GNA Trinucleotide Loop Sequences Producing Extraordinarily Stable DNA Minihairpins†

Satoko Yoshizawa,‡,§ Gota Kawai,‡ Kimitsuna Watanabe,‡ Kin-ichiro Miura,|| and Ichiro Hirao*,⊥

Department of Chemistry and Biotechnology, Faculty of Engineering, The University of Tokyo, Hongo, Tokyo 113, Japan, Institute for Biomolecular Science, Gakushuin University, Mejiro, Tokyo 171, Japan, and Department of Chemistry, Indiana University, Bloomington, Indiana 47405

Received July 15, 1996; Revised Manuscript Received February 3, 1997⊗

ABSTRACT: d(GCGAAAGC) and d(GCGAAGC) fragments form extraordinarily stable DNA minihairpins containing only two G-C base pairs and a GAAA or GAA loop, respectively, with a T_m of 76 °C. These sequences are frequently found in some important regions such as replication origins and promoter regions for transcription. We examined all 64 possible DNA fragments, d(GCNNNGC), in which the triloop region of the d(GCGAAGC) minihairpin was randomized and found that only four fragments, d(GCGNAGC) (N = A, G, C, or T), formed extraordinarily stable minihairpins as shown by their gel mobility and resistance to a single-stranded DNA-specific exonuclease. Structural and thermodynamic analyses suggest that the extraordinary stability is caused by a unique structural property of the trinucleotide sequences corresponding to the GNA loop.

Hairpins are a basic unit of nucleic acid structure and have been shown to play important roles in many biological processes. Hairpins consist of a single-stranded loop region closed by a base-paired stem. Normally the stability of a hairpin structure increases with the number of the base pairs in the stem region. Thus, hairpin structures can be predicted when long complementary regions of sequence are identified.

Recent studies, however, have shown that some DNA or RNA hairpins having loops with certain sequences are unusually stable (Varani, 1995). DNA fragments such as d(GCGAAAGC) or d(GCGAAGC) form extraordinarily stable DNA minihairpins (T_m = 76 °C)[1] with only two G-C base pairs (Hirao et al., 1988, 1989, 1992). The 3-D structure of the d(GCGAAGC) fragment determined by NMR spectroscopy revealed that the hairpin structure is folded back between A_4 and A_5 and is stabilized by one extra non-Watson−Crick G-A base pair in the GAA loop as well as by extensive base-stacking interactions (Figure 1) (Hirao et al., 1994).

The d(GCGAAGC) and d(GCGAAAGC) minihairpin sequences occur frequently in biologically important regions, such as replication origins and promoter regions. For example, the GAAA-loop hairpin exists in the replication origins of G4 phage single-stranded DNA (Godson et al., 1978; Hirao et al., 1990); the GAA-loop hairpin exists in the replication origins of phage ϕX174 (Arai et al., 1981), RSF1010 (Miao et al., 1993), and herpes simplex virus DNA (Elias & Lehman, 1988) and in the promoter regions of phage N4 double-stranded DNA (Glucksmann et al., 1992) and *Escherichia coli* heat-shock gene (Cowing et al., 1985).

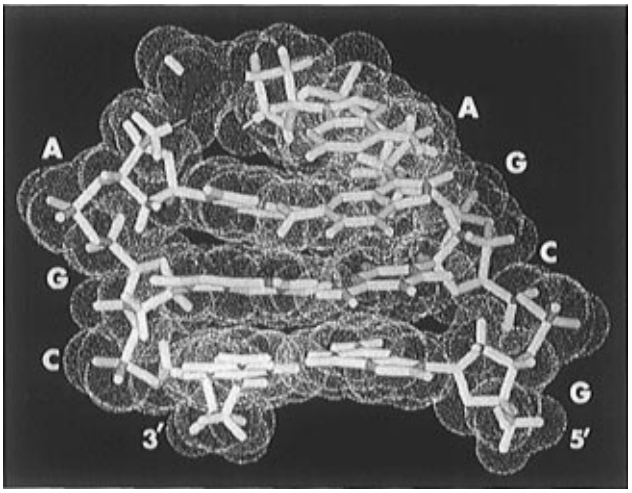

FIGURE 1: Stick model of the d(GCGAAGC) minihairpin as determined in Hirao et al. (1994) with the van der Waals surfaces. The 4′H of A_4 that shows a large upfield shift in the NMR spectrum is colored with red. The three bonds, around which torsion angles differ significantly from usual B-DNA, are shown in blue.

The minihairpin sequences may contribute to form stable hairpin or cruciform structures at proper positions of DNA molecules. For example, the region of the replication origin of G4 phage DNA forms two interconvertible hairpin structures; one consists of a GAAA tetraloop and a four-base-pair stem similar to the GCGAAAGC minihairpin, and the other consists of an AAAGC pentaloop and an eight-base-pair stem (Hirao et al., 1990). Although the secondary structure of this region has been predicted to be the latter hairpin, the GAAA-loop hairpin (T_m = 89 °C), in spite of the shorter stem length, is more stable than the AAAGC-loop hairpin (T_m = 86 °C). Thus, a loop moiety as well as

† This investigation was supported in part by The Mochida Memorial Foundation for Medical and Pharmaceutical Research.

* Author for correspondence. Present address: Yokoyama CytoLogic Project, ERATO, The Institute of Physical and Chemical Research (RIKEN), Hirosawa 2-1, Wako-shi, Saitama 351-01, Japan.

‡ The University of Tokyo.

§ Present address: Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA 95064.

|| Gakushuin University.

⊥ Indiana University.

⊗ Abstract published in *Advance ACS Abstracts,* April 1, 1997.

[1] Abbreviations: T_m, melting temperature; NOESY, nuclear Overhauser effect spectroscopy; DQF-COSY, double-quantum-filtered correlated spectroscopy; HOHAHA, homonuclear Hartmann−Hahn spectroscopy; FID, free induced decay.

S0006-2960(96)01738-2 CCC: $14.00

the stem region potentiates the hairpin formation, and the loop sequence is important for the prediction of the hairpin structure.

