suspicions were well founded in that the GTGGAATGCAATGGAAC heptadecamer indeed forms an extremely stable hairpin (T_m = 63°C) in which the GCA triplet forms a tight-turn loop of one cytidine closed by a sheared G∘A pair, and the first and last TGGAA pentamers form an antiparallel "duplex" containing the intercalative GGGG stack motif. Perhaps the most dramatic result, and a testament to the loop-nucleating power of GCA triplets, is the observation that a single change from GTGGAATGGAATGGAAC to GTGGAATGCAATGGAAC converts a very stable bimolecular duplex to a very stable hairpin. It may be for good reason that TGCAA is the principal pentamer variant found in satellite III, since TGTAA and TGAAA pentamers (at least in NAATGNAATG contexts) form duplex–hairpin equilibrium mixtures (9).

At this point, it is not clear whether $(TGGAA)_n$TGCAA$(TGGAA)_n$ stem–hairpin structures actually form *in vivo* and, if they do, what roles these fold-back structures play—neither is it clear what happens to the complementary $(TTCCA)_n$ strand. However, it is not unreasonable that these noncoding centromeric sequence repeats might form special structures with unique functional properties. Because of their clustered patches of exposed guanine hydrogen bond donors and acceptors, such roles might include direct interaction with the mitotic spindle microtubule filaments in centromere capture or interaction with kinetochore proteins at the protein–chromosome interface, or even interaction with similar proximal repeats in the condensation of the centromere. In any case, it appears that the distribution of tight turn-inducing TGCAA pentamers in TGGAA repeats is likely to "sculpt" such repeats into unique florets of minihairpins.

Regarding the complementary pyrimidine strand, assuming the centromere replicates normally, preliminary imino proton NMR studies indicate that $d(TTCCA)_n$ is an unstructured random coil at neutral pH but takes on a discrete secondary structure under acidic conditions (data not shown). Efforts have been made by others to identify the global structure of these centromeric DNA repeats by using different experimental methods; recently fold-back structures in centromeric dodeca-satellite G-strands were directly visualized by electron microscopy (32). Nevertheless, except for simple organisms such as budding yeast (33–35), our present understanding of centromere structure is rather limited. To fully understand the functional aspects of centromeres in mitosis and meiosis, it will be increasingly important to elucidate the types of structural motifs that can be formed by their constituent satellite repeat units.

This work was supported by National Institutes of Health Grants GM32681 and GM52883 to B.R.R.

