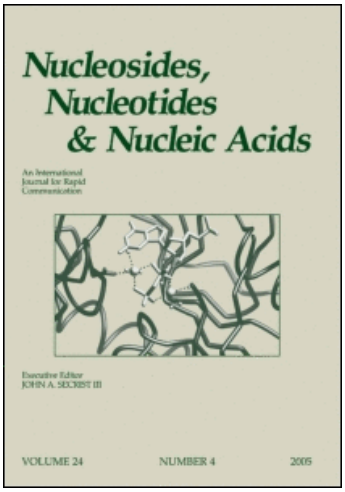


Nucleosides, Nucleotides and Nucleic Acids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

d(GCGTAGC), an Equilibrium Between a Hairpin Structure and an Unusual Duplex

Evline Lescrinier[a]; Shqun Sheng[b]; Ja Schraml[c]; Roger Busson[a]; Piet Herdewijn[a]
[a] Laboratory of Medicinal Chemistry, Rega Institute for Medical Research and Faculty of Pharmacy, Katholieke Universiteit Leuven, Leuven, Belgium [b] Department of Chemistry, Georgia State University, Atlanta, GA, USA [c] Institute of Chemical Process Fundamentals, Prague 6, Czech Republic

To cite this Article Lescrinier, Evline , Sheng, Shqun , Schraml, Ja , Busson, Roger and Herdewijn, Piet(1999) 'd(GCGTAGC), an Equilibrium Between a Hairpin Structure and an Unusual Duplex', Nucleosides, Nucleotides and Nucleic Acids, 18: 11, 2721 — 2744
To link to this Article: DOI: 10.1080/07328319908044637
URL: http://dx.doi.org/10.1080/07328319908044637

NUCLEOSIDES & NUCLEOTIDES, 18(11&12), 2721–2744 (1999)

d(GCGTAGC), AN EQUILIBRIUM BETWEEN A HAIRPIN STRUCTURE AND AN UNUSUAL DUPLEX

Eveline Lescrinier,[a] Shuqun Sheng,[b] Jan Schraml,[c] Roger Busson[a] and Piet Herdewijn[a]*

[a] *Laboratory of Medicinal Chemistry, Rega Institute for Medical Research and Faculty of Pharmacy, Katholieke Universiteit Leuven, Minderbroedersstraat 10, B-3000 Leuven, Belgium.*

[b] *Department of Chemistry, Georgia State University, University Plaza, Atlanta, GA 30303-3083, USA*

[c] *Institute of Chemical Process Fundamentals, Rozvojova 135, 165 02 Prague 6, Czech Republic*

ABSTRACT: The DNA heptamer d(GCGTAGC), as determined by NMR spectroscopy, occurs in aqueous solution in a temperature- and salt-dependent equilibrium between a monoloop hairpin (where the unpaired thymidine, forming the loop, is stacked on top of the guanine (G3) base) and a duplex with the unpaired thymidine residues from opposite strands intercalating and costacking between shared G:A pairs.

Introduction

Single-stranded nucleic acids can adopt a variety of secondary structures. The information of this secondary structure is often hidden in the base sequence of the nucleic acid strand. Before using an oligonucleotide for physicochemical and biological studies, information is needed on its structure and thermodynamic stability. The following example highlights the unexpected results that can be obtained.

While our work on the structural analysis of the DNA heptamer d(GCGTAGC) was progressing, Hirao and coworkers[1] reported that this oligomer forms a monoloop hairpin with T on the top of the loop followed by a sheared G:A base pair and two G:C/C:G Watson-Crick base pairs. However, we found evidence that this oligomer

exists in solution (0.150 M NaCl and 10 mM sodium phosphate buffer) as a temperature-dependent equilibrium between a monoloop hairpin and an unusual duplex. In the absence of NaCl and phosphate buffer, d(GCGTAGC) occurs only in its hairpin form, even at low temperature (0°C). The duplex form most likely contains a G:A-bracketed unpaired thymidine stack motif. The pyrimidine stack motif is comparable with the previously observed purine stack,[2] although it may be considered less stable since purine stacking is energetically more favorable than pyrimidine stacking.

Oligonucleotide hairpins consist of a single-stranded loop closed by a base-paired stem and can vary in constitution and stability.[3] They occur not only in single-stranded RNA and DNA but also in dsDNA, and are considered to have important biological functions. In particular, unusually stable hairpins have been studied in which tetra-nucleotide loops are included.[4,5,6,7] These structures are found in ribosomal RNAs,[7] hammerhead ribozymes,[8] catalytic introns,[9] bacteriophage mRNA[6] and the replication origins of single-stranded phage DNA.[10]

While hairpin loops in nucleic acids generally require two or more residues to span the nucleotides of the opposite strands in the stem region, a few examples of single nucleotide loop structures are known.[1,11,12] Recently, three-residue hairpin loops in DNA have drawn considerable attention, because some genetic diseases have been found to involve the expansion of a run of tandemly repeated triplets in the target gene.[13,14] In all monoloop hairpins described to date, a triad system [GAG, AAA or GNA with N = T,C,A or G] is followed by Watson-Crick base pairs. Stacking interactions are of crucial importance for stabilizing such structures.[15]

Results

NMR analysis: ^1H and ^{31}P resonance assignments at 33 °C. The non-exchangeable protons of d(GCGTAGC) were assigned by conventional methods[16,17] using GDQFCOSY[18] and TOCSY[19] experiments in D_2O. Starting from the H1' resonances in a COSY spectrum, the resonances for H2'1, H2'2 and then H3' were assigned. Starting from the resonances for H2'1 and H2'2 in a TOCSY spectrum, all other ^1H resonances for the sugar moieties were assigned. These assignments were confirmed by a TOCSY experiment[20] modified for ω_1-decoupling. The COSY and TOCSY experiments allowed us to assign T4:H6 (7.12 ppm) and T4:Me (1.43 ppm) resonances that are coupled by a

long range spin-spin coupling which causes off-diagonal crosspeaks between both protons. These experiments also indicated the position of the two C:H6 protons (7.11 and 7.37 ppm). However, sequential assignment of these aromatic cytosine signals is not possible through COSY or TOCSY spectra. In a NOESY experiment with a short mixing time (100 ms), the stronger NOE interactions of H2'1 with H1' compared to H2'2 with H1' of the same deoxyribose unit were used to distinguish stereochemically the two 2' protons. In B-type DNA, with sugars in a 2'-*endo* conformation, the H2'1 proton is the pro-S hydrogen.

The sequential assignment of the ^{1}H signals was based on an aromatic to H2'1/H2'2 proton walk. In a NOESY spectrum obtained with a mixing time of 300 ms (Fig. 1), the C:H6 signal at 7.11 ppm has off-diagonal crosspeaks to two nicely separated H2'1/H2'2 pairs. With one of these two pairs, the considered C:H6 has a strong crosspeak to the H2'1 (2.72 ppm) and a weak crosspeak to the H2'2 proton (2.50 ppm). This is typical for sequential cross-signals where the H2'1/H2'2 protons involved belong to the previous residue in the DNA strand, *G(n-1)*. With the second pair, the C:H6 has a strong crosspeak to the H2'2 (1.65 ppm) and a weak crosspeak to H2'1 (2.34 ppm). This is the H2'1/H2'2 pair that belongs to the same residue. The H2'1 of this pair has also a strong crosspeak to the signal at 8.13 ppm, this is a sequential NOE to the next residue in the DNA strand, *G(n+1)*. Since only C2 is positioned between two deoxyguanosine residues, the signals at 7.11, 8.13 and 7.37 ppm can be assigned to C2:H6, G3:H8 and C7:H6, respectively, while G1:H2'1/H2'2 are those at 2.72 and 2.50 ppm. This in agreement with the assignment of T4:Me that has NOE to G3:H8. At 7.94 ppm there is a proton which has a strong NOE to G1:H2'2 and a weak one to G1:H2'1, this has to be G1:H8. Starting from C7:H6, the G6:H8 proton can be assigned to the resonance at 7.98, since both aromatic protons have a crosspeak to the same G6:H2'1. The last signal with a NOE to the H2'1/H2'2 protons has to be A5:H8.

These NOEs enabled us to trace the G1-C2-G3-T4 and A5-G6-C7 portions of the heptamer sequence. However, no NOE contact was detected between the T4 and A5 units in the heptamer; thus, it was not possible to establish the connectivity between these residues by the ^{1}H-^{1}H NOE pathway.

