Nucleic Acids Research, Vol. 19, No. 19 5159–5167

The solution structure of a DNA hairpin containing a loop of three thymidines determined by nuclear magnetic resonance and molecular mechanics

Y.Boulard, J.Gabarro-Arpa[1], J.A.H.Cognet[1], M.Le Bret[1], A.Guy[2], R.Téoule[2], W.Guschlbauer and G.V.Fazakerley*
Service de Biochimie et de Génétique Moléculaire, Département de Biologie Cellulaire et Moléculaire, Centre d'Etudes de Saclay, 91191 Gif-sur-Yvette Cedex, [1]Laboratoire de Physicochimie Macromoléculaire, Institut Gustave Roussy, F-94800 Villejuif, [2]Laboratoires de Chimie/Radiobiochimie, Département de Recherche Fondamentale, Centre d'Etudes Nucléaires de Grenoble, BP 85X, 38041 Grenoble Cedex, France

Received July 31, 1991; Revised and Accepted September 12, 1991

ABSTRACT

We have determined by two-dimensional nuclear magnetic resonance studies and molecular mechanics calculations the three-dimensional solution structure of a 21 residue oligonucleotide capable of forming a hairpin structure with a loop of three thymidine residues. This structure is in equilibrium with a duplex form. At 33°C, low ionic strength and in the presence of $MgCl_2$ the hairpin form dominates in solution. Six Watson–Crick base pairs are formed topped by the loop structure. The residues 1–3 and 18–21 are not complementary and form dangling ends. Distance constraints have been derived from nuclear Overhauser enhancement measurements. These, together with molecular mechanics calculations, have been used to determine the structure. We do not observe stacking of thymidine residues either over the 3′ or the 5′ end of the stem.

INTRODUCTION

That the extrusion of palindromic sequences in negatively supercoiled closed circular DNA would result in relaxation of torsional stress has been recognized for a long time (1–5). This results in a cruciform structure of two hairpins with a loop region of variable length. These palindromic sequences are not randomly distributed (6) but occur frequently near regulation and promotor sites (6–8). Evidence for the existence of such cruciform structures comes from the enhanced chemical and enzymatic reactivity of the bases at the centre of the palindromic sequence (9–11). This enhanced reactivity is explained by the greater exposure of the residues in the loop region to attack by these molecules.

In a study of cruciform structures formed in the plasmid pBR322 (12) we note the presence of GATC sequences in the stem region of most of the cruciforms identified. An adenine methylase, the Dam methylase, specifically recognizes this sequence and its action is implicated in certain cellular control functions and during mismatch repair (13–15). We have previously reported NMR studies on oligonucleotide sequences containing unmethylated, hemimethylated and fully methylated GATC sequences (16–20). We have thus chosen to include two GATC sequences in the oligonucleotide studied here. Our aim is to study a cruciform structure with a mobile junction. This will be composed of two non-identical oligonucleotides each of 21 residues. As a first step, reported below, we have examined the solution behaviour and structure of one of the two strands. The sequence is capable of forming, in a hairpin structure and excluding possible mismatched base pairs, 6 base pairs, a loop of three thymidines and is terminated at each end by three non-selfcomplementary bases or dangling ends necessary for the formation of our cruciform structure. The hairpin structure could also be in equilibrium with a duplex structure containing 12 base pairs.

MATERIALS AND METHODS

The oligonucleotide studied has the sequence 5′d(C^1 G^2 T^3 G^4 G^5 A^6 T^7 C^8 G^9 T T T C^{13} G^{14} A^{15} T^{16} C^{17} C^{18} G^{19} A^{20} G^{21}).

The synthesis was performed using standard cyanoethyl diisopropyl phosphoramidite methodology with 10 μmol of dG^{ibu} grafted silica support. Cleavage from the solid support and deprotection of the oligonucleotide was carried out with (4×500μl) aqueous NH_3 solution for four hours at room temperature. A further 24 hours at 60°C were necessary to fully remove the isobutyryl protecting groups. The oligonucleotide was purified by HPLC on an anion exchange column (0.75×30 cm, Partisil SAX 10 μm) using a linear gradient of 0.3 M KH_2PO_4 buffer (pH 6.8) with 30% CH_3CN over 50 min. The pure product (700 O.D.) was desalted by dialysis.

* To whom correspondence should be addressed

NMR spectra. The oligonucleotide was 3 mM in strand concentration, dissolved in 10 mM phosphate buffer, pH 7.0 (unless otherwise stated), 10 mM $MgCl_2$ and 0.2 mM EDTA. Chemical shifts were measured relative to the internal reference tetramethylammonium chloride, 3.18 p.p.m.

NMR spectra were recorded on a Bruker AM600 spectrometer. NOESY spectra were recorded with mixing times of 50, 60, 80, 100 and 400 ms in ^2H_2O in the phase sensitive mode (21). Typically 256 t1 increments were recorded. After zero filling the data were multiplied by a 5−15 degree shifted sine bell function in both dimensions except for the short mixing time experiments. For these the data were multiplied by a sine bell shifted by Π/2 prior to Fourier transformation. Distance determination from the initial NOE build up curves was done as described previously (22). This method takes into account the observed differences in effective correlation time for different classes of interproton vectors. This phenomenon has recently been treated theoretically (23 and references therein). For 1-dimensional spectra in H_2O the observation pulse was replaced by a jump and return sequence (24) with the pulse maximum placed at 15 p.p.m..

The TOCSY spectra were recorded in the phase sensitive mode (25) with a 25 and 80 ms mixing times.

UV melting curves were recorded on a Perkin Elmer Lamda 15 spectrophotometer equipped with a Peltier effect thermostated cell holder with a temperature gradient of 1°C per minute for both heating and cooling.

Molecular Mechanics Calculations. Molecular models were constructed with the program RTC, written in this laboratory which enables us to explore a very wide range of conformations of the oligonucleotide. The principal idea is the following: from a starting set of main chain torsion angles, new conformations were generated by incrementing stepwise each angle, by a fixed value, in the first stage by 10°. In this way the program generates systematically all possible conformations.

The conformations thus generated were kept, if they conformed to a set of external distance constraints, derived from the proton-proton NOE derived distances, and if the atoms were further apart than the sum of their van der Waals radii. Having determined a local minimum the program RTC was rerun within this minimum with increments of 2° for all angles. Pairwise interactions between atoms were calculated with the potential functions and parameters of AMBER (26−28).

Energy calculations were carried out with the program AMBER to create the molecular topology file, and the program MORMIN (to be published), which uses the energy parametrization of AMBER, to do the minimizations. The parameters used were as described elsewhere (26,28). All hydrogen atoms were treated explicitly. To simulate the screening effect of the solvent a gas phase potential was employed where the dielectric constant D_{ij}, is proportional to the distance d_{ij} separating a pair of atoms: $D_{ij} = C\ d_{ij}$ (29−30). C was taken as 4 $Å^{-1}$. All atom pairs were included in the calculations of non-bonding interactions. In this work the minimizations were run with the 1−4 interatomic interactions divided by two in agreement with other workers (30). The negative charges of the phosphate groups of the oligonucleotide were neutralized by spheres of a positive unit charge. These counterions had a van der Waals radius of 2 Å and a well depth of 0.3 kcal/mol. They were originally set at 3 Å from the phosphate oxygen atoms.

