Radboud Repository

Radboud University Nijmegen

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

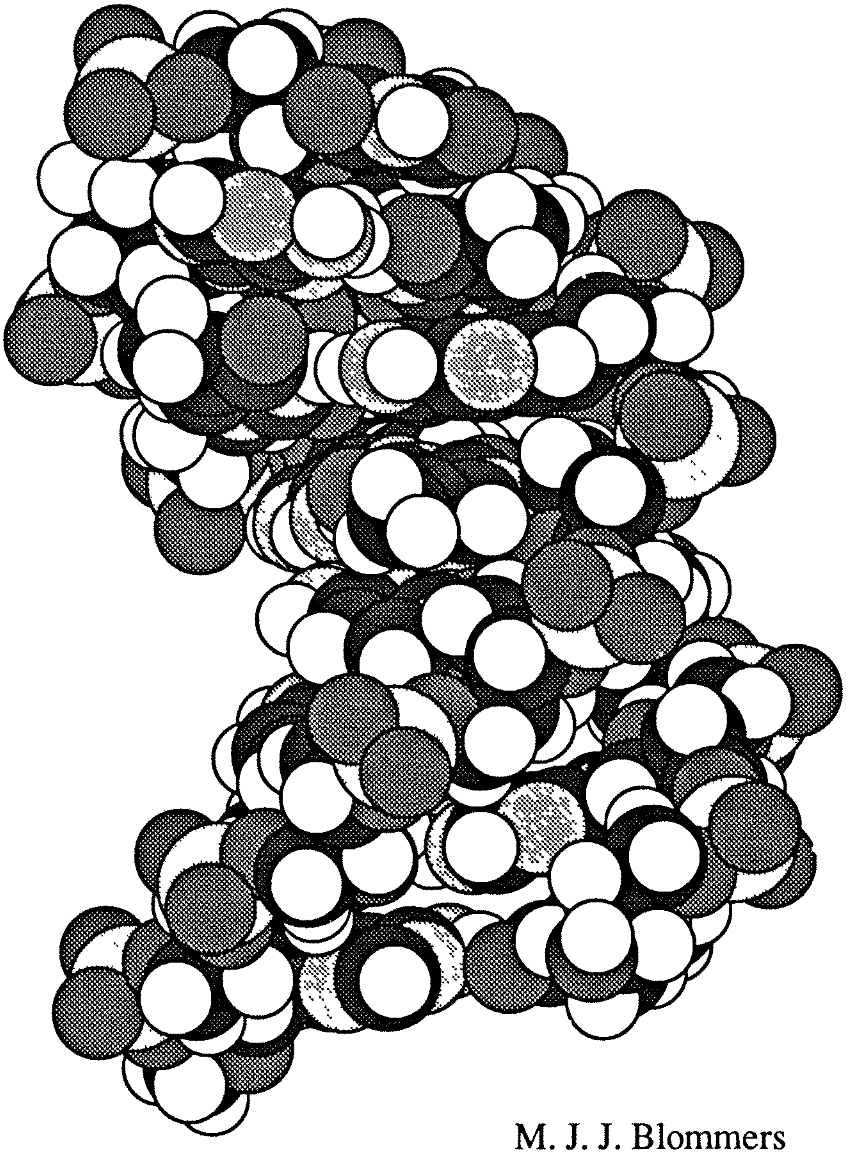
The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/113853

Please be advised that this information was generated on 2017-12-06 and may be subject to change.

ASPECTS OF LOOP FOLDING IN DNA HAIRPINS

Nuclear Magnetic Resonance, Multi-Conformational Analysis, UV-Melting and Molecular Mechanics Calculations



M. J. J. Blommers

ASPECTS OF LOOP FOLDING IN DNA HAIRPINS

Nuclear Magnetic Resonance, Multi-Conformational Analysis, UV-Melting and Molecular Mechanics Calculations

CIP DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Blommers, Marcel Jules José

Aspects of loop folding in DNA hairpins. Nuclear magnetic
resonance, multi-conformational analysis, UV-melting and
molecular mechanics calculations. / Marcel Jules José
Blommers. - [S.l. : s.n.] (Meppel: Krips Repro). - Ill.
Thesis Nijmegen. - With ref. - With summary in Dutch.
ISBN 90-9003640-7
SISO 573.2 UDC 543:577.21(043.3)
Subject headings: nucleic acids / DNA / nuclear magnetic
resonance.

ASPECTS OF LOOP FOLDING IN DNA HAIRPINS

NUCLEAR MAGNETIC RESONANCE, MULTI-CONFORMATIONAL ANALYSIS, UV-MELTING AND MOLECULAR MECHANICS CALCULATIONS

Een wetenschappelijke proeve op het gebied van
de natuurwetenschappen, in het bijzonder de chemie.

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan
de Katholieke Universiteit te Nijmegen,
volgens besluit van het college van decanen in het
openbaar te verdedigen op
maandag 29 oktober 1990
des namiddags te 1.30 uur precies

door

Marcel Jules José Blommers
geboren op 2 maart 1957
te Amersfoort

Druk: Krips Repro Meppel

Promotor: Prof. Dr. C.W. Hilbers

I.S.B.N. 90-9003640-7

CONTENTS

1. **GENERAL INTRODUCTION** 1

2. **THE SOLUTION STRUCTURE OF THE 3'-5' CYCLIC DINUCLEOTIDE d<pApA>**
A COMBINED NMR, UV-MELTING AND MOLECULAR MECHANICS STUDY 33

3. **AN INVESTIGATION OF THE STRUCTURE OF THE 3'-5' CYCLIC DINUCLEOTIDE d<pApA> BY MEANS OF MULTICONFORMER ANALYSIS**
A MOLECULAR MODELLING STUDY 55

4. **NMR STUDIES ON LOOPFOLDING IN A DNA HAIRPIN MOLECULE**
AN INVESTIGATION OF THE SOLUTION STRUCTURE OF d(ATCCTA-TTTT-TAGGAT) 79

5. **CONFORMATIONAL ASPECTS OF HAIRPIN LOOPS IN DNA OLIGONUCLEOTIDES**
A MOLECULAR MODELLING STUDY 107

6. **EFFECTS OF THE BASE SEQUENCE ON THE LOOPFOLDING IN DNA HAIRPINS**
A NMR AND UV-MELTING STUDY 121

7. **THREE-DIMENSIONAL STRUCTURE OF A DNA HAIRPIN IN SOLUTION**
2-DIMENSIONAL NMR STUDIES AND MULTICONFORMATION ANALYSIS ON d(ATCCTA-TTTA-TAGGAT) 143

SUMMARY 195

SAMENVATTING 201

CURRICULUM VITAE 207

NAWOORD 209

CHAPTER 1

GENERAL INTRODUCTION

Nucleic acids play an important role in the biochemistry of the cell. The most outstanding nucleic acid species are the DNA molecules which carry the genetic information and the RNA molecules which are involved in the translation of the base-sequence (genetic code) into the amino acid sequence (protein). In addition, nucleotides and their derivatives function in many other cellular processes. The DNA and RNA molecules consist of a chemically linked sequence of nucleotides; each nucleotide is composed of a purine or pyrimidine base, a deoxy- or ribose sugar ring and a phosphate group. Usually two types of purines are found in nucleic acid chains, i.e., adenine and guanine; the pyrimidines, which are occurring are thymidine or uracil and cytosine. The bases of two strands can form base pairs leading to the formation of a double-helical duplex structure (vide infra). When a sequence of bases is followed by a complementary sequence in the same chain, the polynucleotide may fold back on itself to form a hairpin structure. It consists of a base-paired double-stranded region and a loop of unpaired bases.

The three-dimensional structure or conformation of nucleic acids is expected to be crucial for their biological function. Therefore, the structure of nucleic acids has been studied with various physico-chemical techniques. Among these, X-ray-diffraction is widely used for the determination of the conformation of compounds in the crystalline form. Studies by means of this technique revealed that DNA has a double-stranded structure and that DNA duplexes can adopt at least three different helical conformations: the right-handed A- and B-helices and the left-handed Z-helix. X-ray studies at atomic resolution gave much insight in the local conformation, i.e., base-sequence effects on the structure. However, the crystal structure can deviate considerably from the conformation in solution which depends on the temperature, the ion-strength, the pH and the presence of DNA-binding compounds. Here, various spectroscopic methods are very important in the study of the structure and dynamics of nucleic acids in solution.

Nuclear magnetic resonance

Nuclear magnetic resonance (NMR) is an important spectroscopic method which can be used to study the conformational properties of nucleic acids in solution. For this purpose 2-dimensional NMR techniques are widely used and are subject of discussion in several reviews [1-4]. The interpretation of the 2D-NMR data can practically be divided into three stages: (1) the assignment of the resonances; (2) the examination of the global conformation; (3) the study of the structural details.

(1) Assignments of the resonances: Both NOESY (nuclear Overhauser enhancement spectroscopy) and COSY (correlated spectroscopy) data sets are used to determine the spectral position of the resonances of the protons. Protons which are close in space (i.e., less than 4.5 Å apart) can exchange magnetization via dipolar relaxation [5]. This dipolar interaction of the protons is a through space interaction and is manifested via cross peak connectivities in a NOESY spectrum. In contrast, the cross peaks observed in COSY spectra [6] originate from coherence transfer which is induced by J-couplings. The efficiency of this coherence transfer is dependent on the value of the J-coupling of the coupled spins. The assignment of the resonances of nucleic acids using both NOESY and COSY [7-9] is described in more detail in Chapter 4.

In addition to NOESY and COSY other NMR experiments can be used to interpret the spectrum. The pulse schemes of some of these experiments are given in Fig. 1. The TOCSY experiment (total correlation spectroscopy) [10,11], also referred to as spin-lock COSY or HOHAHA experiment, can be used. Usually, the spin-lock pulse is replaced by a pulse train ,e.g., MLEV-17 [12]. During the spin-lock pulse (see Fig. 1) transfer of magnetization takes place between coupled spins. However, the magnetization transfer is not restricted to directly coupled spins. Coherence transfer may take place between spins which are not directly coupled, but which are both coupled to a third spin. This results in the occurrence of additional cross peaks. The advantage of this experiment is that the extra cross peaks might be found in a less crowded part of the 2D-spectrum, which simplifies the interpretation. Apart from coherence transfer, in the TOCSY experiment also Overhauser effects (through space magnetization transfer) and chemical exchange might occur. A modification of TOCSY in which these effects are amplified is ROESY

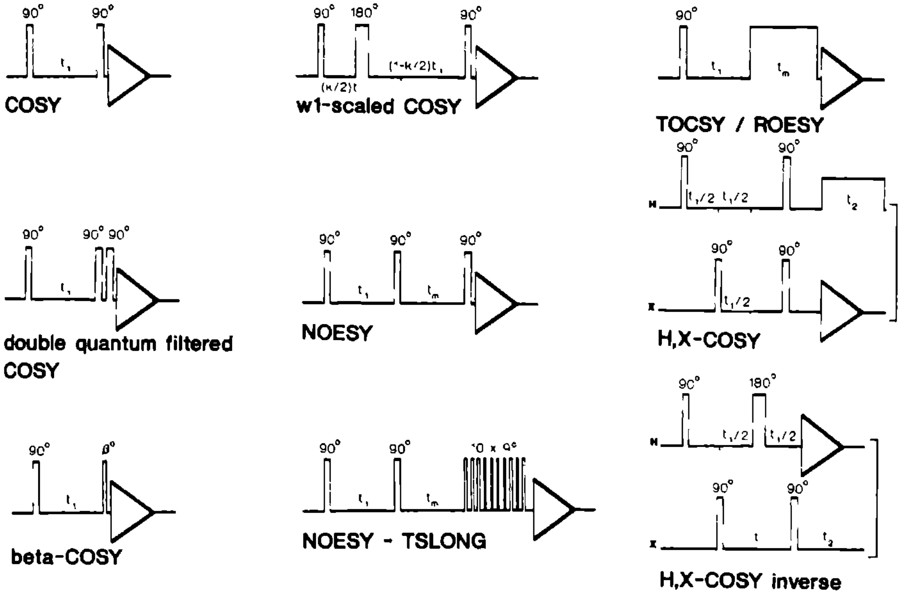

90°
90°
t1
COSY
90°
180°
90°
(1-κ/2)t1
(κ/2)t
w1-scaled COSY
90°
t1
tm
TOCSY / ROESY
90°
90°
H
t1/2 t1/2
t2
90°
90°
X
t1/2
H,X-COSY
90°
90° 90°
t1
double quantum filtered COSY
90°
90°
90°
t1
tm
NOESY
90°
180°
H
t1/2
t1/2
90°
90°
X
t
t2
H,X-COSY inverse
90°
β°
t1
beta-COSY
90°
90°
t1
tm
NOESY - TSLONG

Fig 1. Pulse schemes corresponding to the 2D-NMR experiments which are currently important for the analysis of nucleic acids. The t_1 delay which is increased during the subsequent experiments is indicated. The t_2 time domain corresponds to the free induction decays (FIDs) indicated with triangles. The NOESY, TOCSY and ROESY experiments use a mixing period t_m.

(rotation-frame Overhauser enhancement spectroscopy) [13]. In a ROESY experiment the power of the lock pulse is reduced but its length is increased (Fig. 1). Under these conditions the cross peaks mainly arise from Overhauser effects and chemical exchange. The Overhauser effect in NOESY experiments changes sign in molecules for which $\omega\tau_c \approx 1.1$. At these conditions the NOE cross peaks will be weak or even absent from the spectrum. This is not the case in a ROESY experiment. This results in a higher sensitivity in the measurement of the dipolar interactions. Therefore, the ROESY experiment can be used for the structural determination of small molecules ($\omega\tau_c \approx 1.1$). A second benefit is that possible cross peaks caused by chemical exchange are distinguished by a negative sign from the positive cross peaks originating from dipolar relaxation.

One of the disadvantages of the COSY experiment is the occurrence of the diagonal peaks in the disperse phase. The broad disperse peaks hamper the identification of cross peaks close to the diagonal. In the double-quantum filtered COSY (DQF-COSY) experiment the signals from uncoupled spins (singlets) are canceled [14]. This is brought about by a third pulse which transfers double-quantum coherence into detectable magnetization (Fig. 1). Moreover, the diagonal peaks are now represented in absorption phase. This will increase the resolution near the diagonal.

Another variant of the COSY experiment is the application of ω_1-scaling [18,19]. This improves the efficiency of the coherence transfer between protons that have small couplings. In the standard COSY experiment the intensity of the cross peaks between these protons is weak because the evolution time during the t_1-period is to short to allow optimum coherence transfer. The insertion of a 180° pulse in the t_1-period (Fig. 1) allows a longer evolution period for the same number of t_1-values. As a result, the chemical shifts are scaled down by a factor (1-k) whereas the couplings are unaffected. In addition, a higher resolution of the coupling pattern of the cross peaks is obtained. In this way the fine structure of the H1'-H2' and H1'-H2" cross peaks can be made visible and cab be used to assign stereospecifically the H2' and H2" resonances in S-type sugars [20].

A simplification of the multiplet cross pattern can be obtained by application of E-COSY (exclusive correlation spectroscopy) or β-COSY (small flip-angle correlated spectroscopy) [21]. In these experiments the transfer through directly connected transitions results in the formation of cross peaks, while the signals originating from the other transitions are canceled or relatively small. This reduces the number of peaks within the cross peak multiplet with 50 %.

The investigation of the properties of nucleic acids is often restricted to the analysis of the non-exchangeable protons. Sometimes it is important to study the exchangeable amino- and imino protons as well. One of the methods to measure NOESY spectra of a H_2O sample is the replacement of the 90° acquisition pulse by a semi-selective excitation pulse [22]. This pulse is referred to as TSLONG (time shared long pulse) and consists of ten 9° pulses, separated by a short delay (Fig. 1). The carrier is placed at the edge of the spectrum. In this way the excitation profile will be maximal at the low field part of the spectrum and minimal at the frequency of the H_2O

resonance. No solvent saturation is applied, because in that case the exchangeable protons will be saturated indirectly.

Up to now, the conformational analysis of nucleic acids has been focussed on the use of ^{1}H-NMR studies. For the determination of the structural details of a DNA fragment it is advantageous to study the ^{13}C and ^{31}P nuclei too. Vicinal J-couplings between hetero nuclei are informative with respect to the conformation of the sugar phosphate backbone of the double helix. One of the experiments which can be used to assign the resonances of the ^{31}P or ^{13}C spectrum after assigning the ^{1}H spectrum, is the ^{1}H,X COSY experiment [23], in which X stands for 31P or 13C. In this experiment the magnetization is carried from the proton to the X-nucleus by means of the J-coupling. The transfer of the magnetization is detected on the X nucleus. This experiment has a low sensitivity: The sensitivity of ^{13}C (natural abundance 1,1 %) in a 100% enriched sample is roughly 4 times lower than ^{1}H. Therefore, the experiment is restricted to highly concentrated or ^{13}C enriched samples. An alternative method is given by the so-called inverse experiment [24]. In this experiment the magnetization is transferred from ^{1}H to ^{13}C and subsequently back to the proton. In this case the transfer of magnetization is detected with the sensitivity of ^{1}H-nuclei. The increase of the sensitivity relative to the normal 1-dimensional ^{13}C measurement amounts to 31,6 % for the ideal case.

The introduction of 3D-NMR [25,26] may result in an enhancement of the resolution in comparison with the aforementioned 2D-techniques. The first 3D-NMR experiments applied to biomacromolecules are promising [25-28].

(2) Examination of the global conformation: The presence of weak or strong connectivities in nuclear Overhauser spectra and correlation spectra can be used for the determination of the global conformational features of nucleic acid molecules. Fig. 2 presents various distances between the protons of the mononucleotide unit (Fig. 6). The conformational space available for this mono nucleotide unit is spanned up by the pseudorotational parameters P (pseudorotational phase angle) and ϕ_m (pucker amplitude) and the glycosidic torsion angle χ (see next section); the definition of χ is given in Table I. The sugar to base inter-proton connectivities vary significantly as a function of the sugar conformation and the glycosidic torsion angle. The corresponding simulated NOE intensities have been plotted in Fig. 3 as a

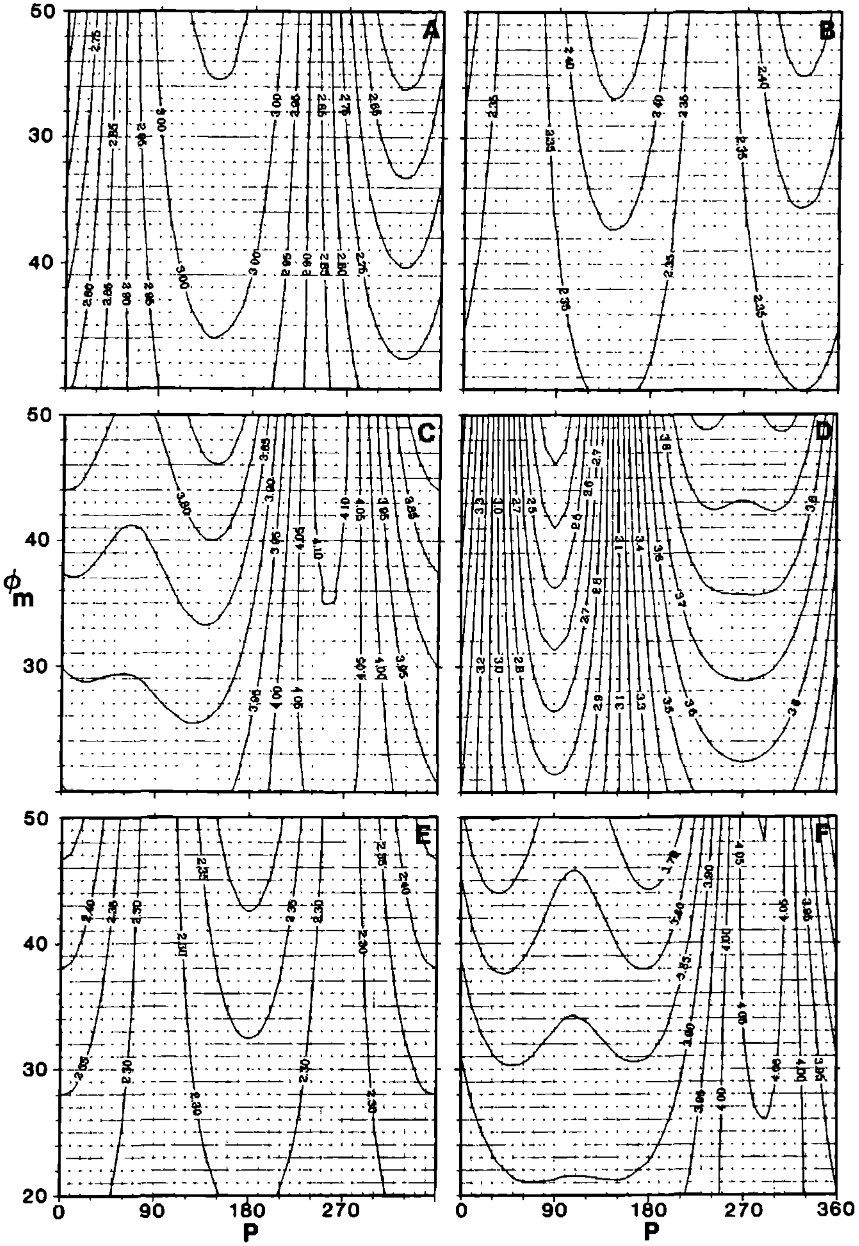
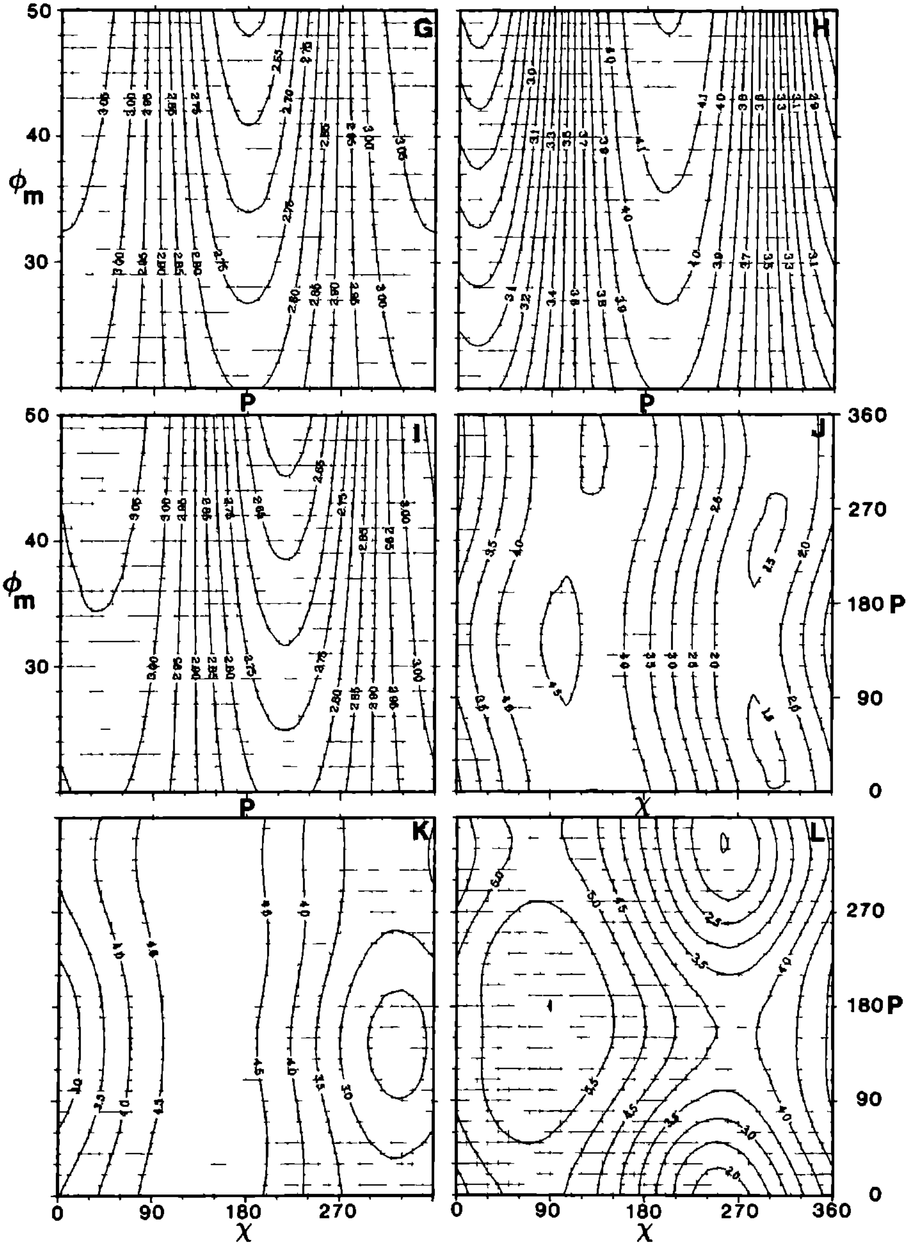
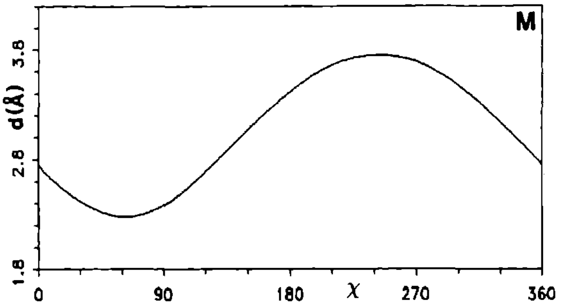


Fig. 2. H-H distances of the conformational space available for the mononucleotide unit of the DNA molecule. Several distances are calculated as a function of the sugar conformation, i.e., the pseudorotational phaseangle, P, and the pucker amplitude, ϕ_m (A-I) and the glycosidic torsion angle χ (J-M): (A) H1'-H2', (B) H1'-H2", (C) H1'-H3', (D) H1'-H4', (E) H2'-H3', (F)



H2'-H4', (G) H2"3', (H) H2"-H4', (I) H3'-H4', (J) H6/H8-H2', (K) H6/H8-H2", (L) H6/H8-H2" and (M) H6/H8-H1'. Note that the contours which connect the equal distances are drawn in steps of 0.05 Å for the H-H distances presented in A, B, C, E, F, G and I. These distances are almost constant and vary only slightly. The H-H distances for H1'-H4' (D) and H2"-H4' (H) are the only



sugar proton distances that apparently vary as function of the sugar conformation. These contours are drawn in steps of 0.1 Å. The most outstanding sugar-base distances are contoured in steps of 0.5 Å (J-L). The H6/H8-H1' distance (M) does not depend on the sugar conformation and is plotted as a function of χ only.

Table I. Definition of the torsion angles in nucleic acids.

α	O3'-P -O5'-C5'
β	P -O5'-C5'-C4'
γ	O5'-C5'-C4'-C3'
δ	C5'-C4'-C3'-O3'
ε	C4'-C3'-O3'-P
ζ	C3'-O3'-P -O5'
χ (Py)	O4'-C1'-N1 -C2
χ (Pu)	O4'-C1'-N9 -C4

ν_0	C4'-O4'-C1'-C2'
ν_1	O4'-C1'-C2'-C3'
ν_2	C1'-C2'-C3'-C4'
ν_3	C2'-C3'-C4'-O4'
ν_4	C3'-C4'-O4'-C1'

function of P and χ. Only the H1'-H4' and the H2"-H4' distances between protons of the sugar ring vary to some extent (cf. Fig. 2) and are useful for the determination of the sugar conformation. J-couplings are much more sensitive for the sugar conformation than most of the NOE intensities (cf. Figs. 2-4). The J-coupling constants of the sugar ring protons have been calculated as a function of P and ϕ_m and are presented in Fig. 4. The J-couplings corresponding to the sugar phosphate backbone are also of interest. Therefore, the ^1H-^1H and ^1H-^{31}P coupling constants, which are a function of the backbone torsion angles β, γ and ε, are presented in Fig. 5.

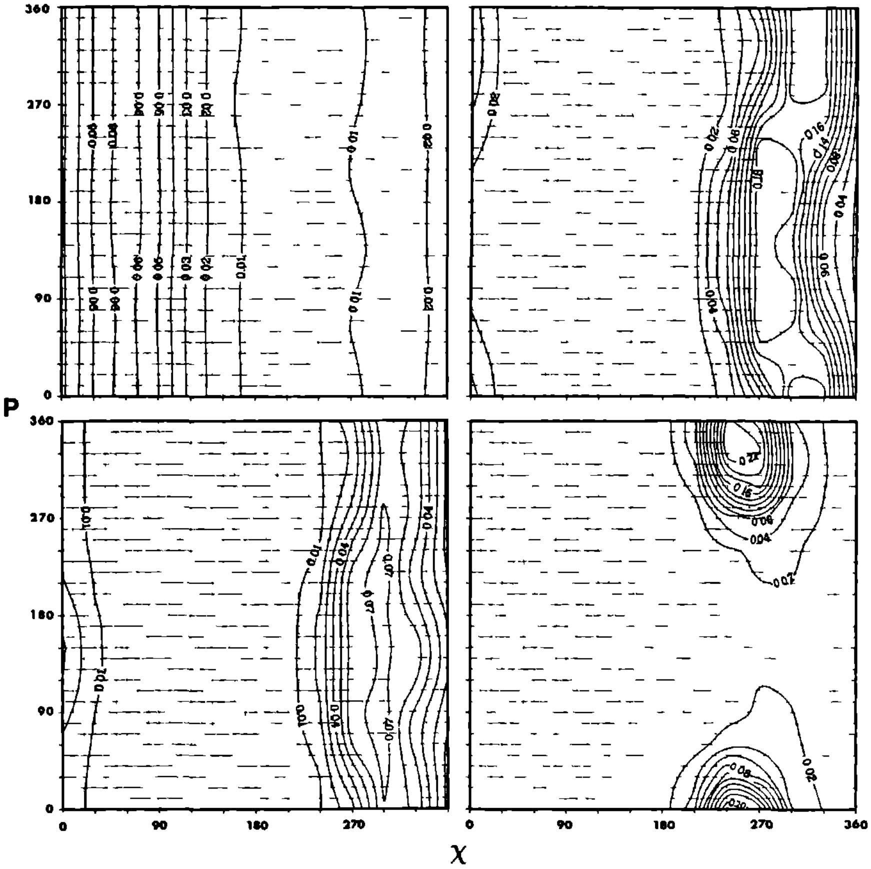


Fig. 3. Contour plots which represent the base-sugar NOE intensities calculated for the conformational space available for the mononucleotide unit. The NOEs have been calculated with the aid of the NOESIM program, which accounts for spin diffusion (see Chapter 7). The NOEs correspond to a molecule with a τ_c of 1.5 ns and corresponds to NOESY spectra recorded with a mixing time of 0.2 s at 500 MHz. (A) H6-H1', (B) H6-H2', (C) H6-H2", (D) H6-H3'. Contours are drawn in steps of 0.01 for A and C. In the plots for H6-H2' (B) and H6-H3' (D) in which the variations of the NOEs are more prominent as compared to A and C, the contours are drawn in steps of 0.02.

The NOEs and J-couplings observed for a particular nucleic acid molecule can be used for a qualitative derivation of the structure. This information permits, for instance, to discriminate between A-, B-, and Z-type DNA. The conformational properties of the A-, B- and Z-helices are summarized in Table II. Using this table in combination with the plots

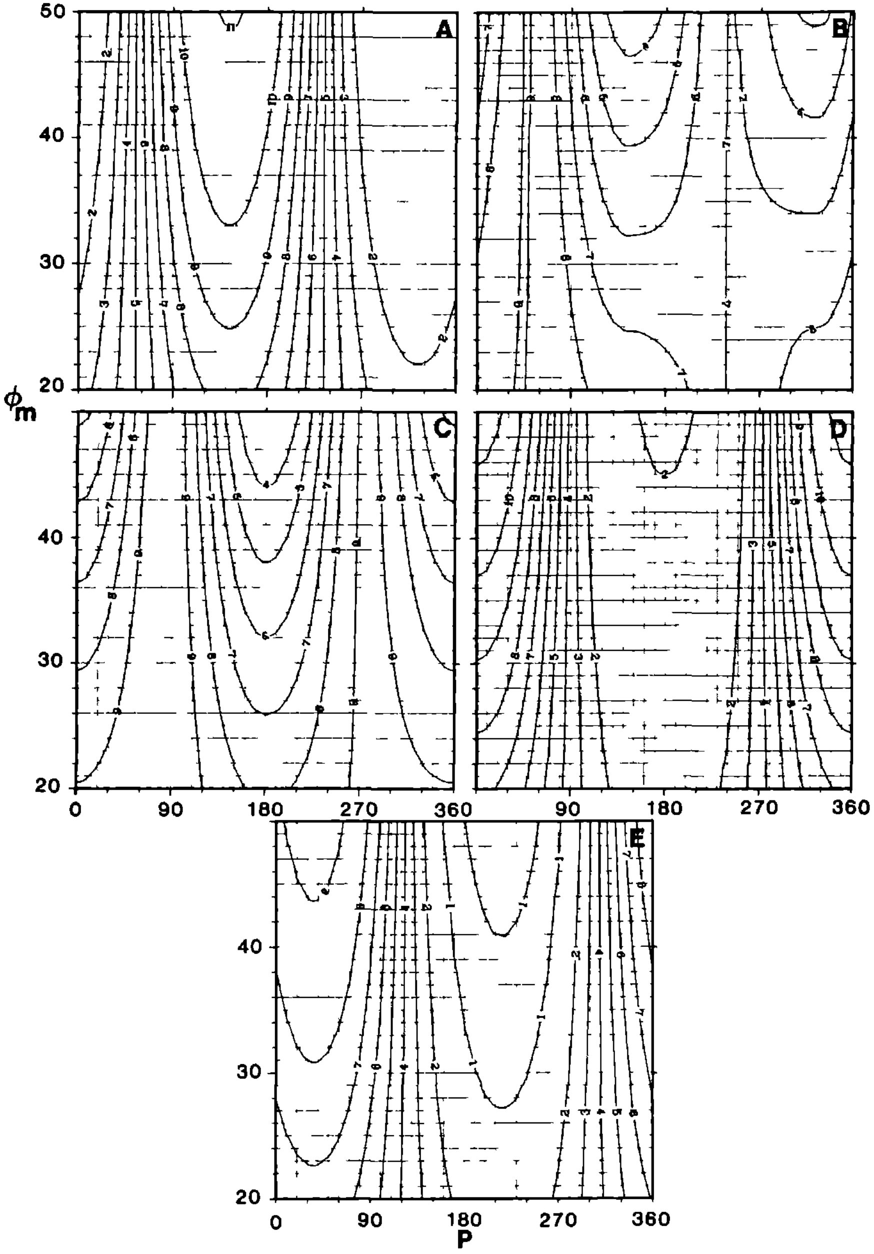


Fig. 4. Sugar ring ^{1}H-^{1}H coupling constants as a function of the conformational space available for the sugar ring. These coupling constants have been calculated with the aid of the EOS Karplus equation and corrected for the Barfield transmission effect (see Chapter 7). Contours are drawn in steps of 1 Hz. (A) $J_{1'2'}$, (B) $J_{1'2''}$, (C) $J_{2'3'}$, (D) $J_{2''3'}$ and (E) $J_{3'4'}$.