The existence of such unusual minihairpins may imply an unknown structural versatility of DNA that is related to biological functions. However, the present limited knowledge of the relationship between the loop sequence and the stability of minihairpins makes it difficult to predict the formation of hairpin structures from base sequences alone. Here, we determine which trinucleotide loop sequence folds into minihairpins and why those hairpins are so stable. We synthesized and characterized all the 64 possible DNA fragments, d(GCNNNGC), to determine which of them form stable trinucleotide-loop minihairpins and studied the structure and thermodynamics of the stable hairpins that we determined.

MATERIALS AND METHODS

T4 DNA Polymerase Digestion Assay. Each 5′-labeled DNA fragment (0.0025 OD_{260nm} unit) was incubated in a buffer containing 125 mM Tris-HCl (pH 8.0), 25 mM $MgCl_2$, and 25 mM 2-mercaptoethanol with or without 0.01 unit of T4 DNA polymerase (Takara) for 1 h at 37 °C. An equal volume of 10 M urea was added to the assay mixture. The DNA fragments were separated on 20% polyacrylamide gel containing 7 M urea and analyzed by a Bio-imaging analyzer BAS2000 (Fuji Film).

Mobility of DNA Fragments. Each 5′-labeled DNA fragment was separated by 20% polyacrylamide gel electrophoresis containing 7 M urea, 1 mM EDTA, and 0.045 M Tris−borate (pH 8.0) at 30 °C with xylene cyanol. The mobility of the fragment was divided by the mobility of the xylene cyanol on the same lane. To eliminate the effect of the difference in base composition of each fragment, each mobility value was normalized by a correction factor $[(Cf)_{GCNNNGC}]$ calculated from the following formula as described by Frank and Köster (1979)

$$(Cf)_{GCNNNGC} = \{(n_a) \times 1.13 + (n_g) \times 1 + (n_c) \times 1.19 + (n_t) \times 1.06\}/7$$

where n_a, n_g, n_c, and n_t show the respective number of the adenines, guanines, cytosines, and thymines in the fragment.

T_m Measurement and Thermodynamic Studies of DNA Fragments. For simple T_m measurements, each DNA fragment (1.5 OD_{260nm} unit) was dissolved in 1 mL of buffer A [50 mM sodium cacodylate (pH 7.0) and 0.1 M NaCl]. Melting profiles were obtained at 260 nm using a Gilford spectrophotometer, Response II, at a heating rate of 0.5 °C/min using 1 cm cuvettes. T_m values were calculated by the first derivatives of the melting curves.

Thermodynamic parameters were determined using the standard method (Puglisi & Tinoco, 1989) from the UV absorbance melting curves. The values listed in Table 2 were obtained from the melting studies performed in buffer B [1 M NaCl, 1 mM EDTA, and 10 mM sodium phosphate (pH 7.0)]. Buffer B was chosen to utilize the nearest-neighbor parameters for the DNA stem determined by Breslauer et al. (1986). Thermodynamic parameters for loop folding were thus calculated by subtracting the nearest-neighbor interactions from those measured for the hairpin.

NMR Studies of d(GCGNAGC) Hairpins. The purified DNA fragments were treated with Dowex cation-exchange resin (pyridinium form), followed by treatment with the resin (Na^+ form), and dissolved in 99.9% D_2O containing 15 mM NaCl and 1 mM sodium phosphate (pD 7.0). NMR experiments were performed on a Bruker AMX-500 spectrometer. All spectra were obtained at a probe temperature of 25 °C. For one-dimensional experiments, free induction decays (FIDs) of 16K data points were accumulated, and spectra of 32K real data points were obtained. A line broadening of 0.2 Hz was applied prior to Fourier transformation. Two-dimensional experiments were performed by accumulating 512 FIDs of 2K data points with the time-proportional phase incrementation method for phase-sensitive detection of the t_1 dimension (Wüthrich, 1986), and spectra of 2K × 1K real data points were obtained. A $\pi/2$-shifted sine-bell window function was applied for each dimension and then two-dimensional Fourier transformation was performed. NOESY (mixing time 200 ms), HOHAHA (99 ms), and DQF-COSY (^{31}P decoupling for both dimensions) spectra of each d(GCGNAGC) fragment were obtained.

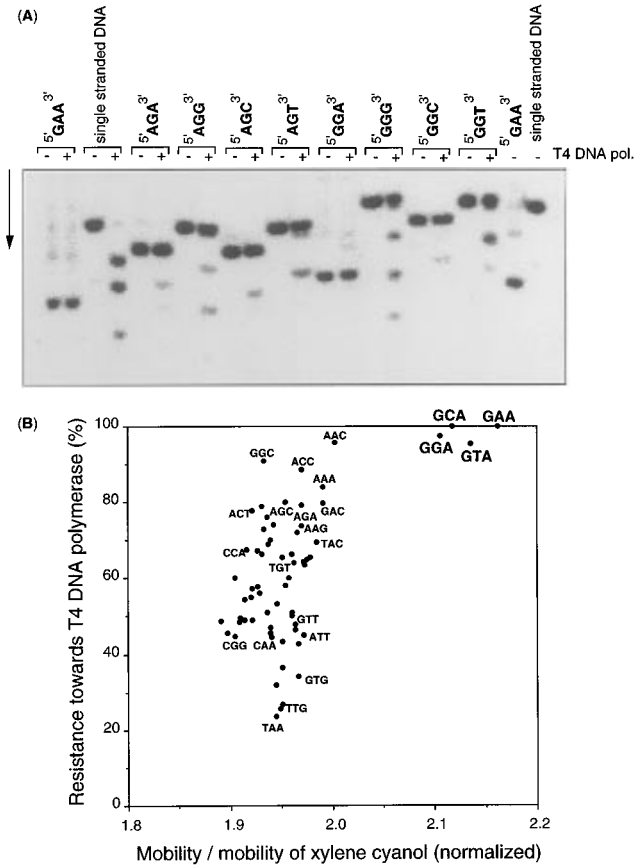

FIGURE 2: (A) Example of the assay for detection of stable hairpins. Autoradiograph of a polyacrylamide gel containing urea showing differences in the mobility and in resistance of d(GCNNNGC) fragments toward T4 DNA polymerase. The sequences of the triloop regions are shown above the panel. 5′GCGAAGC3′ and 5′CGACGAG3′ are used as controls for stable minihairpins and non-hairpin-forming oligomers and denoted as GAA and single-stranded DNA, respectively. A 0.025 OD_{260nm} unit of each DNA fragment was ^{32}P-labeled at the 5′-termini and incubated without (−) or with (+) T4 DNA polymerase. (B) Summary of assays of the type shown in (A). Sequences of the trinucleotide-loop region of the representative fragments are shown. Oligomer resistance against T4 DNA polymerase was estimated by the density of the band corresponding to the intact fragment. The ratio of the mobility of the oligomer to that of xylene cyanol was normalized as shown in Materials and Methods.