To assign the ^{31}P signals of the heptamer and to obtain the missing sequence information from T4 to A5, a ^{1}H-detected ^{1}H-^{31}P heteronuclear correlation (HMBC[21])

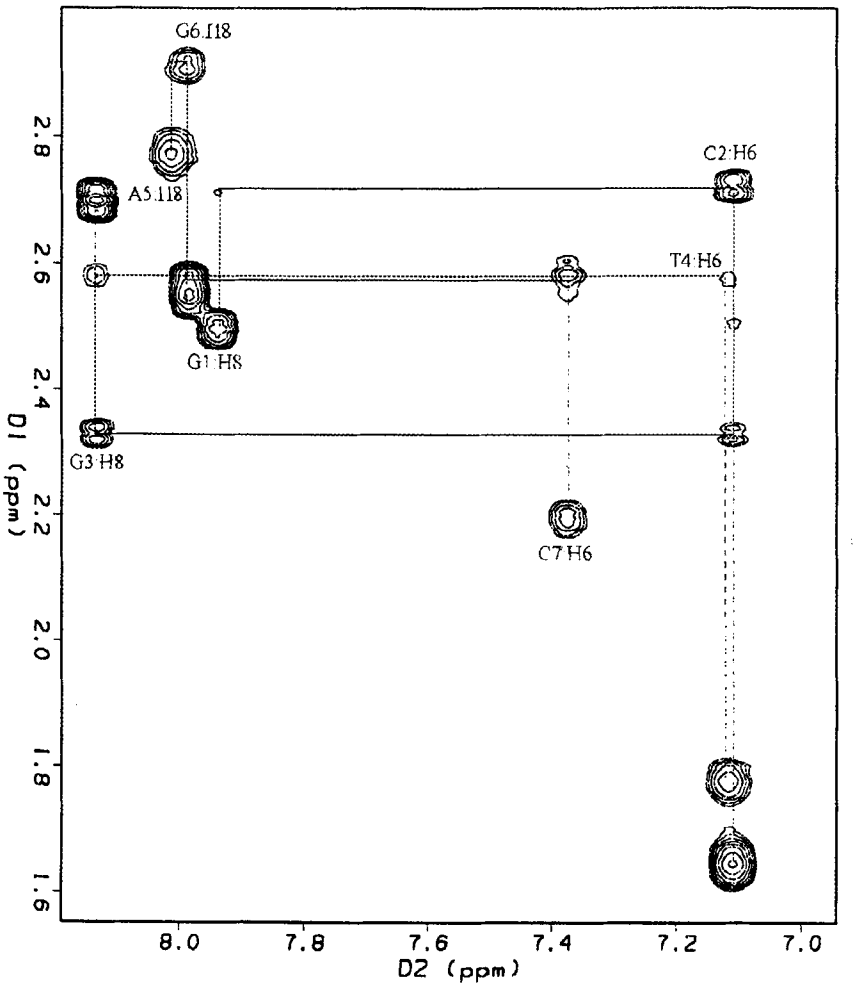


Figure 1: An expansion of a NOESY spectrum (33 °C, mix = 300 ms) showing crosspeaks of H2'1 and H2'2 with the aromatic H8/H6 resonances of the d(GCGTAGC) at 33 °C. The intra-residue H2'2(*n*) with H8/H6(*n*) and inter-residue H2'1(*n*) with H8/H6(*n+1*) crosspeaks are intense compared to the weak or absent intra-residue H2'1(*n*) with H8/H6(*n*) NOE interactions. Exceptions are the weak G3:H2'1 to T4:H6 connectivity and the absent T4:H2'1 to A5:H8 crosspeak.

spectrum was recorded (Fig. 2). This spectrum allowed the verification of the sequential ^{1}H assignments and proved the connectivity between T4 and A5. The complete ^{1}H and ^{31}P assignments for d(GCGTAGC) in D_2O at 33 °C are compiled in Table 1. It is noteworthy that the ^{31}P resonances (Fig. 3) between units G6 and C7, G1

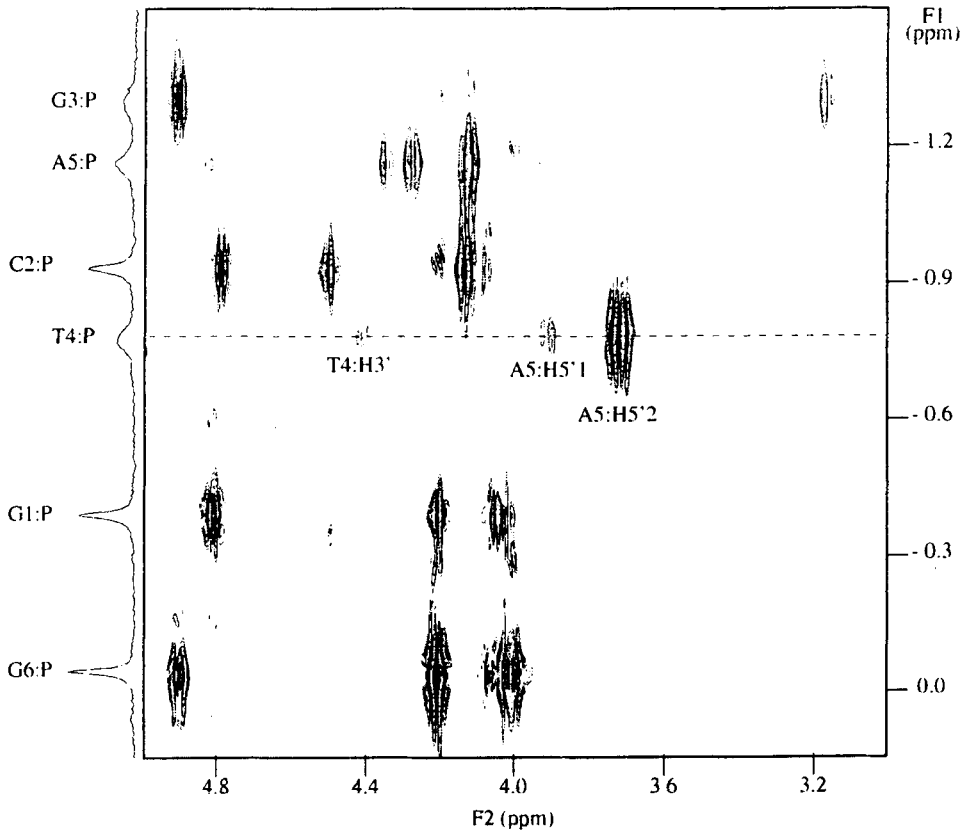
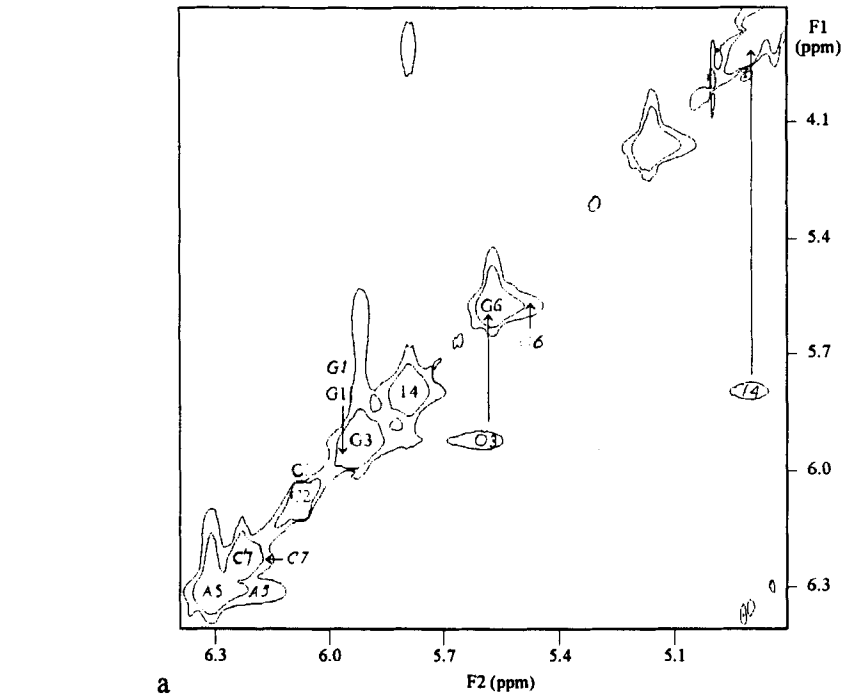
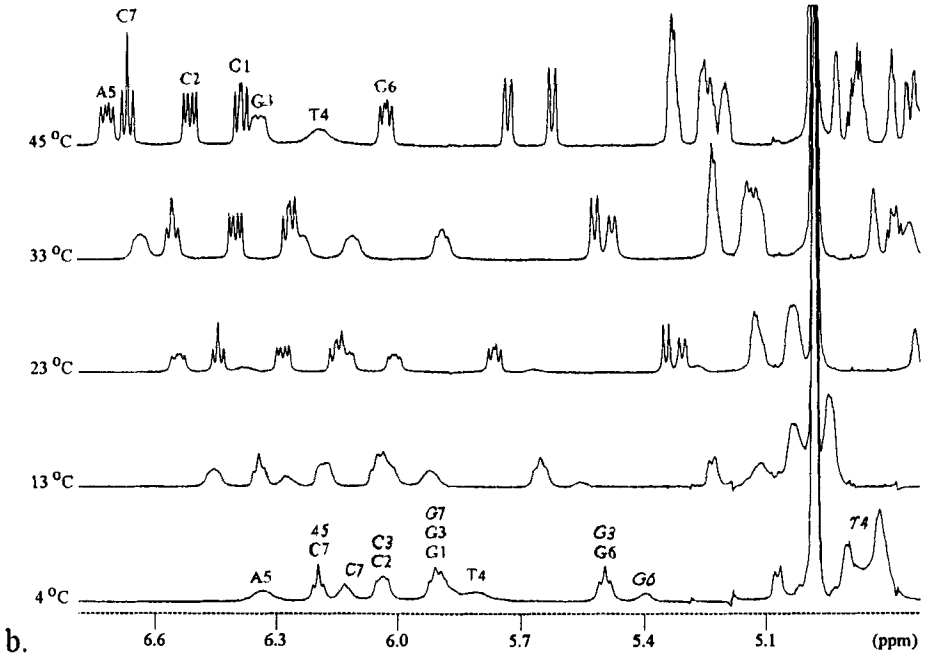


Figure 2: Full ^{1}H-^{31}P HMBC spectrum at 33 °C. The connectivity from T4 to A5 over P is indicated.

