

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/28519>

Please be advised that this information was generated on 2018-10-22 and may be subject to change.

The Transition from a Neutral-pH Double Helix to a Low-pH Triple Helix Induces a Conformational Switch in the CCGG Tetraloop Closing a Watson-Crick Stem

Maria J. P. van Dongen¹, Sybren S. Wijmenga¹
Gijs A. van der Marel², Jacques H. van Boom² and
Cornelis W. Hilbers^{1*}

¹NSR Centre for Molecular Structure, Design, and Synthesis, Laboratory of Biophysical Chemistry University of Nijmegen Toernooiveld, 6525 ED Nijmegen, The Netherlands

²Gorlaeus Laboratories Department of Organic Chemistry, State University of Leiden, P.O. Box 9502 2300 RA Leiden The Netherlands

The CCGG-loop in a DNA fragment, which is capable of forming an intramolecular triple helix as well as a hairpin structure, was investigated by NMR and molecular modeling studies. The structure of this loop is found as a type II conformation, one of the three commonly observed folding patterns of tetraloops, irrespective of the geometry of the underlying helix. In each situation, the loop exhibits a base-pair between the first cytosine and the guanine residue of the loop. The geometry of this base-pair, however, depends upon the circumstances. At neutral pH, in the hairpin form of the molecule, a Watson-Crick C-G base-pair is formed, whereas at low pH, when the strand exists as an intramolecular triple helix, a Hoogsteen C⁺-G base-pair is present.

We used molecular modeling to lay the foundations for understanding the observed conformational switch. A lower amount of strain, related to the short C1'-C1' of the base-pair, and protonation effects of the structure comprising the Hoogsteen base-pair turn out to outweigh the effects of a more stable base-pair, improved stacking and more favorable interactions in the minor groove of the structure comprising the Watson-Crick C-G base-pair. The models also provide an explanation for the general preference of loops meeting the consensus sequence -d(CYNG)- to fold into a type II conformation, i.e. with the base of second loop residue turned into the minor groove.

© 1996 Academic Press Limited

Keywords: conformational switch; Hoogsteen base-pair; Watson-Crick base-pair; pyrimidine-purine-pyrimidine triple helix; nuclear magnetic resonance spectroscopy

*Corresponding author

Introduction

Hairpins are abundant structural elements in nucleic acids. They form an integral part of

Present address: S. S. Wijmenga, Department of Medical Biochemistry and Biophysics, University of Umeå, S-90187 Umeå, Sweden.

Abbreviations used: NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; 1D, one-dimensional; 2D, two-dimensional; P.E.COSY, primitive exclusive correlated spectroscopy; TOCSY, total correlation spectroscopy; RMSD, root mean square deviation; TPPI, time proportional phase incrementation; MLEV, Malcolm Levitt; E-BURP, excitation band-selective uniform response pure-phase.

virtually all larger RNA structures and, in addition, are thought to be transiently present in DNA, playing a crucial role in processes such as gene regulation, recombination and mutagenesis. For instance, it has been postulated that DNA hairpins are formed by the triplet repeats encountered in human genetic diseases such as the fragile X syndrome and myotonic dystrophy (Chen *et al.*, 1995; Santhana Mariappan *et al.*, 1996a,b). Apart from an *in vivo* essay performed by Davison & Leach (1994), which showed that d(CNNG) and d(TNNA) sequences (N = any nucleotide) favor tight loop formation, current knowledge and understanding of the effects of loop size and nucleotide sequence on hairpin stability is based on the results obtained from a vast number of *in vitro*

studies (e.g. Haasnoot *et al.*, 1983; Blommers *et al.*, 1989; Antao *et al.*, 1991; Hilbers *et al.*, 1991; Antao & Tinoco, 1992).

A detailed comparison of the accumulated thermodynamic and structural data, presented by Hilbers *et al.* (1994), shows that the loops of stable and unusually stable DNA and RNA hairpins consist of two to four residues (for a definition see Hilbers *et al.*, 1994), and fold into a limited number of well-defined, compact conformations. Less stable loops, on the other hand, show considerable internal mobility. For example, no more than three distinct folding patterns describe all conformations established thus far for stable and unusually stable tetraloops. In each of these loops, the first and the fourth residue form a (non-canonical) base-pair. In cases where a Watson-Crick base-pair is formed these loops have sometimes been designated two-membered loops (*vide supra*), but here we consider these as a special class of tetraloops, being part of the same three folding patterns: Type I-loops (Figure 1a), exclusively found in DNA hairpins, adopt a conformation in which the first three bases at the 5'-end of the loop form a continuous stack on the 3'-end of the stem. Type II-loops (Figure 1b), found in both DNA and RNA hairpins, adopt a conformation in which the base of the second loop residue from the 5'-end of the loop is turned into or towards the minor groove and the base of the third residue is lying over the base-pair formed by residues one and four. Finally, type III-loops (Figure 1c), solely found in RNA hairpins, adopt a conformation in which the last three bases at the 3'-end of the loop stack onto the 5'-end of the stem. For three-membered loops similar patterns can be distinguished, as has, e.g. very recently been illustrated by Kuklenyik *et al.* (1996).

The classification given above can partly be understood in terms of the simple geometric loop-folding principle described by Haasnoot *et al.* (1986), which implies that the structure of a loop is dictated by the stacking pattern of the bases in the double helical stem of the hairpin. It predicts that a loop closing a B-type helix shows continuous base stacking at the 5'-side and a sharp turn in the backbone at the 3'-side. By contrast, the bases in a loop closing an A-type helix, are predicted to exhibit a continuous stack at the 3'-side of the loop and the turn is predicted to be located at the 5'-side. In this respect, the observation of, on one hand, DNA type I-loops and RNA type III-loops on the other, agrees with this principle. However, the loop-folding-principle does not account for the manner in which type II-loops are folded. This motif seems to depend less on the geometry of the underlying helix, as follows from the fact that this folding pattern is exhibited by both DNA and RNA hairpin loops, but more on the nature of the second residue of the tetraloop. Altona and co-workers (Ippel, 1993) found for DNA miniloops with the consensus sequence G(CNNG)C that if this residue is a pyrimidine, there is a preference for folding this base into the minor groove (type II), whereas if this

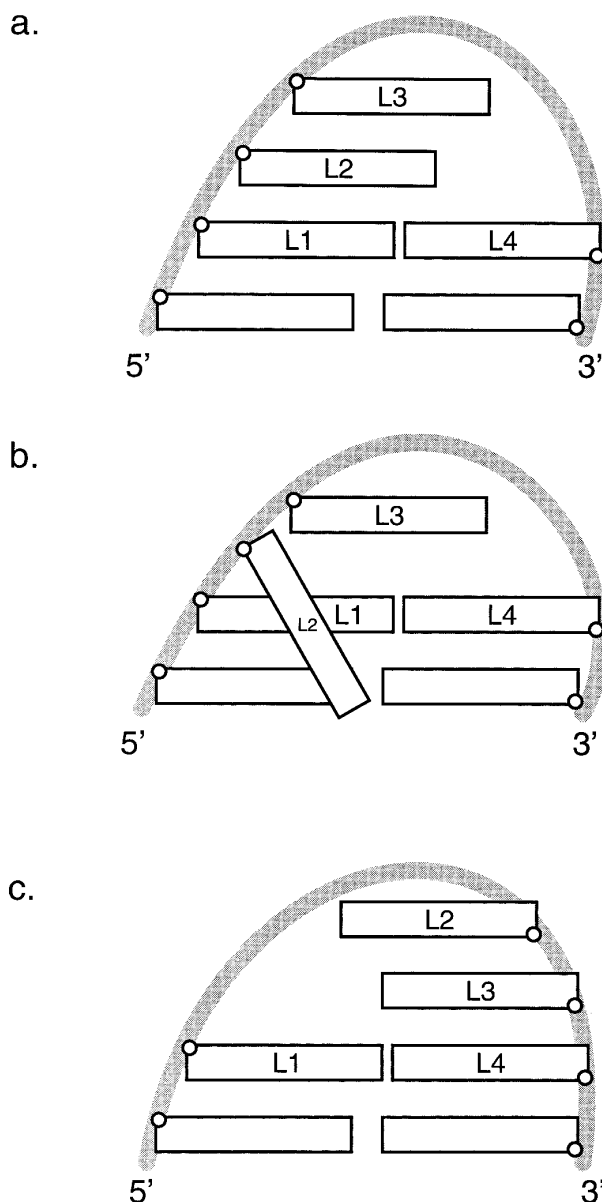


Figure 1. Schematic representation of the three distinct classes of folding for tetraloops: a, type I; b, type II; c, type III.

residue is a purine, a structure with continuous stacking (type I) is favored.