Energy refinements were usually terminated when the root mean square of the energy gradient was less than 0.2 kcal/Å. Sugar pucker conformations were modified by changing the corresponding torsion angle δ (C5′-C4′-C3′-O3′). The torsion angle δ was forced to δ0 by addition of $E = \Sigma\ k(\delta-\delta 0)^2$ to the energy function of AMBER. The constant k was set to 900 kcal/mol.rad^2. For some energy minimizations, the position in space of certain atoms was constrained around fixed coordinates with a harmonic potential, the force constant used was 20 kcal/mol.rad^2.

Comparison with the NMR data. The value $F = \Sigma\ (r_{ij}-d_{ij})^2/d_{ij}$, where r_{ij} is the distance obtained by NMR and d_{ij} is the distance in a proposed structure, was used to give an estimate of the overall agreement between the computed structure and the NMR data. The function, F, although empiric, has the advantage of being a simple sum and of giving more weight to short distances that are normally measured with the best precision.

The molecular structures were displayed on a Silicon Graphics IRIS 4D70GT using the program MORCAD (to be published).

RESULTS

An oligonucleotide sequence which is capable of forming a hairpin structure is, by definition, also capable of forming a duplex structure albeit with mismatched base pairs or loops. We have first searched for the possible existence of both species by carrying out UV melting experiments. Figure 1 shows the melting curve for a solution 3μM in strand concentration in 100mM NaCl at pH 7.0. The melting curve is clearly biphasic showing an initial T_m of 18°C and a second one of 68°C. Although the heating and cooling curves were recorded at the same rate, 1°C per minute, only the first part of the cooling curve is directly superposable on the heating curve. The first T_m is completely absent. We observe that on keeping the solution at 4°C the original optical density is recovered only after ca. 48 hours.

We have investigated the effect of strand concentration and also ionic strength on the melting processes. At an ionic strength of 10mM we have measured these two T_m values as a function of the strand concentration. Figure 2 shows the observed values obtained in the concentration range 0.7−6 μM. As the concentration is increased the first T_m also increases, from 9.3°C to 18°C at the highest concentration. On the other hand that of the second Tm does not change within the limits of experimental error. From this we can conclude that the first T_m corresponds to a bimolecular reaction, that is, the melting and rearrangement or formation of a duplex structure and that the second corresponds to a monomolecular reaction involving the hairpin structure.

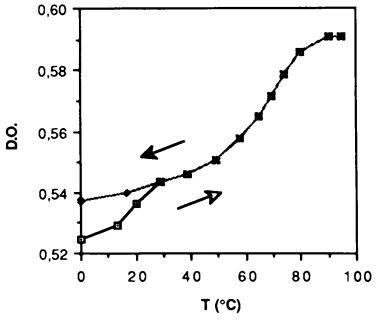


Figure 1. UV absorption as a function of temperature, heating and cooling, for the oligonucleotide, 3μM, in 100 mM NaCl at pH 7.0.

Our aim was not to study the thermodynamics of these processes but rather to find the optimal conditions for NMR studies. Raising the NaCl concentration from 10 to 500 mM increases the T_m of the first transition by ca. 8°C and that of the second by ca. 12°C. On the other hand at low ionic strength, 10 mM, addition of $MgCl_2$ up to 10mM had little effect upon the stability of the duplex structure but significantly, by 15°C, stabilized the hairpin structure which shows a T_m of 73°C.

The UV studies indicated that the sequence studied here may exist in both duplex and hairpin forms and at NMR concentration the duplex form should be even more stable. We have chosen to carry out the NMR experiments at low ionic strength, 10 mM phosphate buffer, and in the presence of 10 mM $MgCl_2$ to stabilize the hairpin form. The remaining variable is that of the temperature. We have followed the chemical shifts of the thymidine methyl groups between 0 and 85°C, four of which are shown in Figure 3 (T^A and T^B of the loop structure, T^{16} and T^7 of the stem, their assignment is described below). In the temperature range 1–30°C we observed shifts for all four resonances. In this temperature range and this concentration changes in proton chemical shifts would be expected, as oligonucleotides tend to aggregate at low temperature. For the methyl resonances of T^A and T^{16} the change in chemical shift on raising the temperature to 30°C shows no clear transition behaviour. On the other hand the methyl resonance found at 1.70 ppm at 1°C, T^B, showed a typical sigmoid curve for the chemical shift with respect to temperature indicating a transition from one species to another. Since the unknown effect of aggregation is superimposed we cannot determine the T_m for this process. Above 50°C all the methyl groups shift, their chemical shifts showing a typical sigmoid curve with temperature as the hairpin structure melts; the T_m is ca. 65°C for the loop residues and ca. 73°C for the stem residues. Certain resonances in the aromatic region of the spectrum are in the slow exchange rate for the structures duplex and hairpin. While these have not been identified, at 30°C the oligonucleotide is ca. 90% in the hairpin form and at 43°C we no longer observe the duplex form.

We have also recorded spectra in H_2O to try to follow the duplex-hairpin transition. At pH 5, 1°C, we observe three sharp imino resonances in the range 10.5–11.5 ppm (not shown) which no doubt correspond to the three central T residue imino protons which are to some extent protected from exchange with the solvent in the duplex form. As the temperature is raised the intensity of these resonances decreases and others appear. But, at the same time, all these resonances become larger such that at 30°C they are no longer resolved. The imino protons of the A.T and G.C base pairs are sharper but are also poorly resolved. It has not been possible to study in detail the hairpin structure in H_2O.

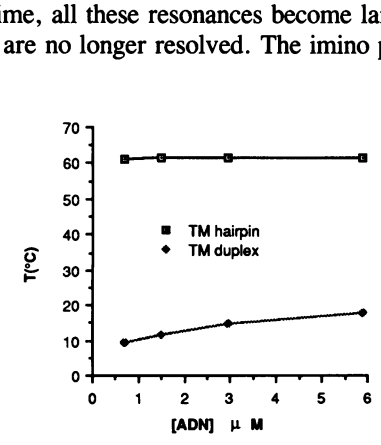


Figure 2. T_m values as a function of oligonucleotide concentration in 100 mM NaCl at pH 7.0.