Table 1: ^{1}H and ^{31}P chemical shifts in ppm for the structure at 33 °C

	H1'	H2'1	H2'2	H3'	H4'	H5'1/H5'2	H8/H6	H2/H5	P
G1	5.95	2.72	2.50	4.82	4.23	3.73	7.94		
C2	6.08	2.34	1.65	4.80	4.24	4.08	7.11	5.15	-0.41
G3	5.92	2.58	2.70	4.92	4.52	4.16/4.11	8.13		-0.94
T4	5.79	2.09	1.78	4.45	2.14	3.39/3.25	7.12	1.43	-1.29
A5	6.31	2.91	2.78	4.84	4.34	3.96/3.78	8.00	8.13	-0.79
G6	5.57	2.57	2.55	4.91	4.38	4.31/4.16	7.98		-1.16
C7	6.23	2.29	2.21	4.48	4.11	4.23/4.06	7.37	5.19	-0.08





and C2, and C2 and G3, representing the connections in the stem region, are sharp, while the three other ^{31}P resonances (belonging to the connections of G3 and T4, T4 and A5 and A5 and G6) are broad. This broadening is in agreement with the broadening of the H1' resonances: the H1' resonances of G3, T4, A5 and G6 are broad and those of G1, C2 and C7 are sharp at 33 °C (see also Fig. 3). This phenomenon was previously ascribed in d(GCGAAGC) to wobbling of A5.[10]

NMR analysis: ^1H resonance assignments at –5 °C. In addition to the aforementioned sequential cross peaks, the NOESY spectrum of the DNA heptamer recorded at 33 °C in D_2O (Fig. 3a) showed strong tails at the diagonal, suggesting the possibility of a fast exchange between two structural types for d(GCGTAGC). Some of the tails point in an upfield direction (*e.g.* T4:H1'), others have a downfield orientation (*e.g.* T4:H4'). Such effect can not be due to shimming or phasing errors. This hypothesis was confirmed by studying the NMR features of the D_2O sample as a function of temperature. When the sample was cooled, signals of a second solution conformer became observable in the one-dimensional proton spectrum; their relative intensities increasing with decreasing temperature (Fig. 3b). Based on exchange cross-peaks and tails of the diagonal peaks in the NOESY experiment at 33 °C, we were able to assign the H1' resonances in the additional structure. The sequential connectivity of the second conformer was assigned by the conventional methods described above in a NOESY experiment in D_2O at –5 °C with mixing time of 100 ms.

At this temperature, no exchange crosspeaks were observed between the signals of both structures. This suggests slow exchange at low temperatures.

NMR Analysis: Imino proton signals at –5 °C. Yoshizawa *et al.*[1] observed, at 25 °C, two signals assigned to the G:C base pairs and a signal at 10.5 ppm assigned to the A:G

Figure 3: *a*: Expansion of the diagonal in the anomeric region of a NOESY spectrum (mix = 100 ms in D_2O at 33 °C) showing exchange crosspeaks of two interconverting struct-ures. Signals of the structure that occurs at low temperature are indicated in *Italic*. *b*: The anomeric region in the ^1H spectrum of d(GCGTAGC) at different temperatures. At 33 °C only one series of anomeric signals is detectable. When the temperature is decreased, new signals occur and others decrease (reference is only valid for spectrum at the bottom). Signals of the structure that occurs at low temperature are indicated in italics.

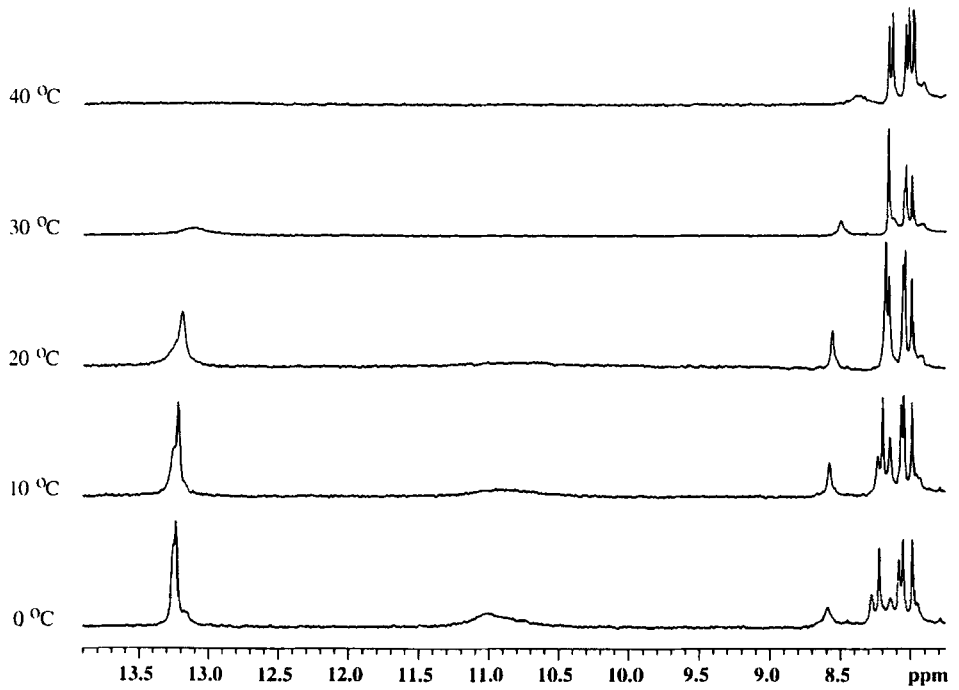


Figure 4: Imino proton region of a 'jump-return' spectrum of d(GCGTAGC) with additional salt at different temperatures (0, 10, 20, 30 and 40 °C).

mispair. At 11.0 ppm we observed only a broad signal at temperatures below 20 °C (Fig. 4). We suspected that this was due to interchange between the two structural types of d(GCGTAGC).

Therefore, we decided to prepare a new sample with the same concentration of d(GCGTAGC) but without additional salts. The pH was adjusted to 6.8 by adding small aliquots of (0.1 mM in H_2O) HCl. In this sample there is only one structural type of the DNA. Even at a temperature of 0 °C, the sample without additional salt had no signals corresponding to the structure that was formed by cooling the sample with additional salt. This indicates that the equilibrium between conformations of d(GCGTAGC) is not only temperature dependent but also dependent on the salt concentration. The imino protons in the low-salt sample are relatively sharp (Fig. 5) and were assigned by a 2D-WATERGATE NOESY in H_2O (150 ms mixing time, 5 °C) (Fig. 6).

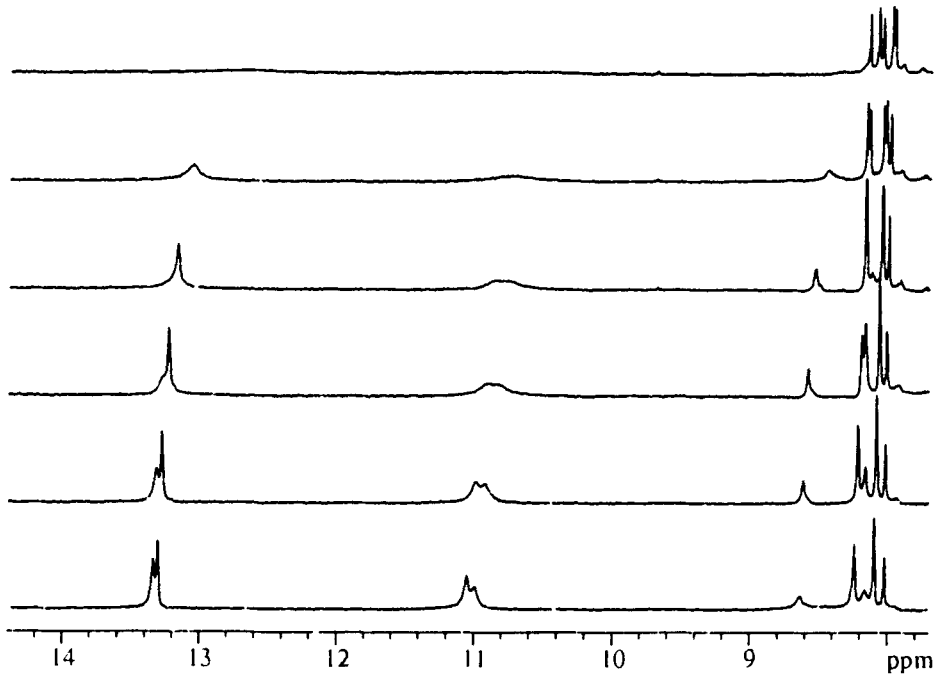


Figure 5: Imino proton region of a 'jump-return' spectrum of d(GCGTAGC) without additional salt at different temperatures (0, 10, 20, 30, 40 and 50°C).