Table II. Torsion angles for regular A-RNA, A-, B- en Z-DNA

	α	β	γ	δ	ε	ζ	χ	sugar
A-RNA	294	189	49	95	202	294	202	N-type
A-DNA	285	208	45	83	178	313	206	N-type
B-DNA	314	214	36	156	155	264	260	S-type
Z-DNA G	52	207	178	76	288	102	89	N-type
Z-DNA C	250	192	54	147	257	269	201	S-type

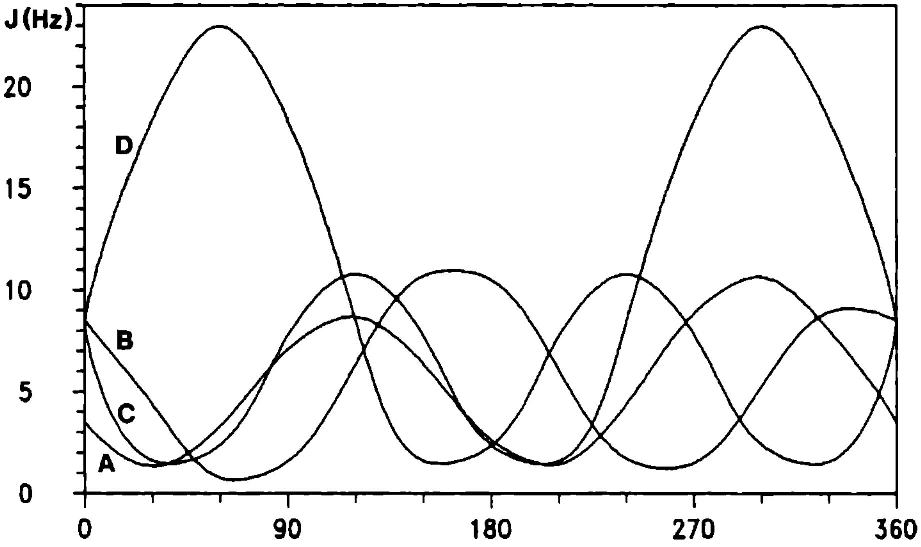


Fig. 5. J-coupling constants calculated for (A) $J_{4'5'}$, (B) $J_{4'5''}$, (C) $J_{5'P}$, (D) $J_{5''P}$ or $J_{3'P}$ as a function of β, γ or ε. $J_{4'5'}$ and $J_{4'5''}$ correspond to γ; $J_{3'P}$ corresponds to ε; $J_{5'P}$ and $J_{5''P}$ correspond to β. The coupling constants were calculated by means of the Karplus equations [131,132] given in Chapter 2 and 7.

presented in Figs. 2-4 one can investigate which particular NOEs and J-coupling constants correspond with an A-, B- or Z-type helical structure.

The A-type helical conformation is characterized by strong intra-nucleotide NOEs between the aromatic H8 or H6 resonances and the H3' resonances (χ^{anti}, N-type sugar; cf. Fig. 3). The coupling constants of the sugar proton resonances $J_{2''3'}$ and $J_{3'4'}$ are strong, while $J_{1'2'}$ is weak (N-type sugar; cf. Fig. 4). The coupling constants of the backbone protons $J_{4'5''}$ are small (γ^{+}; cf. Fig. 5). These data differ from those corresponding to B- or Z-DNA. The inter-nucleotide NOEs between the H6/H8 resonances and the H3' resonances of B-DNA (χ^{anti}, S-type sugar) are much weaker compared with A-DNA (cf. Fig. 3). The cross peaks between the aromatic resonances and the H2' and H2" resonances are strong. The coupling constants of the sugars (S-type) are characterized by small $J_{2''3'}$ and $J_{3'4'}$ coupling constants and a large value for the $J_{1'2'}$ coupling constants (cf. Fig. 4).

The Z-helix follows an alternating pattern for the conformation of the subsequent C and G units (cf. Table II). The G units (χ^{syn}, N-type sugar) are characterized by strong NOEs for the H8-H1' pairs and weak NOEs for the other base-sugar connectivities, whereas for the C units (χ^{anti}, S-type sugar) weak H6-H1' NOEs will be observed. The sugar conformations of the C and G residues can be distinguished on the basis of the $J_{1'2'}$, $J_{2''3'}$ and $J_{3'4'}$ coupling constants (see above). The coupling constant $J_{4'5''}$ has a large value for the G-part (γ^{t}) and a small value for the C part (γ^{+}).

It should be emphasized that most of the DNA and RNA species are not conformationally pure. Therefore the NOE and J-coupling data can only be used for a detailed structure elucidation if all experimental data fit to one structure (in that case the conformation is rigid).

(3) Examination of the details of the structure: The details of nucleic acid structures can be studied after collection of NOE cross peak intensities and vicinal J-coupling constants. At present, several methods are used to translate these data into a structure. In one approach the initial NOE build-up rates are used for a direct determination of the distances. In this determination a well-known molecular distance, r_{ref}, is used as internal reference (r_{ref}). The unknown distance, r_{ij}, between spin i and j is calculated using Eq. 1, in which this distance relates to the reference distance according to [29]:

$$r_{ij} = r_{ref} (\sigma_{ref} / \sigma_{ij})^{1/6} \qquad (1)$$

in which σ_{ij} stands for the cross relaxation rate between proton spin i and j, the slope of the NOE buildup curve at t = 0. The approximations introduced into the relaxation theory to derive Eq. 1 are discussed in more detail in Chapter 7. Here it is just mentioned that the use of Eq. 1 may introduce severe errors in the distance determination [30] because the relaxation in large molecules is not limited to pairs of spins. It is better to include all protons of the molecule in the considerations. This can be achieved by a calculation of the complete relaxation matrix. After diagonalizing this matrix the cross peak intensities can be calculated (cf. Chapter 7, Eqs. 6-10). This procedure can been applied to an isotropically tumbling molecule [31,32], but the effects of internal motions can also be added with the aid of jump-models [33,34] or with a model-free approach [35].

Torsion angles of the sugar-phosphate backbone and sugar conformations can be estimated from an evaluation of coupling constants, using parametrized Karplus equations [36] (Chapter 2, Eq.4; Chapter 7, Eq. 1-4; cf. Figs. 4-5). When the structure is not conformationally pure, the experimental couplings reflect the weighed average of the couplings corresponding to the different conformations involved.

The structural data, i.e., distances and torsion angles, can be used to derive a 3-dimensional model. Currently distance-geometry, restrained molecular dynamics/mechanics [36-38] and simulated annealing are powerful methods to achieve this [39-41]. However, in spite of the widespread application of these methods, the sampling properties of these methods [42] are of concern [30]. In this thesis multi-conformational analysis is applied. This method provides a proper sampling of the conformational space, when the distances between the sample points are not too large. In practice, this is connected with cpu-times which may become prohibitively long. Recently, the explorative properties of the multi-conformational analysis and the exploitative properties of simulated annealing have been combined using the concept of the genetic algorithm [43,44]. Currently this approach is under development and is successful for small molecules up to 500 D or parts of larger molecules.

However, the analysis of DNA hairpins which appeared in the literature has by necessity often been restricted to model building (and subsequently molecular mechanics/dynamics refinement) since sufficient NMR data were mostly lacking. In spite of the applications of the above model building methods, it is important to realize which structural details can be determined by means of the experimental (NMR) data. In this context it must be recognized that ^1H-^1H distance constraints alone, give a poor description of the structural parameters of nucleic acids [30,42].

Description of the conformation of nucleic acids.

DNA is a linear polymer of nucleotides [45]. Each nucleotide unit of this molecule consists of a deoxy-ribose sugar ring, a phosphate group and a base. The base can be adenine, guanine, thymine or cytosine. The unit with an adenine base is shown in Fig. 6, where the numbering of the sugar protons is given. The definition of the torsion angles (Fig. 6) determining the nucleic acid conformation are collected in Table I. The values of these torsion angles are usually expressed in terms of the classical rotamers gauche$^+$, trans or gauche$^-$ (see Figure 7A). Newman projections for the classical rotamers of the torsion angles β, γ and ϵ are given in Fig. 8.

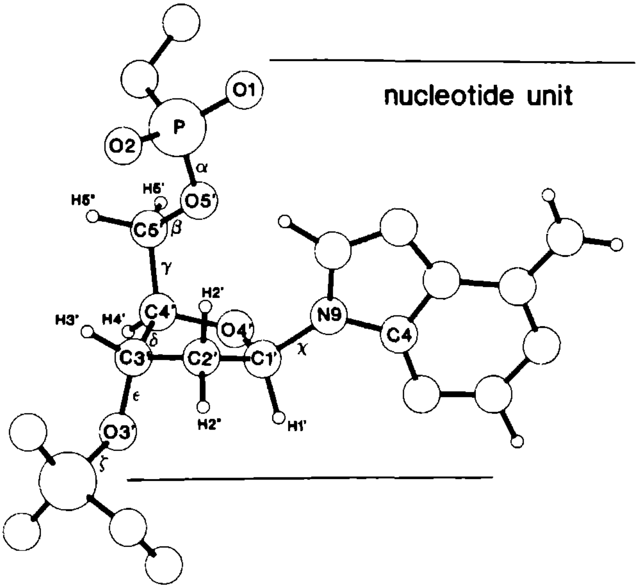


Fig. 6. The nomenclature used for the description of the torsion angles and atoms of the nucleotide unit in DNA. See also Table I.

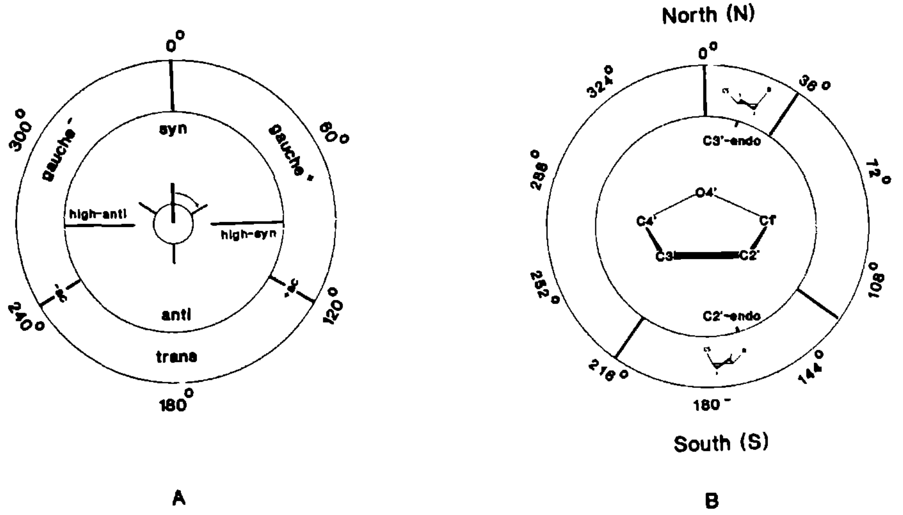


Fig. 7. Schematical representation of the torsion angles (A) and sugar conformations (B). (A) The conformation of a torsion angle is often referred to as gauche$^+$ (around 60°), trans (around 180°) or gauche$^-$ (around 300°). The Newman projection in the center indicates a gauche$^+$ conformation. When the torsion angle is around 120° or -120° the conformation is referred to anticlinal (ac). For the orientation of the base, the syn-trans nomenclature is used, according to the position of the base relative to the sugar. Then the torsion angle χ adopts the values indicated. (B) The sugar conformation can be expressed in terms of pseudorotational phase angle, P, and pucker amplitude, ϕ_m. The sugar conformations are often represented in a pseudorotation wheel shown here. Regions are indicated for which most sugar conformations are experimentally observed. These conformations share the N- and S- domain of the conformation wheel. The twist conformations with P = 0° and P = 180° have been indicated in these regions.

The conformation of the five-membered sugar ring can be described by means of two parameters: the pseudo-rotational phase angle, P, and the pucker amplitude, ϕ_m. P can be calculated from the endocyclic torsion angles ν_i (see Table I) according to [46]:

$$P = \arctan \frac{(\nu_4 + \nu_1) - (\nu_3 - \nu_0)}{2\,\nu_2\,(\sin 36° + \sin 72°)} \qquad (2)$$

The phase angle P = 0° is chosen such that the torsion angle ν_2 has a

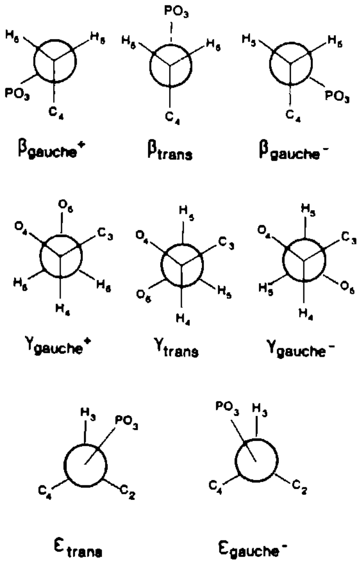


Fig. 8. Newton projections for the allowed staggered conformations apparent for the sugar phosphate backbone torsion angles β (P-O5'-C5'-C4'), γ (O5'-C5'-C4'-C3') and ε (C4'-C3'-O3'-P).

maximum value. The five endocyclic torsion angles are related to the pseudorotational parameter P and pucker amplitude ϕ_m by:

$$\nu_j = \phi_m \cos (P + 144° \, j) \qquad j = 0 - 4 \tag{3}$$

The backbone torsion angle δ, is strongly correlated to the sugar conformation [36]:

$$\delta = 120.6° + 1.1 \, \phi_m \cos (P + 145.3°) \tag{4}$$

The ribose and deoxyribose sugar conformations of a representative number of crystal structures are clustered around C3'-endo and C2'-endo geometry, respectively [47]. The corresponding ranges for the pseudorotational phase angle are indicated in Fig. 7B. Because it concerns conformations in the upper and lower part of the pseudorotation circle, the two classes are often referred to as N-type (North-type) and S-type (South-type) sugars. The furanose sugars have also been subjected to molecular mechanics calculations [48,49]. Two energy minima were found, which correspond to the N-type and S-type conformations observed in the crystal structures. According to the calculations the energy barrier between the two conformers is only a few kcal. This suggests that the two conformations may interconvert rather easily. The dynamics of the (deoxy-)ribose ring has further been examined with the aid of NMR spectroscopy. The determination of accurate values for the J-couplings of the sugar ring protons makes an estimation of the sugar conformation(s) feasible. It appeared that the coupling constants are better described with a two-state model, which accounts for the N-S equilibrium, than by one pure conformation which deviates significantly from the N- or the S-state. This is demonstrated in Fig. 9 where different values of the $J_{3'4'}$ and $J_{1'2'}$ coupling constants for possible values of P (with ϕ_m equal to 30, 35 and 40) are plotted. In addition, the values of a representative number experimental coupling constants [30] are introduced by means of dots. These experimental data scatter around a line, which connects the P-values calculated for a pure N- and S-type sugar. In this thesis the 'two state model' is adhered to although it is not generally accepted.

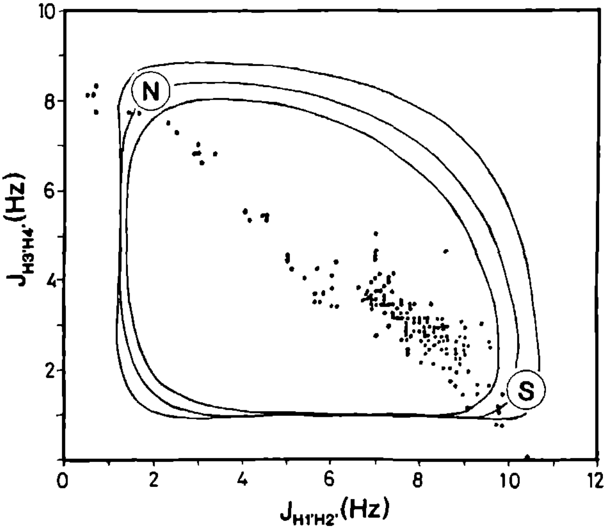


Fig. 9. The coupling constants $J_{1'2'}$ and $J_{3'4'}$ are calculated as in Fig. 4 for all values of P and for ϕ_m = 30, 35 and 40 respectively. This results in the presented three curves. The coupling constants connected to the P = 0° and P = 180° are indicated with N and S. The experimentally observed coupling constants available from the literature are presented here as dots. These dots do not fall on the curves, but scatter around a line which connects the coupling constants of N- and S-type conformations. This strongly suggests that the sugar conformation exhibits an equilibrium between N- and S-type conformations.

The experimental J-coupling constants which monitor the conformation of the backbone torsion angles β, γ and ε also suggest that these angles are not conformationally pure [36]. Therefore, the experimental coupling constant, J_{exp}, represent the sum of the coupling constants corresponding to the various rotamers:

$$J_{exp} = \sum_i x_i J_i \qquad (5)$$

In this equation x_i and J_i represent the molar fraction and the coupling constant of the rotamer i, respectively.

The intrinsic complementarity of the bases in the DNA helix is important for the function of DNA [50,51]. Figs. 10A and 10B show the geometry [52,53] of the most common base pairs: A-T and G-C. These base pairing schemes were first discovered by Watson and Crick [54]. It is noticed that the angle between the drawn dashed lines between both C1' protons of both residues of a Watson-Crick base pair and the C1'-N bonds are almost equal. Therefore a pseudo-axis of symmetry exists, through which inversion of the base pair in the double-helix is possible. It will be shown in Chapter 5 that this pseudo-symmetry axis is lost when the base pair is not flat but buckled. Besides Watson-Crick base pairs other types of base pairing have been found, e.g., the so-called Hoogsteen pairing which was discovered with the aid of X-ray diffraction. The hydrogen bonding scheme of such an alternative A-T pair is shown in Fig. 10C [55]. Because the angles between the line which connects both C1' atoms and the lines through the C1'-N bonds differ substantially [56] from each other, a pseudo-axis of symmetry is not present in this base pair. This type of pairing is found in the hairpin formed by d(ATCCTA-TTTA-TAGGAT) (see Chapter 6 and 7).

DNA hairpins

Hairpins occur frequently in RNA. They are recognized as one of the basic units of the secondary structure of RNA. E.g., the tRNA molecule folds up in a L-shaped tertiary structure in which three loops are formed. In spite of the double-stranded nature of DNA, there is increasing evidence that hairpins may also be formed in DNA and that they have interesting biological implications.

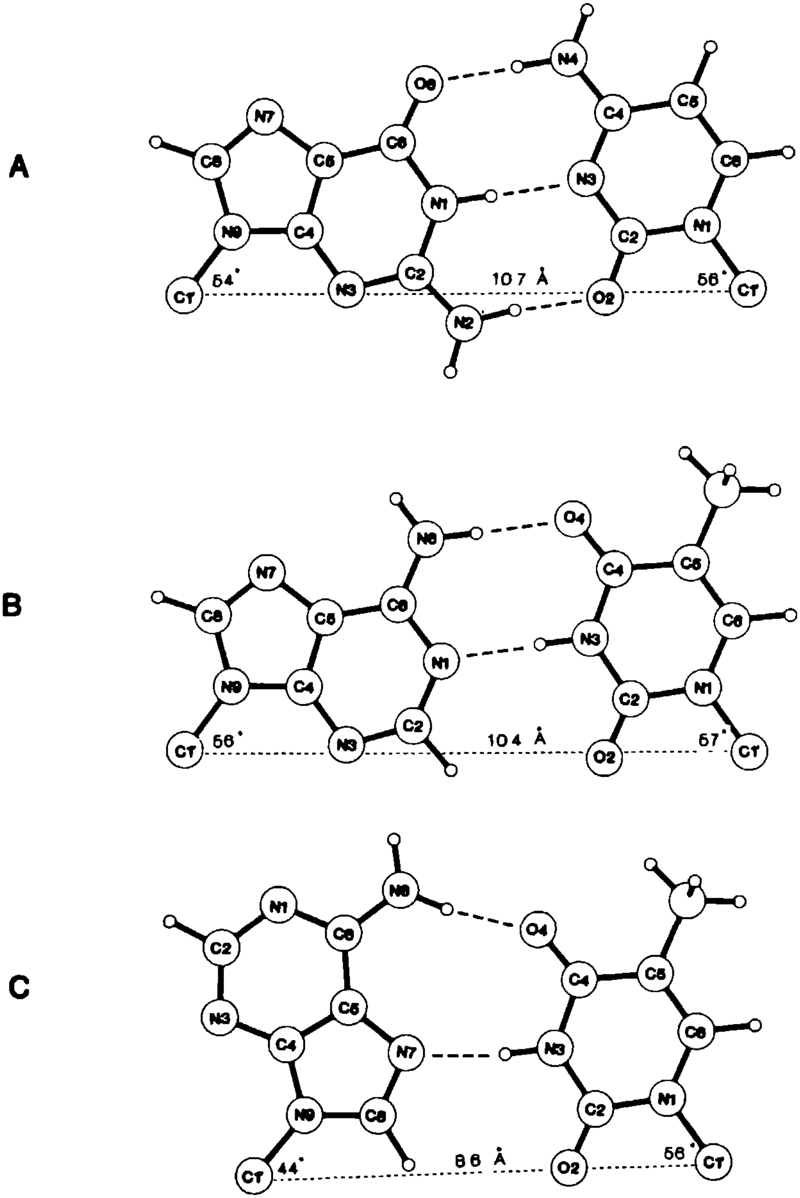

A
O6
N4
N7
C6
C4
C5
C8
C5
N3
C6
N1
N9
C4
C2
N1
C1'
54°
N3
C2
10.7 Å
O2
56°
C1'
N2
B
N6
O4
N7
C4
C5
C8
C5
C6
C6
N1
N3
N9
C4
C2
N1
C1'
56°
N3
C2
10.4 Å
O2
57°
C1'
C
N6
N1
O4
C2
C6
C4
C5
C5
N3
C6
C4
N7
N3
N9
C8
C2
N1
C1'
44°
8.6 Å
O2
56°
C1'

Fig. 10. Standard Watson-Crick A-T (A) and C-G (B) pair and the A-T (C) Hoogsteen pair. The distances between the C1' atoms crossing the minor groove are given as well as the angles between the (dashed) line which connect the C1'-atoms and the C1'-N bond.

Sequence	Reference
d(CGCGCG-TTTT-CGCGCG)	[65-67]
d(CGCGCGCGCG-TTTT-CGCGCGCG)	[68]
d(CGCG-TTTT-CGCG)	[26,69,70]
d(CGCGC-GATC-GCGCG)	[71,72]
d(CGCGC-GTAC-GCGCG)	[71,72]
d(CGCGCG-TTTTT-CGCGCG)	[71,73]
d(CGCGCG-AAAAA-CGCGCG)	[71]
d(CGCG-TATA-CGCG)	[74]
d(CGC-GATTC-GCG)	[75]
d(CGCG-AATT-CGCG)	[76]
d(CGCG-TTGTT-CGCG)	[77,78]
d(CGCG-TTTGT-CGCG)	[79]
d(CG-CGTG-CG)	[81-83]
d(CG-CGAG-CG)	[80]
d(CG-CTAG-CG)	[84,85]
d(CGC-CGCA-GC)	[86,87]
d(CGC-CGTA-GC)	[86]
d(GC-GAAA-GC)	[88]
d(..TATA-TATA-TATA..)	[89,90]
d(ATCC-TATA-GGAT)	[91-95]
d(ATCC-TATTA-GGAT)	[91-95]
d(ATCCT-ATTT-AGGAT)	[91-95]
d(ATCCT-ATTTT-AGGAT)	[81-95]
d(ATCCTA-TTTT-TAGGAT)	[91,92-97]
d(ATCCTA-TTTTT-TAGGAT)	[91-95]
d(ATCCTA-TTTTTT-TAGGAT)	[91-95]
d(ATCCTA-TTTTTTT-TAGGAT)	[91-95]
d(ATCCTA-CTTG-TAGGAT)	[98-100]
d(ATCCTA-GTTC-TAGGAT)	[98-100]
d(ATCCTA-TTTA-TAGGAT)	[98-101]
d(ATCCTA-ATTT-TAGGAT)	[98-100]
d(ATTACG-CTTG-CGTAAT)	[100]
d(ATTACG-GTTC-CGTAAT)	[100]
d(ATTACG-TTTT-CGTAAT)	[100]
d(ATCCTA-TAAA-TAGGAT)	[100]
d(ATCCTA-AAAA-TAGGAT)	[100]
d(ATCCTA-AAAC-TAGGAT)	[100]
d(ATCCTA-AACA-TAGGAT)	[100]
d(ATCCT -AAAT -AGGAT)	[100]
d(ATCCTA-AAAAAA-TAGGAT)	[100]
d(ATCCTA-CTGG-TAGGAT)	[102]
d(ATCCTA-CGTG-TAGGAT)	[102]
d(ATCCTA-CAAG-TAGGAT)	[102]
d(ATCCTA-GAAA-TAGGAT)	[102]
d(ATCCTA-GTTA-TAGGAT)	[102]
d(CGAACG-TTTT-CGTTCG)	[103]
d(CGAACG-AAAA-CGTTCG)	[103]
d(CGAACG-CCCC-CGTTCG)	[103]
d(CGAACG-GGGG-CGTTCG)	[103]
d(CTCCTC-TTGTT-GAGGAG)	[79]
d(CGCGTA-CGCG-TACGCG)	[104]
d(CCA-ATTT-TGG)	[105]
d(CCA-TTTT-TGG)	[105]
d(CCAA-TTT-TTGG)	[105]
d(GAATTC-TTT-GAATTC)	[106]
d(GAATTC-TTTT-GAATTC)	[106]
d(GAATTC-AAAA-GAATTC)	[106]

Table III. DNA hairpins that have been subject to NMR and/or UV-melting studies.

Inverted repeats or palindrome sequences are often found in non-coding regions of the genome. It has been proposed that these sequences in double-stranded DNA may be converted into a cruciform structure, in which two hairpins are present [57]. These cruciform structures might be a mechanism to relax torsional stress in negatively supercoiled plasmids. The formation of cruciforms has indeed been demonstrated in synthetic as well as in natural circular DNAs [58,59].

Hairpin formation may also be functional in species, which carry the genetic information in a single stranded DNA molecule, e.g., in single stranded phages. The DNA of various filamentous phages, among them the phage M13, has been sequenced and the complete intergenic region of these phages carries five distinct hairpin domains. The origin of replication is located in one of these putative hairpin loops. Recently, it has been shown that a 51-mer oligonucleotide, harboring the sequence around the origin of replication of bacteriophage M13 indeed folds into a higher order structure in which two hairpins are apparent [60].

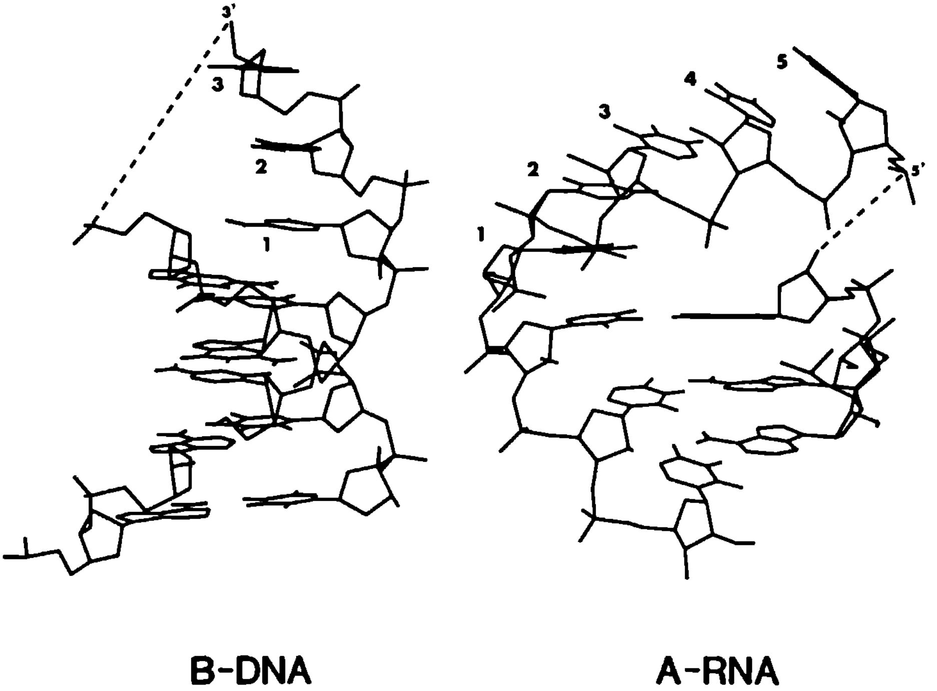
Sequences of purines in one strand and pyrimidines in the other occur frequently in prokaryotes as well as in eukaryotes. One half of the polypurine tract and the center of the polypyrimidine are often associated with S1-nuclease sensitive sites under supercoiled pressure, which indicates that these sequences then adopt a special structure. The formation of this structure is supported when the pH is lowered. This sensitivity for S1-nuclease and pH can be explained when the DNA adopts a triple-helical stem-loop structure, also referred to as H-DNA [61,62]. In this structure the double-stranded polypurine-polypyrimidine is converted into a hairpin, a single-stranded polypurine and a polypyrimidine strand which is associated to the major groove of the hairpin stem.

Guanine-rich sequences, such as d(TT-GGGG-TT-GGGG) are found in the telomeric ends of eukaryotic chromosomes. Oligonucleotides with these sequences dimerize to form a quadruplex consisting of two hairpin loops; the stem of the hairpins are associated to form an `antiparallel quadruplex, containing guanine base tetrads [63,64].

During the last 10 years the structures and thermodynamic properties of DNA hairpins have been studied extensively [30]. Table III gives an overview of the oligonucleotides which have been investigated with the aid of NMR or by means of melting experiments. The availability of synthetic oligonucleotides in mg amounts [107] has given the DNA-hairpin studies an enormous stimulus. The biophysical investigations of DNA hairpins by various groups is currently concentrated on two questions: what is the optimal number of nucleotides in a hairpin loop and how can the folding, i.e., structure, of the hairpin loop been characterized. At this stage a simple answer to both of these questions is not available. This is caused by the fact that the detailed loop structure depends on the loop size and on the nucleotide sequence of the hairpin.