To investigate further the influence of the underlying helix geometries on loop-folding, we performed NMR studies on a 30-base DNA oligonucleotide which can fold into two different conformations, both comprising a CCCG-loop. At neutral pH, the fragment is present as a hairpin structure with a double helical stem, while at low pH an intramolecular triple helix is formed. Much to our surprise, we observed a conformational switch of the CCCG-loop upon the transformation of the hairpin to the triple helix structure, i.e. the nature of the base-pair formed between the first and last residue in the loop was found to change from

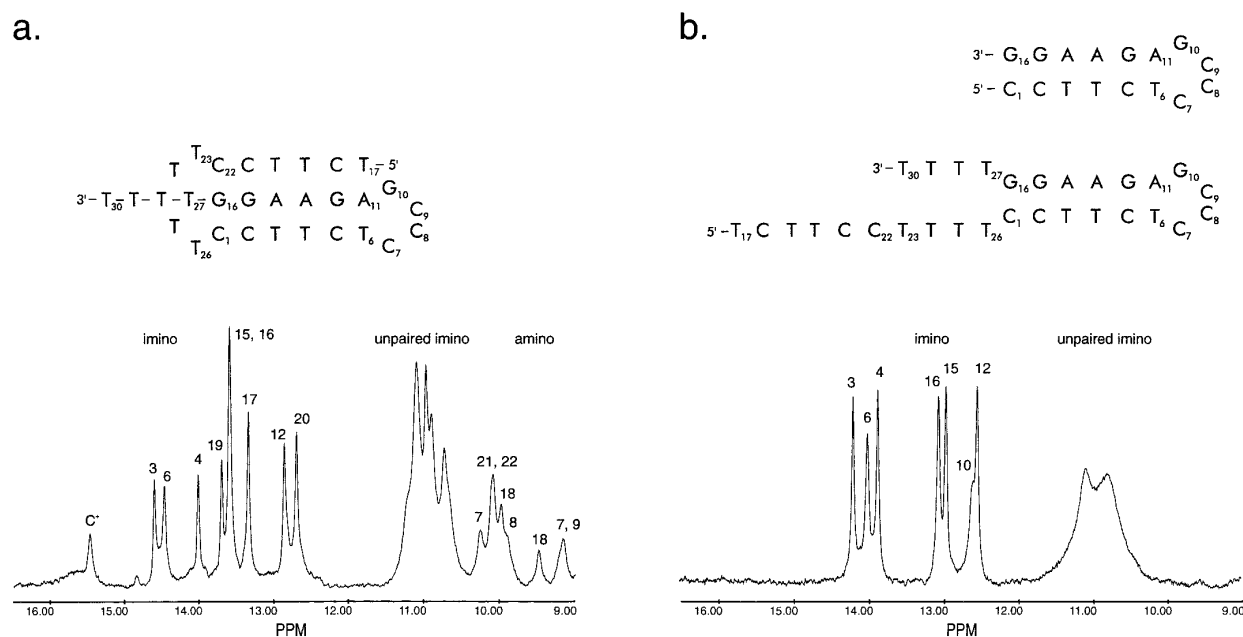


Figure 2. a, Nucleotide sequence and proposed secondary structure of the 30-base DNA strand (top), and exchangeable proton spectrum (bottom) at pH 4.5. b, Nucleotide sequences and proposed secondary structures of the 16-base and 30-base DNA strands (top), and exchangeable proton spectrum (bottom) at pH 7.0.

a Watson-Crick C-G pair at neutral pH to a Hoogsteen C⁺-G pair at low pH, while the overall folding pattern of the loop remained the same. This paper describes the three-dimensional structures of the CCCG-loop under both conditions. In addition, it aims to provide a basis for understanding the conformational switch of the base-pair, and the general preference of tetraloops to fold into type II conformations.

Results

CCCG-loop structure at low pH

Characterization of an intramolecular triple helix

The one-dimensional exchangeable-proton spectrum of the 30-mer at low pH (pH 4.5), recorded for the H₂O-sample, is shown in Figure 2a. Noteworthy are the resonances down-field from 15.0 ppm and between 9.0 and 10.5 ppm, which are typical of the imino and the amino protons, respectively, of protonated cytosine residues. Thorough inspection of the NOE contacts observed in NOE difference and 2D-NOESY spectra unmistakably revealed that the fragment adopts a triple helical structure, as shown in the upper part of Figure 2a, under these conditions. For example, in addition to NOEs demonstrating six Watson-Crick interactions, distinct cross-peaks were observed between thymine imino proton and adenine H8 resonances, as well as between the resonances belonging to the amino groups of protonated and unprotonated cytosine residues, respectively (see e.g. Mooren *et al.*, 1991b; Radhakrishnan *et al.*, 1991; Macaya *et al.*, 1992).

NMR spectra were also recorded for a H₂O-sample of the 16-mer (shown in Figure 2b) at pH 4.5. This fragment, which, at neutral pH, produces the same spectra as the 30-mer (*vide infra*), was designed such that it cannot form a triple helix. Instead, the spectra obtained at pH 4.5 gave evidence of the existence of multiple conformations. Partial protonation of the cytosine residues in the stem, which is prevented in the 30-mer by triple helix formation, might be the cause of this phenomenon.

Resonance assignments

Assignments for the exchangeable proton resonances in the helical parts of the 30-mer directly followed from the analysis of the NOE spectra mentioned above. Subsequently, the non-exchangeable base and sugar ring protons of the molecule were assigned by analysis of spectra recorded in ²H₂O. Residue specific and sequential resonance assignments were carried out by standard procedures, using NOESY, P.E.COSY and TOCSY spectra (Wijmenga *et al.*, 1993), complemented by data obtained from a ¹H-³¹P hetero-correlation spectrum (Pardi *et al.*, 1983) and a selective version of a 3D-TOCSY-NOESY spectrum (Mooren *et al.*, 1991a; Wijmenga *et al.*, 1994). A comparison of *J*-coupling values and relative NOE intensities (Blommers *et al.*, 1991) resulted in the stereospecific assignments for the H5' and H5'' resonances of C8, C9, and G10. The results, as far as the CCCG-loop region is concerned, are summarized in Table 1A. We mention that remarkable down-field shifts were observed for the H5 and H6 resonances of C8 and the H2' resonance of G10. The H2' resonance of C8

Table 1. Proton chemical shifts

Residue	NH ^a	NH ₂ ^a	H6	H5	H1'	H2'	H2''	H3'	H4'	H5'	H5''
			H8	CH ₃ H2							
A. CCGG-loop in the triple helix at 25°C											
T6	14.44	—	7.65	1.61	6.06	2.26	2.53	4.82	4.27		
C7		10.13	9.31	7.70	5.64	5.80	2.10	2.35	4.86	4.48	4.09 ^b
C8		9.83	8.34	8.30	6.39	6.14	1.70	2.21	4.61	4.32	4.26
C9		9.23	8.25	7.54	5.74	5.27	1.65	2.06	4.32	3.14	3.33
G10				6.89	—	5.84	3.63	2.57	4.69	3.75	4.25
A11	—	8.19	7.94	8.10	7.61	6.28	2.95	3.02	5.06	4.43	3.99 ^b
B. CCGG-loop in the hairpin at 25°C											
T6	14.04	—	—	7.39	1.73	6.12	2.07	2.43	4.88		
C7	—			7.69	5.83	6.09		2.47 ^b		4.28	
C8	—			7.84	6.05	6.17	2.00	2.28	4.58	4.23	4.15 ^b
C9	—			7.46	5.56	5.56	1.67	2.14	4.50	3.45	3.58 ^b
G10	12.66			7.76	—	5.49	2.56	2.68	4.99	4.37	3.74 ^b
A11	—			8.12	7.77	5.93	2.64	2.86	5.06	4.43	4.16 ^b

^a Chemical shifts of exchangeable protons at 7°C.
^b Ambiguous stereo-specific assignment.

and all sugar proton resonances of C9, especially H4', were found at exceptional up-field positions.

The amino proton resonances of the cytosine residues in the CCGG-loop, could be assigned *via* their intra-residue NOEs to H5, observable in the NOESY spectrum recorded for the H₂O sample (Table 1A). It was found that the amino proton resonances of C7 are considerably down-field shifted (9.31 and 10.13 ppm) with respect to the positions of unprotonated cytosine residues, which resonate around 6.5 and 8.0 ppm. This observation implies that residue C7 is protonated. The resonances of the amino protons of the other cytosine residues in the loop, C8 and C9, also showed down-field shifts, although less pronounced (8.2 to 9.8 ppm), possibly indicating that these residues are (partially) protonated as well.

Derivation of distance restraints

For the helical parts of the molecule standard 5'-3' sequential connectivities were observed in the NOESY spectrum. These sequential cross-peak patterns, involving the H1'/H2'/H2'' and H6/H8 resonances, were observed to continue partly into the loop region, i.e. regular contacts were found between residues T6 and C7, and between G10 and A11. Furthermore, H1'(C7)-H6(C8) and H2'(C8)-H6(C9) NOE cross-peaks were observed.

In addition to the standard sequential sugar-base connectivities, the NOESY spectra showed a large number of unusual contacts between the various loop residues. Semi-sequential connectivities were observed between C7 and C9, involving H6-H5 and H2'/H2''-H5 inter-residue NOEs. Between the residues C7 and C8, and between the residues C8 and C9 several sugar-sugar contacts were established. No NOEs were detected between C9 and G10, except for very weak H4'/H5'/H5''-H8

cross-peaks. As far as the exchangeable protons are concerned, several cross-peaks were observed resulting from the amino protons of C7, as well as some weak contacts resulting from the amino protons of C8.

The NOE intensities, schematized in Figure 3, were translated into loose distance restraints (type A) for the generation of a set of starting structures (*vide infra*). Alternatively, non-overlapping inter-residue NOEs were converted into more precise distance restraints (type B). These quantitatively derived distance restraints replaced the corresponding loose restraints in the structure refinement calculations. The NOE intensities of the exchangeable protons in the NOESY spectrum of the H₂O sample were used in a very qualitative manner, because of the relatively long mixing time applied in this experiment and the unknown effect of solvent exchange.

Derivation of dihedral restraints

Intra-residue NOEs, which principally reflect the values of the glycosidic torsion angles, χ , and the sugar puckers, showed no abnormal patterns, except for those of G10. For this residue a very intense H8-H1' NOE was observed, while contacts between H8 and other sugar protons were absent. This situation is characteristic of a χ -angle residing in the *syn* domain. Additional support for this idea came from the chemical shift of the H2' proton of G10, which is considerably shifted to a down-field position (3.63 ppm). Quantitative interpretation of the NOEs by the method described by Wijmenga *et al.* (1993) revealed that $\chi(10)$ falls in the region between 60° and 120°. All other residues in the loop were found to adopt glycosidic torsion angles in the *anti*-range.