1. Grady, D. L., Ratliff, R. L., Robinson, D. L., McCanlies, E. C., Meyne, J. & Moyzis, R. K. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 1695–1699.
2. Prosser, J., Frommer, M., Paul, C. & Vincent, P. C. (1986) *J. Mol. Biol.* **187,** 145–155.
3. Chou, S.-H., Zhu, L. & Reid, B. R. (1994) *J. Mol. Biol.* **244,** 259–268.
4. Zhu, L., Chou, S.-H. & Reid, B. R. (1995) *J. Mol. Biol.* **254,** 623–637.
5. Li, Y., Zon, G. & Wilson, W. D. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 26–30.
6. Chou, S.-H., Cheng, J.-W. & Reid, B. R. (1992) *J. Mol. Biol.* **228,** 138–155.
7. Chou, S.-H., Cheng, J.-W., Fedoroff, O. Yu. & Reid, B. R. (1994) *J. Mol. Biol.* **241,** 467–479.
8. Zhu, L., Chou, S.-H. & Reid, B. R. (1995) *Nat. Struct. Biol.* **2,** 1012–1017.
9. Chou, S.-H., Zhu, L. & Reid, B. R. (1996) *J. Mol. Biol.* **259,** 445–457.
10. Hare, D. R. & Reid, B. R. (1986) *Biochemistry* **25,** 5341–5350.
11. Sklenar, V., Miyashiro, H., Zon, G., Miles, H. T. & Bax, A. (1986) *FEBS Lett.* **208,** 94–98.
12. Zhu, L. & Reid, B. R. (1995) *J. Magn. Reson. B* **106,** 227–235.
13. Hare, D. R., Wemmer, D. E., Chou, S.-H., Drobny, G. & Reid, B. R. (1983) *J. Mol. Biol.* **171,** 319–336.
14. Giessner-Prettre, C., Pullman, B., Tran-Dinh, S., Neumann, J.-M., Huynh-Dinh, T. & Igolen, J. (1984) *Nucleic Acids Res.* **12,** 3271–3281.
15. Orbons, L. P. M., van der Marel, G. A., van Boom, J. H. & Altona, C. (1987) *Eur. J. Biochem.* **170,** 225–239.
16. Hirao, I., Kawai, G., Yoshifumi, N., Ishido, Y., Watanabe, K. & Miura, K. (1994) *Nucleic Acids Res.* **22,** 576–582.
17. Sarma, R. H., Mynott, R. J., Wood, D. J. & Hruska, F. E. (1973) *J. Am. Chem. Soc.* **95,** 6457–6459.
18. Altona, C. (1982) *Recl. Trav. Chim. Pays-Bas. Belg.* **101,** 413–433.
19. Blommers, M. J. J., van de Ven, F. J. M., van der Marel, G. A., van Boom, J. H. & Hilbers, C. W. (1991) *Eur. J. Biochem.* **201,** 33–51.
20. Gorenstein, D. G., Schroeder, S. A., Fu, J. M., Metz, J. T., Roongta, V. & Jones, C. R. (1988) *Biochemistry* **27,** 7223–7237.
21. Harbison, G. S. (1993) *J. Am. Chem. Soc.* **115,** 3026–3027.
22. Zhu, L., Reid, B. R., Kennedy, M. & Drobny, G. P. (1994) *J. Magn. Reson. A* **111,** 195–202.
23. Zhu, L., Reid, B. R. & Drobny, G. P. (1995) *J. Magn. Reson. A* **115,** 206–212.
24. Egli, M. & Gessner, R. V. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 180–184.
25. Sussman, J. L., Seeman, N. C., Kim, S.-H. & Berman, H. M. (1972) *J. Mol. Biol.* **66,** 403–421.
26. Suck, D., Manor, P. C. & Saenger, W. (1976) *Acta Crystallogr. Sect. B* **32,** 1727–1737.
27. Frederick, C. A., Coll, M., van der Marel, G. A., van Boom, J. H. & Wang, A. H.-J. (1988) *Biochemistry* **27,** 8350–8361.
28. Blommers, M. J. J., Haasnoot, C. A. G., Walters, J. A. L. I., van der Marel, G. A., van Boom, J. H. & Hilbers, C. W. (1988) *Biochemistry* **27,** 8361–8369.
29. Boulard, Y., Gaballo-Arpa, J., Cognet, J. A. H., Bret, M. L., Guy, A., Téoule, R., Guschlbauer, W. & Fazakerlay, G. V. (1991) *Nucleic Acids Res.* **19,** 5159–5167.
30. Mooren, M. M. W., Pulleyblank, D. E., Wijmenga, S. S., van de Ven, F. J. W. & Hilbers, C. W. (1994) *Biochemistry* **33,** 7315–7325.
31. Vissel, B., Nagy, A. & Choo, K. H. A. (1992) *Cytogenet. Cell Genet.* **61,** 81–86.
32. Ferrer, N., Azorin, F., Villasante, A., Guitiérrez, C. & Abad, J. P. (1995) *J. Mol. Biol.* **245,** 8–21.
33. McGrew, J., Diehl, B. & Fitzgerald-Hayes, M. (1986) *Mol. Cell. Biol.* **6,** 530–538.
34. Ng, R. & Carbon, J. (1987) *Mol. Cell. Biol.* **7,** 4522–4534.
35. Hyman, A. A., Middleton, K., Centola, M., Michison, T. J. & Carbon, J. (1992) *Nature (London)* **359,** 533–536.