Studies on loop folding

Hairpins consist of a double-helical stem closed by a single-stranded loop region. The melting of the hairpin induces hyperchromicity effects. The thermodynamic parameters of the hairpin-to-coil transition can be derived from the UV-absorbance-temperature curve. Pioneering studies have been carried out by Baldwin and co-workers, who investigated the melting behavior of $d(AT)_n$ hairpins [89,90]. It was found that these hairpins form a loop of four nucleotides. Haasnoot et al. studied a series of hairpins formed by $d(ATCCTA\text{-}T_n\text{-}TAGGAT)$, in which n = 1-7 [91]. It was concluded that, at least for this series of hairpins, optimal stability was reached when the loop consists of four or five nucleotides. These results were in contrast with those available at that time for RNA hairpins. For the series $r(AAAAAA\text{-}U_n\text{-}UUUUUU)$ in which n= 4, 5, 6 and 8 [108] and $r(AAAAG\text{-}C_n\text{-}UUUU)$ in which n = 4, 5 and 6 [109] an optimal stability was reported for hairpins with a loop of six or seven nucleotides. To explain these differences in stability, it was assumed that the double-helical stem of a RNA hairpin has an A-type helical structure whereas its counterpart in the DNA hairpin has a B-type helical structure [110]. If one then tries to build a loop on one end of ,e.g., an A-stem and one extends its stacking pattern at the 5'-end with five nucleotides a small gap remains between the 5'-phosphate of the first and the 3'-phosphate of the other strand. This is illustrated in Fig. 11. This gap can easily been closed by one or two nucleotides. This architecture of an A-type hairpin loop structure is indeed found in the crystal structure of the anticodon loop of the tRNA molecule, where the folding is essentially the same (cf. Fig. 12). Extension of the 3'-end of the A-type double helix in an A-type single-helical fashion does not lead to the necessary reduction of the inter-strand phosphate distance (cf. Fig. 1, Chapter 5). The base



B-DNA

A-RNA

Fig. 11. A structural model of an A-RNA hairpin can be obtained when an A-type double helix is extended at the 5'-end with five residues in an A-type helical fashion. In that case, a gap of 11 Å remains to the other strand, indicated with the dashed line which can be closed by two nucleotides. In contrast when the other strand, i.e., the 3'-end, of a B-type double helix is extended with three nucleotides, then the gap (dashed line) to the other strand obtains a minimum value of 11 Å.

stacking extension of the 5'-end, which shortens the inter-strand phosphate distance in A-RNA, does not lead to a reduction of the inter-strand gap in DNA hairpins with a B-type stem. In contrast, there the 3'-end of the B-type double helix has to be extended with (only) two or three nucleotides, to obtain a distance between the 3'-phosphate and the 5'-phosphate of the other strand which is short enough to close the remaining gap with one or two nucleotides [94,110]. This is shown in Fig. 11. The 2D-NMR data obtained at that time for the DNA hairpin formed by d(ATCCTA-TTTT-TAGGAT) corroborate this type of loop folding. Thus, the observed differences between the DNA and RNA hairpins can be explained by a simple loop folding principle, which assumes that the A- and B-type stacking pattern of at least one strand of the double helix is propagated into the loop. Hence, the structure of a hairpin loop is dictated by the geometrical constraints imposed by the A-RNA or B-DNA helix [94]. The folding principle is discussed more extensively in Chapter 5, where it is proposed that the details of the folding may depend on the nucleotide sequence of the loop or even of the stem. Investigations which were conducted to verify and further refine the loop folding principle are presented in this thesis.

Along with these investigations two hairpins, i.e., d(CGCG-TTTT-CGCG) and d(C_mGC_m-GT-GC_mG), have been studied in other laboratories with the aid of NMR spectroscopy [69,81] and detailed models for these hairpins have been proposed.

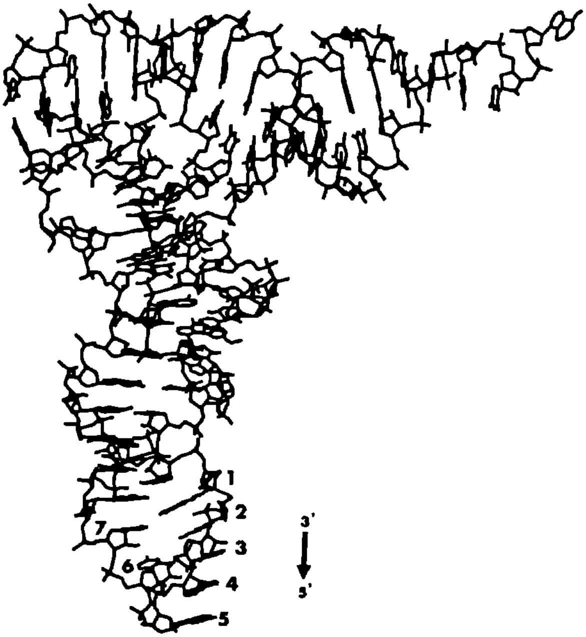


Fig. 12. The structure of yeast $tRNA^{Phe}$ determined with the aid of X-ray diffraction [130]. The anticodon loop, at the bottom of the presented model, represent an example of loop folding in an A-RNA hairpin. Five bases in the anticodon loop, i.e., -$AYAAG_m$- which are numbered 1-5, follow the stacking pattern of the stem on top of the 5'-end of the double-helical stem. Then the remaining gap is bridged by the two remaining nucleotides, i.e., -UC_m- which are numbered 6 and 7.

The hairpin formed by d(CGCG-TTTT-CGCG) has been investigated by means of NOESY [69]. The resonances of the base and sugar protons, excluding the H4', H5' and H5" resonances, were assigned. Buildup rates of the cross peak intensities were obtained. Inter-proton distances were calculated using Eq. 1 of this chapter, using the cytidine H5-H6 proton pair (for which the distance is known) as a reference. It is noted that this procedure does not account for spin diffusion. The obtained inter-proton distances were used subsequently for a determination of the hairpin structure with the aid of distance-geometry calculations. The structural details are subject to discussion because some of the structural parameters are not realistic on physical grounds, but the global folding of the loop may be correct. The third and fourth thymidine in the loop stack on the 5'-end of the stem. In the same way the stacking-pattern of the stem is continued at the 3'-end with one nucleotide. The sugar of the second loop thymidine makes the turn and joins the first and third residue. The stacking of the bases in the loop is different from that proposed for the aforementioned hairpin formed by d(ATCCTA-TTTT-TAGGAT). Therefore the folding seems to deviate from the prediction. However, as a result of the studies described in this thesis, the formulation of the loop folding principle is further extended and refined. Then, the structure of the hairpin formed by d(CGCG-TTTT-CGCG) can be explained.

The oligonucleotide d(C_mGC_m-GT-GC_mG), in which the cytidines are methylated at C5, forms a 'mini-hairpin' with the base stacking pattern, which is predicted by the loop folding principle [81]. The hairpin forms a loop of only two nucleotides. The existence of two-membered loops seems to be at variance with the earlier statement that DNA hairpins have an optimum loop size of four or five nucleotides. This loop can be considered as a four-membered loop in which the first loop base (cytidine) and the fourth loop base (guanine) have adjusted their position to enable base pairing. This interpretation is supported by the observation that some hairpins with the potential to form a two-membered loop adopt a four-membered loop structure (cf. Chapters 5 and 6). The ^1H spectrum of the mini-hairpin was assigned completely. The conformations of the sugar moieties were derived from the coupling constants of the sugar ring protons. All sugars have predominantly a S-type conformation, even those of the loop. The cross peak connectivities in the NOESY spectrum were used to derive a structural model.

The stacking pattern of the B-type stem is propagated at the 3'-end of the stem. The turn in the propagation of the phosphate backbone between the residues T5 and G6 is accomplished with a change in the torsion angles β and γ. It could be derived from the coupling constants that β and γ adopt a gauche$^+$ and trans conformation, respectively, instead of the γ^+ and β^t conformation usually occurring in B-DNA.

As was pointed out before the details of the loop structure will depend on the structural optimizations which are allowed within the predicted loop folding, and thus depend on the nucleotide sequence. This explains for instance the small loop sizes for some B-DNA [81,100] and A-RNA [111-114] hairpins. A more extended picture of the folding of hairpins follows from the studies described in this thesis.

Outline of this thesis.

The principle aim of the investigations presented in this thesis is acquiring of insight in the structure and thermodynamic properties of DNA hairpins.

The hairpin studies are introduced after a thorough investigation of the cyclic dinucleotide d<pApA>, in which the 5'-end of the dinucleotide is connected to the 3'-end to obtain a cyclic sugar-phosphate backbone (Chapter 2 and 3). Cyclic dinucleotides are synthetic compounds, which were synthesized initially with the aim to study loops, i.e., turns in nucleic acids. It was discovered in a later stage that one of the dinucleotides, r<pGpG>, is the biological activator of the enzyme cellulose synthase in Acetobacter xylinum. Moreover, this class of dinucleotides can inhibit the DNA-dependent RNA-polymerase at the initiation stage. NMR homo- and hetero coupling constants were used for the structural analysis of this molecule. The conformation could be derived with atomic resolution. The results were compared with that of a parallel X-ray diffraction study. It appeared that the results corroborate each other. The cyclic dinucleotide is particularly interesting in relation to loop folding, because the phosphate backbone makes a complete turn. Part of the turn in the phosphate backbone appears to coincide with that of the hairpins studied in Chapter 4 and 7. Most intriguing is the preference for N-type deoxyribose conformations in this molecule. Chapter 3 describes an investigation of the conformation of the cyclic dinucleotide, studied in Chapter 2, with multiconformer analysis. By means of molecular mechanics calculations several local energy conformations were generated, the conformers were compared and the structural properties of the molecule were discussed in detail.

Chapter 4 describes the analysis of the hairpin formed by d(ATCCTA-TTTT-TAGGAT) with the aid of NMR spectroscopy. Information from NOEs and J-coupling constants were used as starting points for model building studies. In the presented model the stacking pattern of the bases in the stem is propagated at the 3'-end of the stem into the loop. The fourth base is turned inside to form a wobble T-T pair with the first base in the loop. Using multi-conformational analysis the detailed conformation of the sharp turn between the third and fourth residue could be derived.

In Chapter 5 a general loop folding principle is discussed, which gives a good explanation for the overall loop structure of the DNA hairpin studied in the previous chapter. Moreover, the folding principle explains also the difference in structure and loop size of DNA- versus RNA-hairpins. Molecular modelling studies, in which the influence of the base-sequence on loop folding was examined, presented in this chapter, formed the basis for the experiments performed in Chapter 6. It was found that when complementarity is introduced in a loop of four nucleotides, an additional base pair is sometimes formed. This happens to be the case, e.g., for the loop sequence -CTTG- and -TTTA- and not for the sequence -GTTC- or -ATTT- (Chapter 6). Introductory 2D-NMR experiments on the hairpin formed by d(ATCCTA-TTTA-TAGGAT) showed that the base pair between the first loop thymidine and the fourth loop adenine is of a Hoogsteen type.

In Chapter 7 a method is described along which the experimental NOEs and coupling constants generate constraints for structure determination by means of multi-conformation analysis. This was applied to the hairpin formed by d(ATCCTA-TTTA-TAGGAT). A detailed conformation of the hairpin loop was obtained at atomic resolution. The structure is discussed in detail with respect to the folding principle described in Chapter 4 and 5.

REFERENCES

1 Ernst, R.R, Bodenhausen, G. & Wokau, A. (1987) "Principles of nuclear magnetic resonance in one and two dimensions." 1th ed., Clarendon press, Oxford.

2 Wühtrich, K. (1986) "NMR of proteins and nucleic acids." 1th ed. Wiley and Sons, New York

3 Ernst, R.R. (1987) Chimia **41**, 323-340.

4 Kessler, H., Gehrke, M. & Griesinger, C. (1988) Angew. Chem. **100**, 507-554.

5 Macura, S & Ernst, R.R. (1980) Mol. Phys. **41**, 95-117.

6 Aue, W.P., Bartholdi, E. & Ernst, R.R. (1976) J. Chem. Phys. **64**, 2229-2246.

7 Scheek, R.M., Russo, N., Boelens, R., Kaptein, R. & van Boom, J.H. (1983) J. Am. Chem. Soc. **105**, 2914-2916.

8 Hare, D.R., Wemmer, D.E., Chou, S.-H., Drobny, G. & Reid, B.R. (1983) J. Mol. Biol. **171**, 319-336.

9 Haasnoot, C.A.G., Westerink, H.P., van der Marel, G.A. & van Boom, J.H. (1983) J. Biomol. Struct. Dyns. **1**, 131-149.

10 Braunshweiler, L. & Ernst, R.R. (1983) J. Magn. Res. **53**, 521-528.

11 Davies, D.G. & Bax, A. (1985) J. Am. Chem. Soc. **107**, 7197-7198.

12 Bax, A. & Davies, D.G. (1985) J. Magn. Res. **65**, 355-360.

13 Bothner-By, A.A, Stephens, R.L., Lee, J., Warren, C.D., Jeanloz, R.W. (1984) J. Am. Chem. Soc. **106**, 811-813.

14 Piantini, U., Sørensen, O.W. & Ernst, R.R. (1982) J. Am. Chem. Soc. **104**, 6800-6801.

18 Hosur, R.V. (1990) Progress in NMR spectroscopy **22**, 1-53.

19 Rance, M., Wagner, G., Sφrensen, O.W., Wühtrich, K. & Ernst, R.R. (1984) J. Magn. Res. **59**, 250-261.

20 Chazin, W.J., Wühtrich, K., Hyberts, S., Rance, M., Denny, W.A. & Leupin, W. (1986) J. Mol. Biol. **190**, 439-453.

21 Griesinger, C., Sφrensen, O.W. & Ernst, R.R. (1989) J. Chem. Phys. **85**, 6837-6852.

22 Haasnoot, C.A.G. & Hilbers, C.W. (1983) Biopolymers **22**, 1259-1266.

23 Bodenhausen, G. & Freeman, R. (1977) J. Magn. Res. **28**, 471-476.

24 Sklenar, V. & Bax, A. (1987) J. Magn. Res. **71**, 379-383.

25 Griesinger, C., Sφrensen, O.W. & Ernst, R.R. (1989) J. Magn. Res. **84**, 14-63

26 Vuister, G.W. & Boelens, R. (1987) J. Magn. Res. **73**, 328-333.

27 Griesinger, C., Sφrensen, O.W. & Ernst, R.R. (1987) J. Magn. Res. **73**, 574-579.

28 Vuister, G.W., Boelens, R. & Kaptein, R. (1988) J. Magn. Res. **80**, 176-185.

29 Gronenborn, A.M., Clore, G.M. & Kimber, B.J. (1984) Biochem. J. **221**, 723-736.

30 van de Ven, F. J.M., & Hilbers, C.W. (1988) Eur. J. Biochem. **178**, 1-38.

31 Boelens, R., Koning, T.M.G., van der Marel, G.A., van Boom, J.H. & Kaptein, R. (1989) J. Magn. Res. **82**, 290-308.

32 Borgias, BA. & James, T.L. (1988) J. Magn. Res. **79**, 493-512.

33 Tropp, J. (1980) J. Chem. Phys. **72**, 6035-6043.

34 Keepers, J.W. & James, T.L. (1982) J. Am. Chem. Soc. **104**, 929-939.

35 Lipari, G. & Szabo, A. (1983) J. Am. Chem. Soc. **104**, 4546-4570.

36 Altona, C. (1982) Recl. Trav. Chim. Pays-Bas, **101**, 413-433.

37 Weiner, S.J., Kollman, P.A., Nguyen, D.T. & Case, D.A. (1986) J. Comp.

Chem. **7**, 230-235.

38 van Gunsteren, W.F., Berendsen, H.J.C., Hermans, J., Hol, W.G.J. & Postma, J.P.M. (1983) Proc. Natl. Acad. Sci. USA **80**, 4315-4319.

39 Reid, B.R. (1987) Q. Rev. Biophys. **20**, 1-34.

40 Patel, D.J., Shapiro, L. & Hare, D. (1987) Q. Rev. Biophys. **20**, 35-112.

41 Patel, D.J., Shapiro, L. & Hare, D. (1987) Annu. Rev. Biophys. Biophys. Chem. **16**, 423-454.

42 Metzler, W.J., Hare, D.R. & Pardi, A. (1989) Biochemistry **28**, 7045- .

43 Lucasius, C.B. & Kateman, G. (1989) International conference on Genetic Algorithms. **3** Editor Shaffer, J.D., San Mateo, CA., 170-176.

44 Lucasius, C.B., Blommers, M.J.J., Buydens, L.M.C. & Kateman, G. (1990) "A genetic algorithm for conformational analysis of DNA." Handbook of the Genetic Algorithm. chapter 18 (in press)

45 Saenger, W. (1984) "Principles of nucleic acid structure." 1th ed., Springer Verlag, New York.

46 Altona, C. & Sundaralingam, M. (1972) J. Am. Chem. Soc. (1972) **94**, 8205-8212.

47 de Leeuw, H.P.M., Haasnoot, C.A.G. & Altona, C. (1980) Isr. J. Chem. **20**, 108-126.

48 Olson, W.K. (1982) J. Am. Chem. Soc. **104**, 270-267.

49 Olson, W.K. (1982) J. Am. Chem. Soc. **104**, 278-286.

50 Lewin, B. (1983) "Genes." 1th ed. Wiley and Sons, New York.

51 Watson, J.D. (1976) "Molecular biology of the gene." 3th ed. Benjamin Inc., Menlo Park, California.

52 Seeman, N.C., Rosenberg, J.H., Suddath, F.L., Kim, J.J.P. & Rich, A. (1976) J. Mol. Biol. **104**, 109-144

53 Rosenberg, J.H., Seeman, N.C., Day, R.D. & Rich, A. (1976) J. Mol. Biol. **104**, 145-167.

54 Watson, J.D. & Crick, F.H.C. (1953) Nature (London) **171**, 737-738.

55 Hoogsteen, K. (1963) Acta Crystallogr. **16**, 907-916.

56 Frey, M.N., Koetzle, T.F., Lehmann, M.S. & Himilton, W.C. (1973) J. Chem. Phys. **59**, 915-924.

57 Platt, J.R. Proc. Natl. Acad. Sci. USA **41**, 181-183.

58 Hariford, D.B. & Pulleyblank, D.E. (1985) Nucl. Acids Res. **13**, 4343-4363.

59 Naylor, L.H., Lilley, D.M.J. & van de Sande, J.H. (1986) EMBO J. **5**, 2407-2413.

60 van Belkum, A, Blommers, M.J.J., van den Elst, H., van Boom, J.H. & Hilbers, C.W. (1990) Nucl. Acids Res (in press)

61 Hanvey, J.C., Klysik, J. & Wells, R.D. (1988) J. Biol. Chem. **263**, 7386-7397.

62 Voloshin, O.N., Mirkin, S.M., Lyamichev, V.I., Belotserkovskii, B.P. & Frank-Kamentskii, M.D. (1988) Nature **333**, 475-476.

63 Sundquist, W.I. & Klug, A. (1989) Nature **342**, 825-829.

64 Htun, H. & Dahlberg, J.E. (1988) Science **241**, 1791-1796.

65 Benight, A.S., Wang, Y., Amaratunga, M., Chatopadhyana, R., Henderson, J., Hanlon, S. & Ikuta, S. (1989) Biochemistry **28**, 3323-3332.

66 Ikuta, S., Chattopadhyana, K., Ito, H., Dickerson, R.E. & Kearns, D.R. (1986) Biochemistry **25**, 4840-4849.

67 Chattopadhyaya, R., Ikuta, S., Grzeskowiak, K. & Dickerson, R.E.(1988) Nature (london) **334**, 175-179.

68 Wolk, S.K., Hardin, C.C., Germann, M.W., van de Sande, J.H. & Tinoco, I. (1988) Biochemistry **27**, 6960-6967.

69 Hare, D.R. & Reid, B.R. (1986) Biochem. **25**, 5341-5350.

70 Flynn, P.F., Kintanar, A., Reid, B.R. & Drobny, G. (1988) Biochemistry **27**, 1191-1197.
71 Xodo, L.E., Manzini, G., Quadrifoglio, F., van der Marel, G.A. & van Boom, J.H. (1988) Biochemistry **27**, 6321-6326.
72 Xodo, L.E., Manzini, G., Quadrifoglio, van der Marel, G.A. & van Boom, J.H. (1988) J. Biomol. Struct. Dyns. **6**, 139-152.
73 Xodo, L.E., Manzini, G., Quadrifoglio, F., van der Marel, G.A. & van Boom, J.H. (1986) Nucl. Acids Res. **14**, 5389-5398.
74 Wemmer, D.E., Chou, S.H., Hare, D.R. & Reid, B.R. (1985) Nucl. Acids Res. **13**, 3755-3772.
75 Summers, M.F., Byrd, R.A., Gallo, K.A., Samson, C.J., Zon, G. & Egan, W. (1985) Nucl. Acids Res. **13**, 6375-6386.
76 Marky, L.A., Blumenfield, K.S., Kozlowski, S. & Breslauer, K.J. (1983) Biopolymers **22**, 1247-1257.
77 Williamson, J.R. & Boxer, S.G. (1989) Biochemistry **28**, 2819-2831.
78 Williamson, J.R. & Boxer, S.G. (1988) Nucl. Acids Res. **16**, 1529-1540.
79 Williamson, J.R. & Boxer, S.G. (1988) Biochemistry **28**, 2831-2836
80 Orbons, L.P.M., van der Marel, G.A., van Boom, J.H. & Altona, C. (1987) Eur. J. Biochem. **170**, 225-239.
81 Orbons L.P.M., van Beuzekom, A.A. & Altona, C. (1987) J. Biomol. Struct. Dyns. **4**, 965-987.
82 van Beuzekom, A.A. (1989) Ph.D. Thesis, State University Leiden
83 Orbons, L.P.M., van der Marel, G.A., van Boom, J.H. & Altona, C. (1986) Nucl. Acids Res. **14**, 4187-4196.
84 Pieters, J.M.L. Thesis (1989) State University Leiden, The Netherlands.
85 Pieters, J.M.L., de Vroom, E., van der Marel, G.A., van Boom, J.H., Koning, T.M.G., Kaptein, R. & Altona, C. (1989) Biochemistry **29**, 788-799.
86 Gupta, G., Sarma, M.H. & Sarma, R.H. (1987) Biochem. **26**, 7715-7723.
87 Garcia, A.E., Gupta, G., Sarma, M.H. & Sarma, R.H. (1988) J. Biomol. Struct. Dyns. **6**, 525-542.
88 Hiaro, I, Nishimura, Y., Naraoko, T., Watanabe, K., Arata, Y. & Miura, K. (1989) Nucl. Acids Res. **17**, 2223-2231.
89 Scheffler, I.E., Elson, E.L. & Baldwin, R.L. (1968) J. Mol. Biol. **36**, 291-304.
90 Scheffler, I.E., Elson, E.L. & Baldwin, R.L. (1970) J. Mol. Biol. **48**, 145-171.
91 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. (1983a) J. Biomol. Struct. Dyns. **1**, 115-129.
92 Haasnoot, C.A.G. den Hartog, J.H.L., de Rooij, J.F.M., van Boom, J.H. & Altona, C. (1980) Nucl. Acids Res. **8**, 169-181.
93 Haasnoot, C.W. & Hilbers, C.W. (1983) Biopolymers **22**, 1259-1266.
94 Haasnoot, C.A.G., de Bruin, S.H., Hilbers, C.W., van der Marel, G.A., van Boom, J.H. (1985) Proc. Int. Symp. Biomol. Struct. Interactions, Suppl. J. Biosci. **8**, 767-780.
95 Hilbers, C.W., Haasnoot, C.A.G., de Bruin, S.H., Joordens, J.J.M., van der Marel, G.A. & van Boom, J.H. (1985) Biochimie **67**, 685-695.
96 Blommers, M.J.J., Haasnoot, C.A.G., Hilbers, C.W., van Boom, J.H., van der Marel, G.A. (1987) Structure and Dynamics of Biopolymers, NATO ASIS Series E:Applied Sciences **133**, 78-91.
97 Hilbers, C.W., Blommers, M.J.J., van der Ven, F.J.M., van Boom, J.H. & van der Marel, G.A. (1990) CECAM Workshop (in press).
98 Hilbers, C.W., Blommers, M.J.J., Haasnoot, C.A.G., van der Marel, G.A.

& van Boom, J.H. (1987) Fresenius Z. Anal. Chem. **327**, 70-71.

99 Haasnoot, C.A.G., Blommers, M.J.J., Hilbers, C.W. (1987) Springer Series in Biophysics 1: Structure, Dynamics and Function of biomolecules (Eds. A. Ehrenberg, R. Rigler, A. Gräslund, L. Nilsson), 212-216.

100 Blommers, M.J.J., Walters, J.A.L.I., Haasnoot, C.A.G., Aelen, J.M.A., van der Marel, G.A., J.H. van Boom & Hilbers, C.W. (1989) Biochemistry **28**, 7491-7498.

101 Blommers, M.J.J., van de Ven, F.J.M., van der Marel, G.A., van Boom, J.H. & Hilbers, C.W. Eur. J. Bioch. (submitted).

102 Hilbers, C.W., Blommers, M.J.J., van der Ven, F.J.M., van Boom, J.H. & van der Marel, G.A. (1990) Nucleosides and nucleotides (in press)

103 Senior, M.M., Jones, R.A. & Breslauer, K.J. (1988) Proc. Natl. Acad. Sci. USA **85**, 6242-6246.

104 Raap, J. (1987) Thesis. State University Leiden, The Netherlands.

105 Pramanik, P., Kanhouwa, N. & Kan, L.-S. (1988) Biochemistry **27**, 3024-3031.

106 Germann, M.W., Kalisch, B.W., Lundberg, P., Vogel, H.J. & van de Sande, J.H. (1990) Nucl. Acids Res. **18**, 1489-1498.

107 van Boom, J.H., van der Marel, G.A., Westerink, H., 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) Cold Spring Harbor Symposia on Quantitative Biology **47**, 403-409.

108 Uhlenbeck, O.C., Borer, P.N., Dengler, B. & Tinoco, I. (1973) J. Mol. Biol. **73**, 483-496.

109 Gralla, J. & Grothers, D.M. (1973) J. Mol. Biol. **73**, 497-511.

110 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) J. Biomol. Struct. Dyns **3**, 843-857.

111 Heus, H.A., van Kimmenade, J.M.A., van Knippenberg, P.H., Haasnoot, C.A.G., de Bruin, S.H. & Hilbers, C.W. (1983) J. Mol. Biol. **170**, 939-956.

112 Groebe, D.R. & Uhlenbeck, O.C. (1988) Nucl. Acids Res. **16**, 11725-11735.

113 Puglisi, J.D., Wyatt, J.R. & Tinoco, I. (1990) Biochemistry **29**, 4215-4226.

114 Heus, H.A., Formeny, L.J. & van Knippenberg, P.H. (1990) Eur. J. Biochem. **188**, 275-281.

115 Haasnoot, C.A.G., de Leeuw, F.A.A.M., de Leeuw, H.J.M. & Altona, C. (1979) Recueil, J. Roy. Netherl. Chem. Soc. **98**, 576-577.

116 Lankhorst, P.P., Haasnoot, C.A.G., Erkelens, C. & Altona, C. (1984) J. Biomol. Struct. Dyns. **1**, 1387-1405.

117 Westhof, E. & Sundaralingam, M. (1986) Biochemistry **25**, 4868-4878.

CHAPTER 2

THE SOLUTION STRUCTURE OF THE 3'-5' CYCLIC DINUCLEOTIDE d<pApA>
A COMBINED NMR, UV-MELTING AND MOLECULAR MECHANICS STUDY

ABSTRACT

The 3'-5' cyclic dinucleotide d<pApA> was studied by means of ^{1}H- and ^{13}C-NMR experiments, UV-melting experiments and molecular mechanics calculations. The ^{1}H- and ^{13}C-NMR spectra were analyzed by means of 2-dimensional NMR experiments. J-coupling analysis of the 1D- and 2D- ^{1}H and ^{13}C spectra was used to determine the conformation of the ring systems in the molecule. It appeared that at low temperature (283 K) the deoxyribose sugars adopt a N-type conformation. The geometry is best described by an intermediate between the $^{3}_{2}$T and ^{3}E forms. In addition we were able to derive all other torsion angles in the phosphate backbone ringsystem, i.e., α^{+}, β^{t}, γ^{+}, δ (=89°), ϵ^{t} and ζ^{+}. When the molecule is subjected to an energy minimization procedure (using the program AMBER) the sugar ring system retains practically speaking the torsion angles found from the NMR experiments, while the torsion angles around the glycosidic bond adopt a value of 175° in the minimum energy conformation. UV-melting experiments indicate that two molecules can form a dimer in which the adenine bases are intercalated. The feasibility of this structure is indicated by molecular mechanics calculations. At higher temperatures the dimer is converted into separate monomers. In the monomer form the sugars exhibit S-pucker during 20 % of the time. Concomitantly with the conversion of the N- to the S-conformation the torsion angles α and γ change.

INTRODUCTION

Circular DNA and RNA molecules play an important role in molecular biology. Sometimes such molecules are formed as intermediates as is known for instance for cyclic RNA molecules [1,2] and in other cases the circularity is the natural appearance of DNA and RNA, e.g., for plasmids, the DNA of some viruses and viroids [3,4]. The smallest conceivable cyclic oligo-DNA or RNA molecules are the cyclic dinucleotides in which the nucleotide units are connected by 3'-5' linkages as is illustrated in Fig. 1. Recent studies have indicated that also these small molecules may play an important role in cellular reactions. Interesting examples are the compounds r<pUpU> and r<pApU>, which are effective inhibitors of the DNA-dependent RNA polymerase

---

Blommers, M.J.J., Haasnoot, C.A.G., Walters, J.A.L.I., van der Marel, G.A., van Boom, J.H. & Hilbers, C.W. (1988) Biochemistry 27, 8361-8369.

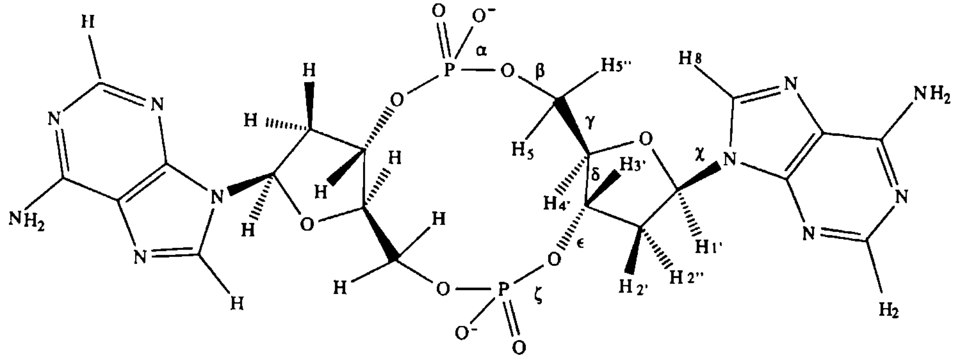


Fig. 1. Structure of the 3'-5'-deoxy cyclic dinucleotide d<pApA>. The ring system of the dinucleotide is closed by linking the 5'-end of the first adenosine to the 3'-end of the second. Protons are numbered and the torsion angles are characterized by α, β, etc. according the IUPAC IUB convention for nucleic acids.

of Escherichia coli during the initiation phase of transcription [5,6]. A rather different activity is exhibited by r<pGpG> for which it was reported that it acts as the natural activator of the enzyme cellulose synthase in Acetobacter Xylinum [7]. The analogous cyclic deoxyribonucleotide d<pGpG> was shown to be active as well during the cellulose synthesis [8].

In the present paper the solution structure of d<pApA> is studied by means of high-resolution NMR spectroscopy, UV-melting experiments and molecular mechanics calculations. The study of the deoxyribose was preferred over the ribose compound because the deoxyribose rings are more amenable to structural studies; adenine bases were chosen instead of guanine bases because the latter may easily form tetramers in solution, which would complicate the structural analysis. It was found that at low temperatures the deoxyribose rings of d<pApA> adopt a pure N conformation, which is quite unusual for the deoxy type of sugar. In addition, at NMR concentrations, i.e., at about 5-20 mM concentrations, the d<pApA> molecules form dimers in which the adenine bases most likely are intercalated between one another. These results and the temperature dependent behavior of this molecule are discussed. The presented study was set up and carried out independently from a crystal structure determination of the molecule by Frederick et al. at M.I.T.[9]. It is satisfying to see that both studies lead to closely similar results.

MATERIALS AND METHODS

The cyclic dinucleotide d<pApA> was synthesized according to published procedures [8]. NMR samples were prepared by dissolving 4 mg of the sodium form of a freeze-dried sample of the title compound in 400 µl of D_2O , this corresponds to a 17 mM d<pApA> concentration. ^1H, ^{13}C and ^{31}P NMR spectra were recorded on a Bruker WM500 NMR spectrometer equipped with an ASPECT 2000 computer. 1-dimensional ^1H spectra without $J(^1H-^{31}P)$ splittings were obtained by means of ^{31}P decoupling. ^1H-^{13}C hetero correlated spectra were recorded on a WM200 Bruker spectrometer equipped with an ASPECT 2000 computer.

In addition, an ω_1-scaled double quantum filtered COSY spectrum (ω_1-scaled DQF-COSY) [10] was measured on a Bruker AM-500 spectrometer equipped with an ASPECT 3000 computer, using the pulse sequence and phase-cycling scheme depicted in Fig. 2. For each t_1 value 512 free induction decays were stored in 2K of computer memory. The t_1 values were varied in steps of 100 µs up to a total value of 0.5 s. The scaling factor k was equal to 0.5. The carrier was placed in the middle of the spectrum and quadrature detection was applied. Sine-bell multiplication was carried out to improve the resolution of the multiplet splitting patterns. Zero-filling up to 1K in t_1 yields after Fourier transformation a 2D-spectrum with a resolution of 2.4 Hz/pnt in ω_1 and 2.4 Hz/pnt in ω_2.

PHASELIST	+X	+X	-Y	-Y	-Y
	+X	-X	-Y	Y	-Y
	+X	+X	-Y	+X	X
	+X	-X	-Y	+X	-X
	+X	+X	-Y	+Y	+Y
	+X	-X	-Y	+Y	+Y
	+X	+X	-Y	-X	+X
	+X	-X	-Y	-X	+X

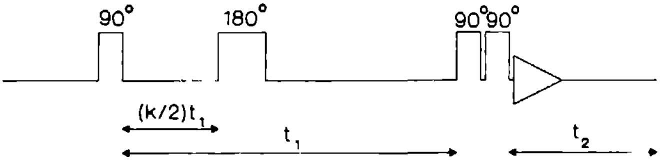


Fig. 2. Pulse scheme for acquisition of the phase sensitive ω_1-scaled DQF-COSY spectrum. The scaling in the ω_1-direction is achieved by a τ-π-τ pulse during the t_1-period to refocus the magnetization of the spins. The degree of ω_1-scaling is determined by the factor k which was chosen equal to 0.5. Phase cycling applied in this experiment is indicated. In addition, this phase-cycling scheme was modulated with the CYCLOPS scheme; thus a complete number of 32 phase cycles was applied.