The sugar pucker parameters, i.e. the fraction S-pucker, f_s , as well as the predominant angle of

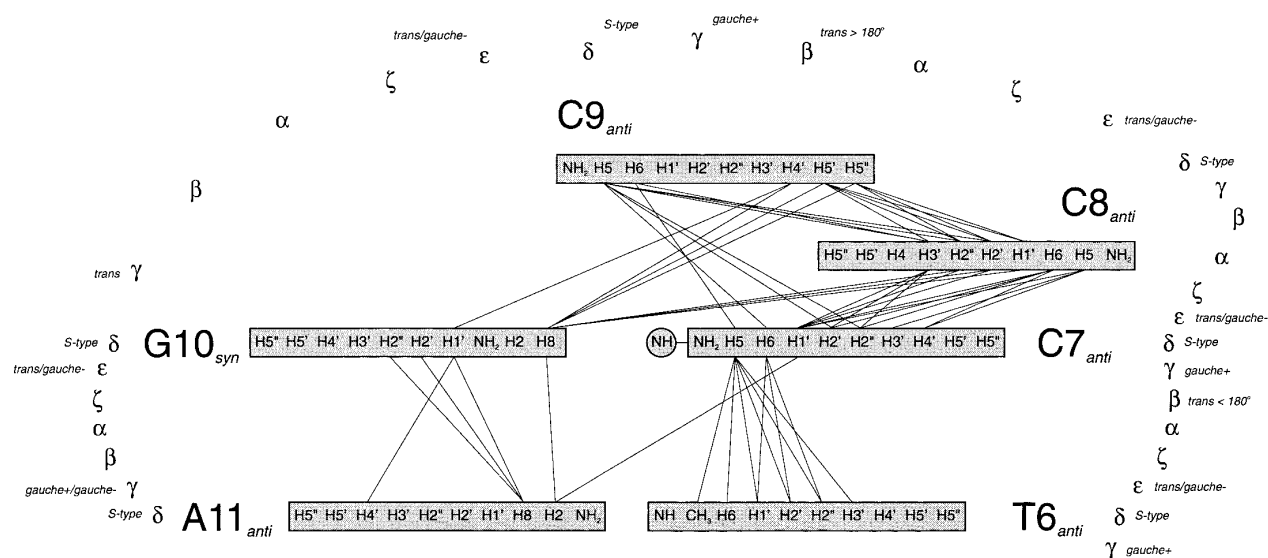


Figure 3. Schematic representation of the inter-residue distance and dihedral restraints for the structure calculations of the CCCG-loop at low pH.

pseudorotation, P , were derived from J -coupling constants observed in a P.E.COSY spectrum (Van Wijk *et al.*, 1992), supplemented by quantitative TOCSY data (Van Duynhoven *et al.*, 1992). It was shown that all residues have sugars with predominantly S-type puckers. Investigation of intra-residue NOE cross-peak intensities (Wijmenga *et al.*, 1994) resulted in the same conclusions.

Information on the dihedral angles of the backbone largely originated from the ^{31}P - ^1H heteronuclear correlation spectrum. Although large parts could not be assigned due to extensive overlap, three phosphorous resonances of the CCCG-loop region could be identified. The resonance of C7pC8 is situated at somewhat higher field, at -2.98 ppm, than the majority of resonances, whereas C8pC9 resonates at relatively low field, at -4.61 ppm. Both of them show a cross-peak to H4', which is caused by a J -coupling over four bonds. This established that the P-O5'-C5'-C4'-H4' bonds in these residues are all in one plane, i.e. they form a W-shaped configuration, which implies that $\beta(8)$ and $\beta(9)$ are in the *trans* and $\gamma(8)$ and $\gamma(9)$ are in the *gauche*⁺ region (Sarma *et al.*, 1973). It was furthermore deduced from the relative values of the P-H5' and P-H5'' J -coupling constants that $\beta(C8) < 180^\circ$ and $\beta(C9) > 180^\circ$. The third phosphorous resonance which could be assigned, C9pG10, resonates at -3.35 ppm. No cross-peak to H4' could be observed for this resonance. In accordance with this finding, the H4'-H5'/H4'-H5'' J -couplings and intra-residue NOE-intensities of G10 indicated that the γ -angle of this residue resides in the *trans*-region. For the γ -angle of A11, on the other hand, the *trans*-domain could be excluded on the basis of relative H3'-H5' and H3'-H5'' NOE-intensities.

Finally, it is known that for the ϵ -angles the *gauche*⁺ region is forbidden for steric reasons (Mooren, 1993). Figure 3 gives an overview of the dihedral restraints in the CCCG-loop structure.

Structure calculations of the CCCG-loop at low pH

The lists of loose distance restraints and permitted torsion angle domains, supplemented by hydrogen bonding restraints for the T6-A11 base-pair, served as input for the distance geometry calculations (Table 2). During subsequent cycles of DIANA and CORMA (as described in Materials and Methods), repulsive restraints between proton pairs for which no NOE could be detected were added. A final DIANA run, using the complete input data set, generated 69 structures with an average pairwise RMSD of 1.40 Å. The subsequent restrained molecular dynamics refinement procedure by X-PLOR (see Materials and Methods) mainly affected the local backbone conformation. Structures with unfavorable energies were discarded from further analysis. The procedure resulted in a set of 22 refined structures, shown in Figure 4a, with an average pairwise RMSD of 0.42 Å.

Analysis of the dihedral angles of the set of refined structures revealed that three distinct families of structures could be distinguished, represented by two, six, and 14 structures, respectively. For each of these sets an average structure was calculated, which subsequently was subjected to a short run of restrained energy minimization. It followed that the total energy of the set of most abundant conformations is most favorable. The alternative sets either possess a total

Table 2. Statistics of structure determination of the CCCG-loop in the triple helix

	DIANA	X-PLOR
Structure restraints:		
Number of distance restraints	118	118
Intra-residue	24	24
Inter-residue	60	60
Type A	53	11
Type B	0	42
Exchangeable protons	7	7
T-A base-pair	9	9
Lower limit	25	25
Number of dihedral restraints	42	54
Planarity restraints present (+/-)	-	+
Precision of the structures:		
Average RMSD relative to average structure for all heavy atoms (Å)		
69 structures	0.99	—
22 structures	—	0.30
14 structures	—	0.11
Average pairwise RMSD for all heavy atoms (Å)		
69 structures	1.40	—
22 structures	—	0.42
14 structures	—	0.14

energy more than 20 kcal mol⁻¹ higher, or turn out to be unstable in the energy-minimization procedure. Figure 4b shows the set of 14 structures, with an average pairwise RMSD of 0.14 Å, belonging to the most favorable conformation of the CCCG-loop at low pH. Different views of the average structure calculated from this set are shown in Figure 5a and b. From these Figures it can be seen that the first cytosine and the guanine residue are positioned in such a way that a Hoogsteen C⁺-G base-pair is formed (Figure 6a). The base of the second cytosine is turned into the minor groove, whereas the third cytosine stacks on the C⁺-G base-pair.

CCCG-structure at neutral pH

Structure characterization

Figure 2b shows the one-dimensional proton spectrum of the 30-base DNA strand, dissolved in H₂O, at neutral pH (pH 7.0). An identical spectrum was obtained for a shorter molecule (16-mer), which is indicative of the formation of similar hairpin structures by the two fragments. The seven

imino proton resonances observed in the spectrum were assigned by NOE difference and 2D-NOESY experiments. In these spectra a continuous chain of NOE connectivities between imino protons in adjacent base-pairs in the stem was established, running from G16 via G15, T3, T4, G12, and T6 all the way up to G10. The assignments of the non-exchangeable proton resonances (Table 1B) were obtained from a NOESY spectrum recorded for the ²H₂O-sample of the 16-mer.

The position and intensity of the imino proton signal of G10 in the spectrum is characteristic of the formation of an additional Watson-Crick type base-pair in the loop. The χ -angle of G10, like that of all other residues, is in the *anti*-domain, as could be concluded from the intensities of intra-residue H8-H1'/H2'/H2'' contacts in the NOESY spectra, and from the chemical shift of the H2' proton resonance of this residue. In addition, an NOE was found between the imino proton of G10 and the H5 proton of C7. These observations indicate the formation of a Watson-Crick C7-G10 base-pair (Figure 6b). Various NOEs between G10 and A11 and between T6 and C7 confirm that residue G10 as well as residue C7 stack onto the helix stem in a more or less regular fashion.

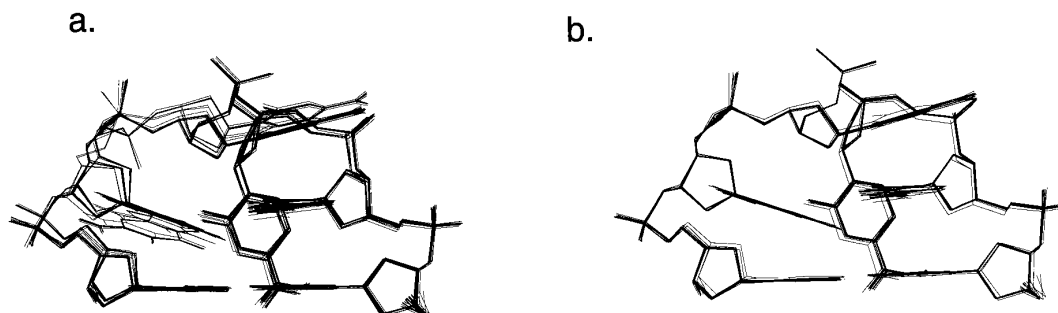


Figure 4. Superposition of calculated structures of the CCCG-loop at low pH. a, The set of 22 accepted refined structures after the calculations in X-PLOR. b, The 14 low-energy structures selected from the set shown in a.