Table 1: ^{1}H Chemical Shifts (ppm) of Nonexchangeable Protons of the d(GCGNAGC) Hairpins Measured at 0.54 mM in 0.15 M NaCl and 10 mM Sodium Phosphate (pD 7.0) at 25 °C

	H1′	H2′/H2″	H3′	H4′	H5′/H5″	H8/H6	H2/H5
				(a) GCGAAGC Hairpin			
G_1	6.01	2.57 2.77	4.85	4.26	3.75	8.00	
C_2	6.10	1.69 2.37	4.85	4.29	4.13	7.08	5.08
G_3	5.42	2.73 2.60	4.96	4.51	4.16	8.13	
A_4	6.03	2.38 2.31	4.63	**2.20**	3.14 3.46	8.17	8.14
A_5	6.35	2.97	4.89	4.41	3.87 4.04	8.08	8.20
G_6	5.61	2.66	4.96	4.44	4.24 4.35	8.07	
C_7	6.34	2.20 2.32	4.51	4.19	4.12 4.28	7.55	5.45
				(b) GCGTAGC Hairpin			
G_1	6.04	2.58 2.79	4.87	4.25	3.76	8.02	
C_2	6.14	1.68 2.38	4.84	4.29	4.14	7.18	5.21
G_3	5.97	2.75 2.63	4.96	4.59	4.16 4.23	8.20	
T_4	5.84	1.81 2.14	4.47	**2.12**	3.31 3.44	7.16	(1.43)[a]
A_5	6.38	2.85 2.96	4.88	4.39	3.80 3.98	8.07	8.18
G_6	5.63	2.65	4.96	4.43	4.21 4.36	8.07	
C_7	6.33	2.19 2.29	4.52	4.16	4.10 4.28	7.54	5.47
				(c) GCGCAGC Hairpin			
G_1	6.04	2.58 2.79	4.86	4.24	3.74	8.02	
C_2	6.14	1.79 2.48	4.86	4.30	4.15	7.15	5.15
G_3	5.86	2.77 2.56	4.95	4.55	4.18 4.22	8.21	
C_4	5.90	1.66 2.20	4.46	**2.04**	3.28 3.45	7.38	5.35
A_5	6.36	2.86 2.96	4.86	4.39	3.80 4.00	8.07	8.18
G_6	5.63	2.65	4.95	4.42	4.21 4.35	8.07	
C_7	6.33	2.18 2.27	4.51	4.15	4.11 4.24	7.55	5.46
				(d) GCGGAGC Hairpin			
G_1	6.03	2.56 2.76	4.85	4.23	3.74	8.01	
C_2	6.12	1.64 2.30	4.84	4.28	4.12	7.13	5.16
G_3	5.52	2.65 2.46	4.92	4.51	4.13	8.04	
G_4	5.83	2.32	4.55	**2.10**	3.16 3.38	7.84	
A_5	6.38	2.90 2.96	4.88	4.39	3.84 4.00	8.08	8.20
G_6	5.63	2.65	4.96	4.44	4.20 4.35	8.08	
C_7	6.33	2.17 2.27	4.51	4.15	4.12 4.28	7.56	5.47

[a] The value in parentheses for T4 is showing the chemical shift of the methyl group protons of T.

RESULTS

Determination of Stable Triloop Hairpin Sequences. Our previous studies have shown that short DNA fragments such as d(GCGAAGC) and d(GCGAAAGC) form extraordinarily stable hairpins with T_m values of 76 °C. These hairpins are so stable that they show higher mobility not only in native polyacrylamide gels but also in those containing 7 M urea (Hirao et al., 1988, 1989). We have also shown that these minihairpins are resistant to the exonuclease activity of enzymes such as T4 DNA polymerase that specifically digests single-stranded DNAs (Hirao et al., 1990; Yoshizawa et al., 1994). Thus, the formation of the minihairpin was assayed by two distinct methods using two characteristic properties of the minihairpin: (i) faster mobility on a gel containing 7 M urea compared with single-stranded DNA of the same chain length and (ii) resistance to the exonuclease activity of T4 DNA polymerase. Each ^{32}P-labeled d(GCNNNGC) fragment was incubated with or without T4 DNA polymerase and then analyzed by a polyacrylamide gel electrophoresis containing 7 M urea. Figure 2A illustrates an autoradiograph of some of the fragments, and Figure 2B summarizes the results for all 64 fragments. Four fragments containing GTA-, GCA-, GGA-, and GAA-loop sequences, [d(GCGNAGC)], show distinctly fast mobility and strong nuclease resistance. The T_m values of these fragments in a 0.1 M NaCl solution (pH 7.0) (buffer A) are 76, 73, 71, and 70 °C for GAA-, GGA-, GCA-, and GTA-loop minihairpins, respectively. These T_m values are independent of the oligomer concentration between 1.6 and 24 μM. Other than GNA-loop hairpins, some fragments, for example, AAC-, AGA-, AAG-, GGC-, ACC-, and GAC-loop-containing fragments, form minihairpins, although these are less stable (T_m = 47−59 °C).

NMR Studies of d(GCGNAGC) Minihairpins. The structure of the d(GCGAAGC) hairpin has previously been determined using NMR spectroscopy (Hirao et al., 1994) (Figure 1). Here, structures of all the d(GCGNAGC) hairpins are analyzed by NMR experiments.