Determination of the hairpin structure. Since there was only one structure detectable in the proton spectrum of d(GCGTAGC) at 33 °C in the presence of 150 mM NaCl, we started a conformational analysis of the heptamer under these conditions.

The pattern of the COSY connectivities (strong $^3J_{H1'-H2'1}$ $^3J_{H1'-H2'2}$ and $^3J_{H3'-H2'2}$ peaks and weak or absent $^3J_{H3'-H2'1}$ and $^3J_{H3'-H4'}$ peaks) combined with the magnitudes of $^3J_{H1'-H2'1}$ (5.3–6.0 Hz) and $^3J_{H1'-H2'2}$ (7.8–10.0 Hz) indicate that the deoxyribose ring conformations are predominantly C2'-*endo* for the residues G1, C2, G3 and A5, G6 and C7. The coupling constants for the T4 deoxyribose sugar could not be determined in the phosphorus-decoupled phase-sensitive GDQFCOSY spectrum.

NOESY data at 33 °C for six different mixing times (50, 100, 150, 200, 250 and 300 ms) were processed using the FELIX software package (BioSym Technologies, San Diego, CA). Based on the build-up curves, interproton distances were categorized as strong, medium or weak, using the isolated two-spin approximation method[22]. The

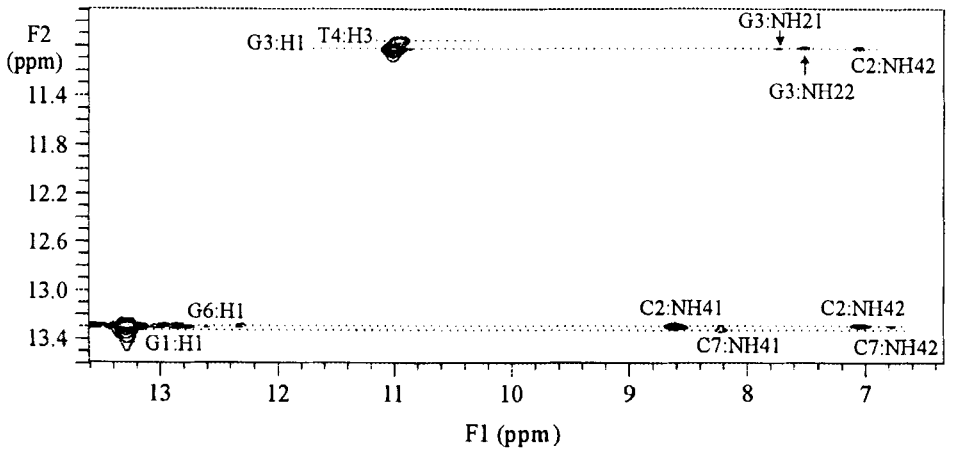


Figure 6: Expansion of the 2D-WATERGATE NOESY in H_2O (150ms mixing time, 5°C) depicting the assignment of the imino-signals in the sample without additional salt. Under these conditions, d(GCGTAGC) adopts only one conformation.

cytosine H5 to H6 distance of 2.45 Å was used as an internal reference. This resulted in 91 NOE distance restraints for DGII calculations.

The C2 and G3 sequential NOESY connectivity and the NOE between G3:H8 and G3:H3' define the G3 base conformation as anti. Three different G3 to A5 base pair hydrogen bonding motifs are consistent with an "anti" G3 conformation[7,23] (Fig. 7).

Distance geometry calculations for the hairpin d(GCGTAGC) based on the NOE-derived distance restraints, supplemented by G1 to C7 and C2 to G6 Watson-Crick hydrogen-bonding restraints, did not define the G3 to A5 base-pair hydrogen-bonding motif. In order to find out which of the three possible G3:A5 mismatch base pairs were compatible with the NOESY-derived distances, three separate sets of DGII calculations were performed. In each set, the NOE distance restraints were supplemented by the hydrogen bonding restraints of one of the three possible G3:A5 base pairs. Chirality restraints at the C1', C3' and C4' deoxyribose positions were routinely employed. The first and second set, incorporating the restraints belonging to the G:A mismatches A and B, respectively, gave no reasonable structures. This corresponds to WATERGATE-NOESY spectrum of the low-salt sample, where there were no observable crosspeaks from G3:H1 to A5:H2 (as expected for type A) or G3:H1 to A5:H8 (as expected for

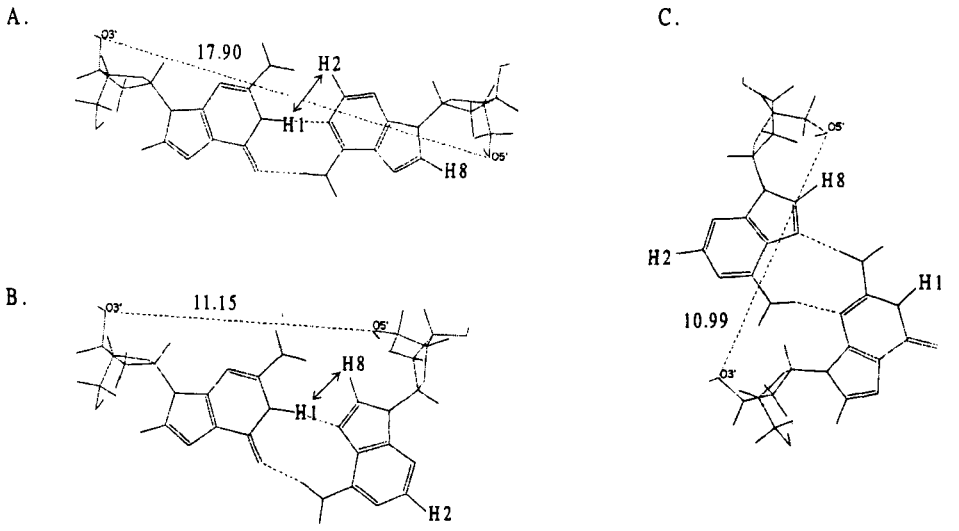


Figure 7: Three possible AG base pairing motifs. Distances between G3:O3' and A5:O5' are indicated in Å. (*cf.* Katahira *et al.* [23] for discussion of literature occurrences of these structures). Expected NOEs between A and G are indicated by arrows.

type B). To have good, planar G:A pairing in type A, the G3:CO3' to A5:O5' distance is simply too large to be bridged by one residue. However, in the sheared geometry (type C) the distance is only 11Å and can be easily spanned by one nucleotide. This is comparable to the distance in type B, but this type yielded no reasonable structures and does not correspond to our data from the WATERGATE-NOESY spectrum of the low-salt sample. Out of an ensemble of 20 structures generated by DGII calculation, using the third restraint set, 10 structures showed near-planar G3:A5 base pairing and good planar base stacking. These were further refined following the IRMA/RMD method as implemented in the BioSym software package.

In general, the calculations converged structurally in three IRMA/RMD cycles. An ensemble of 10 DGII generated IRMA/RMD refined structures is presented in Fig. 8. It shows a well defined stem with the T4 residue in the loop wobbling above G3. The stacking of T4 is possible because of the sharp turn between the T4 and A5 residue. This turn is caused by altered α (O3'-P-O5'-C5'), γ (C3'-C4'-C5'-O5') and ζ (C3'-O3'-P-O5') dihedral angles compared to those in normal duplex DNA. Unlike α and ζ, the γ

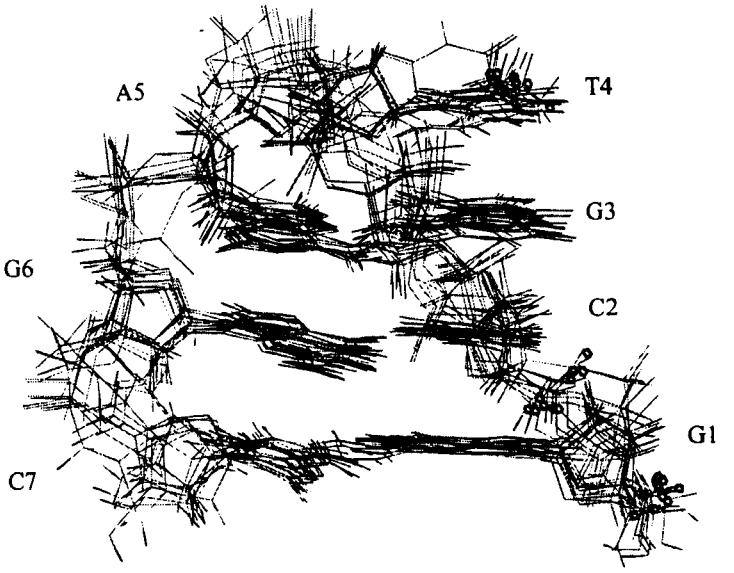


Figure 8: Ensemble of 10 DGII-generated IRMA/RMA-refined structures.