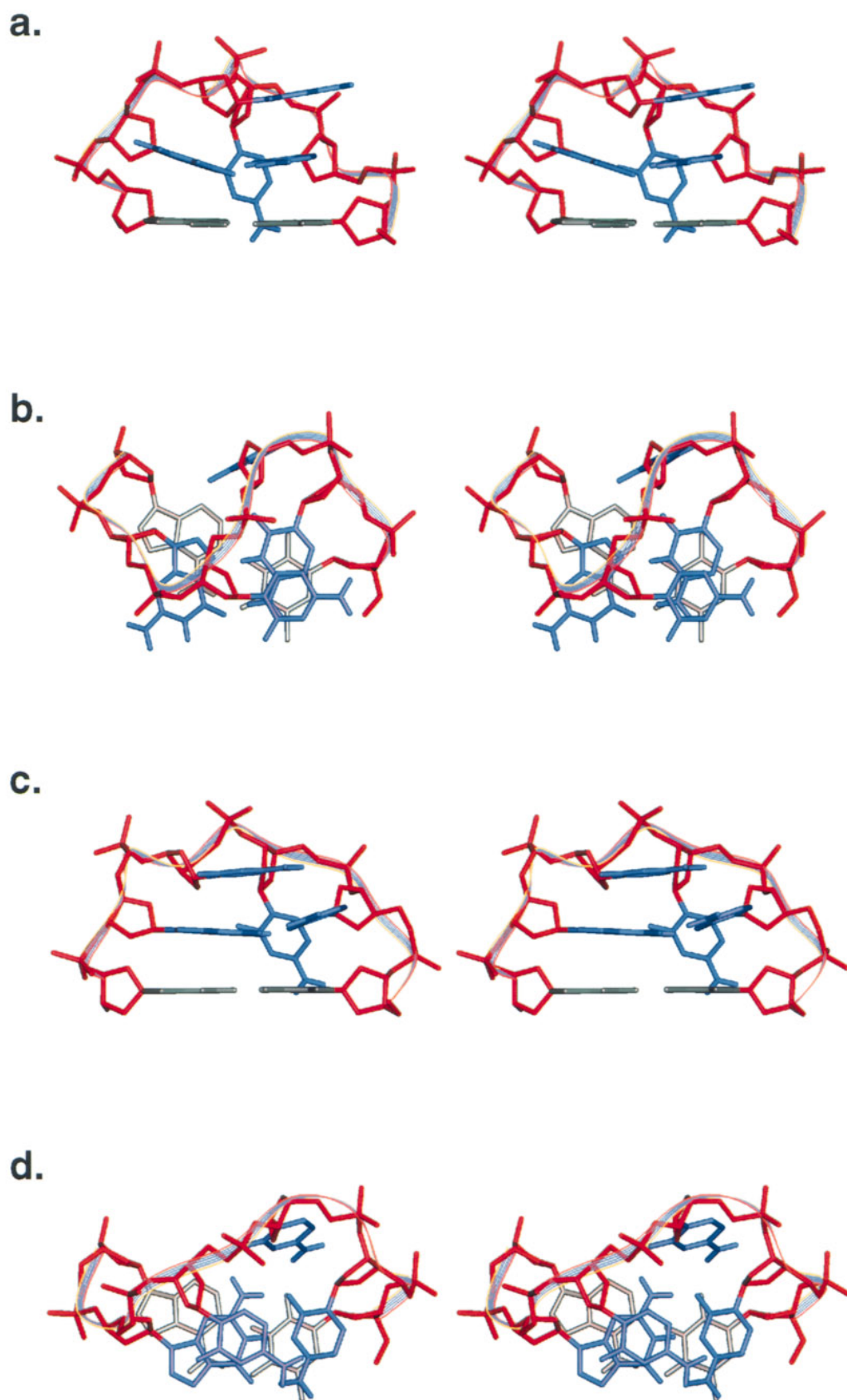


Figure 5. Stereo presentations of the minimized average structure of the CCCG-loop at low pH: a, side view from the major groove; b, top view. Structural model for the CCCG-loop at neutral pH, c, side view from the major groove; d, top view.

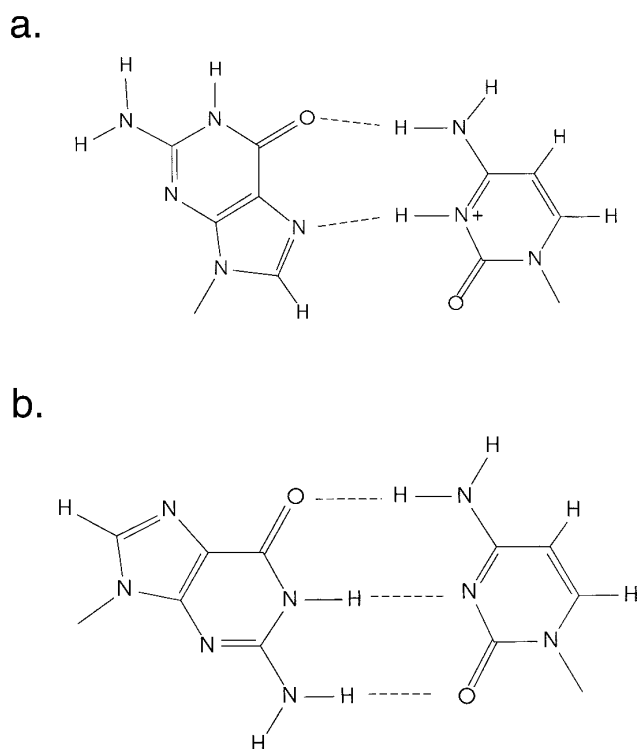


Figure 6. a, Hoogsteen C⁺-G, and b, Watson-Crick C-G base-pair.

The conformation of the loop at neutral pH was further characterized by NOEs typical of a tetraloop structure in which the second base is folded into the minor groove: H1'(C7)-H2'(C8), H4'(C7)-H5(C8), and H2''(C8)-H5'(C9) NOE contacts were observed (Ippel, 1993). In addition, the H5 and H6 proton resonance positions of residue C8 are down-field shifted to 6.05 and 7.84 ppm, respectively, characteristic of a base situated in the minor groove (Orbons *et al.*, 1987; Blommers *et al.*, 1991; Ippel, 1993). The base of C9 may be more or less stacked on the C7-G10 base-pair plane, as indicated by a weak H2''(C7)-H5(C9) NOE (Ippel, 1993). In general, the chemical shifts and NOE patterns of the CCCG-loop at neutral pH turned out to be very similar to those of the -d(CCTG)- and -d(CTCG)-miniloops described by Ippel (1993).

Structural modeling for CCCG-loop at neutral pH

The data thus showed that the CCCG tetraloop is similarly folded at neutral and low pH, with the second residue turned into the minor groove and a base-pair formed between residues one and four (*vide supra*). However, the nature of the base-pair in the loop is different for the two situations. At low pH, C7 and G10 constitute a Hoogsteen C⁺-G base-pair, with the glycosidic torsion angle of G10 in the *syn*-domain. At neutral pH these residues form a Watson-Crick C-G pair. The glycosidic torsion angle of G10 then resides in the *anti*-domain.

Table 3. Dihedral angles of the CCCG-loop

	T6	C7	C8	C9	G10	A11
A. Low pH structure (average structure after restrained energy minimization)						
α		291.4	203.4	292.4	91.9	302.6
β		190.3	167.9	211.3	199.2	178.4
γ	59.7	53.5	52.3	75.1	176.1	50.0
δ	141.0	139.3	138.8	137.5	142.5	140.7
ϵ	178.1	178.6	212.3	248.8	178.4	
ζ	267.2	272.7	280.1	67.4	263.4	
χ	240.8	252.9	214.5	205.2	72.6	269.1
B. Neutral pH structure						
α		262.4	204.3	299.1	274.8	311.0
β		53.3	150.5	205.3	57.4	136.3
γ	59.9	168.5	62.8	63.4	182.6	63.0
δ	136.2	137.4	136.7	137.3	135.6	137.6
ϵ	238.7	206.5	192.2	180.9	184.4	
ζ	299.7	269.2	278.4	276.2	210.3	
χ	242.6	235.4	202.4	199.5	281.1	244.0

We decided therefore not to perform an *ab initio* structure calculation for the CCCG-loop at neutral pH. Instead, the average structure with lowest overall energy for the CCCG-loop at pH 4.5, but with the guanine turned in the *anti*-position, served as a starting point for the generation of a structural model for the loop at pH 7.0. Alternatively, a starting-structure with dihedral angles as in the -d(CCTG)- and -d(CTCG)-miniloops (Ippel, 1993) was constructed. Subsequently, these structures were subjected to molecular dynamics calculations and energy minimization (see Materials and Methods for details). From the two resulting structures, which are both consistent with the obtained NOE and chemical shift data as described above, the one resulting from the miniloop definitions turned out to be energetically most favorable (Figures 5c and d).

Discussion

Structural features of the CCCG-loop at low pH

Figure 5a and b shows that the CCCG-loop at low pH exists as a well-defined, compact structure. The values of the dihedral angles defining this loop are given in Table 3A. It can be gleaned from this Table that C7 exhibits only minor deviations from the values commonly observed for a B-type helix. In other words, the geometry of the sugar-phosphate backbone is preserved and the stacking interactions of the stem are continued into the loop. Similarly, at the 3' end of the loop, G10 stacks on A11 in a regular fashion, except that its glycosidic torsion angle lies in the *syn*-domain. The guanine base forms a Hoogsteen base-pair, involving two hydrogen bonds, with the protonated C7 base (Figure 6a).

At this point it is emphasized that this base-pair configuration directly followed from the experimentally derived restraints. Neither electrostatic terms nor explicit distance restraints enforcing hydrogen bonds were introduced during the structure calculations. Nevertheless, the distances between the donor and acceptor atoms, H3(C7)-N7(G10) and H4(C7)-O6(G10), turned out to be 2.3 and 2.0 Å, respectively, not very different from the ideal distance of about 1.95 Å (Saenger, 1984). Subsequent energy minimization, using a force field including a term for the electrostatic interactions, further reduced these distances to about 1.6 Å each, without significant alteration of the structure as a whole.

