The Transition from a Neutral-pH Double Helix to a Low-pH Triple Helix Induces a Conformational Switch in the CCCG 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. Hilbers¹ *

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 CCCG-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.

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

Introduction

Hairpins are abundant structural elements in nucleic acids. They form an integral part of 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
A Conformational Switch in the CCCG Tetraloop

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

<table>
<thead>
<tr>
<th>Residue</th>
<th>NH*</th>
<th>NH2*</th>
<th>H5</th>
<th>H6</th>
<th>H7</th>
<th>H2</th>
<th>H1'</th>
<th>H2'</th>
<th>H2&quot;</th>
<th>H3'</th>
<th>H4'</th>
<th>H5'</th>
<th>H5&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>T6</td>
<td>14.44</td>
<td>—</td>
<td>7.65</td>
<td>1.61</td>
<td>6.06</td>
<td>2.26</td>
<td>2.53</td>
<td>4.82</td>
<td>4.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7</td>
<td>10.13</td>
<td>9.31</td>
<td>7.70</td>
<td>5.64</td>
<td>5.80</td>
<td>2.10</td>
<td>2.35</td>
<td>4.69</td>
<td>4.26</td>
<td>4.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8</td>
<td>9.83</td>
<td>8.34</td>
<td>8.30</td>
<td>6.39</td>
<td>6.14</td>
<td>1.70</td>
<td>2.21</td>
<td>4.61</td>
<td>4.32</td>
<td>4.26</td>
<td>4.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C9</td>
<td>9.23</td>
<td>8.25</td>
<td>7.54</td>
<td>5.74</td>
<td>5.27</td>
<td>1.65</td>
<td>2.06</td>
<td>4.32</td>
<td>3.14</td>
<td>3.33</td>
<td>3.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G10</td>
<td>6.89</td>
<td>—</td>
<td>5.84</td>
<td>3.63</td>
<td>2.57</td>
<td>4.69</td>
<td>3.75</td>
<td>4.25</td>
<td>4.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A11</td>
<td>8.19</td>
<td>7.94</td>
<td>8.10</td>
<td>7.61</td>
<td>6.28</td>
<td>2.95</td>
<td>3.02</td>
<td>5.06</td>
<td>4.43</td>
<td>3.99</td>
<td>4.10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A. CCCG-loop in the triple helix at 25°C

B. CCCG-loop in the hairpin at 25°C

*Chemical shifts of exchangeable protons at 7°C.

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 CCCG-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, \( \chi \), 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 \( \chi \)-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
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-$^1$H 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 $\gamma$-angle of this residue resides in the trans-region. For the $\gamma$-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 $\epsilon$-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

<table>
<thead>
<tr>
<th>Structure restraints:</th>
<th>DIANA</th>
<th>X-PLOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of distance restraints</td>
<td>118</td>
<td>118</td>
</tr>
<tr>
<td>Intra-residue</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Inter-residue</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Type A</td>
<td>53</td>
<td>11</td>
</tr>
<tr>
<td>Type B</td>
<td>0</td>
<td>42</td>
</tr>
<tr>
<td>Exchangeable protons</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>T-A base-pair</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Lower limit</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Number of dihedral restraints</td>
<td>42</td>
<td>54</td>
</tr>
<tr>
<td>Planarity restraints present (+/-)</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

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.

CCC‘G-structure at neutral pH

Structure characterization

Figure 2b shows the one-dimensional proton spectrum of the 30-base DNA strand, dissolved in H$_2$O, 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 H$_2$O-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 $\gamma$-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.
A Conformational Switch in the CCCG Tetraloop

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

We decided therefore not to perform an \textit{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 \textit{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 3. 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 \textit{syn}-domain. The guanine base forms a Hoogsteen base-pair, involving two hydrogen bonds, with the protonated C7 base (Figure 6a).