The chemical shifts of all the protons were assigned (Table 1) by analyzing NOESY, ^{31}P-decoupled DQF-COSY, and HOHAHA spectra. The chemical shifts of the sugar protons of d(GCGTAGC), d(GCGCAGC), and d(GCGGAGC) are quite similar to those of the d(GCGAAGC) hairpin. The chemical shift of H4′ of the fourth residue in each fragment shows a large upfield shift (Table 1) compared to other H4′ signals. The large upfield shift of H4′ of A_4 in the d(GCGAAGC) hairpin can be explained from its 3-D structure, in which H4′ of A_4 (colored red in Figure 1) is located just above the adenine base of A_5.

Similar patterns of cross peaks were also observed in NOESY and DQF-COSY spectra of these d(GCGNAGC) hairpins. No cross peaks were observed between H2″ of the fourth nucleoside and H8 of A_5 in the NOESY spectra (Figure 3), indicating that all fragments are folded back between N_4 and A_5. The NOESY cross peaks of H8−H1′, H8−H2′, and H8-H2″ show that the conformations around the glycoside bonds of all the residues, except for A_5, in d(GCGNAGC) fragments are in the *anti* form. For A_5, strong NOE's were observed for A_5 H8−H1′, H8−H2′, and H8−H2″, suggesting that the conformation around this base wobbles around *anti* and *high-anti* conformations. The coupling constants $J_{H1',H2'}$ and $J_{H1',H2''}$ together with the results

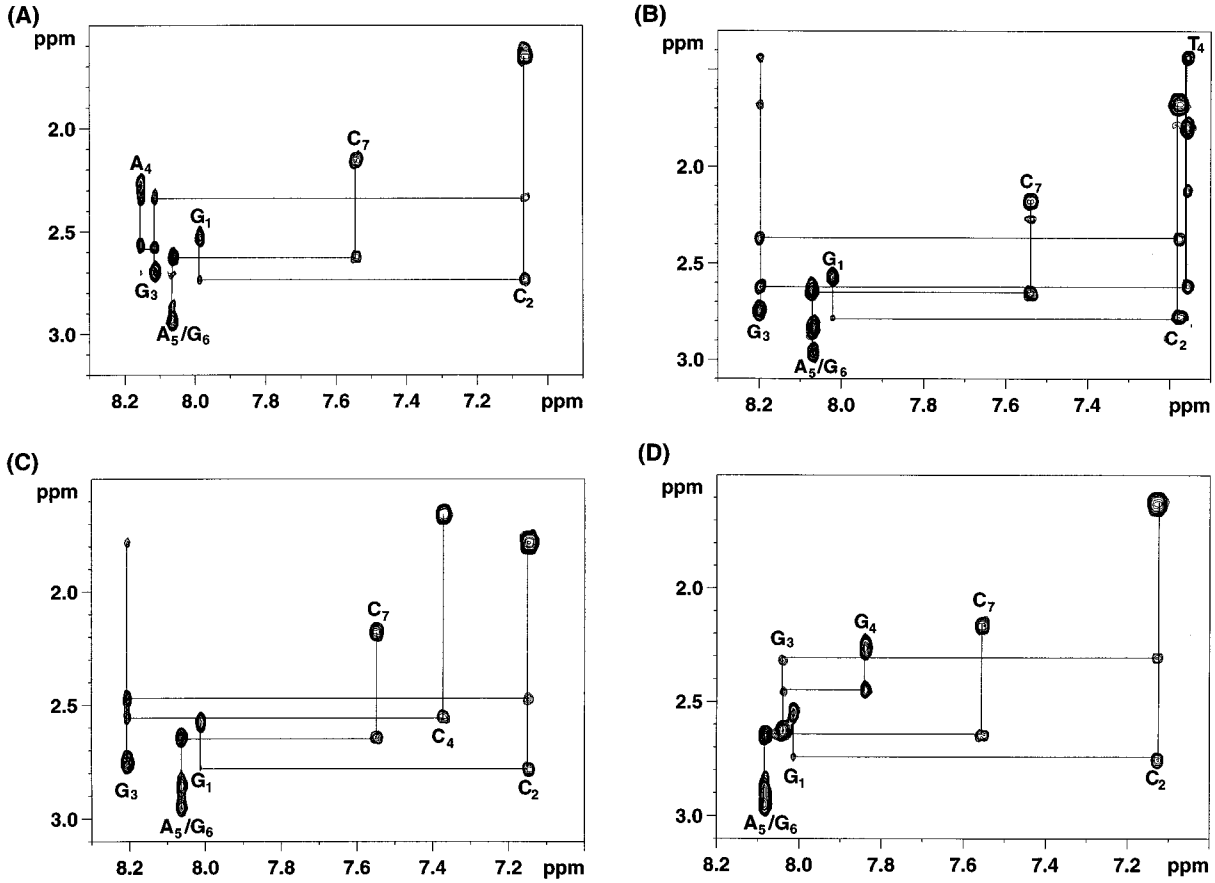

FIGURE 3: Portions of the NOESY spectrum showing the connectivity between sugar proton H2′/H2″ and aromatic proton H8/H6 resonances of the (A) d(GCGAAGC), (B) d(GCGTAGC), (C) d(GCGCAGC), and (D) d(GCGGAGC) hairpins.

from the NOESY spectra show that all residues in d(GCGNAGC) fragments adopt the C2′-*endo* sugar pucker conformation. The coupling constants $J_{H5',P}$, $J_{H5'',P}$, and $J_{H4',P}$ show that the conformations around the C5′−O5′ bond for all residues, except for A_5, adopt the *trans* form. The $J_{H5'',P}$ of A_5 was ∼20 Hz in every d(GCGNAGC) fragment, indicating that the conformation around the C5′−O5′ bonds adopts the *gauche*$^+$ form. ^{31}P-decoupled DQF-COSY spectra show that $J_{H4',H5''}$ of A_5 in every d(GCGNAGC) fragment was ∼10 Hz, indicating that the conformation of H4′−C4′−C5′−H5″ is *trans* and, thus, the conformation of C3′−C4′−C5′−O5′ is also *trans*. Therefore, all the d(GCGNAGC) hairpins are folded back between N_4 and A_5 with a sharp turn.