The base of C8, the second residue of the loop, is folded into the minor groove, which brings about typical down-field shifts for the H5 and H6 resonances of this residue (Orbons *et al.*, 1987; Blommers *et al.*, 1991; Ippel, 1993). The position of C8 is determined by two angular adjustments: $\alpha(8)$ -*trans*; $\epsilon(8)$ -(-)*anticlinal*. Although these modifications drastically reduce the rise from C7 to C9, the general course of the phosphate backbone is not altered. Thus, they enable C9 to stack onto the C⁺-G base-pair, manifested by several semi-sequential NOE contacts between C7 and C9. The sugar-ring of C9 is positioned directly over the base plane of G10, resulting in large up-field shifts of the sugar proton resonances from C9, especially of H4' (3.14 ppm). Between the residues C9 and G10 a sharp turn occurs, after which the chain takes the course of the opposite strand. Reasonably well-defined torsion angles were found for this step, despite the fact that the NMR data provided only few sequential NOEs between residues C9 and G10. The low energy conformation shows $\epsilon(9)$, $\zeta(9)$, $\alpha(10)$ and $\gamma(10)$ altered into (-)*anticlinal*, *gauche*⁺, *gauche*⁺, and *trans*, respectively, while $\beta(10)$ resides in the *trans* domain as usual. This combination of torsion angle values, especially the γ -*trans* angle, is frequently observed in DNA loop structures. The ζ -*gauche*⁺, α -*gauche*⁺ combination classifies the turn as a so-called π 1-turn (Hilbers *et al.*, 1994).

The CCCG-loop structure at low pH shows striking similarities with the loops formed by -d(TTTA)- (Blommers *et al.*, 1991) and -r(UUCG)- (Varani *et al.*, 1991; Allain & Varani, 1995). In all these loops the base of the second residue is folded into or towards the minor groove (type II), and all these loops possess a non-canonical base-pair, with the fourth loop residue in the *syn*-position. In addition, the values observed for the dihedral angles in the backbone are largely identical in the three loops.

Structural features of the CCCG-loop at neutral pH

Figure 5c and d shows the structure derived for the CCCG-loop at neutral pH based on the available experimental evidence (see Results). The structure shows a Watson-Crick base-pair formed

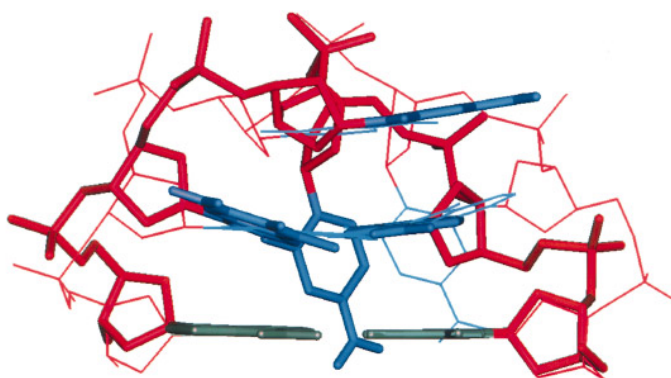


Figure 7. Best-fit superposition of the CCCG-loop structures at low (thick lines) and neutral pH (thin lines).

by residues C7 and G10 (Figure 6b). The base of C8 is turned into the minor groove, while residue C9 stacks on the Watson-Crick C-G base-pair. Table 3B gives the values of the dihedral angles of this loop structure. It shows that many of the angles in the backbone are not very different from those in the low pH structure. Still, the turn in the backbone, situated between C9 and G10 is effected in this case by $\beta(10)$ -*gauche*⁺, $\gamma(10)$ -*trans*. Additional deviations from the B-helix values occur for the dihedral angles in the T6-C7 and G10-A11 steps, i.e. $\beta(7)$ -*gauche*⁺, $\gamma(7)$ -*trans*, similar to the results on the -CTCG- and -CCTG- loops (Ippel, 1993), and $\zeta(10)$ -*trans*, $\beta(11)$ -(+)*anticlinal*.

The overall appearance of the neutral pH structure of the CCCG-loop is consistent with those of other DNA loops meeting the consensus sequence -CYNG- (Y = pyrimidine, N = any nucleotide). This class of loops has been extensively studied by Altona and collaborators, all at neutral pH (Orbons *et al.*, 1987; Ippel *et al.*, 1992; Ippel, 1993). Common features are the Watson-Crick C-G pair, and the position of the Y base in the minor groove (type II). Very recently, Jucker & Pardi (1995) described the structure of the loop formed by -r(CUUG)-. It was found that also this RNA tetraloop shows a type II folding pattern quite similar to the neutral pH structure of the DNA CCCG-loop.

Loop folding patterns

The CCCG tetraloop was studied under two different circumstances. At neutral pH the loop closes the B-type double helix stem of a hairpin. At low pH a triple helix is formed, in which the Watson-Crick double helix is supposed to adopt a conformation somewhere between A and B-type (Liquier *et al.*, 1991; Macaya *et al.*, 1992; Radhakrishnan & Patel, 1994). As is shown in Figure 7, the loop structures are remarkably similar in each situation. The neutral as well as the low pH form of the structure can be described as type II-loops, characterized by the location of the second loop residue in the minor groove (Figure 1). These

Table 4. Energy contributions of the CCCG-loop

	1. CCCG-loop (low pH) ^a	2. CCCG-loop (low pH) ^b	3. CCCG-loop (neutral pH) ^c
Total energy (kcal mol ⁻¹)	166.3	142.4	154.5
Bonds	17.9	15.0	15.0
Bond angles	168.6	154.5	155.8
Dihedral angles	108.8	106.6	110.2
Impropers	44.4	42.9	43.1
van der Waals	-71.2	-71.6	-70.1
Electrostatics	-102.2	-105.0	-99.6
Restrained dihedrals	8.8	—	—
Restrained distances	24.5	—	—
Planarity	0.4	—	—
Interaction energy (kcal mol ⁻¹)			
C7-G10 H-bonds	-8.6	-7.2	-12.8
T6/A11-C7/G10 stacking	-9.0	-11.8	-11.8
C7/G10-C9 stacking	-2.9	-5.1	-6.8
C8 folding	-13.2	-14.6	-18.4

^a Average structure of 14 low energy-structures of the CCCG loop at low pH (see Figure 6c), after restrained energy-minimization.
^b CCCG-loop structure as characterized in column 1, after 1000 additional steps of unrestrained energy minimization.
^c CCCG-loop structure at neutral pH, obtained from mini-loop definitions, after 20 ps free molecular dynamics and 1000 steps of unrestrained energy minimization.

results correspond with the hypothesis put forward for type-II folding, namely that this type of folding is independent of the conformation of the underlying helical stem, in contrast to the type I and III foldings, which are significantly influenced by the stem geometry.

Conformational switch

Changing from neutral to acidic conditions, the base-pair configuration in the CCCG-loop alters. Residues C7 and G10 form a Watson-Crick C-G base-pair at neutral pH, while in an acidic environment these bases form a Hoogsteen C⁺-G pair. This transformation is quite surprising at first sight, because it involves the disruption of three hydrogen bonds, in favor of only two new ones.

Together with the conformational switch of the base-pair in the CCCG-loop, the conformation of the stem changes from a B-type helix to a triple helix. Therefore, the transition of the C-G base-pair into a C⁺-G pair may have been induced by two distinct effects. In the first place, it is possible that lowering the pH directly causes the formation of a C⁺-G base-pair. The second possibility is that the conformational change of the underlying stem region dictates the base-pair configuration in the loop. NMR studies on the 16-base DNA fragment, corresponding with the hairpin forming part of the 30-mer, at neutral as well as at low pH were set up to clarify this issue. As the loop of this fragment only experiences the direct influence of pH, the absence or presence of the imino proton signal of G10 at low pH would support the first or second mechanism, respectively. Unfortunately, the molecule did not fold into a single conformation at low

pH, which prevented definitive conclusions to be drawn from these data.

In a further attempt to uncover the driving force of the switch between the Watson-Crick and Hoogsteen base-pair in the CCCG-loop, an analysis of the conformational energies was performed for the two situations. The contributions of the various components of the force field were compared, as well as some distinct interaction energies (Table 4). In accordance with the qualitative argumentation (*vide supra*), these calculations indicate that the C7-G10 interaction energy of a Watson Crick base-pair is more favorable than that of a Hoogsteen pair ($\Delta\Delta H = -5.6$ kcal mol⁻¹). Still, the low pH structure possesses a lower overall energy than the structure at neutral pH (12.1 kcal mol⁻¹), necessary to explain the experimental observation of the conformational switch.

Some well-defined non-bonded energy terms, i.e. interactions concerning the base situated in the minor groove and the stacking interactions, were subsequently analyzed. It was found that the position of residue C8 in the minor groove is more advantageous in the structure comprising the Watson-Crick C-G base-pair ($\Delta\Delta H = -3.8$ kcal mol⁻¹). Besides a hydrogen bond between the amino group of C8 and the carboxyl-2 group of T6, which can be formed in both situations, a Watson-Crick C-G base-pair enables C8 to form a second hydrogen bond, i.e. between its carboxyl group and the amino protons of G10. Also the stacking of C9 onto the C-G/C⁺-G base-pair is more favorable in case of the C-G pair (-1.7 kcal mol⁻¹). Interestingly, the value of the stacking energy felt by the third residue is in each case substantially smaller than that of the interaction energy experienced by C8, possibly

explaining the general preference of this type of folding by -CYNG-loops. The stacking of the C-G/C⁺-G base-pair onto the stem results in an interaction energy which is about equal for the C-G and C⁺-G base-pair.

The (over)compensation of the interactions favoring the Watson-Crick base-pair is accomplished by two distinct effects. The first one is the protonation of C7. The additional imino proton forms one of the two hydrogen bonds in the Hoogsteen C⁺-G base-pair and, at the same time, introduces a net positive charge on this residue. In our structural model these protonation effects contribute favorably to the total energy by 14.6 kcal mol⁻¹. However, care should be taken with the interpretation of the exact value of this number. It strongly depends on the effective charge on the residue as a result of potential shielding effects. No shielding of the charge at residue C7 results in a contribution of 30.8 kcal mol⁻¹, whereas complete shielding would reduce its effect to only 7.1 kcal mol⁻¹.