Spectra of non-exchangeable protons

We first recorded two NOESY spectra with a mixing time of 400 ms at 33°C and 43°C. The two spectra are qualitatively similar but for the same accumulation time the signal to noise was significantly better at 33°C and we do not observe cross peaks involving the minor duplex species. All subsequent experiments were carried out at 33°C. The region for the interactions H8H6/H1'H5 of the 400 ms spectrum is shown in Figure 4. Usually, resonance assignment can be started from the ends of a helix but in this case we do not know whether the terminal non-hydrogen bonded trinucleotides will give rise to cross peaks. We can differentiate the two strands of the stem by starting from the downfield A H8 resonances. While this is an assumption it is confirmed by the fact that we are able to make the complete assignment and is consistent with other regions of the spectrum. From A^6 we should find G^5 and then another G residue whereas from A^{15}, G^{14} and then a C residue. From the resonance at 8.23 ppm we can follow to a G residue at 7.94 ppm and then to a C residue at 7.55 ppm. This must be the fragment A^{15}–C^{13}. For C^{13} we do not observe a cross peak with a T residue in this region of the spectrum. In the other direction we can follow the connectivities to T^{16}, through C^{17} and C^{18} although the cross peaks are weak. But the chain does not stop at the last base which can be hydrogen bonded. From C^{18} we can continue through G^{19} to A^{20} at 8.07 ppm. Even though these bases are not hydrogen bonded significant base stacking must occur such that we observe interresidue cross peaks. From the AH8 resonance at 8.19 ppm we can follow the connectivities T, C to G^9. Here the chain continues to a T residue, labelled T^A

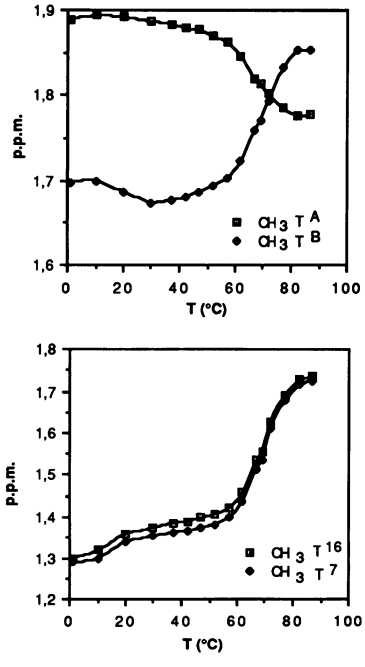


Figure 3. Chemical shift as a function of temperature for the methyl groups of T^A and T^B (upper) and T^{16} and T^7 (lower). The oligonucleotide was in 10 mM phosphate buffer, pH 7.0, and with 10 mM $MgCl_2$.

and then via a weak interresidue cross peak to another T residue, labelled T^B where the chain stops. In the other direction we can follow the connectivities from A^6 to G^4. A dinucleotide involving a T residue can be identified from an interresidue cross peak and this must be G^2-T^3. The base protons of the three remaining residues can be identified by elimination. A strong cross peak between resonances at 7.47 and 5.83 must arise from the interaction H6-H5 of C^1 (also seen in TOCSY spectra). The missing T is found at 7.49 ppm (from the intense H6-CH_3 cross peak, see below) and thus G^{21} is assigned to the peak at 7.86 ppm from its intraresidue cross peak. The A^{15}H2 shows three cross peaks with H1′ protons, peaks A−C as does the A^6H2, peak D, and two overlapping ones labelled peak E. Two normal interbase cross peaks are observed for the interactions TH6-CH5, peaks F and G. One other cross peak is found, peak H, which appears to be between G^2H8 and C^1H5.

Figure 5 shows another region of the same spectrum, that for interactions between base protons and H2′H2″CH_3 protons. For the stem part the intra and interresidue cross peaks correspond to what is expected for a normal B form DNA. Usual interbase H8-TCH_3 cross peaks are observed for the steps A^6-T^7 and A^{15}-T^{16}, peaks A and B respectively. Another, much weaker, is observed between G^2 and T^3, peak C. Concerning the loop we observe that both G^9H8 and T^AH6 give a cross peak with the $T^B$$CH_3$ resonance peaks D and E respectively. T^AH6 shows one interresidue cross peak with the H2′ or H2″ of G^9, the other would be overlapping with an intraresidue cross peak. Similarly T^BH6 shows a cross peak with one of the H2′/H2″ protons of T^A. On the other hand only intraresidue cross peaks are observed for T^C.

Complete assignment of the H2′H2″H3′H4′ and partial assignment of the H5′H5″ (but not the relative assignment) has been obtained through analysis of TOCSY spectra (not shown). The relative assignment of the H2′ and H2″ resonances was obtained from analysis of a short mixing time NOESY spectrum. The observed chemical shifts are given in Table 1.

The loop structure may be expected to give rise to interactions which are not normally observed for B DNA. All regions of the 400 ms NOESY spectrum have been carefully analysed to search for such close contacts. Obviously, at such a long mixing time, these may be strongly affected by spin diffusion. Figure 6 shows another region of this spectrum. We observe that the two T CH_3 resonances of the stem residues, T^{16} and T^7, in the region 1.3−1.4 ppm show cross peaks with the H3′ resonances of the A residue in the 5′ direction, peaks A and B respectively. These are likely to arise via spin diffusion involving the A H2′H2″ protons. No other cross peaks are observed in this region for these methyl groups. The cross peak pattern observed with the methyl groups of T^A and T^B is quite different. The $T^B$$CH_3$ resonance at 1.67 ppm shows cross peaks with the H3′, H4′ resonances of T^A,peaks D and E respectively. Both of these cross peaks are weak and are of similar intensity. The $T^A$$CH_3$ resonance at 1.88 ppm shows a large cross peak with the H4′ resonance of G^9, peak F and a much smaller one with the H3′ resonance, peak C. Due to spectral crowding we cannot say if

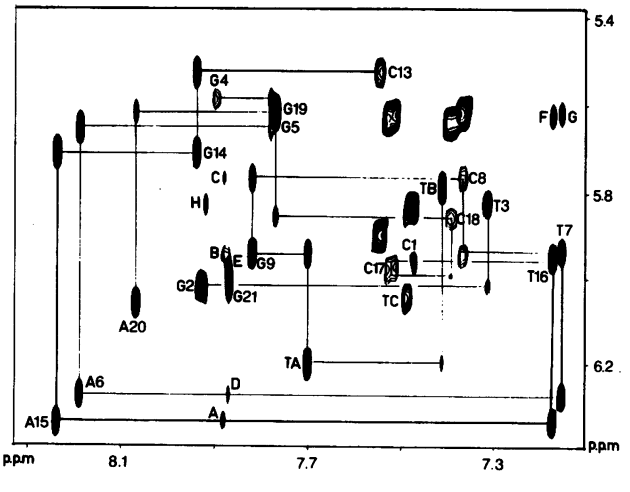


Figure 4. Part of the NOESY spectrum recorded with a 400 ms mixing time, at 33°C, corresponding to the interactions between the H8/H6 protons and the H1′/H5 protons.