In the 90% H_2O solution, we observed two sharp imino proton signals at 13.0−13.2 ppm corresponding to two G-C base pairs for all the d(GCGNAGC) fragments at 25 °C. At 5 °C, G_3 was observed as a broad signal around 10.5 ppm (Figure 4). Imino protons of the G in the sheared G-A base pair are found around 9.5−10.5 ppm (Heus & Pardi, 1991; Katahira et al., 1993). The G_3 and N_4 imino protons are not observed at a temperature above 5 °C, because the exchange of the imino protons with the solvent protons becomes faster with the increase in temperature.

Thus, we have concluded that all d(GCGNAGC) fragments form structures similar to that of the minihairpin of d(GCGAAGC) (Figure 1), where the fragment is folded back between A_4 and A_5 that is caused by the rotation around the three bonds of N_4 O3′−P, A_5 O5′−C5′, and A_5 C5′−C4′. These four minihairpins form a sheared G-A base pair in addition to the two G-C Watson−Crick base pairs. Each strand of $G_1C_2G_3N_4$ and $A_5G_6C_7$ forms a B-form helical structure with tight base-stacking interactions. The stability

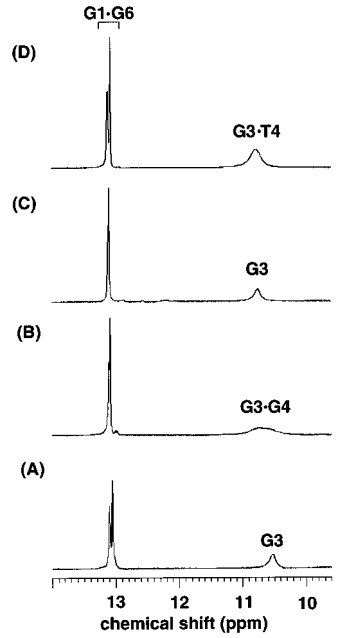

FIGURE 4: Imino proton spectra of d(GCGNAGC) at 5 °C in 15 mM NaCl and 1 mM sodium phosphate (pD 7.0): (A) d(GCGAAGC), (B) d(GCGGAGC), (C) d(GCGCAGC), and (D) d(GCGTAGC) hairpins.

differences among d(GCGNAGC) hairpins correlate well with the stacking stability between the G_3 and the fourth base.

Thermodynamic Studies of d(GCGNAGC) Hairpins. Thermodynamic parameters, ΔG, ΔH, and ΔS, of each of these minihairpins were obtained through UV absorbance melting profiles (Puglisi & Tinoco, 1989), and those of the loop

Table 2: Thermodynamic Parameters for Hairpins and Loop Formation[a]

	hairpin				loop		
sequence	T_m (°C)[b]	ΔH (kcal/mol)	ΔS (eu)	$\Delta G_{37°C}$ (kcal/mol)	ΔH (kcal/mol)	ΔS (eu)	$\Delta G_{37°C}$ (kcal/mol)
GCGAAGC	72.3	−30.8	−89.2	−3.2	−19.7	−62.5	−0.4
GCGGAGC	70.5	−30.6	−89.0	−3.0	−19.5	−62.3	−0.2
GCGCAGC	67.0	−30.4	−89.6	−2.7	−19.3	−62.9	+0.1
GCGTAGC	66.3	−29.1	−85.8	−2.5	−18.0	−59.1	+0.3

[a] The thermodynamic parameters for the hairpins were obtained from the UV absorbance melting profiles measured in 1 M NaCl, 1 mM EDTA, and 10 mM sodium phosphate at pH 7.0 according to Puglisi and Tinoco (1989). The parameters for the loop formation were calculated by subtracting the nearest-neighbor parameters for the 5′GC3′/5′GC3′ stem (Breslauer et al., 1986). [b] The T_m listed in this table is showing the temperature where the calculated proportion of the hairpin and the random form is equal.

folding were calculated by subtracting the effect of the nearest-neighbor interactions estimated for (5′GC3′/5′GC3′) in the stem region (Antao et al., 1991; Breslauer et al., 1986) (Table 2). The $\Delta G_{37°C}$ values for the d[C(GNA)G] loop folding are between −0.4 and 0.3 kcal/mol. It has been shown that the $\Delta G_{37°C}$ values for the loop folding of stable RNA hairpins such as r(UNCG) loop hairpins are much smaller than those of usual hairpins (Antao & Tinoco, 1992; Antao et al., 1991). The $\Delta G_{37°C}$ value for the d(GAA) loop folding determined in this study, however, is much smaller than that of the loop folding of the r(UNCG) loop (1.0 kcal/mol). Normal hairpin loops have positive $\Delta G_{37°C}$ values for the loop folding that require longer base pairs in the stem region to form hairpin structures. The small $\Delta G_{37°C}$ values for the loop folding of the d(GNA) loop suggest that these loops can form stable hairpins with fewer base pairs compared to normal loops and that the most stable d(GAA) loop ($\Delta G_{37°C} = -0.4$ kcal/mol) could serve to fold back the GCGAAGC fragment by itself.

Unusual Nature of a d(GAA) Trinucleotide Fragment. The results of the thermodynamic study suggest that d(GAA) loop itself may form a folded structure that is advantageous for minihairpin formation. To test this possibility, we performed native polyacrylamide gel electrophoresis of tri- or tetranucleotide fragments with sequences that correspond to the loops of minihairpins (Figure 5A). Trideoxyribonucleotide d(GAA) shows faster mobility than d(AGA) and d(AAG) fragments. The abnormal gel mobility of d(GAA) is also observed even if an extra C is attached to the 5′-termini of the trinucleotides. Other trinucleotides, such as d(GTT), d(TGT), and d(TTG), exhibit similar mobility to each other. Among these fragments, only the d(GAA) sequence serves as the loop sequence for forming the stable minihairpin (T_m = 76 °C), while d(GCAGAGC) and d(GCAAGGC) form much less stable minihairpins (T_m = 50−60 °C) (Figure 2B). The unusual mobility of the d(GAA) fragment is not observed when it is run on a gel containing 7 M urea (Figure 5B) or on a gel at 60 °C (data not shown). Thus, the unusual mobility is caused by a conformation different from those of fragments with other sequences.