The second compensating effect arises from a heterogeneous term which we shall denominate as strain. It consists of the contributions of the geometric terms of the force field, e.g. bond-angle and dihedral angle energy ($\Delta\Delta H = 5.4$ kcal mol⁻¹), as well as of non-bonded interaction energies ($\Delta\Delta H = 3.1$ kcal mol⁻¹). Strain occurs when the structure has to deviate from the ideal helix geometry to accomplish the formation of a loop. It has been argued before (Blommers *et al.*, 1989) that non-canonical base-pairs with a shortened C1'-C1' distance (8.0 to 9.0 Å) contribute to loop closing in tetraloops, whereas Watson-Crick base-pairs, with a longer C1'-C1' distance (10.0 to 11.0 Å), do not. A direct comparison of the global implications of the two different base-pairs, C⁺-G and C-G., for the CCCG-loop structures is presented in Figure 7. The image appears that the more compact C⁺-G pair, confines the amount of strain experienced by the two remaining loop residues, resulting in a more stable structure.

Conclusions

It is found that the structure of the CCCG-loop of the 30-mer depends on pH. Under neutral conditions, when the molecule is present as a hairpin with a stem of Watson-Crick base-pairs, an additional Watson-Crick C-G base-pair is present between the first and the last residue in the loop. In the low pH form of the molecule, an intramolecular triple helix, this base-pair is converted into a Hoogsteen C⁺-G base-pair. In both situations the course of the backbone in the stem is continued at the 5'-side of the loop and the second loop residue is folded into the minor groove (type II), supporting the idea that this folding motif does not depend upon the geometry of the underlying helix. The position of this residue is stabilized by the formation of one or two hydrogen bonds.

Some uncertainty remains as to what extent

variations in the amount of strain determine the conformational switch of the base-pair. This term appears to be more favorable in the case of the compact Hoogsteen C⁺-G base-pair, relative to the Watson-Crick C-G pair, and even compensates for the net loss of one hydrogen bond. However, the effect of protonation of the cytosine residue(s) is an important issue as well, an effect which is hard to quantify accurately with the methods used in this study. Therefore, with the present data at hand, a detailed molecular dynamics free energy comparison, to investigate the type II-loop structures in more quantitative sense, becomes viable. The C1'-C1' distance in the base-pair in the loop, we expect, will then turn out to be an important factor.

Materials and Methods

Sample preparation

The oligonucleotides d(TCTTCC-TTTT-CCTTCT-CC-CG-AGAAGG-TTTT) and d(CCTTCT-CCCG-AGAAGG) were synthesized on a Pharmacia LKB Geneassembler special *via* the phosphotriester method (van Boom *et al.*, 1983) and desalted by passing through a G15 column (Pharmacia). NMR samples were typically 1 to 4 mM DNA in 500 μ l 90% H₂O/10% ²H₂O or 99.98% ²H₂O containing 0.1 M NaCl and adjusted to the appropriate pH (meter reading) by the addition of small amounts of HCl. The ²H₂O sample was prepared by dissolving the DNA in 99.98% ²H₂O, followed by three lyophilization steps, prior to the final preparation of the sample under N₂.

NMR spectroscopy

All NMR experiments, unless indicated otherwise, were conducted on a Bruker AMX-600 spectrometer. Phase-sensitive detection of the indirectly observed frequencies was achieved by the TPPI method (Marion & Wüthrich, 1983). The spectral width for the ¹H-dimensions was set to 10 ppm and 25 ppm for the experiments on the ²H₂O- and H₂O sample, respectively. 1D-spectra were processed and analyzed with the nmr1 program of the NMRi software package (New Methods Research, Inc., Syracuse, N.Y.), running on a SUN Sparc workstation. For the 2D-spectra the MNMR software package (PRONTO Software Development and Distribution, Copenhagen, Denmark) and the XEASY package (Xia & Bartels, 1994), which were both running on a Silicon Graphics Indigo workstation, were used for processing and analysis, respectively.

Series of 1D-spectra for the 16-base and 30-base DNA strands, dissolved in H₂O and adjusted to different pH values in the range between 4.5 and 8.0, were acquired at various temperatures using a 1 (-1) spin echo sequence to suppress the H₂O signal (Sklenář & Bax, 1987). NOE difference measurements were performed for the 30-mer as described by Wagner & Wüthrich (1979), using a 1 (-1) sequence (Plateau & Guéron, 1982). For the same sample, a NOESY experiment (Jeener *et al.*, 1979), with a mixing time of 400 ms, was performed at a temperature of 7°C. Suppression of the H₂O signal was achieved by replacing the last pulse by a jump-and-return sequence with the excitation maximum set at approximately 13.0 ppm. The spectrum was acquired with 2048 points in *t*₂ and 1024 points in the *t*₁-dimension.

For the $^2\text{H}_2\text{O}$ -sample of the 30-mer (pH 4.5, temperature 25 °C), two NOESY spectra (mixing times 80 and 200 ms) were recorded on a Varian UNITYplus spectrometer operating at 750 MHz with 2048 and 640 data points in the t_1 and t_2 -dimensions, respectively. A series of TOCSY-experiments (Bax & Davis, 1985) with clean-MLEV17 mixing sequences (Griesinger *et al.*, 1988), varying in time from 15 to 260 ms, were recorded with 512 t_1 increments. A P.E.COSY experiment was carried out by the method described by Bax & Lerner (1988) with 2048 data points in t_2 and 512 data points in t_1 . The 1D reference spectrum was acquired with 20 times as many transients as the 2D spectrum. A ^{31}P - ^1H hetero-correlation spectrum was acquired on a Bruker AM-400 spectrometer, operating in the inverse mode, with 2048 points in t_2 and 256 points in t_1 and with the spectral width in the ^{31}P -dimension set to 2000 Hz. The pulse scheme proposed by Sklenář *et al.* (1986) was used, leaving out the ^1H 180° presaturation pulses, and with the phase cycle reduced to two steps for axial peak suppression. A three-dimensional TOCSY-NOESY experiment (Oschkinat *et al.*, 1989a,b) was recorded with $256(t_1) \times 86(t_2) \times 1024(t_3)$ data points. To optimize the spectral resolution, the experiment was performed such that the second dimension was selective for the H2'/H2" region, which was achieved by the combination of a non-selective 90° pulse and an E-BURP-2 excitation pulse (Gean & Freeman, 1991; Freeman, 1991), centered at 2.55 ppm, after the 63 ms clean-MLEV17 mixing period. The NOE mixing time in this experiment had a duration of 200 ms.

In addition, experiments were performed on the 16-mer at pH 7.0. One-dimensional spectra at various temperatures and a NOESY-spectrum (mixing time 400 ms) at a temperature of 1 °C, were recorded for this fragment dissolved in H_2O . For the $^2\text{H}_2\text{O}$ -sample a NOESY spectrum (200 ms) was acquired at 25 °C on a Varian UNITYplus spectrometer operating at 500 MHz.

Structure restraints for the CCCG-loop at pH 4.5

For the initial structure calculations, NOE intensities in the 80 ms NOESY spectrum were translated into loose distance restraints (type A), i.e. 3.5 to 7.0 Å (weak), 2.5 to 5.0 Å (moderate), and 1.8 to 3.5 Å (strong). For the subsequent structure refinement, 42 non-overlapping inter-residue NOEs were quantitatively converted into distance restraints (type B) by means of the isolated spin pair approach, with the intra-residue H6-H5 NOEs (2.41 Å) from the three cytosine residues in the loop and the intra-residue H1'-H2" NOEs (2.25 Å) serving as internal references. For well resolved cross-peaks error bounds of -10% and +15% of the calculated distances were used for lower and upper limits, respectively. Spin diffusion was relatively unimportant here, as was validated by repeating the latter procedure for the 200 ms NOESY, which did not produce distances deviating more than 10% from the results obtained from the 80 ms NOE intensities and by involving back calculations of the NOESY spectrum (*vide infra*). The NOEs in the spectra recorded for the H_2O -spectra were used in a very qualitative manner (lower and upper bounds of 1.8 and 7.0 Å, respectively). In addition, the T6-A11 base-pair closing the loop was preserved by explicit restraints on the lengths of the hydrogen bonds and the C1'-C1', N1-N9 and C6-C8 distances (Saenger, 1984).

On the basis of the intra-residue H6/H8-H1'/H2'/H2" NOE intensities (Wijmenga *et al.*, 1993) restraints for the glycosidic torsion angles χ were derived. J -coupling constants, i.e. $J_{\text{H1}'\text{H2}'}$ and $J_{\text{H1}'\text{H2}''}$ and sums of J -couplings, i.e. $\Sigma J_{\text{H1}'} = J_{\text{H1}'\text{H2}'} + J_{\text{H1}'\text{H2}''}$, $\Sigma J_{\text{H2}'} = J_{\text{H2}'\text{H1}'} + J_{\text{H2}'\text{H2}''} + J_{\text{H2}'\text{H3}'}$, and $\Sigma J_{\text{H2}''} = J_{\text{H2}''\text{H1}'} + J_{\text{H2}''\text{H2}'} + J_{\text{H2}''\text{H3}'}$, which were determined from the cross-peak splittings observed in the P.E.COSY spectrum, allowed for the calculation of bounds for the phase angles of pseudorotation, P , and the percentages of S-pucker, ξ , for the loop residues (van Wijk *et al.*, 1992). Alternatively, the approach of the quantitative analysis of TOCSY intensities (Van Duynhoven *et al.*, 1992) was used. The TOCSY transfer of the H1' proton was measured at several mixing times, and fitted to curves calculated for different values of ξ and P . Subsequently, permitted ranges for the endocyclic torsion angles and δ were deduced according to the relationships (Altona & Sundralingam, 1972):

$$\delta = \nu_3 + 125^\circ$$

$$\nu_j = \phi_m \cos(P + 144(j - 2))$$

with ϕ_m equal to 37° and P in the range between -20° and 36° for predominant N-puckers, and between 144° and 180° for predominant S-puckers.