The conformational analysis of the molecule was performed on the basis of the J-couplings derived from the one-dimensional spectra and the ω_1-scaled DQF-COSY spectrum. To this end approximate J-couplings were first deduced directly from the experimental spectra. Subsequently, 1D- and 2D-spectra were simulated with the mentioned J-couplings as input data. Further refinement of these parameters was obtained by adjusting the J-couplings and chemical shifts so as to obtain optimal agreement between experimental and simulated spectra. For the 1D-spectra this was achieved in an interactive manner using the program PANIC [Bruker library], for the ω_1-scaled DQF-COSY spectrum the program COSIM, which was written by M. Blommers, was applied. The J-couplings between the sugar ring protons served as an input for the computer program PSEUROT [11] by means of which sugar ring conformations were calculated. Based on these and other NMR data structures of the molecule d<pApA> were build with the aid of the molecular modelling program Chem-X [12] in combination with some homewritten routines. These programs were executed on a VAX 11/785 computer which was connected to a Sigmex graphical terminal. The resulting structures were subjected to energy minimizations which were carried out using the program AMBER (version 2.0) [13]. The all atom version of the Force Field was used with hydrated counterions added to the system. For the minimization of the d<pApA> dimer structure MacroModel, version 1.5 [14] with the united atom Force Field of AMBER was used.

RESULTS AND DISCUSSION

NMR experiments

Fig. 3 shows an ω_1-scaled DQF-COSY spectrum of the cyclic dinucleotide d<pApA> recorded at 298 K, together with the 1-dimensional spectrum of this compound measured at the same temperature. In the 1-dimensional spectrum each pair of protons of the same type (e.g., each pair of H1' protons) is represented by a single set of isolated resonances (except for the H4'/H5' resonances which are partly overlapping). As we will show in the section 'Conformational analysis',at low temperature the sugar ring as well as the 12-membered sugar-phosphate backbone ring is rigid. Thus, the dinucleotide has a two fold axis of symmetry which renders the two halves of the molecule magnetically equivalent. It is noted in passing that the apparent symmetry was preserved in all spectra recorded at different temperatures, i.e., from

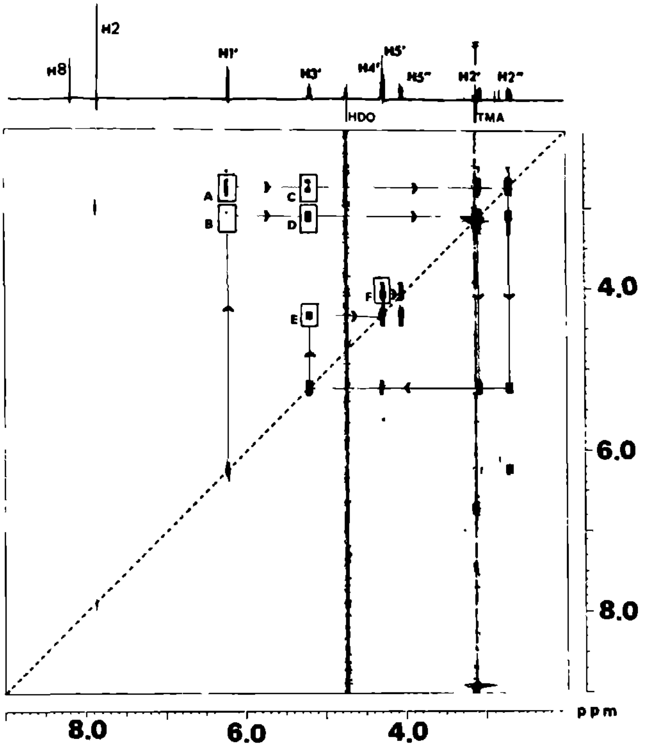


Fig. 3. The 500 MHz ^{1}H ω_1-scaled DQF-COSY spectrum of d<pApA>. The dashed line is drawn through the diagonal peaks. Note that the singlet peaks arising from the H8 and H2 protons in the 1-dimensional spectrum added on top do not give rise to signals in the 2-dimensional spectrum due to the double-quantum filter. The solid line with arrows connects via the cross peaks the different resonances of the ringsystem. The assignments following from this connectivity pattern are indicated in the 1D spectrum. An enlarged view of the cross peak contained in box D is presented in Fig. 5B.

281 to 344 K. The spectral assignment indicated in the 1D-spectrum was obtained from the connectivity pattern which can been drawn for the cross peaks in the ω_1-scaled DQF-COSY spectrum in the standard manner (see Fig. 3). Discrimination between the H2' and H2" signals and between the H5' and H5" resonances was achieved on the basis of the following reasoning. The J-coupling of one of the C2' sugar protons (resonance position 3.1 ppm) with the C1' proton is small, i.e., 1.5 Hz. This means that this can only be the $J_{H1'H2'}$ coupling [15] and consequently the multiplet at 3.1 ppm is assigned to the H2' spins. Using the reasoning of Remin and Shugar [16] it can be

concluded that the H5' signal resonates downfield from the H5" signal. The small values of the coupling constants $J_{H4'H5'}$ and $J_{H4'H5''}$ indicate that γ is in a gauche$^+$ conformation. In that case, H5' and H5" experience a different shielding; the H5" is in the vicinity of the 3'-phosphate and therefore shifts upfield relative to H5'. The magnitudes of these couplings is in agreement with this assignment, because due to the O4'-O5' repulsion, the torsion angle between H5" and H4' is somewhat smaller than 60°, so that $J_{H4'H5''}$ is somewhat larger than $J_{H4'H5'}$. The J-couplings were deduced by comparison of the experimental and simulated spectra as described in the section Materials and Methods. For the ω_1-dimensional ^1H spectra this was done with and without the presence of ^1H-^{31}P J-couplings. The results are depicted in Fig. 4; examination of the spectra shows that excellent agreement between experimental and simulated spectra is obtained. The

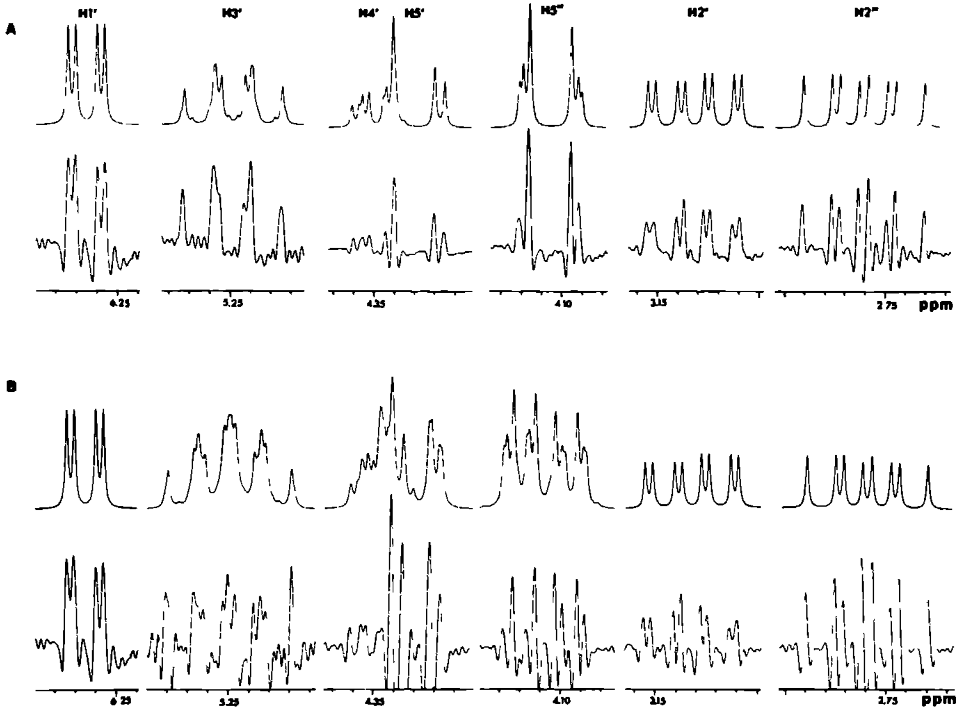


Fig. 4. Experimental and simulated multiplets of the sugar protons of d<pApA> at 298 K (17 mM). The ^1H-spectrum (A) as well as the ^{31}P-decoupled ^1H-spectrum (B) are shown. The experimental spectra were resolution enhanced with Gaussian multiplication before Fourier transformation.

ω_1-scaled DQF-COSY spectrum, which was recorded with the $J(^1H-^{31}P)$ splitting preserved, was simulated with the couplings included. As an example the experimentally observed and simulated cross peaks between the resonances of H2' and H3' are shown in Fig. 5. These cross peaks contain all J-couplings of the sugar ring spin system.

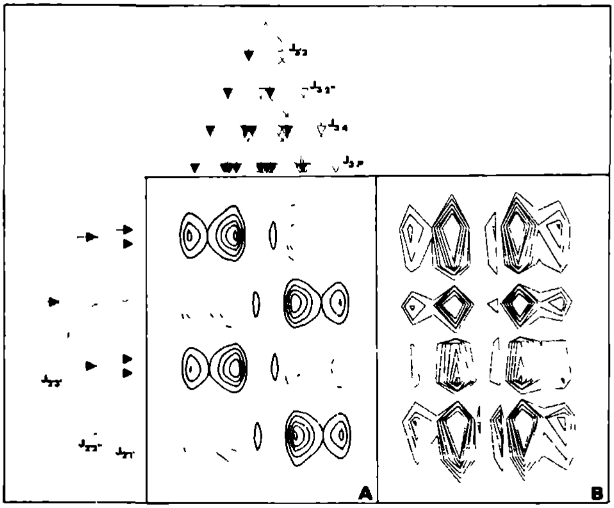


Fig. 5. Extended view of the measured (B) and simulated (A) cross peaks between the H2' and H3' resonances in the ω_1-scaled DQF-COSY spectrum. The interpretation of the splitting pattern is indicated on top and left of the simulated cross peak. The cross peak is split by the active J-coupling, $J_{2'3'}$, in positive and negative absorption peaks. All other splittings do not change the sign of the absorption lines as is indicated by the filled and open arrows. Because of the opposite signs of the absorptions, some of the signals almost cancel as is clear in the center of the experimental cross peak (B). In the simulated spectrum the negative peaks are dotted; although not indicated, it is noted that the signs of the experimental peaks agree with the simulation. The cross peak in (A) was simulated with a Lorentzian line shape in both directions, LB = 2.0 Hz, Hz/pnt = 2.4. The following J-couplings used were: $J_{2'3'} = 7.2$, $J_{2'2''} = -13.8$, $J_{2'1'} = 1.5$, $J_{3'2''} = 9.1$, $J_{3'4'} = 7.7$ and $J_{3'P} = 6.6$.

The J-couplings appeared to be temperature as well as concentration dependent. Therefore the coupling constants were measured at different temperatures and at a low concentration of the title compound (i.e., the concentration was reduced 20-fold with respect to the 17 mM used in the experiments discussed above). The results are summarized in Table I, where the values of the J-coupling constants are given for the different conditions. It is interesting to note that the coupling constants observed for the

Table I. J-couplings determined for d<pApA> (concentration 17 mM) at different temperatures by means of simulation of the multiplet splitting patterns of the ^{1}H-spectra. a) indicates the 31P-decoupled spectrum; b indicates the low concentration sample (0.8 mM).

J-coupling	T(K)					
	281.0	295.5	295.5^a	308.5^a	326.5	298.0^b
$J_{1'2'}$	0.7	1.5	1.7	2.5	3.4	3.0
$J_{1'2''}$	7.1	7.0	7.1	7.1	7.2	7.3
$J_{2'2''}$	-13.5	-13.8	-13.7	-13.7	-14.0	-14.1
$J_{2'3'}$	7.5	7.2	7.2	7.2	7.3	7.3
$J_{2''3'}$	9.5	9.1	9.0	8.5	7.6	8.0
$J_{3'4'}$	7.7	7.7	7.7	7.3	6.8	7.0
$J_{3'P}$	5.8	6.3	---	---	7.2	7.0
$J_{4'5'}$	2.2	2.5	2.3	3.0	3.8	3.9
$J_{4'5''}$	1.5	1.6	1.8	1.9	2.2	2.0
$J_{4'P}$	3.2	2.8	---	---	2.2	2.2
$J_{5'5''}$	-12.1	-12.0	-12.1	-11.8	-11.7	-11.7
$J_{5'P}$	0.4	0.7	---	---	2.7	2.9
$J_{5''P}$	5.8	5.4	---	---	4.9	4.9

low concentration sample and those obtained at the highest temperature for the high concentration sample are virtually the same.

The ^{13}C spectra were assigned by means of hetero correlated spectroscopy (not shown). Using the proton decoupled ^{13}C spectra, the ^{31}P-^{13}C couplings were derived in the same manner as described for the ^{1}H spectra; the results are collected in Table II.

Table II. ^{13}C-^{31}P coupling constants determined for d<pApA> from ^{13}C-spectra.

J-coupling	T (K)			
	305	321	334	347
$J_{C2'P3'}$	<1.0	<1.0	<1.0	<1.0
$J_{C3'P3'}$	4.9	5.2	5.3	5.3
$J_{C4'P3'}$	10.9	10.3	9.9	9.9
$J_{C4'P5'}$	10.9	10.3	9.9	9.9
$J_{C5'P5'}$	4.6	5.1	5.2	5.2

UV-melting experiments

The concentration dependence of the J-couplings at 298 K suggests that intermolecular interactions between cyclic d<pApA> molecules may play a role at concentrations needed for NMR experiments. The most likely interaction that comes to mind is a stacking of the adenine bases. Measurement of the chemical shifts of the various proton resonances as a function of temperature does not yield unambiguous evidence for the occurrence of stacking-destacking processes and therefore UV-melting experiments were conducted. The results are collected in Fig. 6. Indeed, appreciable hyperchromicity effects are observed when the temperature of the d<pApA> sample is raised as expected for destacking processes. At a concentration of 17 mM, corresponding to the concentration used in the NMR experiments, a melting curve is obtained, characterized by a melting temperature of 27°C. The most straightforward explanation for this observation is the presence of a monomer-dimer equilibrium. Indeed, on the basis of a two state equilibrium the melting profile can be simulated by fitting a melting curve to the data points using a least square refinement procedure. On this basis a transition enthalpy of 21 kcal/mol for the melting process is estimated. At fourth-fold and sixteen-fold lower concentrations, temperature increase leads to absorption

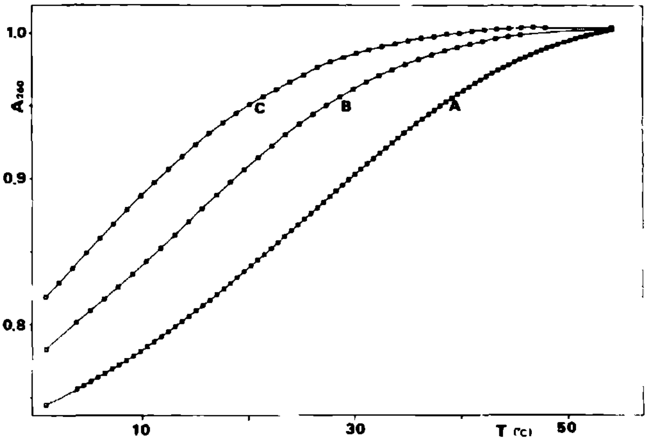


Fig. 6. UV-melting profiles of d<pApA> recorded at three different concentrations. Along the y-axis the relative absorbance at 260 nm is indicated. Curve A corresponds to 17 mM d<pApA>, curve B to 4 mM and curve C to 1 mM. NaCl up to 200 mM was added to the samples. The melting temperatures are 27°, 15° and 7°C, respectively.

profiles which are shifted to lower temperatures as expected for a monomer-dimer equilibrium. From these experiments we therefore conclude that dimers are formed at NMR concentrations (17 mM) at 281 K (8° C) (see Fig. 6). By raising the temperature and/or diluting the sample the equilibrium is shifted in favor of increased monomer concentrations.

Conformational analysis

Deoxyribose conformation

The deoxyribose conformation was deduced from a pseudorotational analysis of the sugar ring on the basis of the available coupling constants (see Table I) [17,18], using the program PSEUROT [11]. The analyses indicate that at the lowest temperature, i.e., 281 K, and d<pApA> concentrations of 17 mM, the sugar ring adopts a pure (100 %) N-type conformation characterized by the pseudorotational parameters P = -6.6 and ϕ_m = 34.9, where P is the phase angle of pseudorotation and ϕ_m is the amplitude of the pucker. The sugar pucker is best described as an intermediate between the envelope form C2'-exo, ^2E, and the twist form $^3{}_2T$. At higher temperatures a mixed population of N and S sugar conformations builds up (see Table III), although at 326.5 K the N-conformer is still present during 78 % of the time while at this temperature d<pApA> is virtually in the monomeric form (see Fig 6).

Table III. Pseudorotational parameters, i.e the pseudorotational phase angle, P, and the pucker amplitude, ϕ_m, for the sugar rings obtained at different temperatures. The fraction of the N-conformers x(N) is presented as well. a) indicates the twenty times diluted sample (0.8 mM).

T(K)	P(N)	ϕ_m(N)	P(S)	ϕ_m(S)	x(N)
281.0	-6.6	34.9	-	-	1.00
295.5	-5.7	35.9	130.0	38.0	0.96
308.5	-1.9	35.6	130.0	38.0	0.87
326.5	0.9	34.1	127.9	38.7	0.78
298.0^a	0.9	34.1	130.7	37.7	0.83

As has been mentioned already the J-couplings do not only depend on temperature but on the d<pApA> concentration as well. Therefore, also a J-coupling analysis was carried out at the low concentration (for a twenty-fold diluted sample, vide supra). It was found that at 298 K when d<pApA> is predominantly in the monomeric form, 83% of the sugar rings are N-conformers with P(N) = 0.9 and ϕ_m = 34.1. The S-type conformer is characterized by the pseudorotational parameters P = 130.7 and ϕ_m = 37.7 (Table III). From these results it can be concluded that the d<pApA> monomer has a tendency to form S-type sugar puckers. Comparing the population of N-sugars of the monomers at 298 K (83 % N) and at 326.5 K (78 % N) indicate that the N-S equilibrium is not very temperature sensitive.

Conformation around C4'-C5'

The torsion angle γ (C3'-C4'-C5'-O5') is known to occur in three staggered conformations namely gauche$^+$, trans and gauche$^-$. The percentage of the population present in the gauche$^+$ conformer can be estimated with the help of the empirical formula [17]:

$$x_{g^+} = (13.75 - \Sigma) / 10.05 \qquad (1)$$

where $\Sigma = J_{H4'H5'} + J_{H4'H5''}$. It follows from Table I that at 281 K Σ = 3.7 Hz, so that the gauche$^+$ population x_{g+} = 100 %. Thus we may assume that at this temperature the torsion angle γ is purely gauche$^+$. If this situation obtains it follows from Eq. 3 (see below) that γ is close to 53° [19]. Further examination shows that at elevated temperature (326.5 K) x_{g+} = 78 %. We therefore conclude that the reduction of the gauche$^+$ population occurs concomitantly with and to the same extent as the reduction of the N-conformer of the sugar ring. This conclusion is corroborated by the results obtained at low concentration, i.e., 0.85 mM d<pApA>: for the monomer x_{g+} = 78 %, while we found that at these conditions the population of the N-conformer is 83 % (vide supra). The population around γ can also be estimated from the J-couplings in terms of the three classical rotamers:

$$\begin{aligned} J_{4'5'} &= x_{g+}J'_{g+} + x_t J'_t + x_{g-}J'_{g-} \\ J_{4'5''} &= x_{g+}J''_{g+} + x_t J''_t + x_{g-}J''_{g-} \end{aligned} \qquad (2)$$

The J-couplings of the individual rotamers were calculated using the generalized Karplus equation [20]:

$$^{3}J_{HH} = 13.22 \cos^2 \phi - 0.99 \cos \phi + \Sigma(0.87 - 2.46 \cos^2 (\xi_i\phi + 19.9 \mid \Delta\chi_i \mid)) \Delta\chi_i \qquad (3)$$

where ϕ represents the proton-proton torsion angle and $\Delta\chi_i\phi$ is the difference in electronegativity between the substituents of the H-C-C-H fragment and hydrogen. ξ_i is +1 or -1 depending on the orientation of the substituent [20]. Calculation of the population of the three rotamers at 326.5 K yielded the following distribution: 77 % gauche$^+$, 3 % trans and 20 % gauche$^-$. Practically speaking the population of the trans conformer is negligible. The result for the gauche$^+$ population is in good agreement with that obtained from Eq. 1. Moreover, we are able to discriminate between a trans and the gauche$^-$ rotamer because the 5' and 5'' proton resonances could be assigned individually. It is noted that it follows from both analyses that the gauche$^-$ population of γ and the population of the S-conformation of the sugar ring have about the same value at elevated temperatures.

Conformation around C5'-O5'

An estimation of the torsion angle β (C4'-C5'-O5'-P) can be obtained from the coupling constants $J_{H5'P}$, $J_{H5''P}$ and $J_{C4'P}$.We used the Karplus equations, parametrized for these couplings by Lankhorst et al. [21]:

$$J_{CCOP} = 6.9 \cos^2 \phi - 3.4 \cos \phi + 0.7 \qquad (4a)$$

$$J_{HCOP} = 15.3 \cos^2 \phi - 6.1 \cos \phi + 1.6 \qquad (4b)$$

to derive the torsional angle β. The $J_{C4'P}$ coupling constant amounts to 10.9 Hz (see Table II); this corresponds with the largest value observed for a ^{13}C-^{31}P vicinal coupling [21] and leads to the conclusion that β adopts a trans conformation (Eq. 4a). An independent estimation of the percentage of the β^t conformer can be obtained with the following equation [21]:

$$x_t = (25.5 - J_{H5'P} - J_{H5''P}) / 20.5 \qquad (5)$$

which yields β^t = 94 %. Since the very small $J_{H5'P}$ coupling constant (approx. 0.4 Hz) is not determined very accurately because of the inevitable linebroadening at lower temperatures, we estimate that the inaccuracies in $J_{H5'P}$ may amount to 0.2 to 0.3 Hz. Therefore the result obtained from Eq. 5 may well indicate that β adopts a 100 % trans conformation. It is noted that the relative large value of the long range constant $J_{H4'P}$ also indicates that β is fixed in a pure trans conformation. For the classical 180° trans conformation one expects $J_{H5''P}$ and $J_{H5''P}$ to be equal. However, in our case the differences between the coupling constants, $J_{H5'P}$ and $J_{H5''P}$ amounts 5.4 Hz. This suggests that the magnitude of β is higher than the ideal staggered (180°) value [17]. Despite the inaccuracy of the value for $J_{H5'P}$ we can safely conclude that it is incompatible with Eq. 4b. This finding throws some doubts on the overall validity of these Karplus equations. For the time being, we solved this problem by optimizing the value of the torsion angle β in such a way that the square of the differences between the measured coupling constants and the couplings calculated from the Karplus equations (Eqs. 4a and b) was minimal. At 281 K, this yields a value for β of 196°. At higher temperatures the values of the coupling constants change somewhat. However, it can be concluded that the β^t form remains the predominant conformer.

Conformation around C3'-O3'

The torsion angle ε (P-O3'-C3'-C4') is monitored by three J- couplings, namely, $J_{H3'P}$, $J_{C2'P}$ and $J_{C4'P}$. The $J_{C4'P}$ coupling amounts to 10.9 Hz, which in analogy with the torsion angle β shows that ε also adopts a trans conformation [21]. Combination of this information with the result at 281 K, i.e., $J_{H3'P}$ = 5.8 Hz, and the Karplus Eqs. 4a and b yields ε^t = 200°.

Lankhorst et al. [21] have noted that the $^3J_{C4'P5'}$ shows an approximate linear behavior as a function of the percentage of N-conformer of the sugar ring. When the J-coupling approaches a value of 11 Hz the population of the sugar ring approaches a 100 % N conformation. This is also reflected in the behavior of the present molecule. At higher temperatures the $J_{C4'P}$ coupling decreases. This is consistent with an increase of the torsion angle ε. At higher temperature the population of the S-sugar and the γ^- rotamer amounts 20 %. This change in conformation requires an adaption of ε to an increased value by the same amount. The corresponding value of ε is estimated to be 230°; i.e., from the J-couplings at 326.5 K it can be calculated that 80 % ε(200°) is in equilibrium with 20 % ε(230°).

Model building and molecular mechanics

Given the torsion angles determined from the NMR experiments and the bond angles, available from the literature, it is possible to build a model for the cyclic part of the d<pApA> molecule indicated in Fig. 1. It is noticed that the torsion angles α and ζ cannot be determined from a J-coupling analysis. It can be shown in a general way however that for the present ring system which its twofold axis of symmetry sufficient constraints are available to derive α and ζ unambiguously. In practice the calculation was performed with the aid of the goniometric formula [22]:

$$\begin{aligned} x_n &= d_{n-1} - x_{n-1}\cos\theta_{n-2} - y_{n-1}\sin\theta_{n-2}\cos\omega_{n-3} + \\ &\quad z_{n-1}\sin\theta_{n-2}\sin\omega_{n-3} \qquad (6) \\ y_n &= x_{n-1}\sin\theta_{n-2} - y_{n-1}\cos\theta_{n-2}\cos\omega_{n-3} + z_{n-1}\cos\theta_{n-2}\sin\omega_{n-3} \\ z_n &= y_{n-1}\sin\omega_{n-3} + z_{n-1}\cos\omega_{n-3} \end{aligned}$$

In Eq. 6, x_n, y_n and z_n are the cartesian coordinates of the atoms of the 12-membered ring, d_i represent the bond distances, θ_i the bond angles and ω_i the torsion angles, the subscripts refer to the subsequent atoms n to n-3 of the ring backbone. We incorporated these equations into a computer program and varied the torsion angles α and ζ systematically , while keeping all other torsion angles in the ring constant at the value determined by the

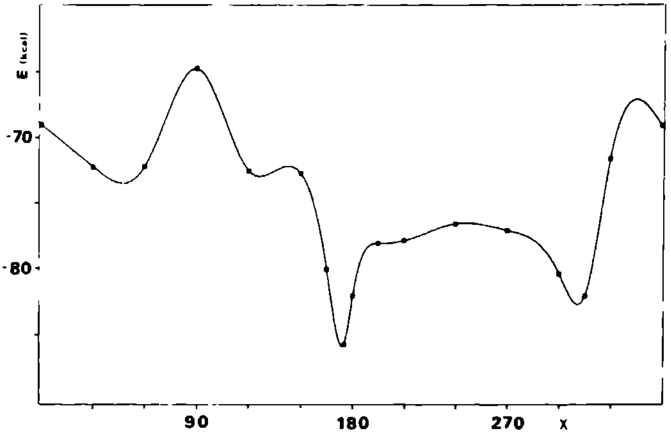


Fig. 7. Energy of d<pApA> as a function of the glycosidic torsion angle χ(O4'-C1'-N9-C4). The torsion angles of structure A in Table IV were used as input for the energy minimization.

J-coupling analysis. In Fig. 8 a contour plot is presented of distances between the n^{th} and n^{th+12} atom for all values of α and ζ. Perfect ring closure obtains when the distance between the n^{th} and the n^{th+12} atom is zero. This approach yields a unique solution for α and ζ, i.e., α as well as ζ were found to have a gauche$^+$ conformation at 281 K. (c.f. Table IV, structure A).

To derive a structure for the complete molecule the value of the torsion angles χ(O4'-C1'-N9-C4) has yet to be established. This problem was approached with the aid of molecular mechanics calculations. Starting with the structure of the cyclic part of d<pApA> described above, molecular energy minimizations were performed for different (constrained) values of the angle χ, using the program AMBER (see Materials and Methods). The energy distribution obtained as a function of χ is presented in Fig. 7. In the plot two energy minima can be discerned corresponding with χ = 175° and χ = 310°. The values of the torsion angles of the sugar and sugar-phosphate backbone

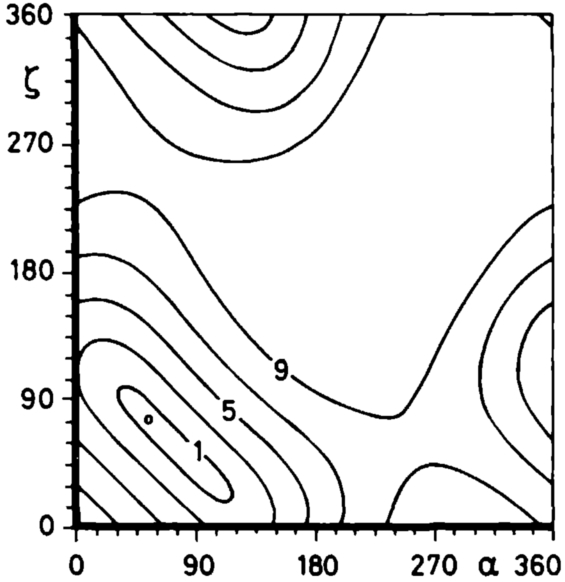


Fig. 8. Contour plot of the distance between the n^{th} and n^{th+12} atom in the sugar-phosphate 12-ring in Å, as a function of the torsion angles α and ζ. In this plot the contour levels of the distance between the n^{th} and n^{th+1} atom vary from 1 to 9 Å; the plotted contour lines are spaced by 2 Å. Unique ring closure occurs when the distance between the n^{th} and n^{th+1} atom is 0 Å, i.e., for the values α = 56° and ζ = 74° indicated by the open dot.

obtained for these structures are virtually the same as those deduced from the NMR experiments (Table IVA). The other structures show some variation in the backbone and/or the sugar conformation. The structure of the molecule with χ = 175° is presented in Fig. 9; the bases are orientated in an antiparallel fashion to maintain the twofold axis of symmetry of the molecule, the distance between them is approximately 7 Å. This distance is ideally suited to accommodate an intercalating base and this directly suggests the possible presence of a dimer structure of the type sketched in Fig. 10A, to explain the occurrence of the stacking-destacking processes in the dynamic monomer-dimer equilibrium deduced from the UV-melting experiments. Indeed, after energy minimization the structure presented in Figs. 10B and C is retained. The displayed dimer is stabilized by three stacking interactions which may explain the high value of the melting enthalpy obtained for the dimer-monomer transition observed in the UV-melting experiment. The energy difference (calculated with the aid of MacroModel using a united atom force field) between two monomer minimum structures and the dimer is 24 kcal; this is in excellent agreement with the experimental melting enthalpy, although with the assumptions made (united atom force field, molecules in vacuum) this result is probably fortuitous. As is indicated in Fig. 10D, in the model each adenine amino group forms two intermolecular hydrogen bonds, one with the free oxygens of the phosphate and one with the oxygen O5' of the partner molecule. Such hydrogen bonds probably hinder the rotation of the amino group, which may form an explanation for the broadening observed for the aminoproton resonances at lower temperatures (not shown). We note that the observation of structures in which the bases are intercalated between one another are not unprecedented. In the three dimensional structure of tRNA this structural motif appears several times. These intercalation interactions are believed to enhance the stability of for instance the corner of the tRNA molecule and in this particular case to maintain the interaction between the T-loop and the D-loop. The results obtained in the present paper indicate that the cyclic dinucleotide d<pApA> and analogues thereof may serve as model system to obtain thermodynamic parameters which characterize the energetics of such intercalation interactions which will undoubtedly be of importance in tertiary structure elements of molecules (e.g., 5S-RNA and other ribosomal RNA's) other than tRNA.

Table IV. Torsion angles of the structures of d<pApA> which are discussed in the text and shown in Figs. 9, 10 and 11. The torsion angles are defined in Fig. 1. A is the monomer with both sugars in the N conformation, B is the dimer form, C is the monomer with both sugars in the S conformation, D is the monomer with one sugar with a N conformation and one sugar with a S conformation. The torsion angles are indicated in degrees or in the notation for the classical conformers: g^- is gauche$^-$, g^+ is gauche$^+$ and t is trans. For the latter situation the torsional angles adopt the idealized classical values.