The unusual folding of the d(GAA) fragment is also represented in the ^{1}H NMR spectra. With increasing temperature, the three H1′ proton resonances in the d(GAA) shift downfield by a similar extent, whereas in the d(AGA) and d(AAG), the H1′ protons of the middle residues show greater downfield shifts compared to other two H1′ protons (data not shown). The observation suggests that the d(GAA) fragment forms a less stacked structure compared with those of the d(AGA) or d(AAG) fragments. The conformation of the d(GAA) fragment may be advantageous for minihairpin formation as a turn structure.

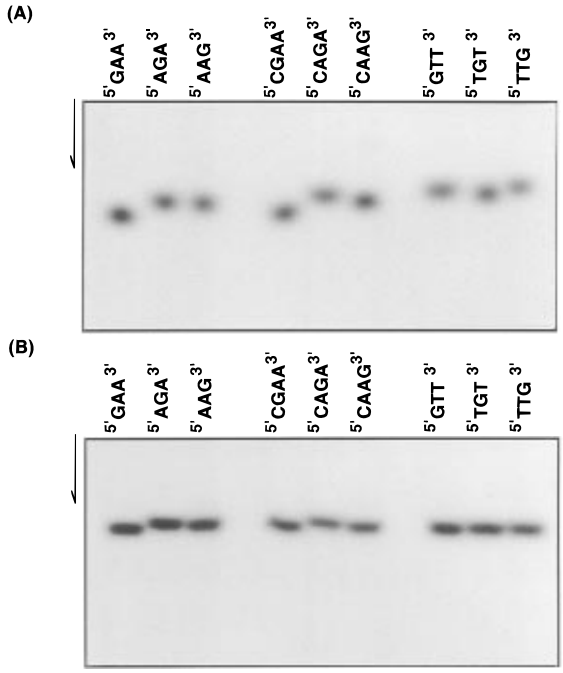

FIGURE 5: Electrophoretic patterns of ^{32}P-labeled trideoxyribonucleotides and tetradeoxyribonucleotides. The 3mers and 4mers were electrophoresed on a 20% polyacrylamide gel (A) and on the gel containing 7 M urea (B) at 37 °C.

DISCUSSION

We have examined the ability of all the sequence variants of the trinucleotide loop in heptamers containing two G-C base pairs to form stable hairpin structures. We found that only the DNA hairpins with GNA loops form extraordinarily stable trinucleotide-loop hairpins. The 3-D structure of the d(GCGAAGC) hairpin determined previously (Hirao et al., 1994) suggests that the extraordinary stability is caused by the additional G-A pair formed between G_3 and A_5 and the extensive base-stacking interaction within $G_1C_2G_3N_4$ and $A_5G_6C_7$. The extensive base stacking is possible because all of the residues in the fragment adopt a B-form DNA conformation except for the three bonds of A_4 O3′−P, A_5 O5′−P, and A_5 C5′−C4′, where the sharp turn occurs. NMR studies show that the structure of each GNA-loop hairpin is quite similar to that of the d(GCGAAGC) hairpin. These results show that the sheared G-A base pair in the loop is required for the formation of stable trinucleotide-loop hairpins. The geometries of the sheared G-A base pair and the Watson−Crick G-C base pair are compared in Figure 6. The sheared G-A base pair allows the distance between the 3′-phosphate of the G and 5′-phosphate of the A to be shortened to 7.5 Å, while the distance of the equivalent phosphates of the Watson−Crick base pair is 16.4 Å. Consequently, the sheared G-A base pair enables the two

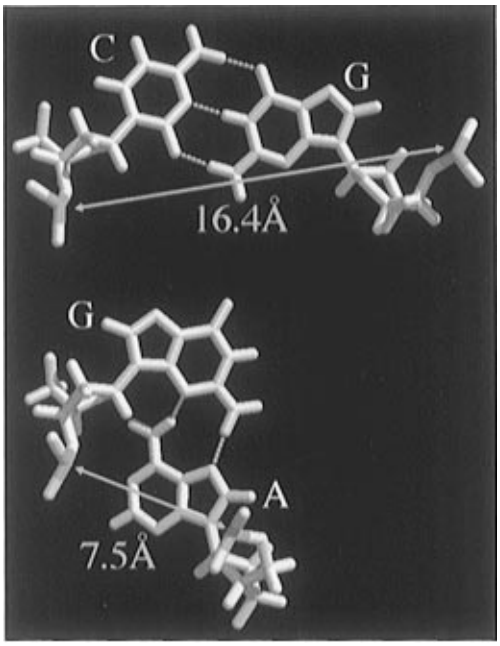

FIGURE 6: Comparison of distances between phosphates in the d(GCGAAGC) hairpin structure. The Watson−Crick base pair formed between C_2 and G_6 (above) and the sheared G-A pair formed between G_3 and A_5 (below) are shown. C_2-3′P, G_6-5′P, G_3-3′P, and A_5-5′P are shown in yellow.

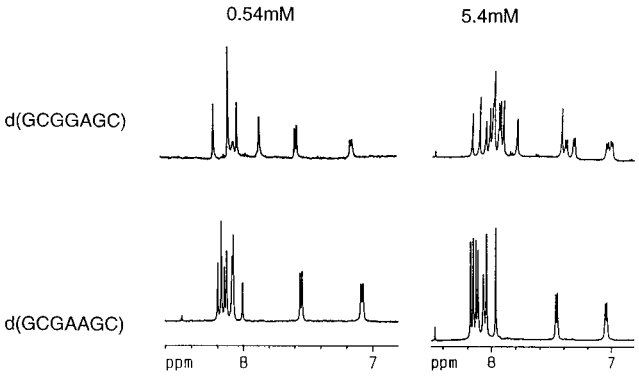

FIGURE 7: H8/H6 regions of the ^{1}H spectra of d(GCGAAGC) and d(GCGGAGC) hairpins measured at concentrations of 0.54 mM in 15 mM NaCl and 1 mM sodium phosphate (pD 7.0) and 5.4 mM in 150 mM NaCl and 10 mM sodium phosphate (pD 7.0).

phosphates to be linked with only one nucleoside, whereas for the Watson−Crick base pairs four to five residues are usually needed to link the base pairs (Haasnoot et al., 1986). The G-A base pair also contributes to the stacking of bases in the structure. Thus, the G-A base pair can connect both strands without destroying the base−base stacking.