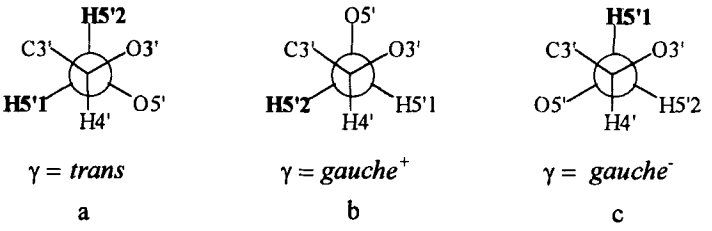
torsional angle can be monitored by coupling constants combined with NOE effects from H3' to H5'1 and/or H5'2.[24,25,26]

In all residues, except A5 and G1, $^3J_{H4',H5'1}$ and $^3J_{H4',H5'2}$ have approximately the same value (Table 2) and there is only one NOE effect from H3' to H5'2. This is typical for a *gauche*$^+$ conformation ($\gamma = +60°$) as shown in Fig. 9. In the A5 residue there is NOE between H3' and both H5'1 and H5'2. This is only possible if γ is *trans*. The large value of $^3J_{H4'-H5'2}$ indicating that H4' and H5'2 are antiperiplanar, confirms the *trans* configuration of γ. According to our model, all α dihedral angles are in the *gauche*$^-$ conformation except the one between T4 and A5, which is *trans*. Unfortunately, this unusual angle cannot be verified by homonuclear and/or heteronuclear coupling constants.

The ζ angle is also difficult to monitor by NMR data. In duplex DNA, ζ can by estimated by its strong correlation to ε ($\zeta = -317-1.23\varepsilon$), as shown by Dickerson *et al.*[27,28] In our hairpin, all ε dihedral angles could be determined based on $^3J_{H3'-P}$ from the relationships $^3J_{H-P} = 15.3\cos^2\theta - 6.1\cos\theta + 1.6$ (ref. 29) and $\varepsilon = -120 - \theta$. The Karplus equation provides four different angle solutions for each value of the coupling constant.

Table 2: $^{3}J_{H4'-H5'1}$ and $^{3}J_{H4'-H5'2}$ are listed in Hz. The values of the coupling constants are estimated from first-order data analysis. (nd = not determined). The γ torsion angle was determined from the relative intra-sugar H3'/H4'↔H5'1/H5'2 NOEs in combination with the $^{3}J_{H4'-H5'1/H5'2}$ couplings.[24,25,26]

	$^{3}J_{H4'-H5'1}$	$^{3}J_{H4'-H5'2}$	γ
G1	4.8	4.8	*gauche*$^{+}$
C2	nd	nd	*gauche*$^{+}$
G3	3.7	3.1	*gauche*$^{+}$
T4	small	small	*gauche*$^{+}$
A5	small	8.0	*trans*
G6	3.8	3.4	*gauche*$^{+}$
C7	3.8	4.3	*gauche*$^{+}$



γ = *trans* a

γ = *gauche*$^{+}$ b

γ = *gauche*$^{-}$ c

Figure 9: Three possible conformations of the γ torsion angle. H5' protons with possible NOE interactions to H3' are in boldface.

Based on our models we considered 40° < θ < 60° as the correct angle (Table 3). However the ζ dihedral angles calculated by the correlation mentioned above do not correspond to those measured from our models in the loop region. This is probably due to the unusual turn in the monoloop hairpin which cannot be compared to the regular backbones in the crystal structures of the duplexes which were used to derive the empirical relationship between ζ and ε.

The coupling constants $^{4}J_{H4'-P}$, $^{3}J_{H5'1-P}$ and $^{3}J_{H5'2-P}$ show the conformation of β (C4'-C5'-O5'-P) is *trans* for all residues. In A5, we observed $^{3}J_{H5'1-P}$ = 3.8 ± 1 Hz and $^{3}J_{H5'2-P}$ = 6.1 ± 1Hz (Fig. 10). The latter value is in contrast to the value reported by

Table 3: $^{3}J_{H3'-P}$ (in Hz) are estimated as described by Blommers *et al.*[30] Possible θ angles are calculated from $^{3}J_{H3'-P}$ by the relationship $^{3}J_{H-P}=15.3\cos^{2}\theta - 6.1\cos\theta + 1.6$ [29]. The θ which was according to our model was used to calculate ε (ε = −120 − θ).

	$^{3}J_{H3'-P}$	θ		ε	
G1	4.8 ± 1	± 103	± 111		
		± 40	± 51	−160	−171
C2	3.8 ± 1	± 98	± 107		
		± 45	± 56	−165	−176
G3	2 ± 1	± 78	± 98		
		± 56	± 78	−176	−198
T4	3.5 ± 1	± 96	± 106		
		± 47	± 59	−167	−179
A5	4.9 ± 1	± 103	± 111		
		± 40	± 51	−160	−171
G6	5.1± 1	± 104	± 112		
		± 38	± 49	−158	−169
C7					

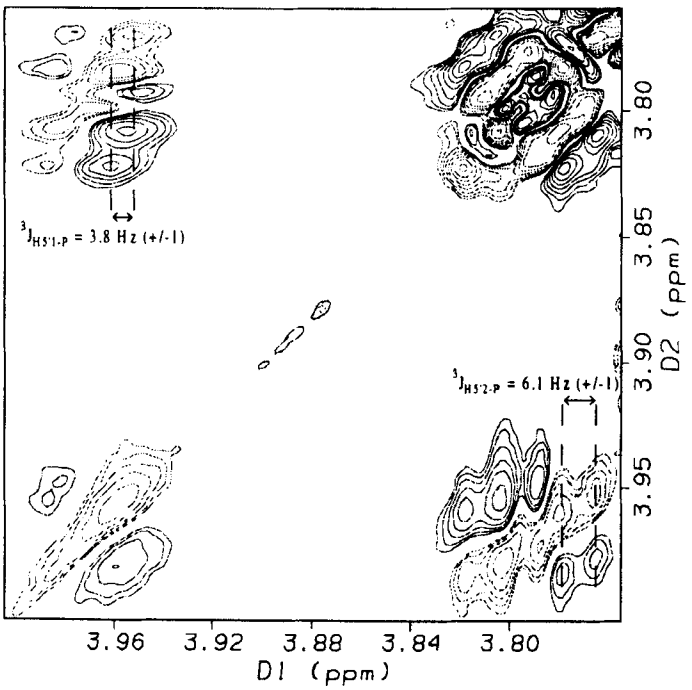


Figure 10: Expansion of the ^{31}P-coupled GDQFCOPS spectrum showing the A5:H5'1 with A5:H5'2 crosspeaks. $^{3}J_{H-P}$ were measured as described by Blommers *et al*[30].

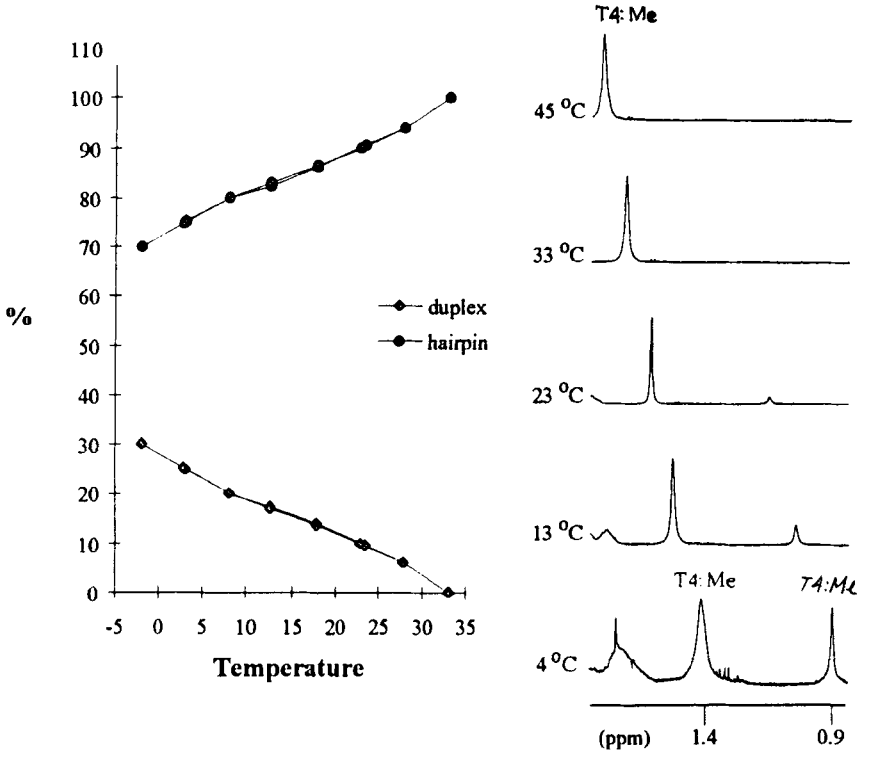


Figure 11: The temperature dependence of the equilibrium between the duplex and hairpin structures. The relative amounts of duplex and hairpin were determined by integration of the isolated T4:Me signals in the high field part of the spectrum. (chemical shift scale is only valid for the bottom-spectrum). Signals of the structure that occurs at low temperature are indicated in *Italic*.