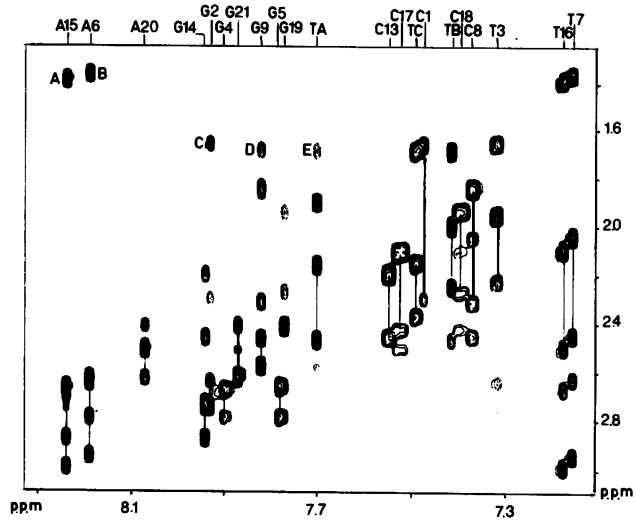


Figure 5. Part of the NOESY spectrum recorded with a 400 ms mixing time corresponding to the interactions between the H8/H6 protons and the H2′/H2″/CH_3 protons.

Table 1. Chemical shifts of non-exchangeable protons at 33°C.

	H8/H6	H5/CH3	H1′	H2′	H2″	H3′	H4′
C^1	7.47	5.83	5.96	1.65	2.28	4.56	4.01
G^2	7.93		6.02	2.62	2.73	4.96	4.34
T^3	7.32	1.64	5.83	1.94	2.21	4.77	4.18
G^4	7.90		5.58	2.67	2.78	4.96	4.32
G^5	7.78		5.66	2.65	2.77	5.00	4.36
A^6	8.19		6.27	2.61	2.93	4.99	4.43
T^7	7.16	1.35	5.94	2.04	2.44	4.84	4.17
C^8	7.37	5.61	5.77	1.82	2.30	4.81	4.13
G^9	7.82		5.94	2.56	2.45	4.97	4.35
T^A	7.70	1.88	6.19	2.15	2.45	4.72	4.31
T^B	7.42	1.67	5.79	1.98	2.23	4.66	3.97
T^C	7.49	1.66	6.04	2.14	2.36	4.71	4.16
C^{13}	7.55	5.89	5.53	2.18	2.43	4.69	4.14
G^{14}	7.94		5.72	2.73	2.86	5.02	4.38
A^{15}	8.23		6.33	2.66	2.96	5.03	4.46
T^{16}	7.18	1.39	5.95	2.09	2.49	4.88	4.21
C^{17}	7.52	5.63	5.98	2.10	2.42	4.82	4.16
C^{18}	7.40	5.65	5.86	1.93	2.25	4.78	4.09
G^{19}	7.77		5.62	2.40	2.40	4.87	4.15
A^{20}	8.07		6.06	2.50	2.61	4.92	4.27
G^{21}	7.86		6.00	2.40	2.61	4.66	4.15

there is a cross peak with the H5′H5″ protons. No cross peaks are found with the G^9 H2′H2″ resonances We do not observe any interresidue interactions with T^C, neither in this region nor in any other.

In the 400 ms NOESY spectrum we observe more than 100 intra or interresidue cross peaks for the loop and the G.C base pair which closes it, Figure 7. Many of these disappear in short mixing time experiments and of those that rest not all contain significant conformational information. The cross peaks for which the NOE build up curves have been measured are shown in Figure 7 and Table 2. It should be noted that several involving the loop are unusual. We observe a close contact between the H1′ proton of G^9 and the H2′ proton of T^A, also contacts between the T^BCH_3 and the H2″ of T^A. We note that the relatively strong cross peak G^9H8-T^BCH_3, peak D in Figure 5, disappears as soon as the mixing time is reduced and must arise from spin diffusion.

The relative intensity of the intraresidue cross peaks between the base H6/H8 protons and the H2′ versus H3′ protons gives a good indication of the sugar pucker (31). We have examined these pairs of interactions for all residues and find no evidence for significant deviation from C2′-endo geometry. This was confirmed in the phase sensitive COSY spectra (not shown) in which we observe that for all sugars J1′2′ + J1′2″ lies in the range 14.5–16.2 Hz. (32).

The assignment of the loop T residues relative to their position in the oligonucleotide sequence is not entirely straightforward.

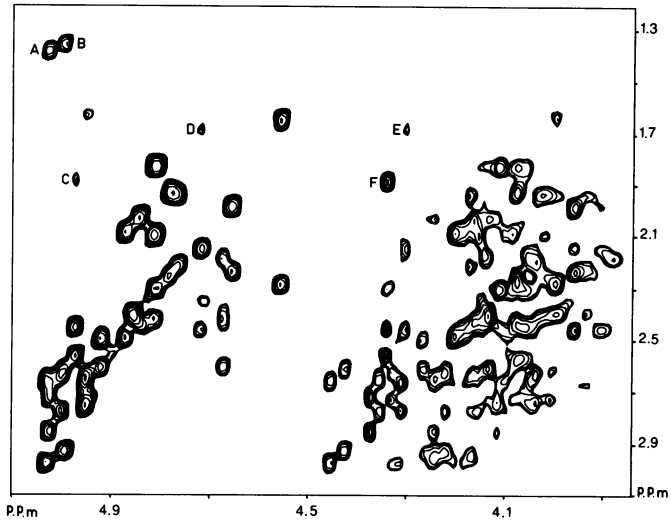


Figure 6. Part of the NOESY spectrum showing interactions between the H2′/H2″/TCH_3 protons and the H3′/H4′/H5′/H5″ protons.

Table 2. Interresidue distance constraints used by the program RTC for creating the loop structure.

a)			
2.3 <	G^9H2″	– T^AH6	< 3.2
3.2 <	G^9H1′	– T^AH2′	< 4.0
3.5 <	G^9H4′	– T^ACH3	< 4.2
	G^9H4′	– T^ACH3	< G^9H3′ – T^ACH_3
b)			
2.3 <	T^AH2″	– T^BH6	< 2.8
3.2 <	T^AH6	– T^BCH_3	< 4.2
3.0 <	T^AH2″	– T^BCH_3	< 4.2
4.1 <	T^AH2′	– T^BCH_3	

We observe unusual interactions between G^9 and T^A, some interactions between T^A and T^B, and no interactions involving T^C and another residue. In view of the interactions between T^A and T^B it seems unlikely that T^C is between these two residues. T^C must be directed away from the loop to account for the absence of interresidue interactions. We could place it at either the 5′ or 3′ end of the loop. It is tempting to place it at the 3′ end in view of the observed interactions but we carried out preliminary modelling studies for both possibilities. It was not possible to create a reasonable model with T^C at the 5′ end of the loop while at the 3′ end the model fits well the distance constraints. Thus the oligonucleotide sequence for the central part is G^9 T^A T^B T^C C^{13}.