The d(GCGCAGC) hairpin structure agrees with the results of a recent study of the structure of a GCA-loop hairpin formed by d(CAATGCAATG) (Zhu et al., 1995). Contrary to our observation, they reported that it forms equilibrium mixtures of duplex and hairpin forms when the loop sequence is substituted by GAA or GTA sequence (Chou et al., 1996). However, our NMR data of d(GCGAAGC) and d(GCGTAGC), measured at 10-fold higher oligonucleotide concentrations (5.4 mM DNA in 0.15 M NaCl and 10 mM sodium phosphate, pD 7.0), obviously do not show any evidence for duplex formation as shown in Figure 7. By contrast, our NMR spectra of d(GCGGAGC), measured at the higher DNA concentration (5.4 mM) and higher salt concentration (0.15 M NaCl), gave extra peaks at the region of the base protons (Figure 7). The spectra of the d(GCGAAGC) hairpin at 5.4 and 0.54 mM are essentially the same, but the number of peaks of base protons in the d(GCGGAGC) spectrum measured at 5.4 mM concentration is double compared to that at 0.54 mM. This result suggests that the fragment containing the GGA-loop sequence exists as a mixture of the hairpin and the duplex, or multiplex, at the high concentration. As the T_m of d(GCGGAGC) and d(GCGCAGC) did not change between concentrations of 1.6 and 24 μM, the transition of the d(GCGGAGC) hairpin to a duplex or a multiplex occurs at concentrations on the order of millimolar.

In our previous work (Hirao et al., 1994), we mentioned that the H1′ protons of the GAA loop gave broad peaks compared to other H1′ protons in the GC stem region due to a small structural fluctuation in the hairpin structure. All 1-D spectra of the GNA hairpins also show the localized broadening of H1′ protons of the GNA loops. Chou et al. suggested that the broadening may be caused by "an equilibrium between a duplex containing a G-A bracketed A-stack $[GAA]_2$ motif and a hairpin containing a GAA-turn loop motif" (Chou et al., 1996). However, the broadening is always observed in the spectra of the GCGNAGC fragments at both 0.54 and 5.4 mM, conditions at which only the hairpin structure (0.54 mM) or the mixture with the multiplex (5.4 mM) exists. Therefore, the observation of the broadening pattern in the loop region is independent of duplex or multiplex formation and might be characterestic of the minihairpin structure. The stabilities of the hairpins are also affected by the closing base pair (Hirao et al., 1992; Serra et al., 1993). Moreover, the conformation of the d(GAAA) loop of the d(GCGAAAGC) hairpin differs from that of the d(CGGCGAAAGCCG) hairpin. In the former hairpin, the loop is folded back between G_3 and A_4 (Hirao et al., 1989), but in the latter, it is folded back between A_6 and A_7 (Tanikawa et al., 1991), demonstrating that the conformation of the loop can be changed by the stem structure. Thus, it may be possible that the differences in the stem sequences of the hairpins impart the dissimilarity in the observations reported by others (Chou et al., 1996) and by us.

Because of the unusual stability and the similarity among the structures of these GNA-loop hairpins, it is reasonable to assume that these hairpins appear in the regions of similar functions. In fact, it is suggested that DNA-dependent RNA polymerase of phage N4 recognizes both GAA- and GGA-loop hairpins for initiation of its function (Cho et al., 1995).

The present thermodynamic studies of d(GCGNAGC) hairpins and studies using the d(GAA) fragment suggest that the d(GAA) fragment can be folded into some structure that is favorable for the formation of an extraordinarily stable hairpin. Other d(GNA) fragments can be expected to behave as the d(GAA) fragment, and it is possible that these d(GNA) sequences act as nucleation sites for unusual structures important for biological function. An NMR study on a centromeric repeat, $(AATGG)_n$ [=$(GGAAT)_n$], shows that the repeat sequence can easily be folded into a hairpin structure consisting of a stem with some A-T and G-A base pairs and a GGA loop (Jaishree & Wang, 1994). Even in the genome or in plasmids, GNA sequences show unusual behavior. For example, the $(CAG)_n$ [=$(GCA)_n$] repeat, which is expanded in human genetic diseases such as Huntington's disease, myotonic muscular dystrophy, and Kennedy's disease, is known to behave as a non-B-DNA structure (Kohwi et al., 1993) as suggested by the reactivity

of the cytidine residues with chloroacetaldehyde. The GAA or GGA repeat sequence found in the rat polymeric immunoglobulin receptor gene (PIGR) is also known to show an unusual S1 nuclease sensitivity (Koch et al., 1995) because of the unusual conformation of the sequence.

The d(GCGAAGC) and d(GCGAAAGC) hairpins are much more stable than their RNA counterparts. Although the r(GCGAAGC) hairpin is less stable than the r(GCGAAAGC) tetraloop hairpin ($T_m = 27 \pm 5$ and 39 °C, respectively) (Hirao et al., 1994, 1992), it is still stable compared with standard RNA hairpins. The abnormal conformation of the d(GNA) trinucleotides discussed in this paper may also be predicted for the r(GNA) trinucleotides. It is known that there are variable arrays of tandem r(GGA) and r(GAA) triplet repeats in the 3′ untranslated regions of rat PIGR-encoded hepatic mRNAs during liver regeneration (Koch et al., 1995).

In addition to increasing thermostability, the minihairpin structure exerts high resistance against nucleases. This nature could be applied to stabilization of oligonucleotides by tagging the minihairpin sequence at the termini (Hirao et al., 1993). We have demonstrated that mRNA can be stabilized by hybridizing it with a small DNA fragment tagging the minihairpin sequence (Hirao et al., 1993; Yoshizawa et al., 1994). Recently, this method has been widely applied to stabilize antisense DNAs (Khan & Coulson, 1993; Poddevin et al., 1994; Tang et al., 1993) and primers for DNA amplification (Caetano-Anolles & Gresshoff, 1994). Also, aptamers obtained by *in vitro* selection (Ellington & Szostak, 1992; Tuerk & Gold, 1990) are expected to be stabilized by the method for gene therapy applications.