The effect of temperature on the structure of the heptamer. As mentioned above, we suspected the occurrence of a second structure at temperatures below 33 °C in our sample with 150 mM NaCl. The percentage of both types of structures at different temperatures was determined by integrating the T4:Me signals, since these did not overlap in the temperature range studied (Fig. 11). This is in contrast to what was claimed by Hirao *et al.*[1], who reported only the hairpin structure.

The occurrence of an equilibrium between a duplex containing a G:A-bracketed unpaired stack and a single-residue loop closed by a sheared G:A pair, was reported for the DNA sequence NAATGNAATG (N = A, G, C, T) by Chou *et al.*[2] They found the stability of the hairpin relative to that the duplex of GNA triplets decreased in the order

Table 4: ^{1}H chemical shifts in ppm of duplex and hairpin at –5°C.

		H1'	H2'1/H2'2	H3'	H4'	H5'1/H5'2	H8/H6	H2/H5
G1	Hairpin	5.88	2.72/2.50				7.91	
	duplex	5.86	2.70/2.49				7.88	
C2	hairpin	6.02	2.28/1.75				7.07	
	duplex	6.02	2.32/1.76				7.07	
G3	hairpin	5.89	2.70/2.57				8.19	
	duplex	5.48	2.69/2.33				8.16	
T4	hairpin	**5.84**	2.07/1.76		**1.91**	3.30/3.14	7.07	1.43
	duplex	**4.86**	1.91/1.66		**2.65**	3.66/3.22	6.83	0.87
A5	hairpin	6.33	2.89/2.71			3.73/3.63	8.11	8.05
	duplex	6.21	2.76/2.30			3.93/3.72	7.74	7.99
G6	hairpin	5.46	2.50/2.33				7.98	
	duplex	5.37	2.46/2.31				7.96	
C7	hairpin	6.17	2.38/2.32				7.21	
	duplex	5.89	2.29/1.76				7.08	

GCA>GAA/GTA>>GGA (in 200 mM NaCl). The chemical shifts of the H1' and H4' signals of the central N residue differed remarkably in duplex and hairpin structures and were used as diagnostic indicators for the conformation the oligonucleotide. The unusual upfield shift of these two signals is explained by ring-current shielding effect from the adenine bases of the flanking G-A mismatch base pairs in the core of the duplex. In the hairpin structure only T4:H4' experiences this effect while H1' not located above an heterocyclic base. In our system, there is a similar shift for the same protons (Table 4, indicated in boldface). Following the results of Chou *et al.*,[2] the structure that is favored at lower temperatures is a duplex with unpaired T residues from opposite strands intercalating and co-stacking between sheared G:A pairs. The high field shift of T4:Me suggests that the methyl group is located between two bases, resulting in a strong ring-current shielding.

A NOESY spectrum recorded at –5 °C gave supplementary data to confirm this hypothesis (Fig. 12). Strong intra-strand NOE interactions are detected between T4:Me and several protons of G3 (H1', H2'1, H2'2). Inter-strand NOEs are visible from A5:H2 to T4:H5'1, T4:H5'2 and T4:H4'. Since there is also a strong NOE from A5:H8 to A5:H2'1, these interactions of A5:H2 cannot be intra-strand.The duplex structure was calculated using restrained molecular dynamics (rMD) in Discover (Biosym

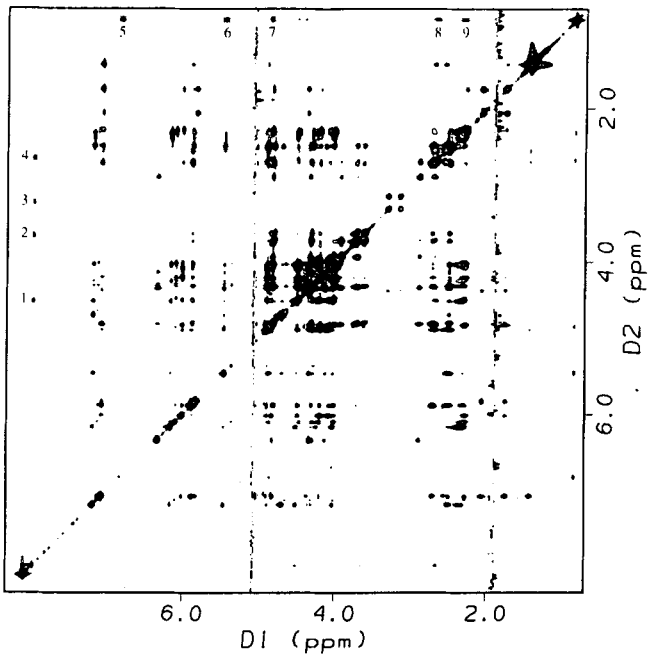


FIG. 12: NOESY spectrum of d(GCGTAGC) at -5 °C (mix = 100ms). All H8 protons of the purine bases are exchanged for deuterium at 80 °C in D_2O. Important NOE interactions of the duplex are indicated: inter-strand A5:H2 with T4:H4'(4) and with T4:H5'1 and T4:H5'2 (2,3); strong interactions of T4:Me with G3:H1'(6) and with G3:H2'1 and G3:H2'2 (8,9).

Technologies), starting from a B-type DNA helix. From a spectrum recorded in D_2O at –5 °C and in 150 mM NaCl, we obtained 64 distance restraints, of which 54 belonged to the GTA region of the duplex (14 inter-strand, 40 intra-strand) that were categorized in strong, medium and weak classes and were supplemented with hydrogen-bonding. Due to the similarity in what was previously observed by Chou *et al.*[2], we assumed sheared base pairing of the G-A mismatch. The resulting structure was further refined by conjugate gradient minimization (Fig. 13)

Materials and methods

Tm determination : The melting Tm of the DNA heptamer d(GCGTAGC) (Eurogentec) was determined in a phosphate-buffered 0.1 M NaCl solution (pH 7.4) by a UV melting curve. The temperature was raised continuously from 25 °C to 90 °C at 0.5 °C/min. The UV absorbance was monitored at 260 nm.