			α	β	γ	δ	ϵ	ζ	χ	P	ϕ_m
A	J-couplings analysis		56	196	60	89	200	74	-	-1	34
	Molecular mechanics		68	191	57	87	200	68	175	13	41
B	J-couplings analysis		59	196	60	92	200	69	-	-7	35
	Molecular mechanics		69	188	57	92	202	64	189	-7	34
C	J-couplings analysis		t	t	g^-	132	230	g^+	-	130	38
	Molecular mechanics		179	169	286	142	262	43	147	147	43
D	J-couplings analysis	N-part	g^+	t	g^+	89	230	g^+	-	1	34
	'' ''	S-part	t	t	g^-	132	230	g^+	-	130	38
	Molecular mechanics	N-part	60	176	51	92	249	61	191	2	36
	'' ''	S-part	177	167	297	127	218	67	214	131	42

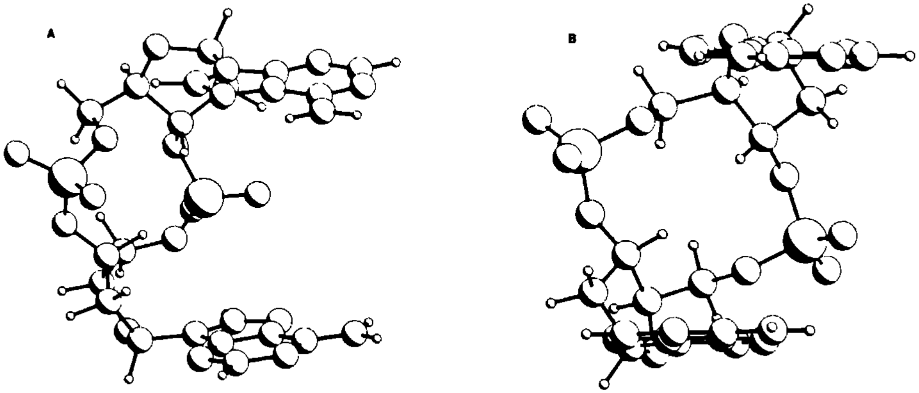


Fig. 9. Ball and stick model of d<pApA>, built according the data in Table IVA, view from two different orientations. A) View which emphasizes the antiparallel orientation of the adenines. B) View, perpendicular to the orientation in (A), in which the form of the cyclic part of the molecule is clearly visible.

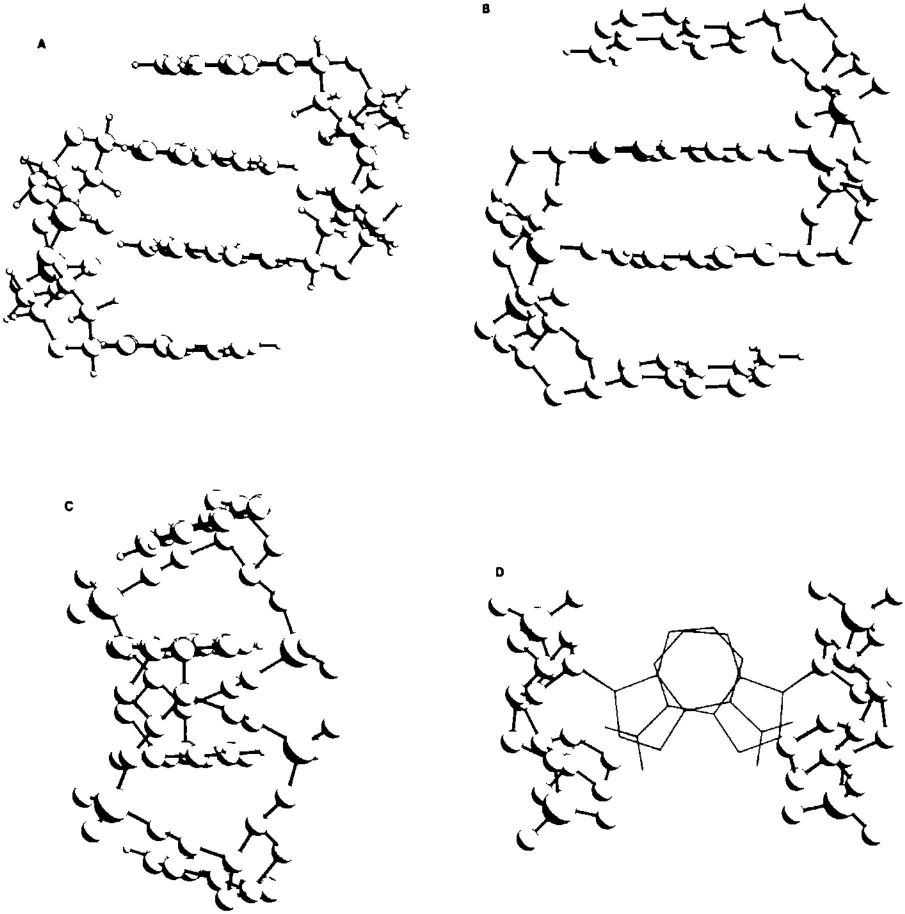


Fig. 10. A) Ball and stick model built for the possible dimer form of d<pApA> in which the adenine bases are intercalated. B) Dimer structure obtained after energy minimizing of the complex in (A) with aid of the program MacroModel. In the calculation a united atom Force Field was used and as a result the non-exchangeable protons are no longer incorporated in the model. In this model the distances between the adenine planes amount to 3.5 Å. C) Perpendicular view of the structure in (B). D) View of the complex in (B) along the twofold symmetry axis running perpendicular to the planes of the central bases. The terminal adenines have been omitted from the drawing. The overlap of the central bases is demonstrated. Moreover, the hydrogen bonding of the amino groups to the phosphate oxygens of the partner molecule (which follows from the MM calculations) is indicated.

It was concluded from the conformational analysis presented in the preceding section that the cyclic dinucleotide is (more or less) conformationally pure at 281 K, while at elevated temperature it is not. This can be understood by considering the proposed monomer-dimer equilibrium. In the dimer form the cyclic dinucleotide is locked in a particular conformation. The monomer, however, can adopt more than one conformation; the most abundant monomer form (78%) appears to be virtually the same as that adopted by d<pApA> in the dimer molecule (Fig. 9). In the second monomer form (the remaining 22 %) the sugars adopt a S-puckered conformation. However, from the NMR data we cannot be sure whether in this monomer form both sugars are S-puckered or one has a S- and the other a N-conformation in this monomer form. Therefore both the symmetric and the asymmetric conformation were

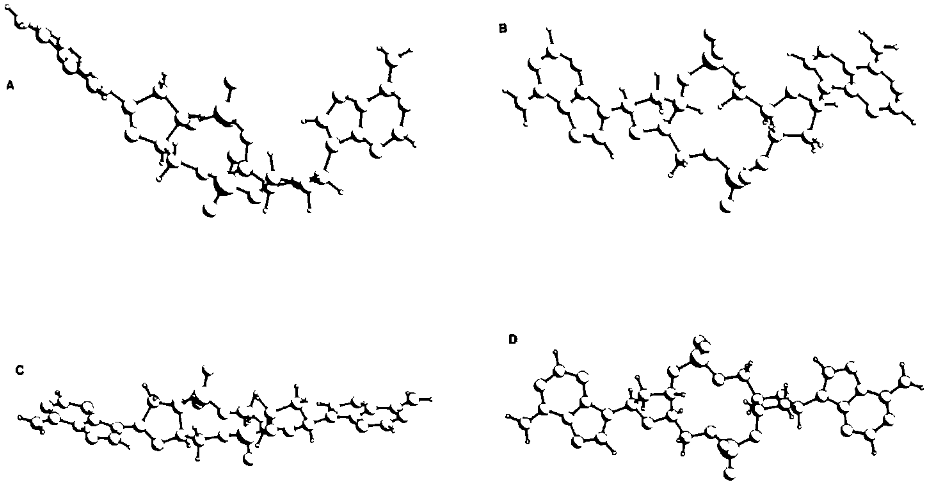


Fig. 11. Energy minimized structures obtained for d<pApA> when one sugar has a S-conformation and the other a N-conformation or both have an S-conformation. A) The asymmetric structure; the sugar in the right part of the molecule is N-puckered. The data listed for structure B and C in Table IV were used as input for the calculations. The results of the energy minimizations are also collected in Table IV. B) View of the asymmetric molecule perpendicular to that exhibited in (A). C) The symmetric structure with both sugars in a S-conformation. Compared to Fig. 9B. the adenine bases have now moved outward. D) View of the symmetric molecule perpendicular to that in (C).

subjected to an energy minimization procedure. To this end a symmetric structure was built with both sugars in the S-conformation and backbone torsion angles represented by structure C in Table IV to conform to the concomitant changes observed in the sugar and backbone conformations (vide supra) when going from a N- to a S- puckered sugar. In addition, an asymmetric structure was built with one side of the cyclic backbone characterized by the torsion angles β^t, γ^-, ϵ^t, and a S-sugar conformation and the other side by β^t, γ^+, ϵ^t, and a N-sugar conformation (structure D in Table IV). The angles α and ζ were calculated according to the procedure outlined above. The structures resulting from the energy minimization (AMBER) are displayed in Fig. 11. The calculated energy of the mixed N/S structure was intermediate between the energy of the NN form and the energy of the SS form. We conclude from these calculations that the mixed form will be favored over the SS form. It is interesting that in these structures the bases connected to a S-type sugar are orientated in such a way that the molecule exhibits a more open form. Thus the changes in sugar pucker have consequences for the conformation of the total molecule. The torsion angles are given in Table IV. The NMR spectra show that a rapid conversion between the different conformations is possible, so that they are most likely not separated by high energy barriers.

REFERENCES

1 Cech, T.R. and Bass, B.L. (1986) Ann. Rev. Biochem. **55**, 599-630.
2 Tabak, H.F., van der Horst, G., Kamps, A.M.J.E., Arnberg, A.C. (1987) Cell **48**, 101-110.
3 Lewin, B. (1983) Genes. John Wiley & Sons Inc., New York.
4 Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A., Weiner, A.M., Molecular biology of the gene (1987) 4th ed. The Benjamin/Cummings Publishing Company Inc., Menlo Park, California.
5 Hsu, C.-Y.J., Dennis, D. (1982) Nucl. Acids Res. **10**, 5637-5647.
6 Hsu, C.-Y.J., Dennis, D., Jones, R.A. (1985) Nucleosides and Nucleotides **4**, 377-389.
7 Ross,P. Weinhouse, H., Aloni, Y., Michaeli, D. Weinberger-Ohana, P., Mayer, R., Braun, S., de Vroom, E., van der Marel, G.A., van Boom, J.H., Benziman, M. (1987) Nature (London) **325**, 279-281.
8 de Vroom, E., Broxterman, H.J.G., Sliedregt, L.A.J.M., van der Marel, G.A., van Boom, J.H. (1988) Nucl. Acids Res. **10**, 4607-4620.
9 Frederick, C.A., Coll, M., van der Marel, G.A., van Boom, J.H. & Wang, A.H.-J. (1988) Biochemistry **27**, 8350-8361.
10 Chazin, W.J., Wuthrich, K., Hyberts, S., Rance, M., Denny, W.A. & Leupin, W. (1986) J. Mol. Biol. **190**, 439-453.
11 de Leeuw, F.A.A.M., Altona, C. (1983) Q.C.P.E. Bull. **3**, 69-70.

12 Davies, E.K. (1986) Chem-X Chemical Crystallography Laboratory. Oxford University. Chemical Design Ltd. Oxford.

13 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C, Alagona, G., Profeta, S. and Weiner, P. (1984) J. Am. Chem. Soc. **106**, 765-784.

14 Still, W.C. (1986) MacroModel. Columbia University. New York.

15 de Leeuw, F.A.A.M., Altona, C. (1982) J. Chem. Soc., Perkin Trans. **2**, 375-384.

16 Remin, M. and Shugar, D. (1972) Biochem. Biophys. Res. Commun. **48**, 636-642.

17 Altona, C. (1982) Recl. Trav. Chim. Pays-Bas, **101**, 413-433.

18 Haasnoot, C.A.G., de Leeuw, F.A.A.M, de Leeuw, H.P.M., Altona, C. (1981) Org. Magn. Reson. **15**, 43-52.

19 Haasnoot, C.A.G., de Leeuw, F.A.A.M., de Leeuw, H.P.M. & Altona, C. (1979) Recueil, J. Roy. Netherl. Chem. Soc. **98**, 576-577.

20 Haasnoot, C.A.G., de Leeuw, F.A.A.M., Altona, C. (1980) Tetrahedron **36**, 2783-2792.

21 Lankhorst, P.P., Haasnoot, C.A.G., Erkelens, C. Altona, C. (1984) J. Biomol. Struct. Dyns. **1**, 1387-1405.

22 Hendrickson, J.B. (1961) J. Am. Chem. Soc. **83**, 4537-4547.

CHAPTER 3

AN INVESTIGATION OF THE STRUCTURE OF THE 3'-5'-CYCLIC DINUCLEOTIDE d<pApA> BY MEANS OF MULTICONFORMER ANALYSIS

A MOLECULAR MECHANICS STUDY

ABSTRACT

The conformational properties of the cyclic dinucleotide d<pApA> were studied by means of molecular mechanics calculations in which a multiconformational analysis was combined with minimum energy calculations. In this approach models of possible conformers are built by varying the torsion angles of the molecule systematically. These models are then subjected to energy minimization; in the present investigation use was made of the AMBER Force field. It followed that the lowest energy conformer has a pseudo-two-fold axis of symmetry. In this conformer the deoxyribose sugars adopt a N-type conformation. The conformation of the sugar-phosphate backbone is determined by the following torsion angles: α^+, β^t, γ^+, ϵ^t and ζ^+. The conformation of this ringsystem corresponds to the structure derived earlier by means of NMR spectroscopy and X-ray diffraction. The observation of a preference for N-type sugar conformations in this molecule can be explained by the steric hindrance induced between opposite H3' atoms when one sugar is switched from a N- to S-type puckers. The sugars can in principle switch from N- to S-type conformations, but this requires at least the transition of γ^+ to γ^-. In this process the molecule obtains an extended shape in which the bases switch from a pseudo-axial to a pseudo-equatorial position. The calculations demonstrate that, apart from the results obtained for the lowest energy conformation, the 180° change in the propagation direction of the phosphate backbone can be achieved by several different combinations of the backbone torsion angles. It appeared that in the low energy conformers five higher order correlations are found. The combination of torsion angles which are involved in changes in the propagation direction of the sugar-phosphate backbone in DNA-hairpin loops and in tRNA, are found in the dataset obtained for cyclic d<pApA>. It turned out that, in the available examples, 180° changes in the backbone direction are localized between two adjacent nucleotides.

INTRODUCTION

In recent years detailed studies of nucleic acid structures have been carried out [1]. These investigations have triggered questions about the relationships between base sequence and structure and have given rise to a very lively field of research [2]. We have been interested in the folding

---

Blommers, M.J.J., Bruins Slot, H.J., van der Marel, G.A., van Boom, J.H. & Hilbers, C.W. (1990) J. Biomol. Struct. Dyns. (in press)

properties of nucleic acids, particularly those of hairpin molecules. Extensive NMR and optical melting studies of folding occurring in synthetic DNA hairpins with four nucleotides in the loop region have shown that the nucleotide sequence and base composition have a marked influence on the loop structure and the hairpin stability. Nevertheless a rather simple model can be used to describe the results [3].

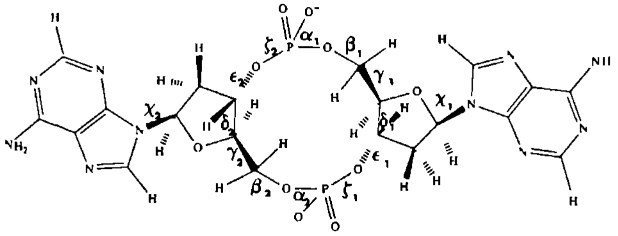


Fig. 1. Structure of the cyclic dinucleotide d<pApA>. The torsion angles of nucleotide 1 and 2 are defined as indicated, according to the IUPAC convention (cf. legend to Table I).

Here we present the results of a study of the dinucleotide d<pApA> in which the 3' and 5' ends of the individual nucleotides are connected to form a circular molecule (Fig. 1). In this molecule the sugar phosphate backbone changes direction by 180° within just a single base step. It is interesting to see by what special structural features this can be accomplished and how these features are related to folding in hairpin loops and to the structure of more 'regular' DNA. The structure of cyclic d<pApA> has been studied before, in independent and parallel investigations, by means of NMR spectroscopy [4] (described in Chapter 2) and X-ray diffraction [5]. In the present chapter we seek to extend the insights, gained in the previous investigations, by studying the molecular structure obtained by minimum energy calculations. In addition, we attempt to correlate these results with those obtained for DNA hairpins in Chapters 4 and 7.

In general, the conformational space of molecules encompasses many local minimum energy structures. Information about the conformations corresponding with the local energy minima including the global energy minimum is essential to obtain insight in the structural features of cyclic d<pApA>. To ensure that these minima are found, proper sampling of the conformational hyperspace is a prerequisite. A powerful method to achieve this goal is the so-called multiconformer analysis [6]. In this approach the internal coordinates of the molecule under investigation are varied systematically to generate a set of different conformations which serve as starting structures for energy minimization. Alterations of the internal coordinates may be restricted to the variation of the torsion angles. The different conformations generated in this manner are subjected to energy minimization to yield local minimum energy conformers which represent a mirror image of the conformational hyperspace provided proper sampling has been conducted during the multiconformation generation step.

The results obtained in this manner show that it is possible to achieve ring closure by more than one combination of the backbone torsion angles. In most of the low energy conformers the deoxyribose sugars adopt a N-type conformation as was found in the earlier NMR and X-ray investigations [4,5]. Comparison of the structures of different conformers shows that S-type conformations can be adopted by the sugars but this requires at least a gauche$^+$ to gauche$^-$ conversion of the torsion angle γ in order to remove the severe steric hindrance between the H3' atoms which occurs upon the transition of N-type to S-type sugars when leaving everything else unchanged. These observations neatly explain the concomitant changes of N-puckered sugars and γ^+-conformation to S-puckered sugars and γ^--conformations in the NMR experiments.

METHODS AND RESULTS

The exploration of the conformational space and the investigation of the minimum energy conformations of the title compound were performed as follows. First, structures were prepared of which the internal coordinates were used as input data for a multiconformer generation. The structures obtained in the latter procedure were then subjected to energy minimization, using a 'united atom' approach, to find the minimum energy conformations. Subsequently, these minimum energy structures were further studied by an 'all atom' energy minimization. The approach is outlined in more detail in Fig. 2 and the results will be discussed according to the scheme in that figure.

Model building

The structure of the cyclic dinucleotide d<pApA> is drawn in Fig. 1. Apart from the bases, the molecule consists of three ring systems, i.e., one twelve-membered sugar-phosphate ring in which two five-membered deoxyribose rings are incorporated.

In the first stage of the analysis (see Fig. 2) three models were generated, which serve as starting structures in the multiconformer generation (vide infra). In the first model both sugar moieties were constrained in a N-type conformation, in the second they were both fixed in a S-type conformation and in the third model one was constrained in a N-type and the other in a S-type conformation. This approach allows an efficient scanning of the conformational space of our molecule. Before the multiconformer analysis was started, these models were subjected to block diagonal Newton-Raphson and full matrix Newton-Raphson energy minimization, using the MacroModel implementation of the united atom of Force Field AMBER [7,8] to remove possible bad contacts between the atoms. The results of these calculations are summarized in Table I where the torsion angles which characterize the different starting structures are listed. It is noted that after this procedure the combinations of NN, SS and NS sugars are retained in the individual molecules.

Table I. Torsion angles of the three energy minimized structures of d<pApA>, designated NN, NS and SS (see text), which served as starting structures for the multiconformer analysis. The angles are defined according to the IUPAC convention, i.e., α = (O3'-P-O5'-C5'), β = (P-O5'-C5'-C4'), γ = (O5'-C5'-C4'-C3'), δ = (C5'-C4'-C3'-O3'), ε = (C4'-C3'-O3'-P), ζ = (C3'-O3'-P-O5'), χ = (O4'-C1'-N9-C4). The values of the angles in the upper row of a conformer, e.g., the NN-conformer, represent one half of the molecule α_1, β_1, γ_1 etc., the second row the other half α_2, β_2, γ_2, etc. (cf. Fig. 1)

	α	β	γ	δ	ε	ζ	χ
NN	163	196	51	83	198	343	219
	73	186	181	92	263	40	191
NS	72	183	61	89	287	290	104
	54	65	181	158	263	40	191
SS	69	185	47	140	293	286	137
	50	62	176	158	269	56	204

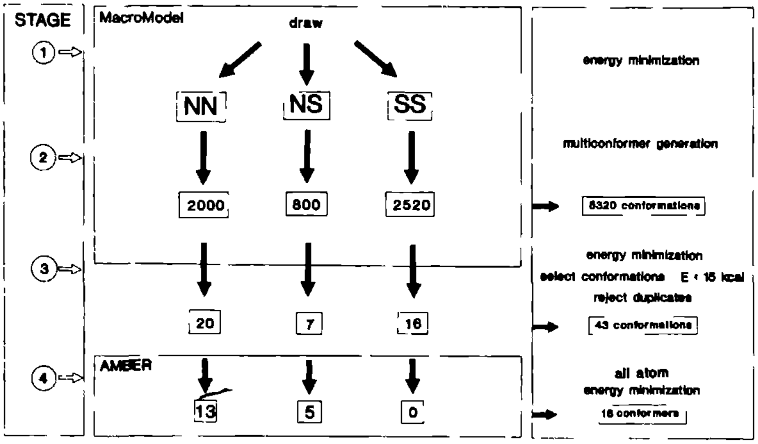

STAGE
①
②
③
④
MacroModel
draw
NN
NS
SS
2000
800
2520
20
7
16
AMBER
13
5
0
energy minimization
multiconformer generation
8320 conformations
energy minimization
select conformations E < 15 kcal
reject duplicates
43 conformations
all atom
energy minimization
18 conformers

Fig. 2. Flow chart indicating the pathways to the low energy conformers. It is noted that the conformers which originate from, e.g., NS starting structures, do not necessarily preserve these sugar conformations. The first stages, which involve the generation of the roughly minimized conformations, were carried out with the aid of MacroModel. The final 'all-atom' refinement was done with the AMBER program (see text).

Multiconformer generation

With the three resulting structures the multiconformer generation was started; this was performed with the MULTIC option of MacroModel (stage 2 in Fig. 2) [7,8]. In the procedure one bond of the cyclic phosphate-sugar backbone, i.e., the O5'-P bond, α_1, is broken temporarily to obtain a pseudo-acyclic molecule (see Fig. 1). During the multiconformer generation this broken bond serves as a closure bond constraint. In this procedure the different conformers are obtained by systematic variation of seven torsion angles in the pseudo-ring, i.e., γ_1, ϵ_1, ζ_1, α_2, β_2, γ_2, ϵ_2, in steps of 60°, and the glycosidic torsion angles χ in steps of 180°, using as starting values the torsion angles of the three, aforementioned, energy minimized structures (see Table I). The torsion angles δ_1 and δ_2 were not changed; they are part of the deoxyribose ring of which the conformation was kept constant during the multiconformer generation step. The three dihedral angles ζ_2, α_1 and β_1, comprising the closure dependent torsion angles were also not varied, but were not kept fixed either. This means that in the different conformers they may adopt any value necessary to achieve ring closure. All torsion angle rotations which result in structures in which the closure bond is outside the range of 0.5-2.5 Å are rejected. In addition, nonbonded rejection constraints of 2.0 Å were applied to eliminate unrealistic conformations. This resulted in a total of 5320 different conformers of which 2000 were obtained from the starting structure with both sugars in the N-conformation, 2520 from the one with both sugars in the S-conformation and 800 from that with one sugar in the S- and the other in the N-conformation (see Fig. 2). These conformers were subjected to 25 steps of block diagonal Newton-Raphson energy minimization using the united atom force field of AMBER as implemented in MacroModel [7] (stage 3 in Fig. 2). Duplicate structures, generated in this step, were then removed from the data set. Conformations were defined as duplicates if all their corresponding torsion angles differed by less than 30°. Also, duplicates arising from the symmetry of the ring system were removed, i.e., if the torsion angles of the first residue in one molecule are equal or nearly equal to the corresponding torsion angles of the second residue in a different molecule and vice versa, these molecules are considered to be duplicates. This reduced our dataset to 2854 conformers. Examination of these roughly refined structures showed that the closure bond retained its normal value, which is an indication that geometrically acceptable conformations are obtained. To the extent that 60° rotations of the ring torsion angles properly sample the conformational space of d<pApA> the present protocol yields a set of different structures, which should provide a good representation of the conformational hyperspace available to our molecule. To give an impression of the dispersion of the internal coordinates of d<pApA>, the values of the different torsion angles α, β, γ, δ, ε, ζ and χ, that are adopted in the different conformers, are plotted in Fig. 3A. It is clear that the values of most torsion angles are dispersed evenly over the available conformational space. This is even true for the torsion angle δ, which cannot adopt values outside the 60-180° range. In some instances the influence of the application of the nonbonded rejection constraint is observable in the torsion angle distribution. For instance, the torsion angle γ is poorly represented around 120°, because of steric hindrance between the C6 and O5' atoms. These structures were removed in the non-bonded cut off tests. The same applies to structures with a S-type sugar conformation for which ε and β are near 120° and 0° respectively. Then the PO_3 group clashes with some of the sugar atoms. For χ the situation is less transparent; we have not systematically checked whether the blank regions in Fig. 3A are solely due to steric hindrance.

From the set of conformers those were selected for further study which had an energy, differing 15 kcal or less from that of the lowest energy structure (Fig. 2). Only 43 out of the total of 2854 conformers fell into this category; the other molecules had energies between 15 and 54 kcal/mol with respect to the lowest energy conformer.

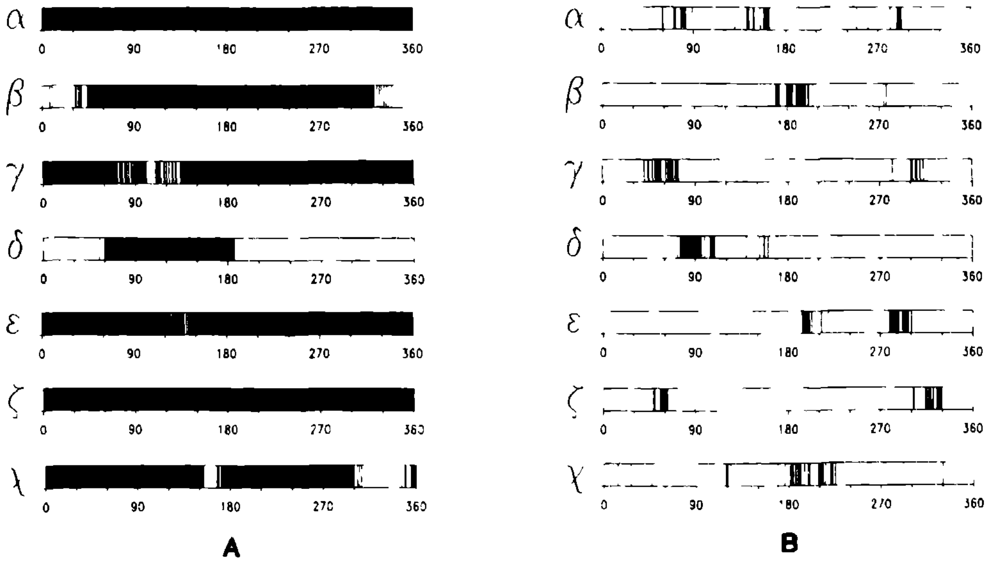


Fig. 3. Distribution of torsion angles present in different conformers of d<pApA> obtained after multiconformer generation and energy minimization. A) Distribution obtained after multiconformer generation and 25 steps of block diagonal Newton Raphson energy minimization (see text). The black regions and vertical bars correspond to torsion angle values adopted in the different conformers. On the given scale most of the individual values of the angles cannot be resolved. B) Distribution obtained for the 18 low energy conformers, i.e., conformers with energies below 15 kcal from the lowest energy conformer. These conformers were obtained with the aid of an additional all atom energy minimization step of 43 low energy conformers obtained under A (see text).

Table II. Torsion angles and energy values (kcal) obtained for the 18 low energy conformations of d<pApA> obtained after an all atom energy minimization. The angles are defined according to the IUPAC convention (see Table I). The conformational parameters of the different conformers are listed in the order of ascending relative energies (E). The energy values are within 15 kcal from the lowest energy state which has been set to zero. The last two columns give the higher order correlation (corr.) of the torsion angles between the sugars defined in Table III and the type of sugar conformations.

nr	E	α_1	β_1	γ_1	δ_1	ϵ_1	ζ_1	χ_1	α_2	β_2	γ_2	δ_2	ϵ_2	ζ_2	χ_2	corr.	sugars
1	0.0	82	195	65	81	194	56	183	79	194	69	77	196	58	221	[1-1]	NN
2	0.0	80	194	68	78	196	57	221	81	194	66	81	195	56	182	[1-1]	NN
3	1.8	71	190	74	91	285	318	211	159	194	54	75	198	49	187	[1-2]	NN
4	3.6	81	197	62	96	295	317	182	290	171	307	85	201	62	210	[1-3]	NN
5	4.1	82	197	64	93	296	318	182	288	170	310	86	201	58	223	[1-3]	NN
6	5.0	77	198	66	92	297	317	225	291	171	306	89	200	59	176	[1-3]	NN
7	5.4	155	179	51	87	279	325	169	148	184	55	86	281	327	201	[2-2]	NN
8	7.4	144	184	57	87	283	325	176	149	181	54	90	279	325	330	[2-2]	NN
9	9.3	249	168	282	94	196	68	188	88	200	68	80	203	62	215	[1-4]	NN
10	9.6	79	200	71	87	293	321	209	289	168	313	83	197	59	14	[1-3]	NN
11	9.8	60	192	54	108	283	302	119	159	277	307	157	213	55	191	[1-5]	NS
12	10.2	71	189	73	91	286	319	47	161	192	54	73	199	51	191	[1-2]	NN
13	10.6	144	185	41	99	293	314	187	142	183	45	94	295	314	199	[2-2]	NN
14	11.0	60	189	49	109	286	303	124	162	274	301	152	219	60	223	[1-5]	NS
15	11.4	282	172	302	104	280	329	213	164	190	45	86	292	325	184	[2-3]	NN
16	11.4	72	205	54	105	294	315	120	293	172	307	161	201	89	190	[1-3]	NS
17	12.5	261	171	283	139	288	321	195	130	192	45	106	300	328	109	[2-3]	SN
18	13.0	59	192	53	109	283	302	121	159	277	307	157	212	56	27	[1-5]	NS

All atom energy minimization

Thus, 43 conformers were prepared for an 'all atom' analysis by addition of hydrogen atoms at the proper sites (it is recalled that the preceding energy minimization involved the united atom approach). Moreover, two 'hydrated' sodium atoms were added, one near each of the two negatively charged phosphate groups of a molecule [8]. The resulting molecular systems were then energy minimized using the AMBER program (version 3.0) [9,10] with the 'all atom' version of the force field [10]. A distance dependent dielectric constant, $\varepsilon = R_{ij}$, was introduced to reduce long range electrostatic interactions. During the minimization some of the molecules form hydrogen bonds between the NH_2 group of adenine and the N1 atom of the other adenine. For this reason we removed the hydrogen bond potential from the Force Field. This resulted in energetically more favorable structures. After 50 cycles of energy minimization the method was switched from steepest descent to conjugate gradient. Minimization was continued until the energy difference obtained after subsequent cycles had reduced to 10^{-7} kcal/mol or the energy gradient was less than 0.1 kcal/mol Å. This procedure yields 18 conformers, that have an energy differing 15 kcal/mol or less from that of the lowest energy conformer. For the other conformers the relative energy values remain higher than 15 kcal. The relative energies and structural parameters of the 18 lowest energy conformers are summarized in Table II.

Torsion angles and their correlations

For the aforementioned 18 low energy conformers the distribution of the torsion angles over the available conformational space is exhibited in Fig. 3B. If we disregard for the moment the angles δ and χ, examination of Fig. 3B shows that for these minimum energy structures all torsion angles adopt values in either the g^+-, t or g^--regions, in other words they occur in the normal energetically allowed regions. This is for instance demonstrated by the values adopted by the torsion angle ε. They occur in the t- and the g^-- but not in the g^+-region just as expected on the basis of earlier observations [11] and energy calculations [12]. On the other hand, it is noted that torsion angle values, which are energetically allowed may not occur in our data set, e.g., the γ^t-conformation.