The biological functions of minihairpin structures are still unknown; sequence data must be accumulated to know their locations in a gene. However, the minihairpin sequences cause strong band compression on sequencing gels (Hirao et al., 1992; Kapelner et al., 1994; Odagiri, 1994; Ruf et al., 1994). The unusual properties of the d(GNA) loop sequence enable formation of an unexpectedly stable hairpin with a short base-paired stem and may cause erroneous missequencing in these regions. Therefore, special attention must be paid during DNA sequencing, especially, of noncoding regions.

ACKNOWLEDGMENT

We thank Mr. T. Kakinuma and Mr. N. Matsumoto for their help in preparation of NMR samples. We also thank Dr. J. D. Puglisi and Dr. I. Matsumura for helpful comments and suggestions for improvement of the manuscript and Dr. A. D. Ellington for helpful discussion.

REFERENCES

Antao, V. P., & Tinoco, I. (1992) *Nucleic Acids Res. 20*, 819−824.

Antao, V. P., Lai, S. Y., & Tinoco, I. (1991) *Nucleic Acids Res. 19*, 5901−5905.

Arai, K., Low, R., Kobori, J., Shlomai, J., & Kornberg, A. (1981) *J. Biol. Chem. 256*, 5273−5280.

Breslauer, K. J., Frank, R., Blöcker, H., & Marky, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746−3750.

Caetano-Anolles, G., & Gresshoff, P. M. (1994) *Bio/Technology 12*, 619−623.

Cho, N. Y., Choi, M., & Rothman-Denes, L. B. (1995) *J. Mol. Biol. 246*, 461−471.

Chou, S.-H., Zhu, L., & Reid, B. R. (1996) *J. Mol. Biol. 259*, 445−457.

Cowing, D. W., Bardwell, J. C. A., Craig, E. A., Woolford, C., Hendrix, R. W., & Gross, C. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2679−2683.

Elias, P., & Lehman, I. R. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2959−2963.

Ellington, A. D., & Szostak, J. W. (1992) *Nature 355*, 850−852.

Frank, R., & Köster, H. (1979) *Nucleic Acids Res. 6*, 2069−2087.

Glucksmann, M. A., Markiewicz, P., Malone, C., & Rothman-Denes, L. B. (1992) *Cell 70*, 491−500.

Godson, G. N., Barrell, B. G., Staden, R., & Fiddes, J. C. (1978) *Nat. New Biol. 276*, 236−247.

Haasnoot, C. A., Hilbers, C. W., van der Marel, G. A., van Boom, J. H., Singh, U. C., Pattabiraman, N., & Kollman, P. A. (1986) *J. Biomol. Struct. Dyn. 3*, 843−857.

Heus, H. A., & Pardi, A. (1991) *Science 253*, 191−194.

Hirao, I., Naraoka, T., Kanamori, S., Nakamura, M., & Miura, K. (1988) *Biochem. Int. 16*, 157−162.

Hirao, I., Nishimura, Y., Naraoka, T., Watanabe, K., Arata, Y., & Miura, K. (1989) *Nucleic Acids Res. 17*, 2223−2231.

Hirao, I., Ishida, M., Watanabe, K., & Miura, K. (1990) *Biochim. Biophys. Acta 1087*, 199−204.

Hirao, I., Nishimura, Y., Tagawa, Y., Watanabe, K., & Miura, K. (1992) *Nucleic Acids Res. 20*, 3891−3896.

Hirao, I., Yoshizawa, S., & Miura, K. (1993) *FEBS Lett. 321*, 169−172.

Hirao, I., Kawai, G., Yoshizawa, S., Nishimura, Y., Ishido, Y., Watanabe, K., & Miura, K. (1994) *Nucleic Acids Res. 22*, 576−582.

Jaishree, T. N., & Wang, A. H. (1994) *FEBS Lett. 347*, 99−103.

Kapelner, S. N., Turner, R. T., Sarkar, G., & Bolander, M. E. (1994) *BioTechniques 17*, 66−67.

Katahira, M., Sato, H., Mishima, K., Uesugi, S., & Fujii, S. (1993) *Nucleic Acids Res. 21*, 5418−5424.

Khan, I. M., & Coulson, J. M. (1993) *Nucleic Acids Res. 21*, 2957−2958.

Koch, K. S., Gleiberman, A. S., Aoki, T., Leffert, H. L., Feren, A., Jones, A. L., & Fodor, E. J. (1995) *Nucleic Acids Res. 23*, 1098−1112.

Kohwi, Y., Wang, H., & Kohwi-Shigematsu, T. (1993) *Nucleic Acids Res. 21*, 5651−5655.

Miao, D. M., Honda, Y., Tanaka, K., Higashi, A., Nakamura, T., Taguchi, Y., Sakai, H., Komano, T., & Bagdaraian, M. (1993) *Nucleic Acids Res. 21*, 4900−4903.

Odagiri, T. (1994) *Virus Genes 8*, 271−274.

Poddevin, B., Meguenni, S., Elias, I., Vasseur, M., & Blumenfeld, M. (1994) *Antisense Res. Dev. 4*, 147−154.

Puglisi, J. D., & Tinoco, I. (1989) *Methods Enzymol. 180*, 304−325.

Ruf, S., Zeltz, P., & Kössel, H. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2295−2299.

Serra, M. J., Lyttle, M. H., Axenson, T. J., Schadt, C. A., & Turner, D. H. (1993) *Nucleic Acids Res. 21*, 3845−3849.

Tang, J. Y., Temsamani, J., & Agrawal, S. (1993) *Nucleic Acids Res. 21*, 2729−2735.

Tanikawa, J., Nishimura, Y., Hirao, I., & Miura, K. (1991) *Nucleic Acids Symp. Ser. 25*, 47−48.

Tuerk, C., & Gold, L. (1990) *Science 249*, 505−510.

Varani, G. (1995) *Annu. Rev. Biophys. Biomol. Struct. 24*, 379−404.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

Yoshizawa, S., Ueda, T., Ishido, Y., Miura, K., Watanabe, K., & Hirao, I. (1994) *Nucleic Acids Res. 22*, 2217−2221.

Zhu, L., Chou, S.-H., Xu, J., & Reid, B. R. (1995) *Nat. Struct. Biol. 2*, 1012−1017.

BI961738P