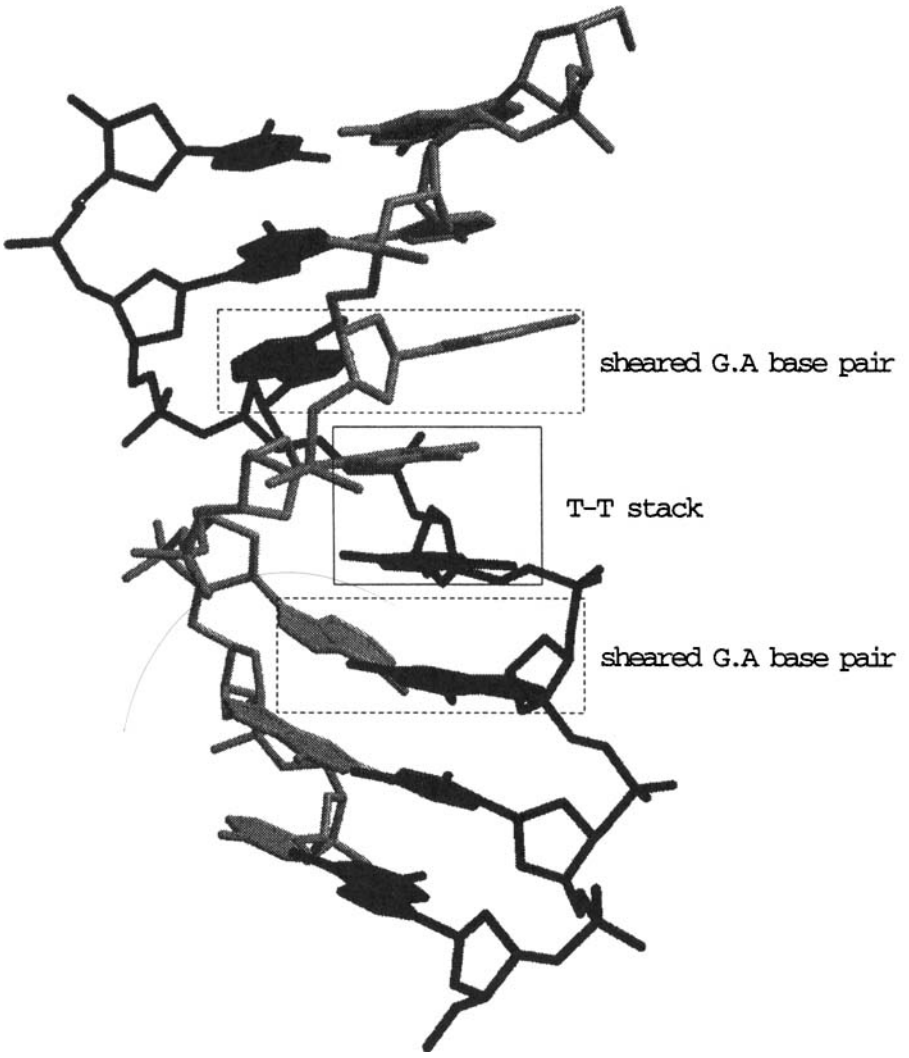


FIG. 13: Duplex structure of d(GCGTAGC) with unpaired staggered thymine residues co-stacking between two sheared G.A base pairs.

NMR sample preparation: Two 5.7 mM solutions of the DNA heptamer d(GCGTAGC) were prepared in 10 mM sodium phosphate buffer (pH 7.0) containing 150 mM NaCl. One of the samples was lyophilized and redissolved twice in 99.996% D_2O (Isotec, Miamisburg, OH). The other sample remained in H_2O. Both samples (0.7 ml) were transferred into 5-mm XXA-5-LP NMR tubes (Campro, Veenendaal, The Netherlands).

A third sample was prepared by dissolving 4.4 mg d(GCGTAGC) in 0.500 ml H_2O and 0.050 ml D_2O (3.8 mM DNA). The pH of this sample was adjusted to 6.8 by adding 0.1 mM HCl in H_2O.

NMR experiments: Unless stated otherwise, NMR experiments were performed on a Varian Unity-500 instrument (operating at 499.668 MHz for ^1H NMR and at 202.276 MHz for ^{31}P NMR), under the control of VNMR software version 5.1A. The instrument was equipped with a 5-mm ID PFG (inverse-detection with pulsed magnetic field gradients) $^1H\{X\}$ probe, tuned for ^{31}P in the "decoupling" channel, and a variable temperature (VT) unit.

The one-dimensional (1D) spectra in H_2O/D_2O (9:1) were recorded using jump-return as the observation pulse.[31] The two-dimensional (2D) NOESY in H_2O (mixing time = 150 ms at 5 °C) was recorded using the WATERGATE sequence.[32] A sweep width of 11000 Hz in both dimensions was used with 64 scans, 4096 data points in t_2 and 512 FIDs in t_1. The data were apodized with a shifted sine bell square function in both dimensions and processed to a 4K × 4K matrix. A series of 1D proton (^1H) and phosphorus (^{31}P) NMR spectra of d(GCGTAGC) in the D_2O solution was recorded at a range of temperatures between –5 °C and +55 °C. For each ^1H experiment, the spectral width was 5000 Hz, 32768 data points were acquired and 8 scans were averaged. ^1H chemical shifts were referenced to the position of the residual HDO signal; the latter was calibrated against 4,4-dimethyl-4-silapentane-1-sulfonate (DSS, 0 ppm) in a separate set of measurements. For the ^{31}P data acquisitions, a spectral width of 2000 Hz was used; 8192 data points were recorded and 48 scans are averaged in each spectrum. During data acquisition, ^1H decoupling (by the WALTZ technique) was performed. ^{31}P chemical shifts were referenced to external 85% H_3PO_4 (0 ppm).

2D gradient double quantum filtered phase sensitive (GDQCOPS[18]) ^1H-^1H correlation spectra in D_2O were recorded at 33 °C, both in the presence and the absence of ^{31}P heteronuclear decoupling (by the WALTZ composite pulse sequence). Data were collected in the phase sensitive hypercomplex mode using 4096 data points in t_2 and 2048 pairs of real and imaginary t_1 increments. A 4000 Hz spectral width was used. The data were apodized with a shifted sine bell square function in both dimensions and processed to a 4K × 4K matrix. Values for ^1H-^{31}P couplings were measured from the non-^{31}P-decoupled GDQFCOSY spectrum as described by Blommers *et al.*[30]

Two-dimensional ^{1}H-^{1}H TOCSY,[19] ω1-decoupled TOCSY,[20] and NOESY[33] spectra of d(GCGTAGC) in D_2O and H_2O were recorded at Georgia State University (Atlanta, GA) on a Varian UnityPlus 600 MHz instrument (operating at 599.897 MHz for ^{1}H NMR and at 242.841 MHz for ^{31}P NMR, equipped with a gradient indirect detection probe and a Varian waveform generator, and running under VNMR software version 5.3B). The TOCSY experiment used the MLEV17 sequence for isotropic mixing of 200 ms duration with a field strength of 7.8 kHz; the mixing time was flanked by two trim pulses of 2 ms each. In the ω1-decoupled TOCSY experiments, a RE-BURP pulse of 11.4 ms was used as semi-selective 180° pulse. The spectral widths were 4000 and 600 Hz in t_2 and t_1, respectively and 160 increments were taken. The strengths of the gradient pulses were 1, –1, 0.8, and –2 G/cm, respectively. The duration of each of the gradient pulses was 1.5 ms.

The ^{1}H-^{1}H NOESY experiments were performed with mixing times of 50, 100, 150, 200, 250 and 300 ms at 33 °C and with a 200 ms mixing time at –5 °C. For all experiments on the DNA heptamer sample in D_2O, the water resonance was selectively suppressed by using a DPFGNOE [34] sequence.

A 2D ^{1}H-detected ^{1}H-^{31}P multiple bond correlation (HMBC[21]) experiment was performed at 33 °C with 512 increments in t_1 and 2048 data points in t_2. The spectral widths of the ^{31}P spectrum in t_1 and the ^{1}H spectrum in t_2 were 800 Hz and 1200 Hz, respectively. For each increment 24 scans were averaged. A delay time of 80 ms was used.

All NMR data were processed on a Silicon Graphics Indigo2 workstation using the Felix95 software package (BioSym/MSI, San Diego, CA).

Structural calculations: Structure generation and refinement were performed on a Silicon Graphics Indigo2 R10000 workstation (IRIX version 6.2) using the InsightII, Discover and NMRrefine software packages licensed from BioSym Technologies (San Diego, CA). Distance geometry calculations were used to generate ensembles of hairpin structures using the DGII program within the BioSym NMRrefine module. The interproton distance constraints between non-exchangeable protons were obtained from NOE build-up rates with the H5-H6 distance of cytidine (2.45 Å) as the internal reference. Hydrogen-bonding restraints were also incorporated into the bounds matrix.

This matrix was smoothed using the triangle and sequential tetrangle inequality relationships. The distances were then embedded in four dimensions to give 20 structures which were then optimized by the simplified DGII simulated annealing procedure. Ten selected hairpin structures were further refined using the iterative relaxation matrix analysis/restrained molecular dynamics (IRMA/RMD) method as implemented in the BioSym software. This process was repeated until the difference between the simulated NOESY spectra and the experimental spectra had converged to a minimum value (as measured by a residual error, the R factor).

The duplex structure was calculated using restrained molecular dynamics (Amber forcefield without electrostatic component, 3000 time steps of 1 fs at 300 K). Using Discover (Biosym Technologies) and starting from a B-type DNA helix. The NOE distance restraints derived from a spectrum recorded in D_2O at −5 °C from the sample with 150 mM NaCl were categorized into strong (1.5-2.5 Å), medium (2.5-3.5 Å) and weak (3.5-5 Å) classes. Hydrogen bonding restraints were treated as NOE distance restraints. Due to the similarity in what was previously observed by Chou *et al.*[2], we assumed sheared base pairing of the G-A mismatch. This procedure resulted in 64 distance restraints, of which 54 belonged to the GTA region of the duplex (14 inter-strand, 40 intra-strand). The resulting structure was further minimized by conjugate gradient energy minimization.

Discussion

Several configurations for purine-purine mismatches have been reported. However, in our monoloop hairpin only the sheared G:A pairing is consistent with our data. This configuration brings the 3'-OH of the descending strand close enough to the 5'-OH of the ascending strand to allow the mismatch base-pair to be spanned by one nucleotide. The sheared G:A pairing, especially when preceded by a C:G pair, has been shown to gain a considerable stabilization by stacking the deoxyribose A5-G6 on one side of the loop and (T4)-G3-C2 on the other side.