The values of the angle δ fall around 90° except for five conformers which give rise to angles around 150° in one half of the molecule. This

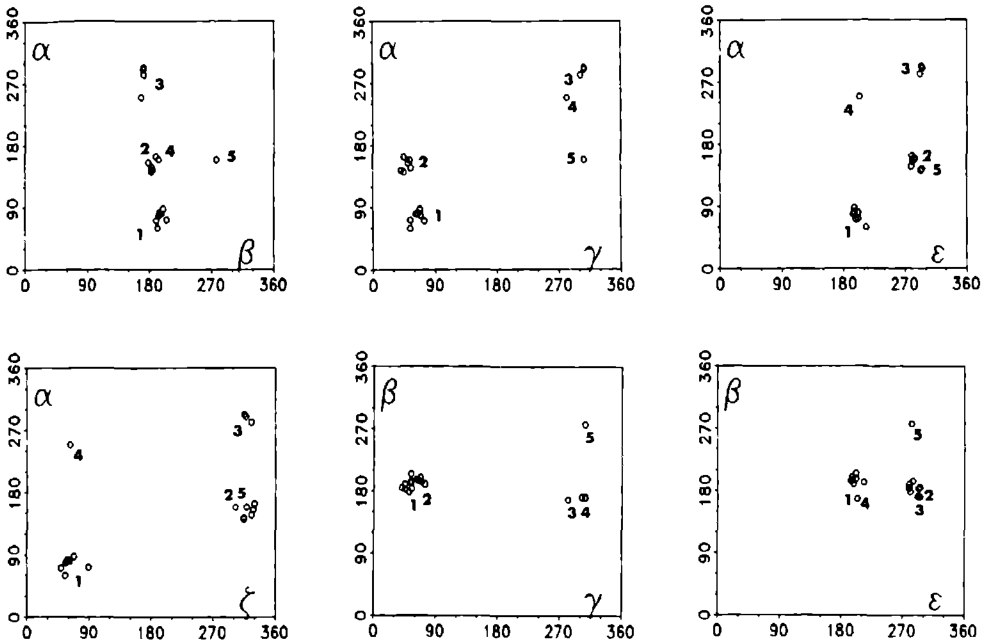
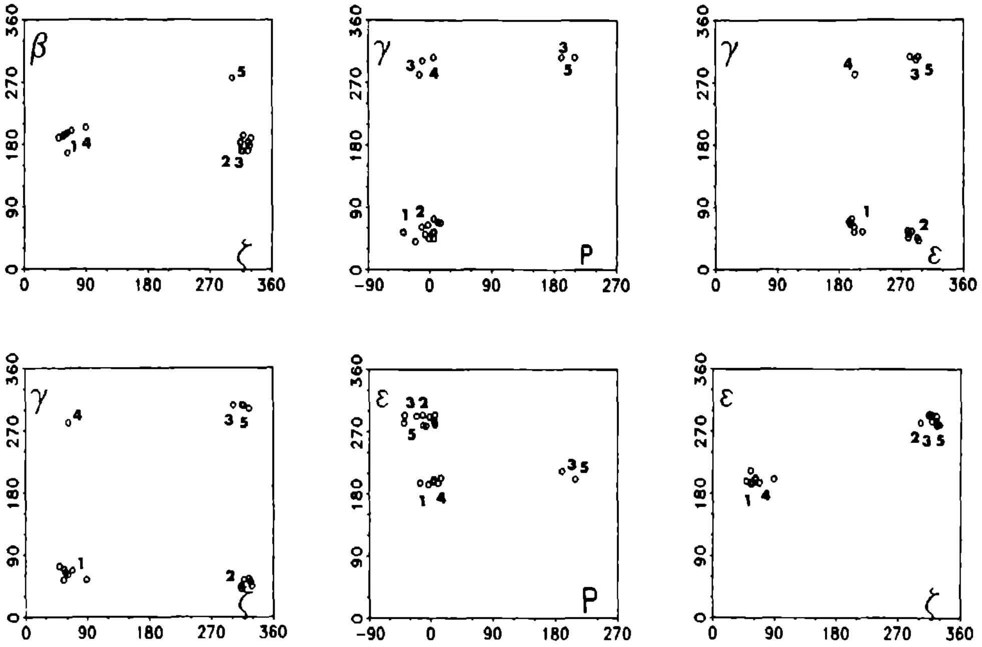


Fig. 4. Correlation plots of various torsion angle combinations obtained for different conformers in Table II, i.e., the conformers numbered 1, 3, 4, 7, 9, 11, 13, 15 and 16. The remaining conformers listed in Table II have not been incorporated because they are essentially duplicates of the ones mentioned above.

means that in the majority of the low energy structures the sugar moieties adopt a N-conformation. The five other conformers contain S-type sugars. An analogous phenomena is observed for the glycosidic torsion angle χ. Most of the χ-values occur near 200°, which means that the bases have a strong preference for the anti-orientation. Four conformers have their base(s) in a syn-orientation, i.e., χ is near 30° for three and at -30° for one molecule. Five bases are found in a high-syn conformation with χ near 120°.

Nucleic acid structures are not only characterized by individual torsion angles but also by the typical combinations of values these angles may adopt or by particular correlations which may exist between these values [12]. For nine of the lowest energy conformations, i.e., for those numbered 1, 3, 4, 7, 9, 11, 13, 15 and 16 (see table II), the torsion angles have been plotted to uncover the possible existence of such relationships.



The other nine conformers are duplicate structures with respect to the 12-membered sugar-phosphate ringsystem, and give therefore no additional information. As can be seen in Fig. 4, e.g., in the plot of α versus β, the latter angle adopts a trans-conformation almost exclusively. Since α has the normally allowed values, no correlation is apparent between α and β. This applies to many other combinations in which β is involved (cf. Fig. 4).

In non-constrained DNA the trans-conformation of β is the lowest energy conformation of this torsion angle [12] and in almost all low energy conformers β adopts this conformation as well. As a result we expect to see a strong correlation between α and γ, because with β in a trans-conformation a crankshaft rotation may be performed in which the distance and the position of the C4' and P5' atoms remain almost unchanged and the P-O5' bond retains the same direction [13]. In A- or B-type helices, switches are observed from α^+,γ^- to α^-,γ^+, while in our case a switch from α^t,γ^+ to α^-,γ^- occurs. In addition, we find the combination α^+,γ^+ and α^t,γ^-. The first is close to the α,γ pairs observed in Z-DNA, the latter is not normally observed. In our case it occurs only when β deviates significantly from the trans-conformation. Within the set of low energy conformers (Table II) these are the conformations with somewhat higher energies.

A distinct correlation is also observed between ζ and ε; when ε is in the t-region, ζ occurs in the g^+-region. On the other hand, when ε is in the g^--region, ζ also falls in the g^--region. We note that the ε^t,ζ^--combination, which is present in A- and B-type DNA does not occur in our dataset.

The combinations found for the phase angle of pseudorotation, P, and γ are also noteworthy. It seems as if the g^+-conformation of γ 'pushes' the sugar exclusively into a N-conformation while the γ,g^--combination is more tolerant, since then both N- and S-type sugars are observed. A similar observation can be made for the ε,P-combination. The ε^--conformation exclusively yields a N-conformation, while ε^t allows both N- and S-conformations.

Further examination of the different conformers shows that also higher order correlations exist for the ring backbone torsion angles. Only five combinations occur for the torsion angles between the sugar residues, i.e., for ε_1, ζ_1, α_2, β_2, γ_2 and for ε_2, ζ_2, α_1, β_1, γ_1. These can be gleaned from Fig. 4 and are listed in Table III . It is interesting to consider the rotations which transfer one combination into the other; they are discussed on the basis of the models drawn in Fig. 5. We start with the lowest energy conformation (Fig. 5 top left; see also Table II). It is a symmetrical

Table III. Sets of torsion angles, e.g., ε_2, ζ_2, α_1, β_1, γ_1, between two sugar residues (cf. Fig. 1) found in the low energy conformers listed in Table II. Only five of these combinations are observed; they are numbered in the most left column. t = trans, + = g^+, - = g^- and -ac = anti clinal.

nr	ε	ζ	α	β	γ
[1]	t	+	+	t	+
[2]	-	-	t	t	+
[3]	-	-	-	t	-
[4]	t	+	-ac	t	-
[5]	-	-	t	-	-

structure of which both sides of the 12-membered ring are characterized by the set of torsion angles in line 1 of Table III. This conformer is therefore designated structure [1-1]. (Henceforth the two halves of the conformers shall be characterized by the line numbers of Table III; each number represents a set of torsion angles). In this molecule the cavity formed by the 12-ring is occupied by the H3' and H5" protons (see Fig. 6). By a concomitant rotation of $\varepsilon^t,\zeta^+,\alpha^+$ to $\varepsilon^-,\zeta^-,\alpha^t$ in one half of the molecule we obtain structure [1-2] (see Fig. 5). As a result the oxygen O1 atom of the phosphate group is turned inward into the cavity of the 12-ring. By this operation an asymmetric structure is generated which can be converted into a symmetric molecule by the same set of rotations in the other half of the conformer yielding structure [2-2]. On the other hand, one can also subject structure [1-2] to a crankshaft rotation so that α^t,γ^+ is converted to α^-,γ^-, giving rise to structure [1-3]. In this conformer O5' turns into the ring cavity from above while O1 rotates out of the cavity

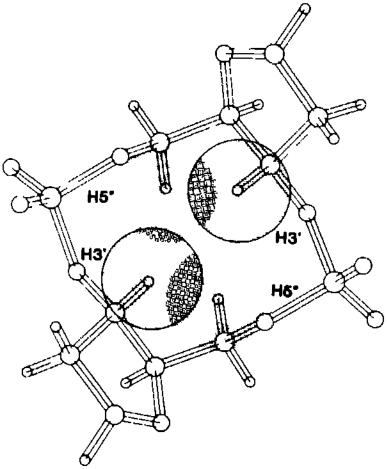
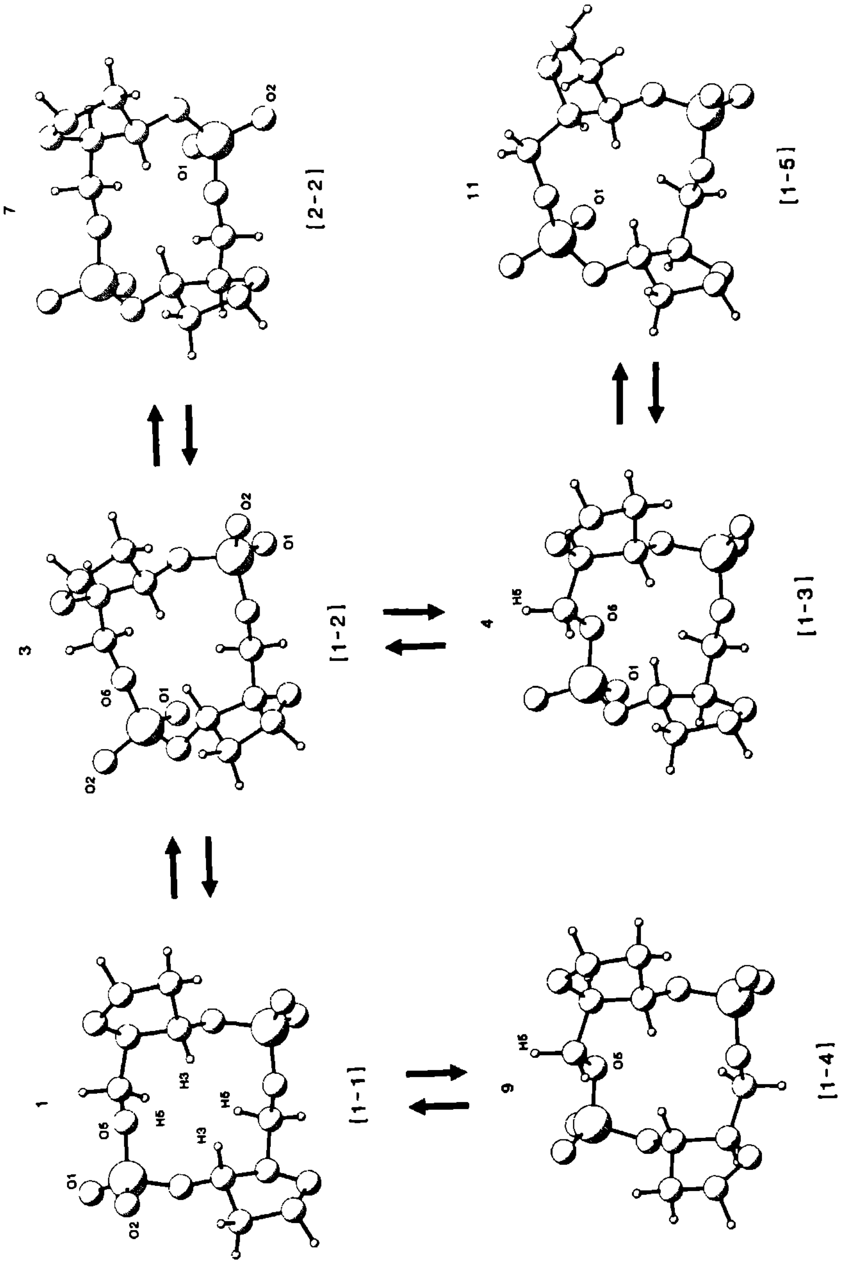


Fig. 6. Ball and stick model of the circular sugar-phosphate backbone of conformer 1 in Table II. The cavity formed by the 12-ring is occupied by the H3' and H5" atoms as is demonstrated by the hatched spheres drawn for these protons; the radius of these spheres was taken equal to the van der Waals radius of the protons. The H3' atoms are directed towards the reader.


1
O1
O2
O5
H5
H3
H3
H5
[1-1]
3
O5
O1
O2
O2
O1
[1-2]
7
O1
O2
[2-2]
9
H5
O5
[1-4]
4
H5
O5
O1
[1-3]
11
O1
[1-5]

from beneath. The conformer [1-3] can be easily converted into the conformer [1-5] by a rotation from α^-,β^t to α^t,β^-. By this operation O1 is reinstalled into the cavity but in comparison with conformer [1-3] a more spacious cavity is created. Finally, it is mentioned that starting from conformer [1-1] the structure [1-4] can be generated by means of a crankshaft rotation involving the transition of α^+,γ^+ to α^t,γ^-; in this manner O5' is turned into the cavity to replace H5".

The three conformers depicted in the upper row of Fig. 5 all have N-type sugars. The formation of S-type sugars is hampered by the fact that the H3' protons, which are situated at the edge of the ring cavity in conformers with N-type sugars, are drawn into its center when the sugars adopt a S-type conformation, thereby giving rise to severe steric hindrance. This can be derived from Fig. 6 where it is indicated how in structure [1-1] the ring cavity is occupied by the H3' and H5" atoms. The H3' atoms protrude somewhat out of the ring in the direction of the reader. Conversion of the sugars into a S-type conformation, i.e., changing δ from ≈ 90° to ≈ 150°, moves these protons into the ring cavity and causes severe clashes. On the other hand, the set of structures in the second row of Fig. 5 provides more room for the H3'-proton. This is the reason that these conformers are found to occur also with S-type sugars; this is for instance the case for conformer [1-3]. The required space is obtained by the rotation of γ^+ to γ^- so that the H5" proton is turned out of the cavity. This conformer allows the presence of N-type as well as S-type sugar conformations. The value of the energy calculated for conformers with a S-type sugar is higher in such conformations since γ has been transferred from g^+ to the energetically less favorable g^-.

Fig. 5. Models of d<pApA> with the different characteristic torsion angle combinations which allow ring closure. The bases, which would be in the reader's direction, have been omitted for reasons of clarity. The number above the models corresponds with the numbering of the conformers in Table II. The numbers below the drawings indicate the combination of torsion angles between the sugars: the individual numbers correspond to the numbering of the rows in Table III (see text). The arrows indicate possible transitions between the conformers as discussed in the text; the atoms changing position in these transitions are indicated.

A transition from N-type to S-type sugars changes the overall appearance of the molecule. In conformers with N-type sugars the bases are oriented more or less in the same direction. This arises from the fact that for this situation the C1'-N9 bond is oriented almost perpendicular to the average plane through the 12-ring. In other words, for a N-type sugar the glycosidic bond has a pseudo-axial orientation. On the other hand, when a sugar is converted from a N- to a S-conformation the glycosidic bond adopts a pseudo-equatorial orientation, which means that the molecule obtains a more extended form. This is illustrated in Fig. 7 where the conformers 4 and 16 from Table II are visualized. The molecules were superimposed so as to obtain an optimal matching between the 12-membered rings. Conformers 4 and 16 belong to the class of structures designated [1-3] which have two N-type or one N-type and one S-type sugar, respectively. Conversion of conformer 4 (thin lines) to conformer 16 (heavy lines) accomplishes the transition from N- to S-type of one of the sugars, leading to extrusion of the attached base. As can been seen in Fig. 7 the base attached to the N-type sugar concomitantly changes from an anti- (182°) to an high-syn (120°) orientation. This change of χ is also observed for the other conformers which have a S-type sugar conformation.

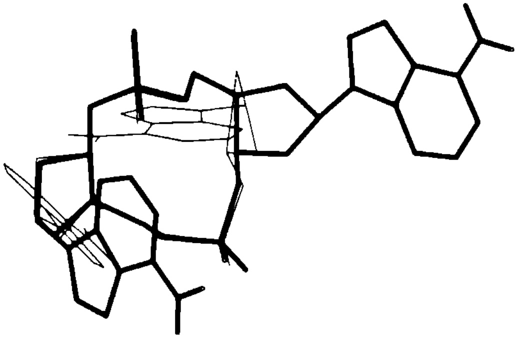


Fig. 7. Visualization of conformer 4 (thin lines) and 16 (heavy lines) from Table II. These molecules correspond to the conformation [1-3] in Fig. 5. The transition of one of the sugars from a N- to a S-conformation leads to an extrusion of the base attached to it.

DISCUSSION

Multiconformer generation

Multiconformer generation methods allow one to scan rapidly through molecular conformational space. Recently, efficient algorithms have been developed to do this systematically for not too large molecules in particular for circular molecules. In this approach torsion angle space is efficiently scanned in a tree search procedure [6,14] and only those conformers are retained which fulfil the input constraints, i.e., in our case the ring closure constraint and the nonbonded rejection constraints. In order to restrict the computer time needed for energy minimization of the conformers obtained after the generation step it is preferred to limit the number of starting structures, i.e., the number of conformers generated by the multiconformer generator, to a minimum and at the same time make sure that proper sampling of the conformational hyperspace is performed. Lipton and Still found that a reasonable compromise between these opposing requirements can be achieved when the torsion angles are varied in steps of 60° and a closure bond distance of 0.5-3.0 Å is introduced [6]. This approach was applied in the present study of d<pApA>. After the conformer generation step the structures were subjected to 25 steps of block diagonal Newton-Raphson energy minimization. Examination of the results showed that the torsion angles in the structures obtained in this procedure were dispersed evenly over torsion angle conformational space (Fig. 3A) suggesting that proper sampling had taken place. In contrast with the earlier studies [6,14] we then proceeded with an all atom minimization of the 43 low energy structures leaving the high energy conformers out of consideration. This saves an appreciable amount of computer time, but implicitly assumes that the energy refinement of these 43 conformers leads to the lowest energy states. We have some indications that, for d<pApA>, this is a reasonable assumption although we can of course not rule out that a continued all atom energy minimization of higher energy conformers does not lead to conformations with energies comparable or lower than those of the 18 conformers following from the aforementioned 43 conformers.

As an example we mention the observation that a conformer is obtained with values for the torsion angles and for the total energy so close to the values obtained for the lowest energy conformer that it can be considered a duplicate of the latter (compare conformer 1 and 2 in Table II) while just prior to stage 3 in Fig. 2 the conformations were entirely different. Another case in point is the NS-conformer built for the monomer form on the basis of available NMR data [4] (vide infra). This structure is not among our lowest energy conformers. Scrutiny of the higher energy structures, i.e., structures with energies higher than those of the aforementioned 43 conformers, showed that a conformation closely resembling this NS-conformer is present among these. It turned out however that continued all atom energy minimization of this conformer leads to a duplicate of structure 9 in Table II, indicating that in the multiconformer generation we did not miss this conformation. It is noted in passing that this convergence did not happen in the earlier calculations carried out in relation to our NMR studies [4]. The only reason we can think of is that the earlier calculations were carried out with the AMBER version 2, while the present calculations were performed with AMBER version 3, which possibly has somewhat different convergence properties.

Structural features

The results show that different conformations, i.e., different combinations of torsion angles, are possible for the twelve-membered ring. Thus, it is more than one set of torsion angles which can accommodate a 180° change in direction of the sugar phosphate backbone through only one nucleotide residue. Comparison of the torsion angles of the lowest energy conformer with those found in the backbone of regular A- or B-type helices shows that they give rise to the same staggered conformations as found in A-type helices except for α and ζ. In the lowest energy conformer these angles fall in the g^+ domain while in A-type helices, or more generally in right-handed helices, both are in the g^- domain. Thus, conversion of α and ζ into the g^+ region is sufficient to induce a complete change in direction of the backbone; this places the bases about 6.8 Å apart, while their planes may remain in a parallel orientation. This was noted before in some model calculations [17].

All possible combinations of ϵ, ζ, α, β and γ, obtained at the very end of stage 2 in Fig. 2, which accommodate a 180° change in the conformation of the phosphate backbone within a dinucleotide step and allow ring closure, are collected in Table IV. Most of these conformers are high energy structures; only five combinations occur for the low energy conformers and these can be converted into one another as has been discussed (vide supra). Several of the combinations listed in Table IV are found in the loop regions of the crystal structure of yeast tRNA^{Phe} [18] and the solution structure of four different synthetic DNA hairpins; the latter were studied by means of NMR spectroscopy [17,19]. Interestingly, the 180° change in backbone direction in the loop regions of these molecules is located in between adjacent residues and the necessary alterations in the torsion angles are not dispersed evenly over the total loop region. Furthermore we note that the turn in the TΨC-loop of the tRNA is characterized by the same combination of backbone angles as conformer [2] of d<pApA> (cf. Table III).

In most of the low energy conformations as well as in the X-ray [5] and NMR structure [4] the deoxyribose sugars adopt a N-conformation which was initially somewhat surprising. Apart from the fact that deoxyribose sugars preferably have S-conformations, the N-conformation also shortens the distance between the phosphate groups from 7 Å (in the S-conformation) to 5.9 Å [20], which adds an unfavorable electrostatic contribution to the energy. Apparently, this is less of a disadvantage than the steric hindrance induced between the H3' atoms when S-conformations are adopted (vide supra).

As discussed already, other combinations of torsion angles that are found in the low energy conformers may accomplish ring closure as well. In this respect it is striking that most of the low energy conformers (Table II) are asymmetric and that in most of these one half of the conformation of the low energy conformer is maintained. For some of the lowest energy conformers (Table II) duplicates are found: e.g., conformer 1 and 2 for [1-1] and conformer 4,5 an 6 for [1-3]. These conformations are practically identical and their existence shows that different starting structures can end in the same energy minimum. For the [1-3] combination, the energy contents increases (about 6 kcal) when one of the bases adopts a syn-conformation (cf. Table II, conformers 4 and 10). An increase of the energy is also obvious for other conformers which change the orientation of one base from anti to syn.

Table IV. Sets of torsion angles as in Table III found for all conformers generated at the end of stage 2 of the multiconformer generation (see Fig. 2). The combinations of torsion angles occurring in turns in tRNA and DNA-hairpin loops (formed by d(C$_m$G-C$_m$GTG-C$_m$G) [17], d(CG-C$_a$TAG-CG) [18], d(ATCCTA-TTTT-TAGGAT and d(ATCCTA-TTTA-TAGGAT) [19]) are indicated. The numbers between the parenthesis indicate the residues between which the turns occurs.

nr	ε	ζ	α	β	γ	observed turn
1	-	+	t	+	+	
2	-	-	-	+	+	junction AC arm D arm (9-10)
3	t	+	+	t	+	d<pApA> (I); conformer [1]
4	t	+	t	t	+	
5	t	t	+	t	+	D-loop (15-16)
6	t	t	-	t	+	
7	t	-	+	t	+	AC loop (33-34)
8	-	+	+	t	+	
9	-	+	t	t	+	
10	-	+	-	t	+	
11	-	t	-	t	+	
12	-	-	+	t	+	
13	-	-	t	t	+	TΨC loop (55-56); conformer [2]
14	t	+	+	-	+	
15	t	+	t	-	+	
16	t	+	-	-	+	
17	t	t	+	-	+	
18	t	t	-	-	+	
19	t	-	+	-	+	
20	t	-	-	-	+	
21	-	+	-	-	+	
22	-	t	t	-	+	
23	-	t	-	-	+	
24	-	-	+	-	+	
25	t	+	+	+	t	
26	t	+	-	+	t	
27	t	t	+	+	t	
28	t	t	-	+	t	
29	t	-	+	+	t	
30	t	-	-	+	t	d(C$_m$G-C$_m$GTG-C$_m$G) (5-6)
31	-	+	+	+	t	
32	-	+	t	+	t	
33	-	+	-	+	t	d(CG-C$_a$TAG-CG) (5-6)
34	-	t	t	+	t	
35	-	t	-	+	t	
36	-	-	+	+	t	
37	-	-	-	+	t	
38	t	+	+	t	t	d(ATCCTA-TTTT-TAGGAT) (9-10)
39	t	+	t	t	t	
40	t	+	-	t	t	
41	t	t	-	t	t	
42	t	-	-	t	t	

nr	ε	ζ	α	β	γ	observed turn
43	-	+	+	t	t	d(ATCCTA-TTTA-TAGGAT) (9-10)
44	-	+	t	t	t	
45	-	+	-	t	t	
46	-	-	+	t	t	
47	-	-	-	t	t	junction T arm AA arm (46-47)
48	t	+	+	-	t	
49	t	+	-	-	t	
50	t	t	+	-	t	
51	t	-	t	-	t	
52	-	+	+	-	t	
53	-	+	t	-	t	
54	-	+	-	-	t	
55	-	-	t	-	t	
56	t	+	+	+	-	
57	t	+	t	+	-	
58	t	+	-	+	-	
59	t	t	t	+	-	
60	t	t	-	+	-	
61	t	-	-	+	-	
62	-	+	t	+	-	
63	-	t	t	+	-	
64	-	-	+	+	-	
65	t	+	+	t	-	
66	t	+	t	t	-	d<pApA> (II); conformer [4]
67	t	+	-	t	-	
68	t	t	t	t	-	
69	t	t	-	t	-	
70	t	-	t	t	-	
71	t	-	-	t	-	
72	-	+	+	t	-	
73	-	+	t	t	-	
74	-	+	-	t	-	
75	-	t	t	t	-	
76	-	+	t	-	-	
77	-	-	t	t	-	D-loop (17-18)
78	-	-	-	t	-	conformer [3]
79	t	+	+	-	-	
80	t	+	t	-	-	
81	t	t	+	-	-	
82	t	-	+	-	-	
83	t	-	t	-	-	
84	-	+	+	-	-	
85	-	+	t	-	-	
86	-	+	-	-	-	
87	-	t	+	-	-	
88	-	t	-	-	-	
89	-	-	t	-	-	conformer [5]

the twelve-membered rings of the two conformers. The positioning of the adenine rings is however quite different. This may result from an enhanced tendency to form hydrogen bonds in the energy minimized structure (double arrows in Fig. 8B): the calculations were carried out in absence of water molecules which may overemphasize the interaction between donor and acceptor groups in the adenines. This interaction may influence the conformation in other parts of the molecule, e.g., may induce the relatively high value of the sugar pucker amplitude, which results in a shift of the C1' atoms towards each other, allowing for a change in the direction of the adenine groups.

In solution d<pApA> is found in a dimer as well as in a monomer form [4]. The monomer occurs in at least two conformations, one of which is almost identical with the conformation of the molecules incorporated in the dimeric species. The other has a more extended shape; one or both sugars are S-puckered, while the torsion angle γ falls in the g^--region and the angle ε has increased its value. From the results obtained in this paper it is now clear that γ has to convert into a g^- value before a S-puckered sugar can be

Table V. The internal coordinates of the structure of d<pApA> determined by means of X-ray crystallography and NMR are compared with those of conformers [1-1] and [1-4]. Because the first conformations have a pseudo-axis of symmetry the torsion angles given are averaged; for the last two conformations the two lines refer to the two different halves of the molecule. a) values of the torsion angles derived directly from experimental data. b) refined structure of NMR(I) obtained with the aid of AMBER 2.0. c) low abundance conformer; torsion angles derived from experimental data.

	α	β	γ	δ	ε	ζ	χ	P	ϕ_m
X-ray	67	196	54	95	202	63	181	-2	31
NMR^a [I]	59	196	60	92	200	69	-	-7	35
NMR^b [I]	69	188	57	92	202	64	189	-7	34
[1-1]	81	195	67	79	195	57	202	4	44
NMR^c [II]	g^+	t	g^+	89	230	g^+	-	1	34
	t	t	g^-	132	230	g^+	-	130	38
[1-4]	88	200	68	80	203	62	215	-15	40
	249	168	282	94	196	68	188	15	41

Comparison with X-ray and NMR structures

The structure of the cyclic dinucleotide d<pApA> had been investigated before with the aid of X-ray diffraction [5] and nuclear magnetic resonance spectroscopy [4]. The conformation of the form predominantly present in solution appeared to be very similar to that derived by means of X-ray crystallography. In Fig. 8A, the crystal structure (thin lines) and the solution structure (heavy lines) are superimposed. It is clear that the conformation of the 12-membered sugar-phosphate backbone of the structures is almost identical. This also follows from a comparison of the torsion angles observed in the two structures (see Table V). The most outstanding difference between the two models is the positioning (indicated by the arrows in Fig. 8A) of the adenine bases, which has been noted before [5]. In Fig. 8B, the lowest energy structure obtained in the present study is represented by thin lines and it is compared with the X-ray structure (thick lines). Again there is good agreement between the conformations of

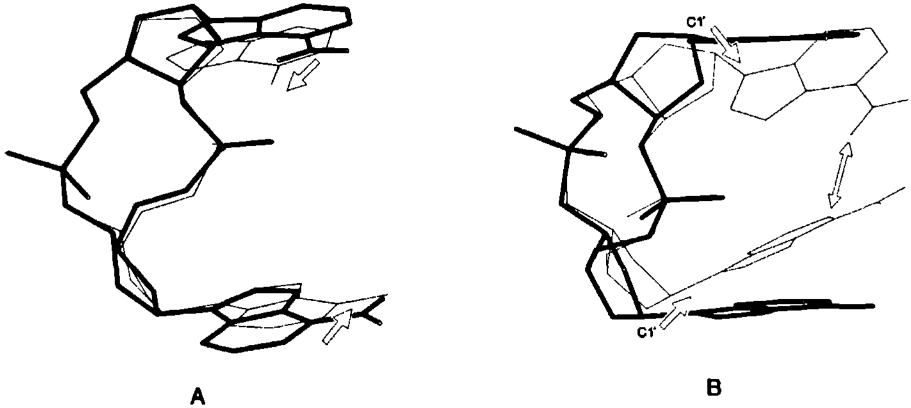


Fig. 8. A) Superposition of the X-ray structure (thin lines) and the NMR structure (thick lines). Coordinates were taken from references 5 and 4, respectively. The structures are almost identical only the exact positions of the bases, indicated with arrows, differ. B) Superposition of the lowest energy conformer (Table II, thin lines) and the crystal structure (thick lines). The conformations of the 12-rings are similar, but the positions of the adenine bases differ resulting from the formation of a hydrogen bond (indicated by a double arrow). The tendency to form such hydrogen bonds is a property of the force field calculations in which solvent molecules are not included.

formed and that a S-type sugar orients the base in a pseudo equatorial position thereby giving rise to the extended shape. If the aforementioned changes occur in only one half of the molecule, we obtain a conformation which resembles that of conformer 9 in Table II (structure [1-4] in Fig. 5). With respect to this extended conformation it is worth mentioning that in the crystal structure of d<pApA> one residue was found to be disordered. The estimated sugar conformation was deviating significantly from a pure C3'-endo conformation. The observed disorder suggests that the equilibrium present between the two conformations in the solution- structure is also of importance in the crystal.

REFERENCES

1. Saenger, W. (1984) Principles of Nucleic Acid Structure, Springer Verlag, Heidelberg.
2. Structure and Expression, **2** (Eds. Sarma, M.H. & Sarma, R.H.) and **3** (Eds. Olson, W.K., Sarma, M.H., Sarma, R.H. & Sundaralingam, M.) (1988) Adenine Press, N.Y.
3. 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) Biochemistry **28**,7491-7498.
4. Blommers, M.J.J., Haasnoot, C.A.G., Walters, J.A.L.I., van der Marel, G.A., van Boom, J.H. & Hilbers, C.W. (1988) Biochemistry **27**, 8361-8369.
5. Frederick, C.A., Coll, M. van der Marel, G.A., van Boom, J.H. & Wang, A.H-J. (1988) Biochemistry **27**,8350-8361.
6. Lipton, M. & Still, W.C. (1988) J. Comp. Chem. **9**, 343-355.
7. Still, W.C. (1986) MacroModel, Columbia University, New York.
8. Singh, U.C., Weiner, S.J., Kollman, P.A. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 755-759.
9. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P. (1984) J. Am. Chem. Soc. **106**, 765-784.
10. Weiner, S.J., Kollman, P.A., Nguyen, D.T. & Case, D.A. (1986) J.Comp. Chem. **7**, 230-235.
11. Lankhorst, P.P., Haasnoot, C.A.G., Erkelens, C. & Altona, C. (1984) J. Biomol. Struct. Dyns. **1**, 1387-1405.
12. Pearlman, D.A. & Kim, S.-H. (1986) J. Biomol. Struct. Dyns. **4**, 69-98.
13. Olson, W.K. (1982) Nucl. Acids Res. **10**, 777.
14. White, D.N.J. & Kitson, D.H. (1986) J. Mol. Graphics **4**, 112-116.
15. Sarma, R.H. (1980) in: 'Nucleic Acid Geometry and Dynamics', Ed. Sarma, R.H., Pergamon Press, NY, 1-45.
16. Westhof, E. & Sundaralingam, M. (1986) Biochemistry **25**, 4868-4878.
17. Orbons, L.P.M., van Beuzekom, A.A. & Altona, C. (1987) J. Biomol. Struct. Dyns. **4**, 965-987.
18. Pieters, J.M.L. (1989) Thesis. State University Leiden, The Netherlands.
19. Chapter 4 and 7 of this thesis.
20. Rich, A., Quigley, G.J. & Wang, A.H.-J. (1980) in: 'Nucleic Acid Geometry and Dynamics', Ed. Sarma, R.H., Pergamon Press, NY, 273-288.