Restrictions for the combination of the dihedral angles β and γ were derived on the basis of the observation of a P-H4' cross-peak in the ^{31}P - ^1H correlation spectrum. Further restraints for β and γ followed from the relative H3'-H5', H3'-H5'', H4'-H5', and H4'-H5'' NOE intensities and relative H4'-H5', H4'-H5'', H5'-P, and H5''-P J -couplings according to the intensities in TOCSY-spectra and the ^{31}P - ^1H correlation spectrum (Blommers *et al.*, 1991). The dihedral angles ϵ of all residues were prevented from adopting values in the sterically forbidden *gauche* domain (Mooren, 1993).

Generation of starting structures

All restraints (i.e. 93 upper and lower distance limits and 42 dihedral angle constraints) were taken as input for the DIANA variable target distance geometry algorithm, version 2.8 (Güntert *et al.*, 1991; Güntert, 1995), which was run on a Silicon Graphics Indigo²™ work station to calculate structures of the CCCG-loop region, including the closing T·A base-pair. On each level of DIANA calculations, 100 iterations were performed, with the weighting factor for the van der Waals repulsion set to 0.2. At the fifth level three times 300 additional iterations were performed, during which the van der Waals repulsion-term was increased to 5.0. The weighting factors for upper and lower limits of the distance restraints and the weighting factor of the dihedral angle constraints were set to 1.0, 1.0, and 50, respectively, throughout the calculation. For 9 out of 12 methylene groups, stereospecific assignments were used.

After the first round of DIANA calculations, which produced 50 structures in total, the coordinates of the structures with the lowest target functions were used as input for the CORMA program (Keepers & James, 1984). It followed from the back-calculated NOESY spectrum that several relatively short distances were present in the structures calculated by DIANA, although no experimental NOEs were observed for the corresponding proton pairs. After verifying that these distances involved assigned, non-exchangeable protons, and that the absence of the NOE was not caused by flexibility of that part of the structure, a lower bound of 3.5 Å was added to the experimental restraints for these proton pairs, and

DIANA was run again. Subsequent cycles of DIANA and CORMA were carried out until all calculated NOEs could be accounted for. In a final DIANA run, producing 500 structures in total, 118 upper and lower bounds and 42 dihedral angle restraints were included. The 69 structures with smallest target function values ($<0.5\text{\AA}^2$) of this set were used for the structure refinement calculations.

Structure refinement

The structures derived after the final DIANA calculations were refined using version 3.1 of X-PLOR (Brünger, 1992), which was run on a Silicon Graphics Indigo^{2TM} work station. A patch was constructed to protonate residue C7, resulting in a residue with a net positive charge of 0.33e, distributed over the base as described by Live *et al.* (1991). During the refinement, of which the protocol is described below, the force field consisted of geometric terms accounting for bond stretching, bond angle bending, and maintaining planar ring structures and (pro)chiral centers. A soft repulsive term was used for the non-bonded interactions. The NMR-derived distance and dihedral angle restraints were added as (soft) square-well potentials, with force constants which were kept constant during the protocol at $50\text{ kcal mol}^{-1}\text{\AA}^{-2}$ and $200\text{ kcal mol}^{-1}\text{rad}^{-2}$, respectively. The T6-A11 base-pair was kept flat using planarity restraints, with a small force constant of $10\text{ kcal mol}^{-1}\text{\AA}^{-2}$. Two consecutive runs of restrained molecular dynamics calculations were performed during which the system was cooled from 1000 K to 100 K during 9 ps in steps of 3 fs. During these calculations the force constant for the non-bonded interactions energy term, which was present as a pure repulsive potential, was increased from 0.003 to 4.0 kcal mol^{-1} , while the van der Waals radii were decreased from 0.9 to 0.75 times the atom radii. The resulting structures were subjected to 200 cycles of restrained conjugate gradient energy minimization and accepted as reasonable conformations if they produced low overall NOE and restrained dihedral energies ($<300\text{ kcal mol}^{-1}$, $<30\text{ kcal mol}^{-1}$, and $<15\text{ kcal mol}^{-1}$, respectively).

Analysis of structures

On the basis of an analysis of the torsion angles in the backbone, the final structures were classified into different sets. For each set an average structure was calculated, which subsequently was subjected to a short run (1200 steps) of restrained energy minimization using X-PLOR. After 200 steps with the force field as used in the structure refinement calculations, terms for electrostatic and dihedral contributions ($2\text{ kcal mol}^{-1}\text{rad}^{-2}$) were added. The soft repulsive non-bonded energy term was replaced by a Lennard-Jones potential. The effect of atomic charges was reduced by the application of partial charges and by choosing a distance dependent dielectric constant $\epsilon = \epsilon_0 \cdot r$. Several specific interaction energy contributions were calculated using the V_{pert} function in X-PLOR.

At various stages of the structure calculations, RMSD values were computed, both as mean RMSDs between an average structure and the converged structures, as well as average pairwise RMSDs, using all heavy atoms for obtaining the best fit superposition. Quanta (Polygen), running on a Silicon Graphics Indigo^{2TM} work station, was used for visualization of the results of the structure calculations.

Generation of a structural model for the CCCG-loop at pH 7.0

The average structure with lowest overall energy for the CCCG-loop at pH 4.5 served as one of the starting points for a molecular modeling study, of the loop structure at pH 7.0. Using Quanta, the glycosidic torsion angle of G10 was altered from *syn* (73°) into *anti* (-90°). Alternatively, a starting-structure with dihedral angles as found in the -d(CCTG)- and -d(CTCG)- miniloops (Ippel, 1993) was constructed. These structures were first regularized by energy minimization (1000 steps), using the same force field as used for the structure refinement calculations (*vide supra*), and subsequently by molecular dynamics calculations (20 ps at 300 K) and energy minimization (2000 steps) in a force field including dihedral and electrostatic contributions. No other restraints other than those for the endocyclic and δ torsion angles, for the distances defining the closing T-A base-pair, and, in the early stages of the calculations, for the distances between hydrogen bond donors and acceptors in the C7-G10 Watson-Crick base-pair were used during these calculations. As a comparison, the low pH structure was subjected to the last stage as well, i.e. 2000 steps of energy minimization in the same force field, using the same limited set of restraints. Additionally, hypothetical variants of the low pH structure with alternative net charges (0 and 1e) on residue C7, as well as structures with all and none of the cytosine residues protonated were constructed and subjected to energy minimization. The overall energies as well as some specific energy contributions of the resulting structures were computed as purely unconstrained total energies.

Acknowledgements

NMR spectra were recorded at the Dutch National hf-NMR Facility (Nijmegen, the Netherlands), supported by SON, and the SON NMR large-scale Facility (Utrecht, the Netherlands). We thank J. Joordens for excellent technical assistance.

References

- Allain, F. H.-T. & Varani, G. (1995). Structure of the P1 helix from group 1 self-splicing introns. *J. Mol. Biol.* **250**, 333–353.
- Altona, C. & Sundralingam, M. (1972). Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Am. Chem. Soc.* **94**, 8205–8212.
- Antao, V. P. & Tinoco, I., Jr (1992). Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucl. Acids Res.* **20**, 819–824.
- Antao, V. P., Lai, S. Y. & Tinoco, I., Jr (1991). A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucl. Acids Res.* **19**, 5901–5902.
- Bax, A. & Davis, D. G. (1985). MLEV-17 based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **65**, 355–366.
- Bax, A. & Lerner, L. (1988). Measurement of ^1H - ^1H coupling constants in DNA fragments by 2D NMR. *J. Magn. Reson.* **79**, 429–438.
- Blommers, M. J. J., Walters, J. A. L. I., Haasnoot, C. A. G., Aelen, J. M. A., van der Marel, G. A., van Boom, J. H. & Hilbers, C. W. (1989). Effects of base sequence on