We observe, as stated above, an interaction between T^BH6 and T^AH2″ which could indicate that T^B and T^A are stacked. However, we observe interactions which are in contradiction with this. For normal stacking the interactions T^BCH_3–T^AH2″H2′ are strong. The interaction with T^AH2″ is on the limit of

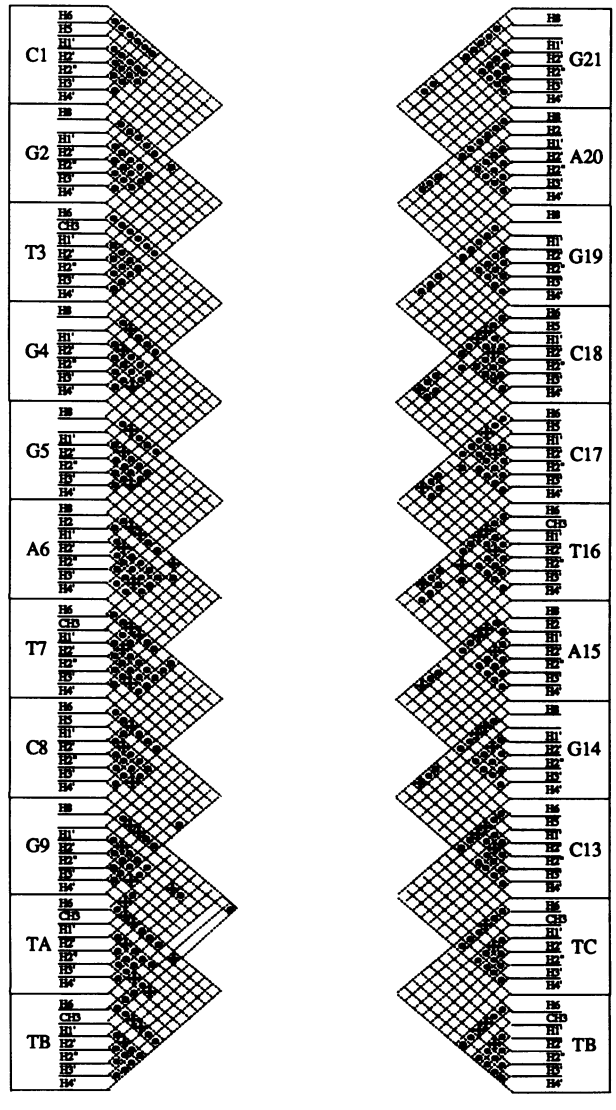
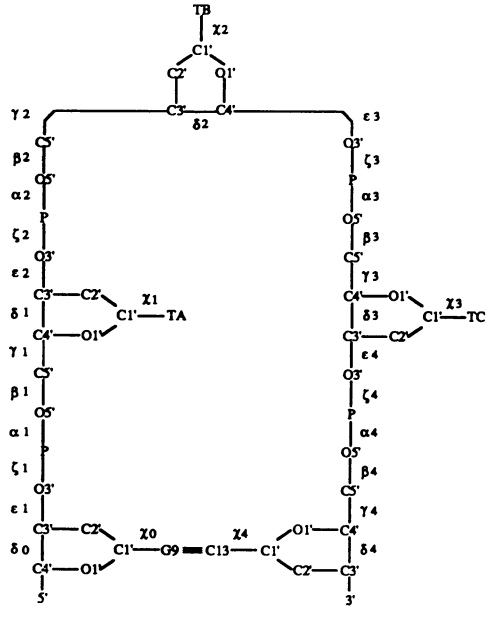


Figure 7. Summary of all NOEs observed. ◙ NOEs observed in the 400 ms spectrum. ⊞ Selected NOEs for which the initial NOE build up curves were measured for distance determinations.

detection at short mixing times and we calculate an interproton distance of ca. 4.0 Å, Table 2, and that with the H2′ proton is undetectable. Qualitatively we can conclude that the arrangement of T^B with T^A is significantly perturbed from that of stacking in B DNA. Similarly, the interactions observed between T^A and G^9 are very different from those observed between adjacent residues in B DNA and T^A cannot be stacked normally on G^9.

Model building and molecular mechanics calculations

The construction of the entire molecular model was started by building the loop structure on the top of the hairpin. The construction of the loop required three steps. Firstly we built the fragment d(GpT).d(C), with the G.C base pair in a standard Watson–Crick conformation. This involves searching for the position of T^A relative to the G.C base pair by systematic variation of the torsion angles $\delta 0$, $\chi 0$, $\epsilon 1$, $\zeta 1$, $\alpha 1$, $\beta 1$, $\gamma 1$, $\delta 1$, and $\chi 1$ (scheme 1). $\delta 0$ and $\delta 1$ were set to values that correspond to a C2′ endo conformation as indicated by the NMR distances. The torsion angles $\chi 0$ and $\chi 1$ were set to values that reproduced the intraresidue distances, Table 2. With the program RTC we searched through all conformations that can be obtained by changing the torsion angles $\epsilon 1$, $\zeta 1$, $\alpha 1$, $\beta 1$ and $\gamma 1$ (scheme 1), as described in Materials and Methods. The program RTC was run to explore all conformations that can be made when each angle is allowed to vary from 30–330° (*cis* conformations were excluded), in steps of 10°. After this initial exploration RTC was run with a smaller step of 2° in the range of best agreement between the extreme angles found in the first run extended by 10° on either side. Experimental interresidue NOE derived distances were not given unique values but rather a fairly wide distance range. The effective correlation time for proton–proton interactions in and next to the loop structure may be influenced by local movement. The fourth distance constraint in Table 2a is that the distance between the methyl group of T^A and H4′ of G^9 be smaller than that to the H3′ of G^9. That this is so is



Scheme 1. Torsion angle notation for the hairpin loop closed by the G.C base pair.

Table 3. Torsion angle ranges found by RTC for a) the sequence d(GpT^A).d(C), b) and c) for the TpT fragment in d(GpT^{A}pT^B).d(C).

a)

$\epsilon 1$		$\zeta 1$		$\alpha 1$		$\beta 1$		$\gamma 1$	
min	max	min	max	min	max	min	max	min	max
−161	−147	−125	−121	31	65	163	−179	175	−159

b)

$\epsilon 2$		$\zeta 2$		$\alpha 2$		$\beta 2$		$\gamma 2$	
min	max	min	max	min	max	min	max	min	max
141	−127	−127	−30	54	104	109	151	112	180

c)

$\epsilon 2$		$\zeta 2$		$\alpha 2$		$\beta 2$		$\gamma 2$	
min	max	min	max	min	max	min	max	min	max
141	−169	−114	−82	−105	−33	−175	−87	29	49

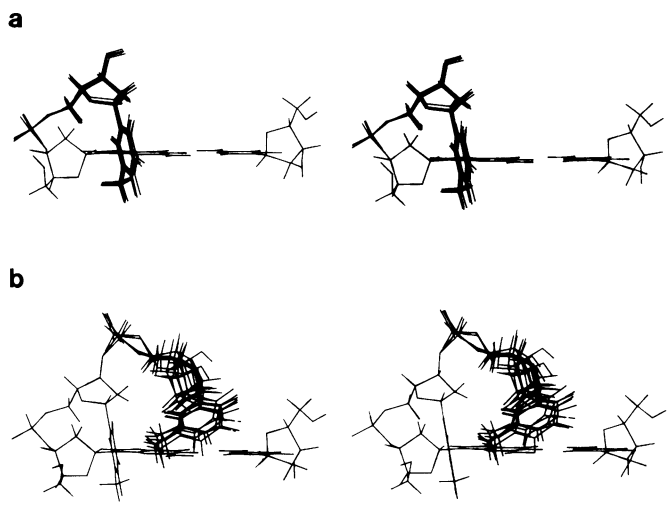


Figure 8. a) A family of conformations satisfying the distance constraints for incorporation of the first T residue on the G.C base pair of the stem. b) as a) with addition of the second T residue.