<table>
<thead>
<tr>
<th>Table 3. Dihedral angles of the CCCG-loop</th>
</tr>
</thead>
<tbody>
<tr>
<td>T6</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>A. Low pH structure (average structure after restrained energy minimization)</td>
</tr>
<tr>
<td>$\alpha$</td>
</tr>
<tr>
<td>$\beta$</td>
</tr>
<tr>
<td>$\gamma$</td>
</tr>
<tr>
<td>$\delta$</td>
</tr>
<tr>
<td>$\epsilon$</td>
</tr>
<tr>
<td>$\zeta$</td>
</tr>
<tr>
<td>$\chi$</td>
</tr>
</tbody>
</table>

B. Neutral pH structure

| $\alpha$ | 262.4 | 204.3 | 199.1 | 311.0 |
| $\beta$ | 53.3 | 150.5 | 205.3 | 63.0 |
| $\gamma$ | 136.2 | 168.5 | 62.8 | 136.7 |
| $\delta$ | 238.7 | 206.5 | 192.2 | 210.3 |
| $\epsilon$ | 299.7 | 269.2 | 278.4 | 281.1 |
| $\zeta$ | 242.6 | 235.4 | 202.4 | 244.0 |
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: \( \alpha(8)\text{-trans} \); \( \epsilon(8)\text{-}(-)\text{anticlinal} \). 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, which closes the B-type double helix stem of a hairpin. At neutral pH a triple helix is formed, in which the Watson-Crick C·G base-pair, and the position of the Y base in the minor groove (type II). Very recently, Jucker & Pardi (1993) 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
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 (\(\Delta H = -5.6\) 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 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 (\(\Delta H = -3.8\) 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

### Table 4. Energy contributions of the CCCG-loop

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

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.
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 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 general 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-CC-AGAAGG-TTIT) and d(CCTTCT-C 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 µl 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 (Sklénář & Bax, 1987). NOE difference measurements were performed for the 30-mer as described by Wagner & Wüthrich (1979), using a 1 (−1) sequence (Plateau & Guérin, 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 H2O-sample of the 30-mer (pH 4.5, temperature 25 °C), two NOEY 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 $t_2$ and $t_3$-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 31P-H hetero-correlation spectrum was acquired on a Bruker AM-400 spectrometer, operating in the inverse mode, with 2048 points in the 31P-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-NOE experiment (Oschkinat et al., 1989a,b) was recorded with 256($t_2$)×86($t_3$)×1024($t_1$) 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 (Geen & 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 NOEY-spectrum (mixing time 400 ms) at a temperature of 1 °C, were recorded for this fragment dissolved in H2O. For the 2H2O-sample a NOEY 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 NOEY 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 NOEY, 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 NOEY spectrum (vide infra). The NOEs in the spectra recorded for the H2O-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 $\chi$ were derived. $J$-coupling constants, i.e. $J_{HH}$ and $J_{HH}$ and sums of $J$-couplings, i.e. $\Sigma_{HH} = J_{HH} + J_{HH}, \Sigma_{HH} = J_{HH} + J_{HH} + J_{HH}$, and $\Sigma_{HH} = J_{HH} + J_{HH} + J_{HH}$, 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, $f_s$, for the loop residues (van Wijk et al., 1992). Alternatively, the approach of the quantitative analysis of TOCSY intensities (Van Dyunhoven 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 $f_s$ and $P$. Subsequently, permitted ranges for the endocyclic torsion angles and $\delta$ were deduced according to the relationships (Altunia & Sundralingam, 1972):

$$\delta = v_1 + 125^\circ$$

with $\phi_n$ 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 $\beta$ and $\gamma$ were derived on the basis of the observation of a H-H cross-peak in the 31P-H correlation spectrum. Further restraints for $\beta$ and $\gamma$ 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 $31P$-H correlation spectrum (Blommers et al., 1991). The dihedral angles $\epsilon$ 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 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 NOEY 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 Å) 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™ 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 kcal mol⁻¹Å⁻² and 200 kcal mol⁻¹rad⁻², respectively. The T6-A11 base-pair was kept flat using planarity restraints, with a small force constant of 10 kcal mol⁻¹Å⁻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 (<30 kcal mol⁻¹, <30 kcal mol⁻¹, and <15 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 kcal mol⁻¹rad⁻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 ε = ε0·r. Several specific interaction energy contributions were calculated using the V_inter 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™ 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


the loop folding in DNA hairpins. Biochemistry, 28, 7491–7498.


Keepers, J. W. & James, T. L. (1984). A theoretical study of distance determinations from NMR. Two-dimen-


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.