- the loop folding in DNA hairpins. *Biochemistry*, **28**, 7491–7498.
- Blommers, M. J. J., van de Ven, F. J. M., van der Marel, G. A., van Boom, J. H. & Hilbers, C. W. (1991). The three-dimensional structure of a DNA hairpin in solution. Two-dimensional NMR studies and structural analysis of d(ATCCTATTTATAGGAT). *Eur. J. Biochem.* **201**, 33–51.
- Brünger, A. T. (1992). *X-PLOR Manual, Version 3.1*, Yale University Press, New Haven.
- Chen, X., Santhana Mariappan, S. V., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoun, A., Smith, S. S., Bradbury, E. M. & Gupta, G. (1995). Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological functions. *Proc. Natl Acad. Sci. USA*, **92**, 5199–5203.
- Davison, A. & Leach, D. R. F. (1994). 2 Base DNA hairpin loop structures in vivo. *Nucl. Acids Res.* **22**, 4361–4363.
- Freeman, R. (1991). Selective excitation in high-resolution NMR. *Chem. Rev.* **91**, 1397–1412.
- Geen, H. & Freeman, R. (1991). Band-selective radio frequency pulses. *J. Magn. Reson.* **93**, 93–141.
- Griesinger, C., Otting, G. & Wüthrich, K. (1988). Clean TOCSY for ¹H spin system identification in macromolecules. *J. Am. Chem. Soc.* **110**, 7870–7872.
- Güntert, P. (1995). *DIANA User's Manual and Instructions, version 2.8*, Zürich, Switzerland.
- Güntert, P., Braun, W. & Wüthrich, K. (1991). Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **217**, 517–530.
- Haasnoot, C. A. G., de Bruin, S. H., Berendsen, R. G., Janssen, H. G. J. M., Binnendijk T. J. J., Hilbers, C. W., van der Marel, G. A. & van Boom, J. H. (1983). Structure, kinetics and thermodynamics of DNA hairpin fragments in solution. *J. Biomol. Struct. Dynam.* **1**, 115–129.
- Haasnoot, C. A. G., Hilbers, C. W., van der Marel, G. A., van Boom, J. H., Singh, U. C., Pattabiraman, N. & Kollman, P. A. (1986). On loopfolding in nucleic acid hairpin-type structures. *J. Biomol. Struct. Dynam.* **3**, 843–857.
- Hilbers, C. W., Blommers, M. J. J., van de Ven, F. J. M., van Boom, J. H. & van der Marel, G. A. (1991). High resolution NMR studies of DNA hairpins with four nucleotides in the loop region. *Nucleosides Nucleotides*, **10**, 61–80.
- Hilbers, C. W., Heus, H. A., van Dongen, M. J. P. & Wijmenga, S. S. (1994). The hairpin elements of nucleic acid structure: DNA and RNA folding. In *Nucleic Acids and Molecular Biology* (Eckstein, F. & Lilley, D. M. J., eds), vol. 8, pp. 56–104, Springer-Verlag, Berlin Heidelberg.
- Ippel, J. H. (1993). Base sequence effects in loop folding of minihairpin structures. In *Loop Folding in DNA Minihairpins. Conformational Studies*, pp. 209–239, Thesis University of Leiden, Leiden, The Netherlands.
- Ippel, J. H., Lanzotti, V., Galeone, A., Mayol, L., van den Boogaart, J. E., Pikkemaat, J. A. & Altona, C. (1992). An NMR study of the conformation and thermodynamics of the circular dumbbell d(pCGC-TT-GCG-TT). *J. Biomol. Struct. Dynam.* **9**, 821–836.
- Jeener, J., Meier, B. H., Bachman, P. & Ernst, R. R. (1979). Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71**, 4546–4553.
- Jucker, F. M. & Pardi, A. (1995). Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry*, **34**, 14416–14427.
- Keepers, J. W. & James, T. L. (1984). A theoretical study of distance determinations from NMR. Two-dimensional nuclear Overhauser effect spectra. *J. Magn. Reson.* **57**, 404–426.
- Kuklenyik, Z., Yao, S. & Marzilli, L. G. (1996). Similar conformations of hairpins with TTT and TTTT sequences; NMR and molecular modeling evidence for T-T base-pairs in the TTTT hairpin. *Eur. J. Biochem.* **236**, 960–969.
- Liquier, J., Coffinier, P., Firon, M. & Taillandier, E. (1991). Triple helical polynucleotide structures: sugar conformations determined by FTIR spectroscopy. *J. Biomol. Struct. Dynam.* **9**, 437–445.
- Live, D. H., Radhakrishnan, I., Misra, V. & Patel, D. J. (1991). Characterization of protonated cytidine in oligonucleotides by ¹⁵N NMR studies at natural abundance. *J. Am. Chem. Soc.* **113**, 4687–4688.
- Macaya, R., Wang, E., Schultze, P., Sklenář, V. & Feigon, J. (1992). Proton nuclear magnetic resonance assignments and structural characterization of an intramolecular DNA triplex. *J. Mol. Biol.* **225**, 755–773.
- Marion, D. & Wüthrich, K. (1983). Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ¹H-¹H spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.* **113**, 967–974.
- Mooren, M. M. W. (1993). Correlations between torsion angles in nucleic acids: Allowed and disallowed conformations. In *On Nucleic Acid Structure Analysis by NMR*, pp. 79–141, Thesis University of Nijmegen, Nijmegen, The Netherlands.
- Mooren, M. M. W., Hilbers, C. W., van der Marel, G. A., van Boom, J. H. & Wijmenga, S. S. (1991a). Three dimensional homonuclear TOCSY-NOESY of nucleic acids. Possibilities for improved assignments. *J. Magn. Reson.* **94**, 101–111.
- Mooren, M. M. W., Pulleyblank, D. E., Wijmenga, S. S., Blommers, M. J. J. & Hilbers, C. W. (1991b). Polypurine/polypyrimidine hairpins form a triple helix structure at low pH. *Nucl. Acids Res.* **18**, 6523–6529.
- Orbons, L. P. M., van Beuzekom, A. A. & Altona, C. (1987). Conformational and model-building studies of the hairpin form of the mismatched DNA octamer d(m⁵C-G-m⁵C-T-G-m⁵C-G). *J. Biomol. Struct. Dynam.* **4**, 965–987.
- Oschkinat, H., Cieslar, C., Gronenborn, A. M. & Clore, G. M. (1989a). Three-dimensional homonuclear Hartman-Hahn-nuclear-Overhauser-enhancement spectroscopy in H₂O and its applications to proteins. *J. Magn. Reson.* **81**, 212–216.
- Oschkinat, H., Cieslar, C., Holak, T. A., Clore, G. M. & Gronenborn, A. M. (1989b). Practical and theoretical aspects of three-dimensional homonuclear Hartman-Hahn-nuclear-Overhauser-enhancement spectroscopy of proteins. *J. Magn. Reson.* **83**, 450–472.
- Pardi, A., Walker, R., Rapoport, H., Wider, G. & Wüthrich, K. (1983). Sequential assignments for the ¹H and ³¹P atoms in the backbone of oligonucleotides by two-dimensional nuclear magnetic resonance. *J. Am. Chem. Soc.* **105**, 1652–1653.
- Plateau, P. & Guéron, M. (1982). Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences. *J. Am. Chem. Soc.* **104**, 7310–7311.

- Radhakrishnan, I. & Patel, D. J. (1994). Solution structure of a pyrimidine-purine-pyrimidine DNA triplex containing an unusual G-TA triple. *Structure*, **2**, 17–32.
- Radhakrishnan, I., Gao, X., de los Santos, C., Live, D. & Patel, D. J. (1991). NMR structural studies of intramolecular $(Y^+)_n(R^+)_n(Y^-)_n$ DNA triplexes in solution: imino and amino proton and nitrogen markers of G-TA base triple formation. *Biochemistry*, **30**, 9022–9030.
- Saenger, W. (1984). *Principles of Nucleic Acid Structure*, chapt. 6, pp. 116–158, Springer Verlag, New York.
- Santhana Mariappan, S. V., Catasti, P., Chen, X., Ratliff, R., Moyzis, R. K., Bradbury, E. M. & Gupta, G. (1996a). Solution structures of the individual single strands of the fragile X DNA triplets $(GCC)_n(GGC)_n$. *Nucl. Acids Res.* **24**, 784–792.
- Santhana Mariappan, S. V., Garcia, A. E. & Gupta, G. (1996b). Structure and dynamics of the DNA hairpins formed by tandemly repeated CTG triplets associated with myotonic dystrophy. *Nucl. Acids Res.* **24**, 775–783.
- Sarma, R. H., Mynott, R. J., Wood, D. J. & Hruska, F. E. (1973). Determination of the preferred conformations along the C-4'-C-5' and C-5'-O-5' bonds of β -5'-nucleotides in solution. Four-bond ^{31}P - ^1H coupling. *J. Am. Chem. Soc.* **95**, 6457–6459.
- Sklenář, V. & Bax, A. (1987). Spin-echo water suppression for the generation of pure-phase two-dimensional NMR spectra. *J. Magn. Reson.* **74**, 469–479.
- Sklenář, V., Miyashiro, H., Zon, G., Miles, H. T. & Bax, A. (1986). Assignment of the ^{31}P and ^1H resonances in oligonucleotides by two-dimensional NMR spectroscopy. *FEBS Letters*, **208**, 94–98.
- Van Boom, J. H., van der Marel, G. A., Westerink, H. P., van Boeckel, C. A. A., Mellema, J. R., Altona, C., Hilbers, C. W., Haasnoot, C. A. G., de Bruin, S. H. & Berendsen, R. G. (1983). Synthesis and conformational analysis of synthetic DNA fragments. *Cold Spring Harbor Symp. Quant. Biol.* **47**, 403–409.
- Van Duynhoven, J. P. M., Goudriaan, J., Hilbers, C. W. & Wijmenga, S. S. (1992). Quantitative evaluation of TOCSY data. Application to sugar ring conformation analysis. *J. Am. Chem. Soc.* **114**, 10055–10056.
- Van Wijk, J., Huckriede, B. D., Ippel, J. H. & Altona, C. (1992). Furanose sugar conformations in DNA from NMR coupling constants. In *Methods in Enzymology* (Lilley, D. M. J. & Dahlberg, J. E., eds), vol. 211, pp. 286–306, Academic Press Inc., San Diego.
- Varani, G., Cheong, C. & Tinoco I., Jr (1991). Structure of an unusually stable RNA hairpin. *Biochemistry*, **30**, 3280–3289.
- Wagner, G. & Wüthrich, K. (1979). Truncated driven nuclear Overhauser effect (TOE). A new technique for studies of selective ^1H - ^1H Overhauser effects in the presence of spin diffusion. *J. Magn. Reson.* **33**, 675–680.
- Wijmenga, S. S., Mooren, M. M. W. & Hilbers, C. W. (1993). NMR of nucleic acids; from spectrum to structure. In *NMR of Macromolecules, A Practical Approach* (Roberts, G. C. K., ed.), pp. 217–288, Oxford University Press, New York.
- Wijmenga, S. S., Heus, H. A., Werten, B., van der Marel, G. A., van Boom, J. H. & Hilbers, C. W. (1994). Assignment strategies and analysis of cross-peak patterns and intensities in the three-dimensional homonuclear TOCSY-NOESY of RNA. *J. Magn. Reson. B*, **103**, 134–141.
- Xia, T. & Bartels, C. (1994). XEASY. Institute of Molecular Biology and Biophysics, Zürich, Switzerland.

Edited by I. Tinoco

(Received 3 May 1996; received in revised form 29 August 1996; accepted 29 August 1996)



Supplementary material, comprising two figures and two tables, is available from JMB Online.