CHAPTER 4

NMR STUDIES OF LOOP FOLDING IN A DNA HAIRPIN MOLECULE

AN INVESTIGATION OF THE SOLUTION STRUCTURE OF d(ATCCTA-TTTT-TAGGAT)

ABSTRACT

The hairpin formed by the partially self-complementary oligonucleotide d(ATCCTA-TTTT-TAGGAT) was studied by means of high-resolution NMR. Application of 2D-NMR allowed an almost complete assignment of the ^{1}H-NMR spectrum. On the basis of the sequential assignments it is concluded that the stem of the hairpin is a B-type double helix. The stacking of the double helical stem is continued into the loop-region, i.e., starting from the 5'-end the first two bases in the loop continue the stacking pattern of the stem; the third nucleotide partly overlaps with the second. The fourth nucleotide closes the remaining gap; its base is turned inward and forms a non-canonical T-T pair with the first base in the loop. To check whether this loop folding is energetically feasible we performed a series of molecular mechanics calculations. The coordinates of the energy minimized structure were used to predict a 2D-NOE spectrum for the hexadecanucleotide. In this way a qualitative agreement between the cross peak intensities of the simulated and experimental spectrum could be obtained. The conformation of the 180° turn in the sugar-phosphate backbone between the third and the fourth loop residue was further examined by means of a multi-conformational analysis. These calculations corroborate the proposed loop structure. The turn in the phosphate backbone is accomplished with the combination of torsion angles ε^t-ζ^+-α^+-β^t-γ^t. The results show that the earlier formulated loop folding principle which states that the loop structure is dictated by the structure of the stem, is operative in this hairpin.

INTRODUCTION

The last decade has witnessed an outburst of NMR studies of nucleic acid structure and dynamics. Initially such investigations were aimed at the study of exchange properties of imino protons involved in hydrogen bonding in base pairs, at the determination of the melting properties of double helical molecules and at resolving the conformational properties of the sugar moieties in small oligonucleotides. Detailed structural studies were

---

Blommers, M.J.J., Haasnoot, C.A.G., Hilbers, C.W., van Boom, J.H. & van der Marel, G.A. (1987) Structure and Dynamics of Biopolymers, NATO ASIS Series E: Applied Sciences **133**, 78-91.

Hilbers, C.W., Blommers, M.J.J., van de Ven, F.J.M., van Boom, J.H. & van der Marel, G.A. (1990) CECAM Workshop (in press).

hampered by the inability to interpret the complicated spectra which are obtained for medium sized and larger nucleic acid molecules. In the last five years this situation has changed dramatically. The introduction of NMR spectrometers operating at 500 MHz and the development of two dimensional Fourier Transform methods have increased the resolution of the NMR method tremendously. The concomitant development of assignment procedures and the possibility to measure short range distances, albeit in a qualitative way, between protons in a molecule, have made the NMR technique one of the most powerful methods to study the structure of molecules in solution.

Assignments of resonances in nucleic acid molecules have concentrated on the identification of imino proton resonances and on the interpretation of pyrimidine and/or purine H6 and H8 resonances via mutual NOE connectivities with H1' and H2'/H2" resonances. Once one has succeeded in assigning these resonances, identification of the remaining proton signals may follow relatively easy. Assignments of the first category have solved a long standing problem in the NMR spectroscopy of nucleic acids, namely, the identification of the imino proton resonances in tRNA's [1-6]. The second approach [7] has played a crucial role in detailed structural studies of well defined DNA fragments in solution [8,9].

In this Chapter the application of the aforementioned methods will be illustrated for the partly self-complementary oligonucleotide d(ATCCTA-TTTT-TAGGAT). It was established elsewhere that this molecule forms a hairpin structure with a melting temperature of 53 °C at 0.2 M NaCl [10-12]. We will first consider 2D-NOE spectra obtained for this molecule and use these for the interpretation of the spectrum. Subsequently, the assignment of the resonances and NOE cross peaks will be used to derive a preliminary structure for the hairpin. The results will be compared with the folding principle derived for loop formation in hairpin molecules and with the results of a molecular mechanics calculation carried out for d(ATCCTA-TTTT-TAGGAT) [13]. The change in the direction of the sugar-phosphate backbone, which is necessary to form a hairpin structure, is located between the third and fourth residue in the loop region. The conformation of this dinucleotide is analyzed with the methods developed in Chapter 7.

MATERIALS AND METHODS

The hexadecanucleotide was synthesized via the phosphotriester method [26]. NMR samples were prepared by dissolving the oligonucleotide to a concentration of 3 mM in a D_2O solution containing 25 mM sodium phosphate, 1 mM sodium cacodylate and 0.1 mM EDTA at pH 7.0. NaCl was added to 200 mM sodium concentration. In addition, a similar H_2O sample (5 % D_2O) was prepared with pH = 6.0.

500 MHz ^1H-NMR experiments were performed on a Bruker WM500 spectrometer interfaced to an Aspect 3000 computer. NOESY spectra were recorded with a spectral width of 5000 Hz, using a mixing time of 75, 125, 200, 300, 400 and 500 ms respectively. All experiments involving the D_2O samples were performed at 294 K. All spectra were acquired with 512 points in the t_1- and 4K points in the t_2-direction with a spectral width of 5000 Hz. The ω_1-scaled DQF-COSY spectrum was recorded with a scaling factor of 2. The pulse sequence and phase list is given in Chapter 2 (Fig. 2). A NOESY and 1D-NOE difference spectra of the sample dissolved in H_2O were recorded at 277 and 274 K, respectively. Suppression of the H_2O resonance was achieved using a semi-selective acquisition pulse. The carrier was placed at the edge of the spectrum, down-field from the imino proton signals. A spectral width of 20000 Hz was used. The residual signal from the solvent was further reduced by means of data shift accumulation [18].

The multi-conformational analysis performed for the loop residues, comprising the 180°-turn in the phosphate backbone, is described in the Materials & Methods section of Chapter 7.

RESULTS AND DISCUSSION

Assignment of non exchangeable protons in d(ATCCTA-TTTT-TAGGAT).

Assignment of sugar backbone and base ring protons is possible on the basis of NOESY spectra recorded in D_2O. Standard sequential assignment procedures are available, which make use of the fact that the distances between the base ring protons H6 and H8 of pyrimidines and purines and the sugar H1' and H2'/H2" sugar protons are short enough (i.e., less than 4 Å) that NOE effects can be observed between the resonances of these protons [14-16]. This is illustrated in Fig. 1 which displays the distances between the mentioned ring protons and the H1' and H2'/H2" sugar protons of the sugar attached to the base and of the sugar of it's 5'-neighbor found in a

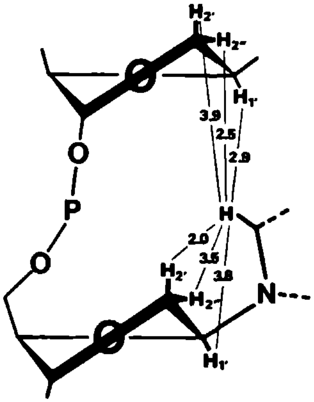


Fig. 1. Average distances between the purine H8 and pyrimidine H6 proton and the H1'/H2'/H2" deoxyribose protons in B-DNA helices [17]. These distances should not be interpreted in an absolute way as the Arnott model is necessarily only an approximation to the true helix conformation at the atomic level. The data serve to show that the purine H8 or pyrimidine H6 proton is not only close to it's intra-nucleotide 1', 2' and 2" protons but also to the 1', 2' and 2" protons of it's 5'-neighboring nucleotide. Therefore, in a 2D NOE experiment intra- as well as inter-nucleotide cross peaks will be observed for the base protons of an oligonucleotide (see Fig. 2).

B-DNA helix. This observation is used in the sequential assignment procedures. Briefly, starting with the cross peak between the base ring proton (H6 or H8) and the sugar H1' proton resonance of the residue at the 5'-end of the sequence, one proceeds by looking for the cross peak between this H1' sugar resonance and the ring proton resonance of the next residue in the 3'-direction. Once this cross peak is found, the procedure is repeated and in this way one proceeds through the nucleotide chain. A similar approach is possible when starting from the 3'-end of the molecule. This procedure allows the assignment of the sugar H1' and base H6 and H8 resonances to particular residues in the sequence.

For d(ATCCTA-TTTT-TAGGAT) this approach is illustrated in Fig. 2. Starting with the H8-H1' cross peak for A1, we can assign the ring and sugar H1' proton resonances until we arrive at the loop residue L3 for which we can assign the H1' resonance, but we are not able to make the connection to the H6 resonance of the loop residue L4. On the other hand, when we start at the 3'-end with the resonances of T1 we are able to proceed to the H6 resonance of L4 mutatis mutandis we are not able to proceed to the H1' resonance of L3. The results suggest that the stacking of the bases present in the stem is transferred to the bases in the loop, i.e., proceeding in the 3'-direction L1 and L2 stack in a more or less helical fashion on A6 and L4 on T6. The situation with respect to L3 is somewhat less clear since the cross peak between the H1' resonance of L2 and the H6 resonance of L3 strongly overlaps with the cross peak between the H1' and the H6 resonance of T2 (c.f. Fig. 2).

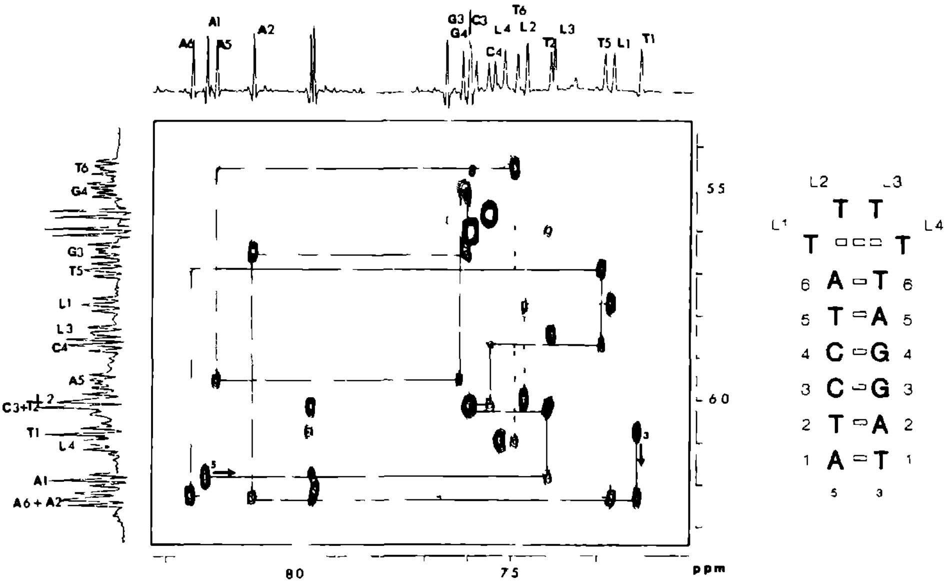


Fig. 2. Part of the 500 MHz ^{1}H NMR NOESY spectrum of d(ATCCTA-TTTT-TAGGAT) recorded with a mixing time of 0.3 s at 27 °C. Cross peaks between aromatic base proton resonances (7.1-8.3 ppm) and H1' proton resonances (5.4-6.3 ppm) are shown. In the assignment of the resonances the numbering is used as indicated for the hairpin structure at the right hand side. The sequential assignment of the cross peak is indicated. Starting with the cross peak between the 5'-terminal A1H8 and A1H1' resonances, a solid line connects all cross peaks belonging to neighboring residues in the stem. The sequential assignment can be extended up to the third base in the loop as indicated by the dashed line. Analogously, the cross peaks from the other strand in the stem can be assigned by starting from the 3'-end as indicated. The sequential assignment can be extended up to the base L4. The connectivity pathway in the loop is again indicated by a dashed line.

A similar approach can be followed with respect to the connectivities between the base ring proton and the H2'/H2" sugar resonances. Because of overlap of cross peaks the analysis cannot be extended into the loop as far as for the H1' resonances, but by a combination of the results with those obtained for the H1' resonances it is nevertheless possible to assign all H2' and H2" resonances. Discrimination between H2' and H2" resonances can be achieved by means of the DQF-COSY (J-splitting patterns of the cross peaks) and the NOESY spectra (intensity differences of cross peaks at shorter mixing times): all H2' resonances are found at higher field compared with the H2" resonances of the same sugar residue.

Having identified the pyrimidine H6 resonances it is now straightforward to analyze the spectrum of the methyl groups of the thymidine residues. In turn these resonances can be used to support the structural features of the hairpin already apparent from the analysis of the H1' and H2'/H2" resonances, i.e., in B-DNA the distance between the thymine methyl groups and the H6 or H8 proton of the residue at the 5'-side is sufficiently short to give rise to cross peaks in a NOESY spectrum. The observed cross peaks corroborate the findings already available from the sequential assignments described above, i.e., the stem of the hairpin forms a B-type double helix, the thymidine L1 is stacked upon A6, L2 is stacked upon L1 and T6 on L4. There is no connectivity between L4 and L3 indicating that the disruption of the regular stacking pattern in the loop is between L3 and L4 as was already indicated by the results of the sequential assignment procedure. The results

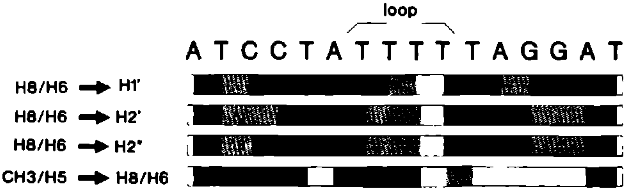


Fig. 3. Summary of the sequential NOEs (dark boxes) observed in the NOESY spectra of d(ATCCTA-TTTT-TAGGAT) between base ring proton resonances and sugar proton or methyl resonances. The hatched boxes indicate situations in which cross peaks between resonances of neighboring nucleotides could be observed but some uncertainty remained due to overlap. Note the absence of cross peaks between the loop residues L3 and L4, which indicates a perturbation of the B-type stacking pattern in this part of the molecule.

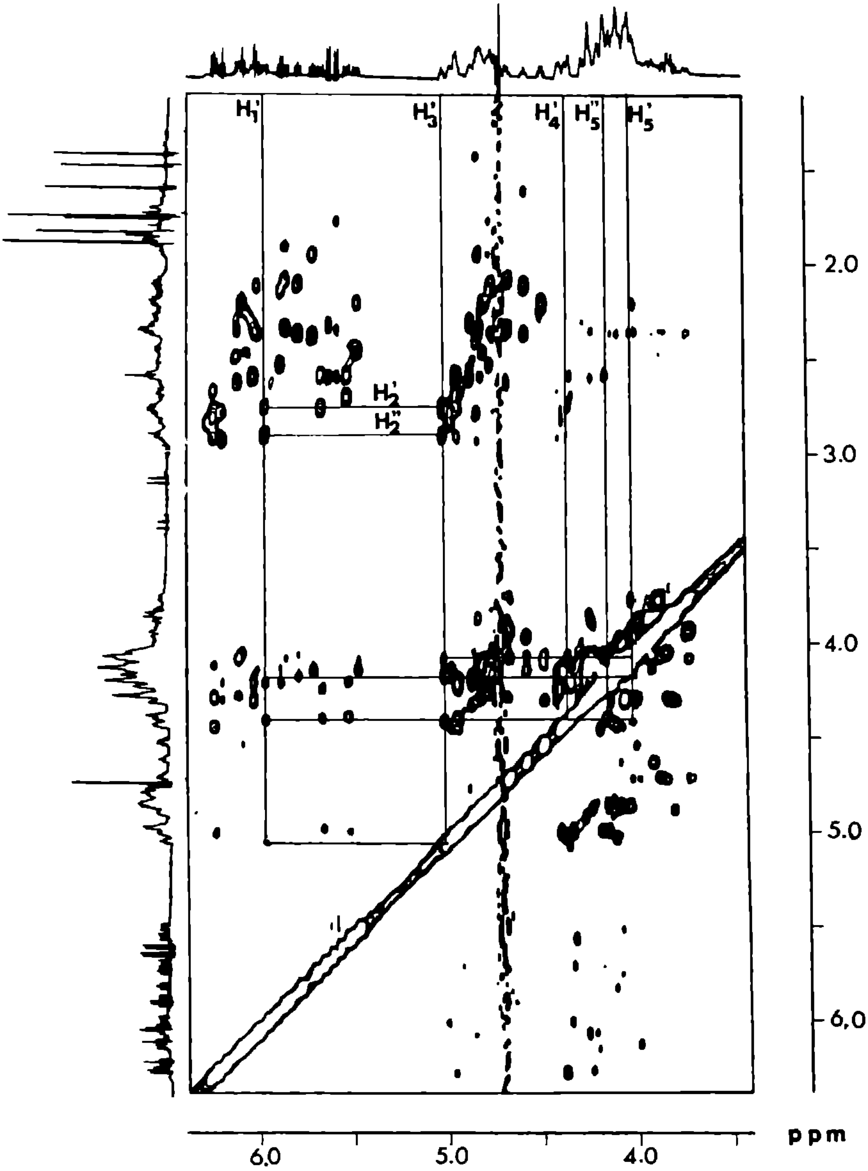


Fig. 4. Part of the 500 MHz ^{1}H NMR NOESY spectrum (ω_1 = 1.15 to 6.28 ppm, ω_2 = 3.43 to 6.37 ppm) of d(ATCCTA-TTTT-TAGGAT) recorded with a mixing time of 0.3 s at 27 °C. The assignment of the sugar protons of residue A5 is indicated.

Table I. Chemical shifts of the ^{1}H-resonances of d(ATCCTA-TTTT-TAGGAT) at 27 °C. Chemical shifts (ppm) are relative to DSS. The shifts between parenthesis are not entirely unambiguous, due to spectral overlap.

Residue	H6 H8	H5 CH_3 H2	H1'	H2'	H2"	H3'	H4'	H5'	H5"
A1	8.23	8.00	6.22	2.73	2.87	4.86	4.25	3.80	3.80
T2	7.44	1.36	6.06	2.24	2.54	4.89	4.23	(4.12)	(4.12)
C3	7.62	5.64	6.05	2.24	2.53	4.83	(4.25)	(3.85)	(3.85)
C4	7.57	5.60	5.91	2.09	2.46	4.79	4.15	(4.17)	(4.08)
T5	7.31	1.69	5.74	1.86	2.32	4.85	4.08	(4.08)	(4.08)
A6	8.28	7.68	6.26	2.70	2.76	5.00	4.38	(4.23)	(4.09)
L1	7.28	1.68	5.82	2.03	2.30	4.78	4.12	(4.15)	(4.01)
L2	7.48	1.78	6.04	2.02	2.31	4.60	4.10	3.92	4.02
L3	7.42	1.54	5.88	2.02	2.26	4.68	4.00	3.89	3.74
L4	7.54	1.84	6.14	2.29	2.56	4.70	4.22	4.00	3.82
T6	7.50	1.83	5.49	2.14	2.40	4.82	4.11	(4.09)	(4.03)
A5	8.23	7.64	6.00	2.69	2.85	5.03	4.38	4.10	4.17
G4	7.65	--	5.55	2.52	2.65	4.97	4.40	(4.16)	(4.16)
G3	7.62	--	5.69	2.52	2.71	4.96	4.35	(4.18)	(4.18)
A2	8.14	7.99	6.27	2.62	2.86	4.97	4.38	(4.18)	(4.18)
T1	7.23	1.42	6.12	2.12	2.15	4.51	3.98	(3.98)	(3.75)

of these assignments are summarized in Table I. The inter-residue connectivities obtained for the hexadecanucleotide are schematized in Fig. 3.

Having obtained an identification for the H1' and H2'/H2" sugar resonances one may attempt to assign the rest of the proton resonances to the individual sugar residues. This is of some interest because eventually this may lead to a definition of the sugar conformation once after the interpretation of the sugar resonances the J-couplings between the proton spins can be determined. In most instances we succeeded in obtaining a complete identification of the sugar proton resonances. An example of the assignment of the sugar resonances of residue A5 is presented in Fig. 4. It is noted that spin diffusion effects between H1' and the other spins such as H3', H4' and H5' or H5" in sugar ring very often facilitate the assignment procedure; this is also clear for the A5 sugar (see Fig. 4). The assignments of the sugar resonances have been confirmed by means of a TOCSY experiment (not shown). The results are given in Table I. The assignments of the resonances which, for the reason of spectral overlap, are not unambiguous are placed between parenthesis. In addition to the above discussed connectivities three other cross peaks are worth mentioning. At ω_1,ω_2 = 1.84, 5.88 ppm a cross peak is observed between the methyl resonance of residue L4 and the H1' resonance of L3. This is the only cross peak observed between a methyl resonance and an H1' resonance and forms an indication of the peculiar structure in the loop formed by the L3 and L4 residues (vide infra). Furthermore we note the occurrence of cross peaks between L3 CH_3 and L2 H3' and between L3 CH_3 and L2 CH_3 indicating that although L3 is stacking on L2 the stacking pattern may deviate from that found in normal B-DNA.

The results obtained so far merely give a qualitative description of the loop structure. For a more detailed characterization additional data are required. In this respect, it is interesting to mention the occurrence of some additionally observed NOE cross peaks between protons of loop

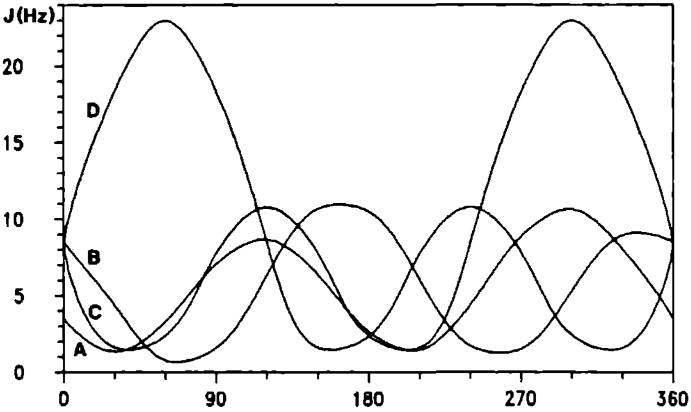


Fig. 5. The torsion angle dependence of the coupling constants J4'5' (A), J4'5" (B), J5'P (C) and J5"P (D), calculated with Eqs. 1 and 2. $J_{4'5'}$ and $J_{4'5''}$ correspond to γ; $J_{5'P}$ and $J_{5''P}$ correspond to β. This plot can be used to derive these torsion angles, when these torsion angles are conformationally pure.

residues. It turned out that the cross peaks between H6 of L4 and its H5'/H5" sugar proton resonances persist even at a mixing time of 75 ms when such effects are no longer observed for the other nucleotides. This indicates that the torsion angle γ of the O5'-C5'-C4'-C3' bond of this sugar moiety corresponds to a gauche$^-$ or trans conformation instead of the normally observed gauche$^+$ conformation. The sugar phosphate backbone can be described in further detail by examining the torsion angles γ and β on the basis of the J-coupling information present in the double-quantum filtered COSY (DQF-COSY) spectrum. The value of the ^{1}H-^{1}H coupling constants $J_{4'5'}$ and $J_{4'5''}$ and the H-P coupling constants $J_{5'P}$ and $J_{5''P}$, which can be obtained from the cross peaks in this spectrum, are a function of the torsion angles γ and β, respectively. They can be calculated using the well known Karplus equations:

$$J_{HCCH} = 13.22 \cos^2 \phi - 0.99 \cos \phi + \Sigma(0.87 - 2.46 \cos^2 (\xi_i\phi + 19.9 \mid \Delta\chi_i \mid)) \Delta\chi_i \qquad (1)$$

and

$$J_{HCOP} = 15.3 \cos^2 \phi - 6.1 \cos \phi + 1.6 \qquad (2)$$

(cf. Chapter 2). For convenience the above mentioned J-couplings are plotted as a function of β and γ in Fig. 5. In standard B-DNA, the torsion angle γ adopts a gauche$^+$ conformation and the $J_{4'5''}$ coupling constant is small. This means that the corresponding cross peak is weak or absent in the DQF-COSY spectrum. For the fourth loop residue, however, an unusually strong cross peak is observed between the H4' and H5" resonances. This is indicated in Fig. 6A (cross peak enclosed in a box, indicated by an arrow), where part of the DQF-COSY spectrum is plotted. This cross peak is shown in more detail in Fig. 6B together with the pattern simulated for this connectivity. The best correspondence between measured and simulated pattern was obtained with the following set of coupling constants: $J_{4'5''}$ = 10.3 Hz, $J_{3'4'}$ = 0.8 Hz, $J_{4'5'}$ = 2.6 Hz, $J_{5'5''}$ = -12.0 Hz and $J_{5''P}$ = 3.1 Hz. These values of the $J_{4'5''}$ and $J_{4'5'}$ coupling constants correspond with a predominance of the γ^t conformation (cf. Fig. 5) in agreement with the conclusions obtained from the NOE measurements (vide supra). In addition, it can be concluded from the observation that the $J_{5'P}$ and $J_{5''P}$ constants are small and that the torsion angle β is strongly biased towards the trans conformation.

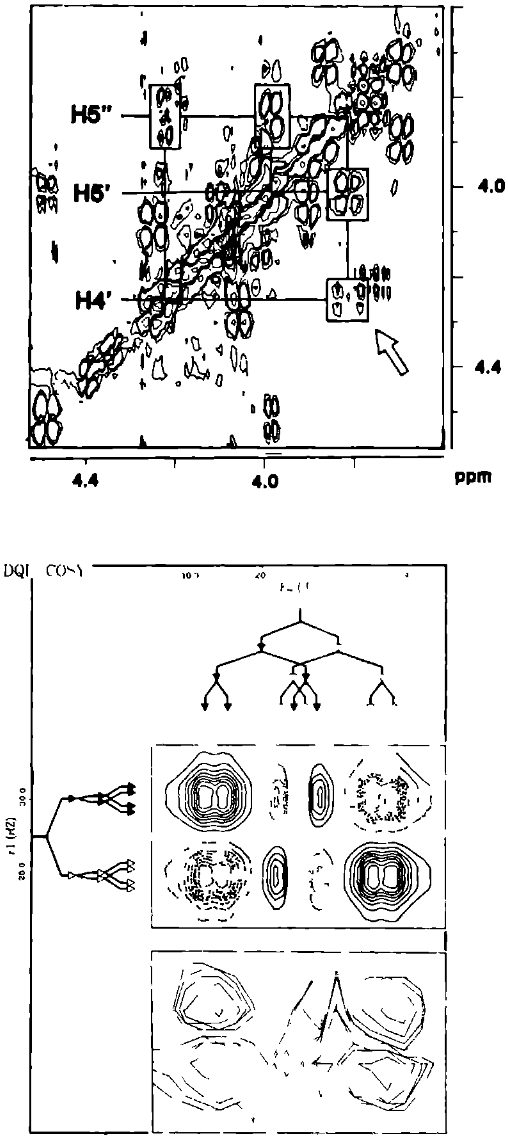


Fig 6. A) Region of the DQF-COSY spectrum in which the H4', H5' and H5" protons resonate. The cross peaks corresponding to the L4 residue are boxed. A strong cross peak is observed between H4' and H5", indicated with an arrow. This cross peak is enlarged (B) and compared with a simulation. The splitting of the signal by the various couplings is indicated. A large active J4'5" coupling splits the cross peak in positive and negative absorption peaks.

Exchangeable protons.

When d(ATCCTA-TTTT-TAGGAT) folds into a hairpin, six canonical Watson-Crick base pairs are formed giving rise to six imino proton resonances when the molecule is dissolved in a H_2O solution. Six such resonances are indeed observed but, in addition, four extra imino proton resonances are present around 11 ppm which must arise from the imino protons of the thymidines in the loop (see Fig. 7). Assignment of the imino protons in the stem were made by means of 1D NOE experiments [18]. Pre-saturation of one particular imino proton resonance should give rise to Overhauser effects at the imino proton resonance of the neighboring base pairs. An example is provided by the experiment presented in Fig. 7. The resonance from GC4 at 12.7 ppm is irradiated and Overhauser effects are observed at 13.7 ppm for the resonance from AT5 and at 12.6 ppm for the resonance from GC3. Using the same approach, we have tried to assign the imino proton resonances of the loop thymidines. Pre-saturation of the high field resonance (intensity corresponding to one proton) at 10.7 ppm gives rise to an Overhauser effect at the imino proton resonance position of base pair AT6 and a larger one at the position of the other loop imino proton resonances (at 10.8 ppm). This suggests that the presaturated resonance is from one of the thymidines (L1 or L4) neighboring the AT6 base pair. The sequential assignments of the non-exchangeable protons, discussed above, indicated that both L1 and L4 are stacked on the AT6 base pair so that indeed an Overhauser effect between the imino proton resonances of AT6 and that of L1 and/or L4 may be expected. L1 and L4 being stacked on base pair AT6 might easily lead to the formation of a T-T base pair involving two hydrogen bonds as indicated in Fig. 7. For such a base pair we expect a large, i.e., 10-20 % Overhauser effect when one of the imino proton resonances is irradiated. This is indeed what is observed in Fig. 7 indicating that such a base pair is present in the loop structure of the hexadecanucleotide. This interpretation is nicely confirmed when an NMR melting experiment is conducted (see Fig. 8). Elevation of the temperature leads to broadening and shifting of the imino proton resonances. The interpretation of these effects has been discussed elsewhere [10]; here we want to point out that at 291 K we obtain two resonances of equal intensity which originate from signals with an intensity ratio of 3:1 at 271 K. In view of the results presented above the most plausible interpretation of this observation is that the imino protons of the thymidines L2 and L3

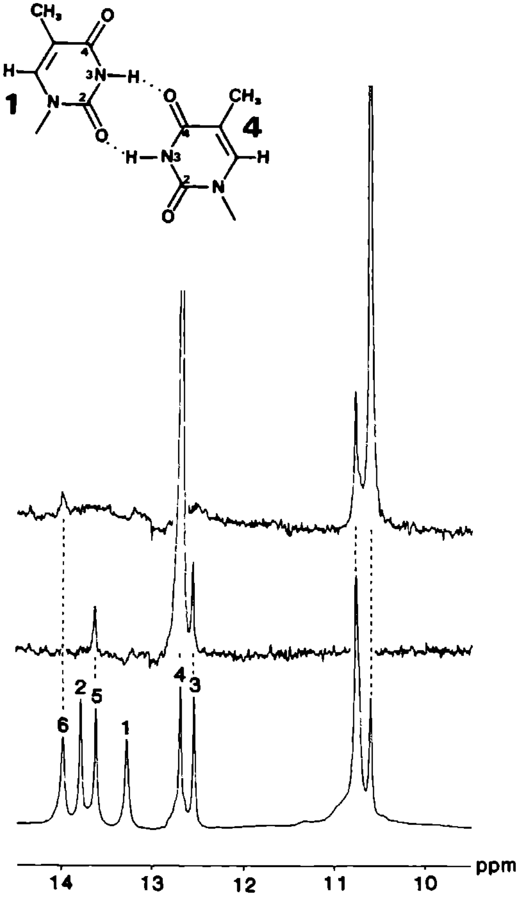


Fig. 7. 500 MHz ^{1}H NMR 1D NOE experiments performed with d(ATCCTA-TTTT-TAGGAT) at 0 °C. The lower trace is the reference spectrum exhibiting the imino proton resonances arising from the base pairs (between 12 and 14.5 ppm) and from the thymidines in the loop (between 10 and 11 ppm). The numbers refer to the base pair numbering given in Fig. 1. The middle trace represents the NOE difference spectrum obtained after presaturation (during 0.3 s) of the imino proton resonance of base pair GC4. NOEs are seen for the imino proton resonances of GC3 and AT5. The upper trace represents the NOE difference spectrum obtained after presaturation of the high field resonance from the loop residues. A weak NOE is observed for base pair AT6 and a much larger NOE (about 15 %) for the imino proton resonance from another loop residue. This result is interpreted to indicate the existence of a T-T base pair between residues L1 and L4 stacked upon the double helical stem of the hairpin (see text).

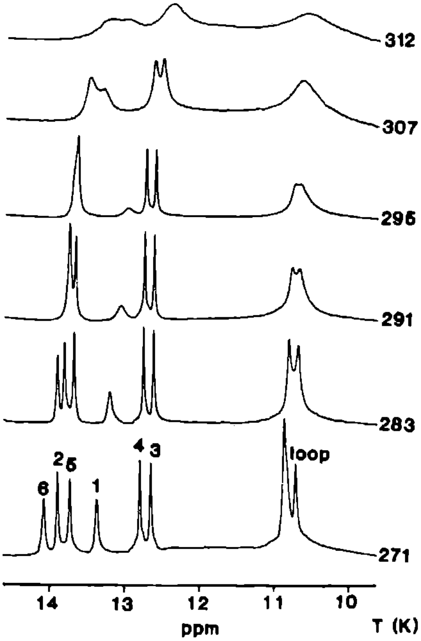


Fig. 8. The 500 MHz imino proton spectrum of d(ATCCTA-TTTT-TAGGAT) recorded as a function of temperature. We note that at 291 K the intensity of the loop imino proton signal has decreased by an amount corresponding to two proton resonances. The remaining signals assigned to the T-T pair disappear more or less concomitantly with the other imino proton signals (see text).

exchange rapidly with H_2O at 291 K while the T-T pair remains intact. As judged from the spectra measured at higher temperatures this base pair melts concomitantly with the base pairs in the stem.