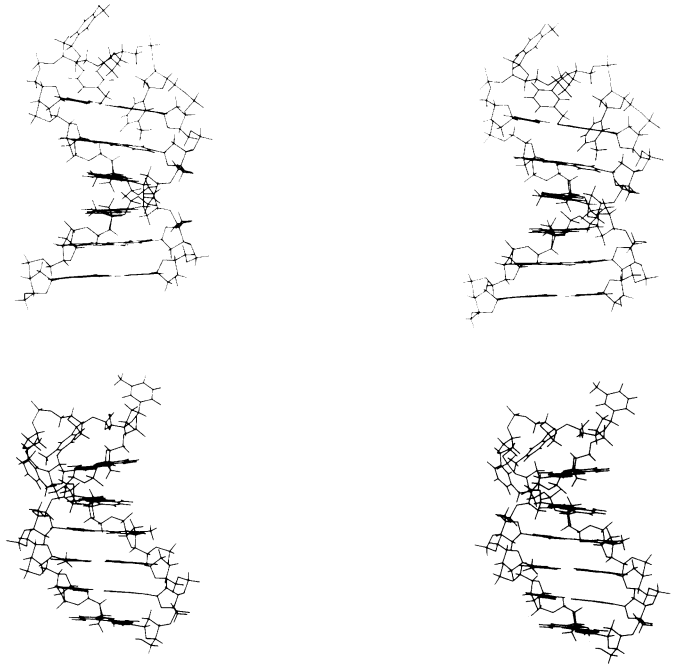


Figure 9. Two stereoscopic views of one structure of a closely related family of the whole hairpin molecule except for the dangling ends.

clearly seen in Figure 6. The conformations satisfying the constraints in Table 2a are found to be confined to narrow ranges, Table 3a and Figure 8a. The torsion angles $\epsilon 1$, $\zeta 1$ and $\beta 1$ remain close to the mean values of a BI conformation (33), while $\alpha 1$ is centered around g^+ and $\gamma 1$ around t. The base plane of T^A is perpendicular to that of G^9 and is positioned halfway between the sugar and the base of G^9 in the minor groove.

In a second stage we have built in the same way the oligonucleotide d(GpT^{A}pT^B).d(C), from the best structures previously found for d(GpT^A).d(C). As previously the angles $\chi 1$ and $\delta 1$ were set to *anti* and C2′ endo conformations as indicated by the NMR results. The program RTC then searched through all possible conformations of T^B, in the same angle intervals and the two stages as described above. In order to position T^B relative to the rest of the structure the angles $\epsilon 2$, $\zeta 2$, $\alpha 2$, $\beta 2$ and $\gamma 2$, scheme 1, were varied. The distance constraint ranges are given in Table 2b. The first three are derived from observed NOEs whereas the fourth is an interaction normally observed for stacked base pairs but unobserved in this system even in NOESY spectra with long mixing times. The absence of an NOE could be due to local motion. We have previously observed that even a cytidine H5-H6 NOE can be nulled when the residue is extrahelical (34). However, here we observe a number of intra and interresidue NOEs involving T^B and their magnitude suggests that the absence of the T^BCH_3-T^AH2′ interaction is not due to local motion but rather to a long internuclear distance. The possible conformations fall in a range a little wider than that observed in the first stage but are sufficient to characterize unambiguously the position of T^B, Figure 8b. The torsion angles fall into two conformation ranges. For the first conformation, Table 3b, α is g^+ and γ is t while in the second conformation α and γ revert to their typical B DNA values of g^- and g^+, Table 3c. This second conformation, for which few structures were found, is less favoured energetically and was discarded in the subsequent model builing. For the first conformation, T^B is found to fill in what T^A had left unoccupied of the minor groove.

The third and last step in the construction of the loop consisited of generating with RTC the conformations of T^C that closed the loop in the oligomer d(GTTTC), built from the best structures obtained in the two previous steps. This was achieved by changing the angles in the right hand branch of scheme 1 ($\epsilon 3$ through to $\gamma 4$). As above $\chi 3$, $\chi 4$ and $\delta 3$, $\delta 4$ were set to values corresponding to *anti* and C2′ endo conformations as indicated by the NMR data. No constraints were imposed other than that the G and C bases were paired in Watson–Crick geometry. The absence of distance constraints involving T^C makes it impossible to find a single conformation for this residue. We observe normal intraresidue NOEs within T^C, Figures 4–6, so the absence of interresidue NOEs must be due to an orientation of T^C outwards relative to its neighbouring residues. In this situation it is not likely to adopt a unique conformation and we cannot say anything about the backbone torsion angles around T^C.

In order to construct the entire hairpin we retained the structures with the lowest mechanical energy, Table 3 a) and b). The stem structure was created from a B DNA helical duplex (35) as the observed interactions show that it must be approximately a B form. After energy minimization of the stem the two structures were joined. This was done by searching for the best overlap of the backbone chains of the homologous G.C pairs as shown in Scheme 2. When the loop structure is mounted on top of the stem the base plane of T^B is found to lie parallel to the sugar phosphate backbone of the second strand. This structure was refined in two steps: first, all atoms, except those of the backbone joining the base pairs C^8.G^{14} and G^9.C^{13}, were constrained by a harmonic potential around their starting values (as explained in Materials and Methods). In the second step the whole system was relaxed.

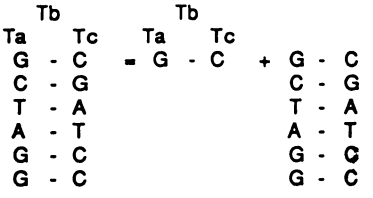
One representative structure of a close family which fit the experimental data is shown in Figure 9. Comparison with the NMR data is given in Table 4. The fit is good albeit less so in the loop region. This is likely to be the result of a less ordered structure in this region. Nevertheless, the agreement between the proposed model and the NMR data is such that, with the exception of the position of T^C, our model corresponds to the solution

Table 4. Interproton distances (Å) determined by NOE measurements and molecular mechanics calculations. The first entry is from the NMR data and the second is the percentage difference between the model distance and the NMR data, (model−NMR)/model. The overall fit criterion for these 54 distances F = 1.3 Although the real NMR derived distances are shown below we consider that they are accurate to only ±15%;