As described in the results section, we did not include backbone constraints in the DGII structure calculations. However, the dihedral angles derived from the combination of NOEs and coupling constants correspond very well with average values measured from the modeled structures. The sharp turn in the molecule is possible because it

requires relatively little change in the backbone torsion angles. While all other residues were found to have ζ(*gauche*$^-$) and α(*gauche*$^-$) phosphate dihedral angles, the A5 nucleotide was found to adopt ζ(*trans*)α(*trans*). Combined with γ(*trans*) in between T4 and A5, this causes the 180° backbone turn in the d(GCGTAGC) monoloop hairpin.

Prior to the present work, monoloop hairpins in DNA closed by a sheared G:A pair had been reported by Zhu *et al.*[12](a GCA turn) and Hirao *et al.*[1,11] (a GNA turn; N = G, C, A, T). Chou *et al.*[2] reported on the DNA sequence NAATGNAATG (N = A, G, C, T), the occurrence of an equilibrium between a single residue loop closed by a sheared G:A pair and a duplex containing a G:A bracketed unpaired stack. DNA monoloops closed by sheared A:A and G:G pairs have been described by Chou *et al.*[2] (an AAA turn and a GAG turn). Hirao *et al.*[11] proposed that the unusual ζ(*trans*), α(*gauche*$^-$) and β (*gauche*$^+$) torsion angle combination to cause the backbone turn and Zhu *et al.*[12] determined a ζ(*gauche*$^+$), α(*gauche*$^+$) and β(~*trans*) loop conformation. This might indicate that other backbone torsion angle combinations can, likewise, cause monoloops.

Unlike Hirao *et al.*,[1] we find that the d(GCGTAGC) oligomer exists in solution as a temperature and salt concentration dependent equilibrium between a monoloop and a duplex structure. At 33 °C with 150 mM NaCl the duplex structure is not detectable in a one-dimensional ^{1}H-spectrum. When this sample is cooled, duplex signals occur in a one-dimensional proton spectrum and increase in strength with decreasing temperature. If the sample has no additional NaCl, then the duplex structure is undetectable, even not at 0 C. In Hirao's paper on d(GCGNAGC) monoloop structures (N = T, C, A, G), most of the NMR study was performed on 0.54mM solutions. Only a few 1 dimensional experiments were reported for samples of 5.4mM heptamer and 0.15M NACl. It was pointed out that for d(GCGGAGC) extra peaks occurred in the region of the base protons and for d(GCGAAGC) the number of base protons is double at 5.4mM compared with the sample at 0.54mM. Since they did not perform experiments at temperatures below room temperature, the extra signals arising from the duplex structure in d(GCGTAGC) could not be detected in their 1 dimensional ^{1}H-spectrum in the 5.4mM sample.

Acknowledgments

This research was supported by grant GOA97/11 from the Katholieke Universiteit Leuven and the Granting Agency of the Czech Republic (203/99/0123). E.

Lescrinier is a research assistant of the Belgian National Foundation for Scientific Research (NFWO). We are indebted to Dr. A Van Aerschot and Mr. G. Schepers for synthesizing the d(GCGTAGC) oligonucleotide, and to Dr. A. De Bruyn and Prof. H. van Halbeek for helpful discussions.

REFERENCES

1. Yoshizawa, S.; Kawai, G.; Watanabe, K.; Miura, K. Hirao, I.; *Biochemistry,* 1997 *36,* 4761-4767.
2. Chou, S.-H.; Zhu, L.; Reid, B. R. I. **1996** *259,* 445-457.
3. Varani, G. Ann. Rev. *Biophys. Biomol. Struct.,* **1995** *24,* 379-404.
4. Hirao, I.; Nishimura, Y.; Naraoka, T.; Watanabe, K.; Arata, Y.; Miura, K. *Nucl. Acids Res.,* **1989** *17,* 2223-2231.
5. Hilbers, C. W.; Blommers, M. J. J.; van de Ven, F. J. M.; van Boom, J. H.; van der Marel, G. A. *Nucleosides Nucleotides,* **1991** *10,* 61-80.
6. Tuerk, C.; Gauss, P.; Thermes, C.; Groebe, D. R.; Gayle, M.; Guild, N.; Stormo, G.; D'Aubenton-Carafa, Y.; Uhlenbeck, O. C.; Tinoco, I.; Brody, E. N.; Gold, L. *Proc. Natl. Acad. Sci.,* **1988** *85,* 1364-1368.
7. Heus, H.; Pardi, A. *Science,* **1991** *253,* 191-194.
8. Heus, H. A.; Uhlenbeck, O. C.; Pardi, A. *Nucleic Acids Res.,* **1990** *18,* 1103-1108.
9. Michel, F.; Westhof, E. *J. Mol. Biol.,* **1990** *1050,* 585-610.
10. Hirao, I.; Ishida, N.; Watanabe, K.; Miura, K. *Biochim. Biophys. Acta,* **1990** *20,* 199-204.
11. Hirao, I.; Kawai, G.; Yoshizwa, S.; Nihshimura, Y.; Ishido, Y.; Watanabe, K.; Miura, K. *Nucleic Acids Res.,* **1994** *22,* 576-582.
12. Zhu, L.; Chou, S.-H.; Xu, J.; Reid, B. R. *Nature Struct. Biol.,* **1995** *2,* 1012-1017.
13. Bates, G.; Lenarch, H. *BioEssay,* **1994** *14,* 277-284.
14. Sutherland, G. R.; Richards, R. I. *Nature Genet.,* **1995** *6,* 114-116.
15. Sandusky, P.; Wooten, E.W.; Kurochkin, A.V.; Kavanaugh, T.; Mandecki, W.; Zuiderweg, E.R.P. *Nucleic Acids Res.,* **1995** *23,* 4717-4725.
16. Pardi, A.; Walker, R.; Rapoport, H.; Wider, G.; Wüthrich, K. *J. Am. Chem. Soc.,* **1983** *105,* 1652-1653.
17. Wüthrich, K. *NMR of Proteins and Nucleic Acids,* Wiley, New York, 1986.
18. Willker, W.; Leibfritz, D.; Kerssebaum, R.; Bermel, W. *Magn. Reson. Chem.,* **1993** *31,* 287-292.
19. Levitt, M.; Freeman, R.; Frenkiel, T. *J. Magn. Reson.,* **1982** *47,* 328.
20. Van Halbeek, H.; Sheng, S.; Lescrinier, E.; De Bruyn, A.; Herdewijn, *P. Semi-Selective F1-Decoupled TOCSY Recorded with Pulsed Field Gradients,* 38th Experimental NMR Conference, Orlando, FL, USA, 1997, 23-27 March.
21. Summers, M. F.; Marzilli, L. G.; Bax, A. *J. Am. Chem. Soc.,* **1986** *108,* 4285-4294.
22. Neuhaus, D.; Williamson, M. in *The Nuclear Overhauser Effect in Structural and Conformational Analysis,* New York, 1989.

23. Katahira, M.; Kanagawa, M. Sato, H.; Uesugi, S.; Fujui, S.; Kohno, T.; Maeda, T. *Nucleic Acids Res.*, **1994** *22*, 1752-1759.
24. Blommers, M. J. J.; Van de Ven, F. J. M.; Van der Marel, G. A.; Van Boom, J. H.; Hilbers, C. W. *Eur. J. Biochem.*, **1991** 201, 33-51.
25. Mooren, M. M. W.; Pulleyblank, D. E.; Wijmenga, S. S.; van de Ven, F. J. M.; Hilbers, C. W. *Biochemistry*, **1994** *33*, 7315-7325.
26. Orbons, L. P. M.; van Beuzekom, A. A.; Altona, C. *J. Biomol. Struct. Dynam.*, **1987** *4*, 965-987.
27. Dickerson, R. E. *J. Mol. Biol.*, **1983** *166*, 419.
28. Dickerson, R. E.; Drew, H. R. *J. Mol. Biol.*, 1981 *149*,761.
29. Lankhorst, P. P.; Haasnoot, C. A.G.; Erkelens, C.; Altona, C. *J. Biomol. Struct. Dyn.*, **1984** *1*, 1387.
30. Blommers, M.J.J.; Nanz, D.; Zerbe, O. *J. Biomol. NMR*, **1994** *4*, 595-601.
31. Plateau, P.; Guéron, M. *J. Am. Chem. Soc.*, **1982** *104*, 7310-7311.
32. Piotto, M.; Saudek, V.; Sklenar, V. *J. Biomol. NMR*, **1992** *2*, 661-665.
33. States, D. J.; Haberkorn, R. A.; Ruben, D. J. *J. Magn. Reson.* **1982** *48*, 286-292
34. Scott, K.; Stonehouse, J.; Keeler, J.; Hwang, T.-L.; Shaka, A. J. *J. Am. Chem. Soc.*, **1995** *117*, 4199-5200.

Received : 3 / 8 / 99
Accepted : 6 / 1 / 99