H6/H8 base i	H2′ intra i		H2″ inter i−1		H2″ intra i		CH_3 inter i+1	
G^4	2.3	−1						
G^5	2.5	−7	2.4	−2				
A^6	2.2	6	2.4	−6	4.0	−6	3.3	0
T^7	2.1	4	2.1	9	3.6	−1		
C^8	2.2	2	2.1	4				
G^9	2.5	−7	2.8	−12	3.3	9		
T^A	2.4	−2	2.8	−15			3.8	20
T^B	2.2	2	2.2	10	3.3	7		
T^C	2.2	2						
C^{13}	2.1	7			3.0	19		
G^{14}	2.3	−1	2.5	−11	3.6	1		
A^{15}	2.2	8	2.3	1	4.2	−9	3.3	5
T^{16}	2.0	6	1.9	13	3.8	−8		
C^{17}	2.0	6	2.2	4				
C^{18}	2.1	6	2.1	3	3.3	10		

H1′ base i	H2″ intra i	
G^4	2.5	−4
G^5	2.4	1
A^6	2.4	0
T^7	2.4	1
C^8	2.4	0
G^9	2.4	1
T^A	2.4	−2
T^B	2.3	6
T^C	2.3	3
C^{13}	2.5	−6
G^{14}	2.4	1
A^{15}	2.4	2
T^{16}	2.4	−1
$C1^7$	2.2	11
C^{18}	2.2	7

T^AH2′ – G^9H1′ 3.4 8, T^ACH_3–G^9H4′ 4.2 −13, T^AH2″ – T^BCH_3 4.0 −6



Scheme 2. The hairpin molecule was created by superposition of the loop structure (centre) on the six base pair stem (right) ensuring a maximum overlap of the backbone chains of the homologous G.C base pairs

Table 5. Torsion angles for the final structure.

	Pucker	Amp.	α	β	γ	δ	ε	ζ	χ
G^4	144	36			60	131	171	−99	−112
G^5	142	37	−63	−177	57	129	173	−101	−113
A^6	143	35	−61	−177	55	129	175	−97	−112
T^7	136	41	−60	176	57	127	178	−105	−115
C^8	141	35	−64	178	57	128	179	−99	−122
G^9	143	39	−63	−176	56	135	−157	−135	−105
T^A	164	34	62	179	173	146	−129	−70	−125
T^B	147	43	73	172	−179	138	−176	−87	−119
T^C	162	37	−65	−126	179	146	−159	76	−113
C^{13}	173	35	96	−155	179	148	−178	−102	−116
G^{14}	144	37	−69	−178	58	132	177	−105	−109
A^{15}	145	34	−63	179	59	127	171	−93	−113
T^{16}	138	40	−60	−179	55	129	178	−104	−114
C^{17}	144	38	−63	177	57	132	179	−106	−116
C^{18}	146	37	−63	175	58	137			−114

structure of the hairpin in so far as it is defined by the measured interproton distances. Few are available for the loop region which has to be fairly extended. The description in terms of torsion angles is given in Table 5. Angles in the stem are close to their classical B DNA values. For T^A and T^B in the loop only α and γ show a significant departure from B DNA values, they adopt a g^+ and t conformation respectively as suggested by the NMR data. This could be expected as the backbone tends to adopt an elongated structure in order to bridge the bases at the extremities of the stem.

DISCUSSION

Qualitatively we can say that the loop structure differs from all other loop structures proposed from solution studies. It has been proposed (36), and many studies support the model, that for a B DNA stem stacking will occur over the 3′ end of the stem. We do not observe the NOE interactions indicating such stacking but already part of a turn. Similarly T^A and T^B do not show normal stacking interactions with one another and this is followed again by a turn through T^C to C^{13} although this region is poorly defined. We are not able to identify the imino proton of the G^9-C^{13} base pair due to very poor spectral resolution but we observe that the intensity for the G.C imino protons is double that for the A.T imino protons (not shown). It seems highly likely that G^9 and C^{13} are base paired. Further the interactions observed between C^8 and G^9 on the one hand and C^{13} and G^{14} on the other show that the terminal base pair lies normally in the helix. This is in contrast to a previous study on a T3 loop (36) for which the terminal base pair was broken to form a loop of five residues. The reason for this difference is presumably that in our system the stem is closed by a G.C base pair whereas in the previous study the stem was closed by an A.T base pair.

Early studies on partly selfcomplementary DNA fragments containing a central Tn (n = 0−7) sequence (37−38) showed that maximum stability is obtained for a loop of four or five nucleotides. More recently it has been shown that in the same sequence with a loop of four thymidines that the terminal thymidines of the loop form a T-T wobble pair (39−40). Further, when base substitution is made for the terminal residues to give a PyrTTPur sequence base pairing is observed for the terminal residues but not observed when the sequence is PurTTPyr (41). This is in agreement with previous studies on an octanucleotide in which three base pairs are formed with a GT loop (42−43). It has also been proposed that a stem comprising only G.C base pairs can be looped by an AT sequence. This trend is also shown by studies on fully selfcomplementary oligonucletides where the base sequence nevertheless could allow the formation of a hairpin structure. When the central sequence is AATT no hairpin structure was observed (44) whereas for TATA a hairpin structure was observed (45). Recently the study of an octanucleotide hairpin structure in which three base pairs form and are looped by a TA sequence has been reported (46,47). Certain similarities exist between this structure and that which we observe here. For the octanucleotide the 5′T of the loop does not show stacking on the 3′ end of the stem but swings into the minor groove. The remaining part of the loop structure is different as would be expected between a loop of two and three residues. For the TA loop four of the five backbone torsion angles, corresponding to ε1−ζ1 and ε2−ζ2 of this present study (between the 3′ base of the stem and the second residue of the loop), are found to be t, presumambly because the backbone has to be very stretched to form a loop of only two residues. The structure that we find is less constrained as the loop is of three residues. As shown (47) the angle γ can be determined from measurement of the H4′/H5′,H5″ coupling constants. In this study these measurements are not possible due to spectral crowding. As stated above the γ^r conformation for T^A and T^B are indicated by the observed NOEs. On the other hand, those for T^C and C^{13} arise from model building in order to close the loop and for which there is no experimental evidence. Our structure is intermediate between these proposed loops of two residues and the study of a four residue loop by NMR and distance geometry calculations (48) and a crystal structure determination of a T4 loop on a Z DNA stem (49). Inclusion of the GATC sequence in the stem, at least in the unmethylated state, does not appear to induce any unusual conformation for the stem.

ACKNOWLEDGEMENTS

M.L.B. is a recipient of grants from l'Association pour la Recherche sur le Cancer and from la Ligue Nationale contre le Cancer and from the Université Pierre et Marie Curie. J.A.H.C., J.G-A and M.L.B. thank Prof. P.A.Kollman for providing us with AMBER 3.0.

REFERENCES

1. Vinograd, J., Leibowitz, J., Radloff, R., Watson, R. and Laipis, P. (1965) Proc. Nat. Acad. Sci. U.S.A. **53**, 1104−1111.
2. Gierer, A. (1966) Nature **212**, 1480−1481.
3. Bauer, W.R. and Vinograd, J. (1974) In Basic Principles in Nucleic Acid Chemistry (Ts'o, P. O. P. , ed.), vol. 2, p. 265 Academic press, New York.
4. Benham, C.J. (1982) , Biopolymers, **21**, 679−696.
5. Mizuuchi, K., Mizuuchi, M. and Gellert, M. (1982) J. Mol. Biol. **156**, 229−243.
6. Muller, U.R. and Fitch, W.M. (1982) Nature **298**, 582−585.
7. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. **13**, 319−351.
8. Wells, R.D., Goodman, T.C., Hillen, W., Horn, G.T., Klein, R.D., Larson, J.E., Muller, U.R., Neuendorf, S.K., Panayotatos, N. and Stirdivant, S.M. (1980) Prog. Nucleic Acid Res. Mol. Biol. **24**, 167−267.
9. Lilley, D.M.J. (1980) Proc. Natl. Acad. Sci. U.S.A. **77**, 6468−6472.
10. Lilley, D.M.J. (1981) Nucleic Acids Res. **9**, 1271−1289.
11. Wells, R.D. and Panayotatos, N. (1981) Nature **289**, 466−470.
12. Sheflin, L.G. and Kowalski, D. (1985) Nucleic Acids Res. **13**, 6137−6153.
13. Modrich, P. (1987) Ann. Rev. Biochem. **56**, 435−466.
14. Radman, M. and Wagner, R. (1986) Ann. Rev. Genet. **20**, 523−538.
15. Marinus, M. (1987) Ann. Rev. Genet. **21**, 113−131.

16. Fazakerley, G.V., Guy, A., Téoule, R. and Guschlbauer, W. (1984) FEBS Lett. **176**, 611–615.
17. Fazakerley, G.V., Guy, A., Téoule, R., Fritzsche, H. and Guschlbauer, W. (1985) Biochemistry **24**, 4540–4548.
18. Fazakerley, G.V., Quignard, E., Téoule, R., Guy, A. and Guschlbauer, W. (1987) Eur. J. Biochem. **167**, 397–404.
19. Quignard, E., Fazakerley, G.V., Téoule, R., Guy, A. and Guschlbauer, W; (1985) Eur. J. Biochem. **152**, 99–105.
20. Quignard, E., Téoule, R., Guy, A. and Fazakerley, G.V. (1985) Nucleic Acids Res. **13**, 7829–7836.
21. Bodenhausen, G., Kogler, H. and Ernst, E.E. (1984) J. Magn. Reson. **58**, 370–388.
22. Cuniasse, Ph., Sowers, L.C., Eritja, R., Kaplan, B., Goodman, M. F., Cognet, J.A.H., Le Bret, M., Guschlbauer, W. and Fazakerley, G.V. (1987) Nucleic Acids Res. **15**, 8003–8022.
23. Withka, J.M., Swaminathan, S. and Bolton, P.H. (1990) J. Magn. Reson. **89**, 386–390.
24. Plateau, P. and Guéron, M. (1982) J. Am. Chem. Soc. **104,** 7310–7311.
25. Davis, G.G. and Bax, A. (1985) J. Am. Chem. Soc. **107**, 2820–2821.
26. Weiner, P.K. and Kollman, P.A. (1981) AMBER: J. Comp. Chem. **2**, 287–303.
27. Weiner, S.J., Kollman, P.A., Nguyen, D.A. and Case, D.A. (1986) J. Comp. Chem. **7**, 230–252.
28. Singh, U.C., Weiner, P., Caldwell, J.W. and Kollman, P.A. (1986) AMBER 3.0, University of California, San Fransisco.
29. Gelin, B. and Karplus, M. (1981) Proc. Natl. Acad. Sci. USA **77**, 801–805.
30. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. Jr. and Weiner, P. (1984) J. Amer. Chem. Soc. **106**, 765–784.
31. Cuniasse, Ph., Sowers, L.C., Eritja, R., Kaplan, B., Goodman, M. F., Cognet, J.A.H., Le Bret, M., Guschlbauer, W. and Fazakerley, G.V. (1989) Biochemistry **28**, 2018–2026.
32. Rinkel, L J. and Altona, C. (1987) J. Biomol. Struct. Dynam. **4**, 621–649.
33. Fratini, A.V., Kopka, M.L., Drew, H.R. and Dickerson, R. E. (1982) J. Biol. Chem. **257**, 14686–14707.
34. Cuniasse, Ph., Fazakerley, G.V., Guschlbauer, W., Kaplan, B.E. and Sowers, L.C. (1990) J. Mol. Biol. **213**, 303–314.
35. Arnott, S., Campbell-Smith, P. and Chanrasekaran, R. (1976) CRC Handbook of Biochem. **2**, 411–441.
36. Haasnoot, C.A.G., Westerink, H.P., van der Marel, G.A., van Boom, J.H., Singh, U.C., Pattabiraman, N. and Kollman, P.A. (1986) J. Biomol. Struct. Dyn. **3**, 843–857.
37. Hilbers, C.W., Haasnoot, C.A.G., de Bruin, S.H., Joordens, J.J. M., van der Marel, G.A. and van Boom, J.H. (1985) Biochimie **67**, 685–695.
38. Haasnoot, C.A.G., Westerink, H.P., van der Marel, G.A. and van Boom, J.H. (1983) J. Biomol. Struct. Dyn. **1**, 115–129.
39. Blommers, M.J.J., Haasnoot, C.A.G., Hilbers, C.W., van Boom, J. H. and van der Marel, G.A. (1987) NATO ASI Ser. E **133**, 78–91.
40. Hilbers, C.W., Blommers, M.J.J., Haasnoot, C.A.G., van der Marel, G.A. and van Boom, J.H. (1987) Fresenius' Z. Anal. Chem, **327**, 70.
41. Blommers, M.J.J., Walters, A.L.I., Haasnoot, C.A.G., Aelen, J. M.A., van der Marel, G.A., van Boom, J.H. and Hilbers, C.W. (1989) Biochemistry **28**, 7491–7498.
42. Orbons, L.P.M., van der Marel, G.A., van Boom, J.H. and Altona, C. (1986) Nucleic Acids Res. **14**, 4187–4196.
43. Orbons, L.P.M., van Beuzekom, A.A. and Altona, C. (1987) J. Biomol. Struct. Dyn. **4**, 965–987.
44. Marky, L.A., Blumenfield, K.S., Kozlowski, S. and Breslauer, K. J. (1983) Biopolymers **22**, 1247.
45. Wemmer, D.E., Chou, S.H., Hare, D.R. and Reid, B.R. (1985) Nucleic Acids Res. **13**, 3755–3772.
46. Pieters, J.M.L., de Vroom, E., van der Marel, G.A., van Boom, J.H. and Altona, C. (1989) Eur. J. Biochem. **184**, 415–425.
47. Pieters, J.M.L., de Vroom, E., van der Marel, G.A., van Boom, J.H., Koning, T.M.G., Kaptein, R. and Altona, C. (1990) Biochemistry **29**, 788–799.
48. Hare, D.R. and Reid, B.R. (1986) Biochemistry **25**, 5341–5350.
49. Chattopadhyaya, R., Grzeskowiak, K. and Dickerson, R. (1990) J. Mol. Biol. **211**, 189–210.