At this point the only resonances that have not been assigned are those from the adenine C2H moiety and those arising from the amino groups. In one dimensional NOE difference spectroscopy these signals can be identified in principle after presaturation of assigned imino proton resonances [1,19]. NOESY experiments of the sample dissolved in H_2O allow a more systematic approach. For the hexadecanucleotide, this is shown in Fig. 9. For the CG pairs, cross peaks can be observed between imino proton resonances and cytidine amino proton resonances. In turn, we see cross peaks between the cytidine amino proton signals and between the cytidine amino protons and the C5H resonance. Note that also a connectivity is observed between the imino protons and the C5H resonance which is caused by spin diffusion. The latter cross peaks provide an independent check for the correctness of the assignments since the position of the C5H resonance was already established (vide infra). The assignment of the three adenine amino proton resonances is also indicated in Fig. 9. It turns out that the assignment of the guanine amino proton signals is not very well possible. In these base pairs the amino group still has an appreciable amount of rotational freedom which at 4°C leads to a strong broadening of the amino proton signals. Apart from the mentioned exceptions we have now come to the point that a complete assignment of the proton spectrum of d(ATCCTA-TTTT-TAGGAT) has been achieved.

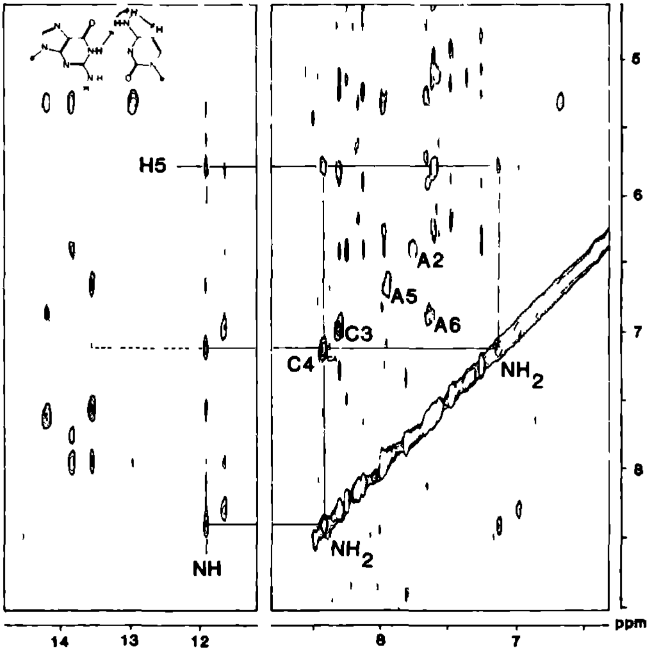


Fig. 9. Part of the 500 MHz ^1H NMR NOESY spectrum of d(ATCCTA-TTTT-TAGGAT) dissolved in H_2O recorded with a mixing time of 0.3 s. at 4°C. Water suppression methods used to obtain this spectrum have been described elsewhere [18]. Two spectral regions are connected, i.e., ω_1 = 4.5 to 9.0 ppm, ω_2 = 11.2 to 14.8 ppm, containing imino proton cross peaks, and ω_1 = 4.5 to 9.0, ω_2 = 6.3 to 8.8 ppm, containing amino and base ring proton cross peaks. The solid lines connect connectivities between the imino proton resonance of G4 (see Fig. 1) and the amino- and C5H resonances of C4. The latter connectivities can only be observed when sufficient spin diffusion is permitted; the spin diffusion pathway is indicated in the base pair (top left). The cross peaks observed for the amino proton resonances of C3, A2, A5 and A6 are also indicated. The resonances of amino protons which are involved in hydrogen bonding are shifted down field from those of the non-hydrogen bonded amino protons.

Conformational features of the hairpin formed by d(ATCCTA-TTTT-TAGGAT).

In earlier studies [10,12] we have found that the stability of DNA hairpin loops as a function of loop size is at its maximum for hairpin molecules with four or five residues (i.e., thymidines) in the loop. This is in contrast to RNA hairpins for which an optimal stability is found for six or seven nucleotides in the loop [20,21]. To provide a structural explanation for this behavior, we have recently advocated that the loop folding in hairpin structures is guided by the propagation of the base stacking pattern of the stem into the loop (Chapter 5). Then the "natural" way to fold a loop in B-DNA hairpins is to extend the stacking of the nucleotides at the 3'-end of the duplex stem. The hexadecanucleotide d(ATCCTA-TTTT-TAGGAT) has been subjected to conformational studies in order to uncover whether this mechanism is operative in this molecule.

It is well established that the conformational characteristics of the sugar moiety may serve to expose the structural genus of the nucleic acid molecule under study, i.e., more generally, whether it has A- or B-type structure. Giving enough J-coupling constants between the proton spins on the sugar ring a detailed conformational analysis of the sugar ring is possible [21]. For the hexadecanucleotide sufficient accurate data for such an analysis are not available. However, in COSY spectra cross peaks between H2" and H3' are lacking for all sugars, which indicates that the J-coupling between these protons is small (about 1 Hz). In addition the sum of the J-couplings, i.e., $J_{1'2'} + J_{1'2''} > 13.6$ Hz for all nucleotides except the terminal ones. This is characteristic of a predominantly S-type sugar conformation which is representative of B-type folding of the DNA fragment [23]. This conclusion is augmented by the observation that in 2D NOE spectra only an intra-nucleotide cross peak between H6/H8 and H2' and not between H6/H8 and H2" is observed at short mixing times (75 ms). Therefore we may conclude that the double helical stem has a B-type structure which is propagated into the loop region.

A more detailed picture is obtained if we incorporate the results of the preceding section. There it was shown that the stacking of the bases in the stem is extended in the 3'-direction into the loop involving L1 and L2 and possibly L3. The fourth thymidine L4 stacks on top of T6 and favors the formation of a T-T pair.

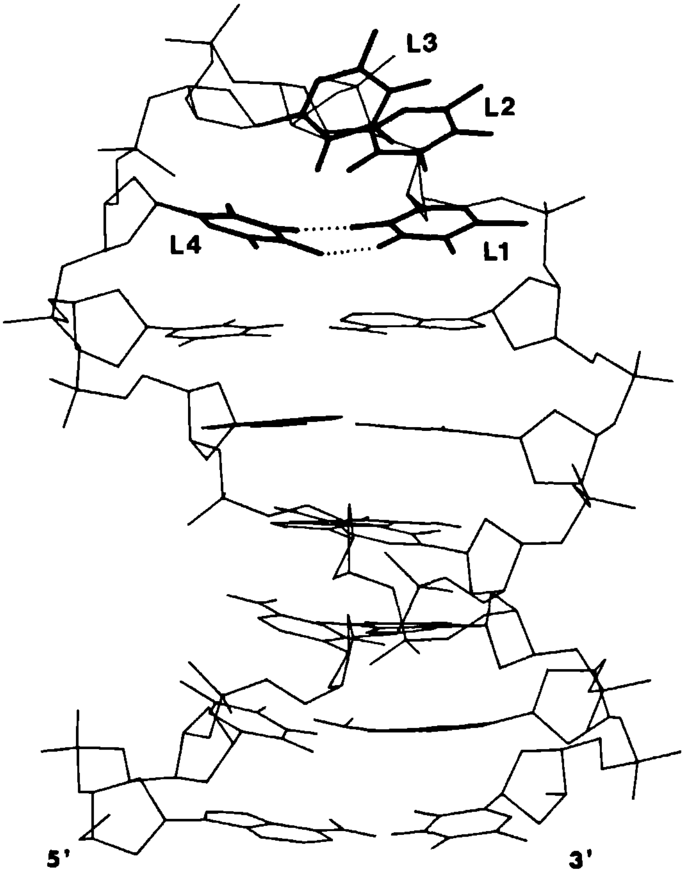


Fig. 10. Structure obtained after energy minimization; use was made of the AMBER program developed by Kollman and collaborators [24,25]. Note that after energy minimization the stacking pattern of the double helical stem is extended into the loop structure in the 3'-direction. The first two bases in the loop, L1 and L2, stack in a more or less B-type helical fashion; the third base L3 is partly stacked on L2 and partly on L4. Base L4 is turned inward into the loop thereby forming a T-T base pair with L1, as exhibited in Fig. 7. The T-T base pair in the simulated structure is somewhat buckled.

In order to check whether this base overlap is energetically feasible performed a series of molecular mechanics calculations. In the hairpin structure, used to start the calculations, the bases L1 and L2 were placed on top of a regular B-DNA stem so that the stacking pattern of the double helical stem was extended in the 3'-direction as prescribed by our folding principle [13]. The dinucleotide formed by L3 and L4 was docked into the remaining gap between L2 and T6. The torsion angles γ and β of L4 were constrained to a trans conformation (see preceding section). After an 'all

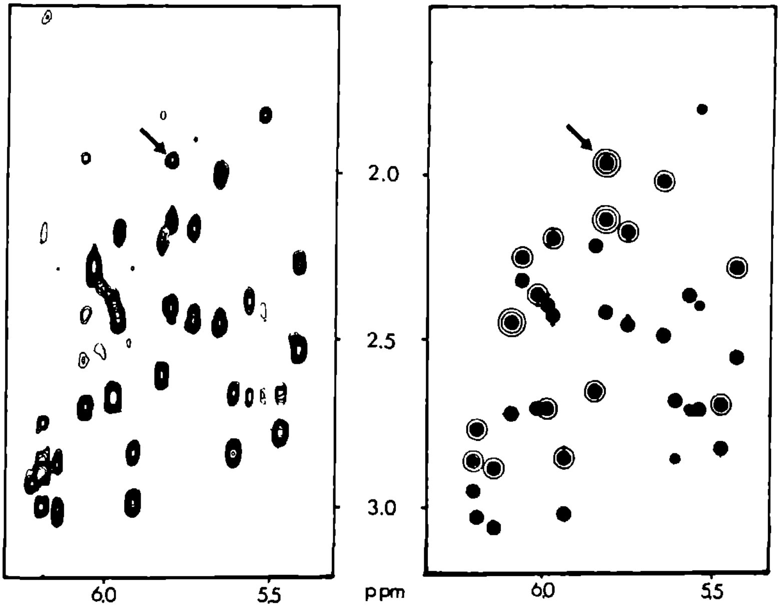


Fig. 11. Part of the 500 MHz ^{1}H NMR NOESY spectrum (ω_1 = 1.5 to 3.2, ω_2 = 5.3 to 6.4 ppm) exhibiting connectivities between H1' and H1'/H2" resonances. The connectivity indicated by an arrow is the exceptional cross peak between the methyl resonance of base L4 and the H1' resonance of base L3 (see text). Cross peaks in the left box are part of the experimental NOESY spectrum, cross peaks in the right box are part of a theoretical NOESY spectrum simulated on the basis of the energy minimized structure presented in Fig. 10.

atom' energy minimization the structure sketched in Fig. 10 is obtained. It is clear that the base stacking scheme derived from the NMR data is also found after the minimization procedure. The bases L1 and L2 continue to stack upon the double helical stem in a B-type fashion. Base L3 is partly stacked upon L2 and partly on L4; the latter base is turned inward to form a T-T pair with L1. This base pair formation nicely reproduces the NMR results. Generally, the structure of the loop (Fig. 10) is rather tight; all bases are turned inward. This is in good agreement with the observation that the exchange of the imino protons is considerably slowed down with respect to the exchange of imino protons of free thymidines. That the loop thymidines are turned inward also follows from binding experiments between d(AAAA) and the hexadecanucleotide [10]. Even at NMR concentrations and low temperature no evidence of binding between these molecules can be obtained. Only after the loop has been extended to 6 thymidines complex formation becomes readily observable.

In order to allow a further comparison between the simulated structure in Fig. 10 and the NMR results, coordinates derived from the energy minimized structure were used to predict a 2D NOE spectrum for the hexadecanucleotide, using the program NOESIM (Chapter 7). Part of the results of this calculation (i.e., the spectral region ω_1 = 1.5 to 3.2 ppm and ω_2 = 5.4 to 6.4 ppm) is presented in Fig. 11 together with the experimental results. Scrutiny of the two spectra shows that the cross peaks are reproduced by the theoretical calculation, albeit a perfect match of the intensities has not yet been obtained. The prediction of a cross peak at ω_1 = 1.84, ω_2 = 5.88 ppm, which arises from the close proximity of the methyl group of L4 and the H1' sugar proton of L3, is particularly interesting. As was mentioned above, in the experimental spectrum a cross peak is observed for this set of proton spins and apparently the energy minimized structure also neatly reproduces this aspect of loop folding.

Multi-conformational analysis

After the results described in the previous sections of this chapter were published [29] the methods described in Chapter 7 were developed. There, it is demonstrated how the available J-couplings and NOE cross peak intensities can be used to derive conformations which satisfy the experimental constraints. In the approach used there, the conformational space of the sugar moieties, the mono-nucleotide and the dinucleotide-units is sampled successively and tested against constraints obtained from the NMR-experiments. Here this type of multi-conformational analysis is applied to the third and fourth residue of the -TTTT- loop [30]. The conformational space of this dinucleotide unit is characterized by the following parameters: the pseudorotational phase angles P(L3) and P(L4), the pucker amplitudes ϕ_m(L3) and ϕ_m(L4), the glycosidic torsion angles χ(L3) and χ(L4) and the backbone torsion angles ε(L3), ζ(L3), α(L4), β(L4) and γ(L4). In order to define the allowed regions of these torsion angles more quantitatively, the experimental NMR parameters were examined. To this end, the sugar conformations of these two residues are investigated with the program MARC (see Chapter 7) using the J-couplings J1'2' and J1'2" derived from the 1D-spectrum and the sums of the couplings of the H2' and H2", derived from the DQF-COSY spectrum. The results of this analysis are collected in Table II. The values of the glycosidic torsion angle χ follow from the analysis of the intensity of the NOE cross peaks between the H6 resonance on the one hand and H1', H2', H2" and H3' resonances on the other hand. (Table II). Because the torsion angles γ and β are both biased to an almost pure (trans) conformation, coupling constants in conjunction with the torsion angle dependence, plotted in Fig. 5, are used to derive these torsion angles quantitatively. An uncertainty of 1 Hz in the estimated coupling constants is assumed. It follows from $J_{4'5'}$ = 2.6 Hz and $J_{4'5''}$ = 10.3 Hz that the torsion angle γ is between 170° and 189° (cf. Fig. 5). Similarly it follows from $J_{5''P}$ = 3.1 and $J_{5'P} < 6$ Hz that β is between 180° and 197°. The torsion angle ε is also derived with the aid of the plot in Fig. 5. The value of $J_{3'P}$ = 6 Hz follows from the fine structure of the 2'-3' cross peak in the DQF-COSY spectrum. This value of $J_{3'P}$ corresponds to three domains: 122°-130°, 192°-198° and 271°-285°. The values of the torsion angles around the phosphorous atom, α and ζ, cannot be derived from J-couplings, but in this case the chemical shift of the phosphorus resonance is informative. A downfield shift of the ^{31}P-resonance is usually associated with a change of one of the torsion angles α or ζ, from a gauche to a trans conformation [27]. The phosphorus spectrum of the hairpin shows that all ^{31}P-resonances fall within a narrow region of 1 ppm [28]. Therefore it is concluded that α and ζ are in a gauche$^+$ and/or gauche$^-$ conformation. At this point of the analysis we have obtained ranges of the torsion angles which can be used as an input for the multi-conformational analysis of the dinucleotide unit (program HELIX). The distance constraints were obtained as follows: the intensity of the unusual cross peak between L3H1' and L4CH$_3$ was fitted with the aid of the NOEFIT program which yielded a distance of 3.4 Å between these protons. This distance is constrained in the HELIX program between 3.1 and 3.9 Å. No cross peaks are observed between the H6 resonance of L4 and the H1' and H2" resonances of L3 . Also in the NOESY spectra obtained for this molecule at different mixing times (75 ms - 700 ms) and ionic strengths (0.1 M, 0.2 M and 1 M NaCl respectively) these cross peaks were absent. Therefore it can be assumed safely that the distance between these protons is larger than 4 Å. The remaining inter-residue distances were constrained between 1.8 and 30 Å in order to prevent the occurrence of conformations in which severe steric hindrance occurs.

On the basis of these data a multi-conformational analysis was carried out, using the program HELIX [30]. Table II summarizes the values of the torsion angles for which models were constructed. In this way 72900 conformations were calculated for this dinucleotide unit and tested against the distance constraints. It appeared that only a minor part, i.e., 160 conformations, satisfy the constraints. Correlation plots for some of the internal parameters can be examined in Fig. 12. The dots in the correlation plots correspond to the aforementioned 160 conformations. It follows that the torsion angles α and ζ are both found predominantly in a gauche$^+$ conformation (cf. Fig. 12A). Conformations are found with ε^t as well as ε^-. When ε adopts a trans conformation only values for β = 197 occur (cf. Fig. 12C).

At this stage, we conclude that the multi-conformational analysis of this dinucleotide unit reveals that the phosphate backbone can be described by two conformations either ε^t-ζ^+-α^+-β^t-γ^t (I) or ε^--ζ^+-α^+-β^t-γ^t (II). We shall refer to these conformations as turn I and II respectively. Both combinations of backbone angles enable the backbone to make a complete turn in the direction of propagation (see Chapter 3). Turn I is in accordance with the model presented in Fig. 10. In this model the three thymidines L1-L3 follow more or less the stacking pattern of the double helical stem. After the third loop nucleotide the turn is made by means of the torsion angle combination I. This results in an orientation of the fourth loop thymidine which enables on the one hand the formation of a wobble T-T pair

Table II. Results of the quantitative analysis of the NMR and derived structural parameters of the dinucleotide unit L3-L4, which is part of the loop section of the hairpin formed by d(ATCCTA-TTTT-TAGGAT).

IIA. Estimated coupling constants, $J_{1'2'}$ and $J_{1'2''}$, and sums of couplings, Σ2' and Σ2" defined as $\Sigma 2' = J_{1'2'} + J_{2'3'} + |J_{2'2''}|$ and $\Sigma 2'' = J_{1'2''} + J_{2''3'} + |J_{2'2''}|$ These data were used to derive the phase angle of pseudorotation, P, and the pucker amplitude, Φ_m, with the aid of the MARC program. An experimental error of 0.2 Hz is assumed for the $J_{1'2'}$ and $J_{1'2''}$ couplings and 1 Hz for Σ2' and Σ2". It followed that the molar fraction of the S-type sugar conformations is between 90 and 100 %. The backbone torsion angle δ, which is related to the sugar conformations was also calculated.

residue	J1'2'	J1'2"	Σ2'	Σ2"	P	Φ_m	δ
L3	9.1	5.3	27	22	166-194	41-43	113-166
L4	10.2	5.1	29	22	157-171	39-42	122-143

IIB. NOE values obtained from cross peak intensities of connectivities between the H6 resonance and the H1', H2', H2" and H3' resonances. The calculations were carried out for a NOESY spectrum recorded with a mixing time of 0.3 s. The rotational correlation time of the molecule was assumed to be 2 ns. The glycosidic torsion angle, χ, was subsequently determined from the NOEs using the program NOESIM.

	NOE				χ
	H6-H1'	H6-H2'	H6-H2"	H6-H3'	
L3	0.016	0.089	0.027	-	235-240
L4	0.027	0.144	0.058	0.037	258-260

IIC. Inter-nucleotide ^1H-^1H distances calculated from NOE intensities by means of the program NOEFIT; conditions were as indicated in IIB.

	NOE	d(Å)	HELIX constraints (Å)
L3H1'-L4CH$_3$	0.025	3.4	3.1-3.9
L3H1'-L4H6	-		4.0-30.0
L3H2"-L4H6	-		4.0-30.0

IID. Sampling values of the torsion angles used in the program HELIX to explore the conformational space of the L3-L4 dinucleotide together with the allowed angles (results) following from this analysis and corresponding with turn I (see text); the angles are given in degrees.

torsion angle	sampling(°)	results(°)
χ(L3)	238	238
P(L3)	171,180,189	171,180,189
ϕ_m(L3)	42	42
χ(L4)	259	259
P(L4)	162,171	171
ϕ_m(L4)	40	40
ϵ	122,130,192, 198,271,285	192,198
ζ	30,45,60,75, 90,105,120,135, 225,240,255,270, 285,300,315,330	30,135
α	30,45,60,75, 90,105,120,135, 225,240,255,270, 285,300,315,330	60,75,90,105,120
β	180,190,197	197
γ	170,180,189	170,180

with the first loop thymidine and on the other hand the stacking upon the base in the 5'-end position of the stem. For the conformations, which were found to fit the distance constraints, the helical rise was calculated and plotted as a function of ε (Fig. 12D). The results are striking: all conformations which satisfy the constraints with ε^t have a helical rise of

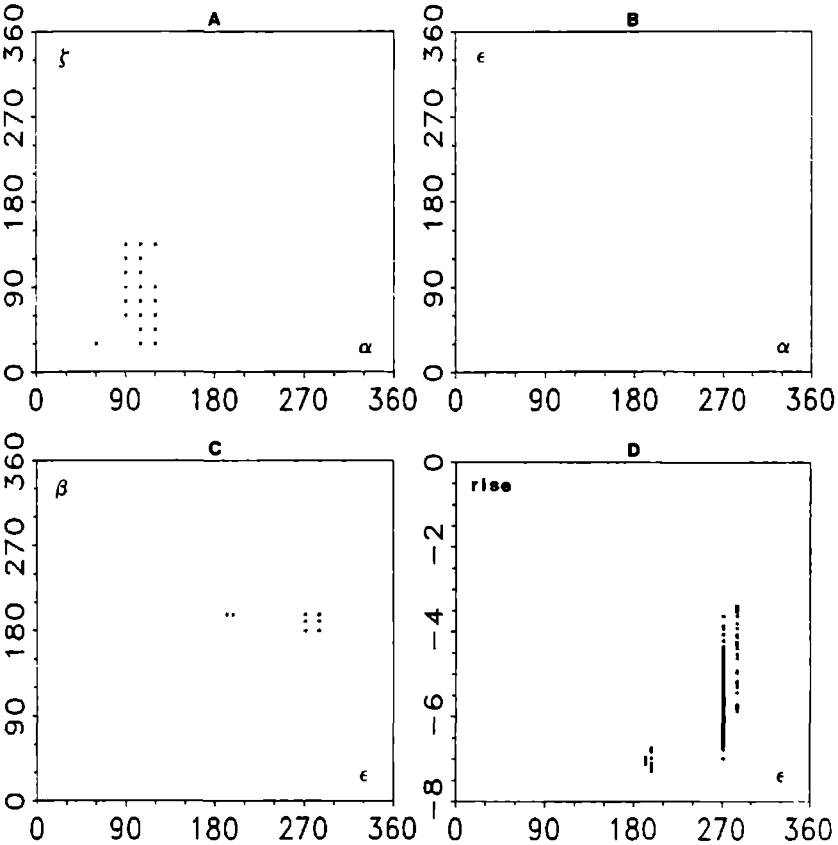


Fig. 12. Results of the multi-conformational analysis of the dinucleotide unit formed by the third and fourth loop residues. The dots in the plots correspond to conformations, which satisfy the constraints. Only the most informative correlations are shown, i.e., α-ζ (A), ζ-ε (B), ε-β (C) and ε-helical rise (D).

about 7 Å, twice that of the helical rise of standard B-DNA. This illustrates that this turn allows L2 to stack between L1 and L3. The conformations with ε^- may be at variance with the model presented in Fig. 10, especially for those conformations with a helical rise around 4 Å. The loop structure, that may satisfy this situation corresponds to a model similar to that derived for the hairpin formed by d(ATCCTA-TTTA-TAGGAT) (see Chapter 7). In this hairpin the turn in the sugar phosphate backbone is also localized between the third and fourth residue in the loop and is described by combination II of backbone torsion angles. The major difference in the folding of both models is the position of the second loop residue: whereas in the first model the thymine is stacked between its neighboring bases L1 and L3, it is bulged out and folded into the minor groove in the second one. A similar model, i.e., with the second loop residue folded into the minor groove, was constructed for the -TTTT- hairpin, with the aid of model building and molecular mechanics calculations and shows the plausibility of such a loop structure. It is noticed, that almost all cross peaks which would be typical for this model would coincide with other cross peaks, e.g., the unusual cross peak between L1H1' and L2H2' expected for the turn II model (see Chapter 7) cannot be observed, because at its predicted position the cross peak between L2H1' and L2H2' is present. However, it has been noted before that a NOE can be observed between the two methyl protons of residue L2 and L3. This NOE is expected for the first loop structure (turn I) and cannot be explained by the second one (turn II). In the latter, the corresponding bases point in different directions and the distance between the methyls is around 9 Å. This particular cross peak is a strong argument in favor of the first model. It should be noted, that the second model cannot be ruled out completely because at longer mixing times a tiny cross peak is found between L3CH_3 and L1H6. The occurrence of such a cross peak can be explained when a minor population is present with a conformation similar to that of the -TTTA- loop (Chapter 7). The major form, however, is described by the folding pattern illustrated in Fig. 10. Eight conformations which correspond to turn I and satisfy the distance constraints are superimposed in Fig. 13. This illustrates that these conformations are very similar. The root mean square difference of the eight conformations was 0.3 Å. To make the analysis complete, Table II summarizes the torsion angles which correspond to conformation I.

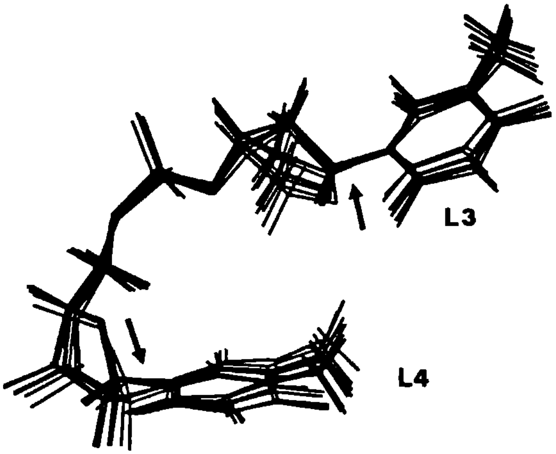


Fig. 13. Superposition of the structures obtained by means of a multi-conformational analysis performed for the dinucleotide step L3-L4 of the loop region of the hairpin formed by d(ATCCTA-TTTT-TAGGAT). The similarity of the conformations indicate that this part of the loop structure is well defined by the experimental data. The particular combination of backbone torsion angles induces a 180° turn in the sugar-phosphate backbone. This is illustrated by the two arrows drawn along the direction of the O4'-C1' bonds, which point in opposite directions. The bases are ≈ 7 Å apart from each other, so that L3 can partly stack on L2 (see Fig. 10).

CONCLUSION

The structural features of the hairpin formed by d(ATCCTA-TTTT-TAGGAT) obtained from the 2D-NMR experiments are in good agreement with the principles of loop folding described earlier [11,12] and in Chapter 5. The stacking pattern of the B-type stem is extended into the loop at the 3'-end of the stem. All sugars, including these in the loop adopt an S-type conformation. A 180° turn of the phosphate backbone is localized between the third and fourth loop nucleotide and could be characterized by means of a multi-conformational analysis. The turn can be described by the following torsion angles ε^t, ζ^+, α^+, β^t and γ^t. The formation of the T-T pair in the loop is interpreted as a local optimization of the hairpin structure within the conformational space delineated by our loop folding principle.

REFERENCES

1 Heerschap, A., Haasnoot, C.A.G.,& Hilbers, C.W. (1982) Nucl. Acids. Res. **10**, 6981-7000.

2 Heerschap, A., Haasnoot, C.A.G.,& Hilbers, C.W. (1983) Nucl. Acids. Res. **11**, 4483-4500.

3 Heerschap, A., Haasnoot, C.A.G.,& Hilbers, C.W. (1983) Nucl. Acids. Res. **11**, 4501-4520.

4 Roy, S. & Redfield, A.G. (1981) Nucl. Acids Res. **9**, 7073-7083.

5 Roy, S. & Redfield, A.G. (1983) Biochem. **22**, 1386-1390.

6 Hare, D.R. & Reid, B.R. (1982) Biochem. **21**, 5129-5135.

7 Reid, B.R., Salisbury, S.A., Bellard, S., Zhakked, Z & Williams, D.H. (1983) Biochem. **22**, 2019-2025.

8 Clore, G.M. & Gronenborn, A.M. (1984) FEBS Lett. **175**, 117-123.

9 Clore, G.M. & Gronenborn, A.M. (1986) Biomol. Stereodynamics (Eds. R.H. Sarma & M.H. Sarma) **4**, 139-155.

10 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) J. Biomol. Struct. Dyns. **1**, 115-129.

11 Haasnoot, C.A.G., de Bruin, S.H., Hilbers, C.W., van der Marel, G.A., & van Boom, J.H. (1985) J. Biosci. **8**, 767-780.

12 Hilbers, C.W., Haasnoot, C.A.G., de Bruin, S.H., Joordens, J.J.M., van der Marel, G.A. & van Boom, J.H. (1985) Biochimie 67, 685-695.

13 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) J. Biomol. Struct. Dyns. **3**, 843-857.

14 Haasnoot, C.A.G., Westerink, H.P., van der Marel, G.A. & van Boom, J.H. (1983) J. Biomol. Struct. Dyns. **1**, 131-149.

15 Scheek, R.M., Russo, N., Boelens, R., Kaptein, R. & van Boom, J.H. (1983) J. Am. Chem. Soc. **105**, 2914-2916.

16 Hare, D.R., Wemmer, D.E., Chou, S.-H., Drobny, G. & Reid, B.R. (1983) J. Mol. Biol. **171**, 319-336.

17 Arnott, S. & Hukins, D.W.L. (1972) Biochem. Biophys. Res. Comm. **47**, 1504-1509.

18 Haasnoot, C.A.G., & Hilbers, C.W. (1983) Biopol. **22**, 1259-1266.

19 Sanchez, V., Redfield, A.G., Johnston, P.D. & Tropp, J. (1980) Proc. Natl. Acad. Sci. U.S.A. **77**, 5659-5662.

20 Gralla, J & Crothers, D.M. (1973) J. Mol. Biol. **78**, 301-319.

21 Uhlenbeck, O.C., Borer, P.N., Dengler, B. & Tinoco, I. (1973) J. Mol. Biol. **73**, 483-496.

22 Altona, C. (1982) Recl. Trav. Chim. Pays-Bas **101**, 413-433.

23 Rinkel, L.J. & Altona, C. (1987) J. Biomol. Struct. Dyns. **4**, 621-649.

24 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P. (1984) J. Am. Chem. Soc. **106**, 765-784.

25 Singh, U.C., Weiner, S.J. & Kollman, P.A. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 755-759.

26 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) Cold Spring Harbor Symp. Quant. Biol., Vol. XLVII, 403-409.

27 Lerner, D.B., Becktel, W.J., Everett, R., Goodman, M & Kearns, D.R. (1984) Biopolymers 23, 2157-2172.

28 Salemink, P.J.M. (1980) Thesis, Chapter 6. Nijmegen, The Netherlands.

29 Blommers, M.J.J., Haasnoot, C.A.G., Hilbers, C.W., van Boom, J.H. & van der Marel, G.A. (1987) Structure and Dynamics of Biopolymers, NATO ASIS Series E: Applied Sciences **133**, 78-91.

30 Hilbers, C.W., Blommers, M.J.J., van de Ven, F.J.M., van Boom, J.H. & van der Marel, G.A. (1990) CECAM Workshop (in press).

CHAPTER 5

CONFORMATIONAL ASPECTS OF HAIRPIN LOOPS IN DNA OLIGONUCLEOTIDES.

A MOLECULAR MODELLING STUDY.

ABSTRACT

In order to investigate the influence of the nucleotide composition and the sequence on the loop folding, a number of DNA hairpins were constructed with aid of molecular modelling in which the nucleotide sequence in the loop was varied systematically. The hairpin structures were refined with the software package AMBER. It appeared from these calculations that an additional base pair within the four-membered loop can be formed in the sequences d(ATCCTA-CTTG-TAGGAT) and d(ATCCTA-TTTA-TAGGAT), whereas in the hexadecanucleotides d(ATCCTA-GTTC-TAGGAT) and d(ATCCTA-ATTT-TAGGAT) such an additional base pair cannot be formed. The calculations are strongly related to a general loop folding principle, which states that the structure of the loop is dictated by the structure of the stem. It is demonstrated that the loop folding can be 'fine tuned' by means of local optimizations. Therefore it is predicted that hairpin structure and stability will depend on base sequence.

INTRODUCTION

Our studies [1-4] towards the structure, kinetics and thermodynamics of DNA hairpins formed in solution by the homologous, (partly-) selfcomplementary DNA fragments d(ATCCTA-T_n-TAGGAT), n=0-7, showed that the inherent stability of DNA hairpins is at its maximum when the loop of the hairpin comprises four or five nucleotides. This finding is at variance with earlier experiments [5] which indicated that in RNA hairpins loop lengths of six to seven residues are the most favorable.

Both observations can be rationalized [6,7] on the basis of a structural model for the architecture of hairpin loops. This model is formulated as a 'folding principle' which states that the folding (i.e., structure) of hairpin loops is dictated by the stacking pattern of the bases in the double

---

Haasnoot, C.A.G., Blommers, M.J.J. & Hilbers, C.W. (1987) Springer Series in Biophysics, 1: Structure, Dynamics and Function of biomolecules (Eds. A. Ehrenberg, R.Rigler, A. Gräslund & L. Nilsson), 212-216.
Hilbers, C.W., Blommers, M.J.J., Haasnoot, C.A.G., van der Marel, G.A. & van Boom, J.H. (1987) Fresenius Z. Anal. Chem. 327, 70-71.
Hilbers, C.W., Blommers, M.J.J., van de Ven, F.J.M., van Boom, J.H. & van der Marel, G.A. (1990) Nucleosides and Nucleotides 9 (in press).

helical stem of the hairpin. For example the 'natural' way to form a loop on top of an A-RNA type duplex stem is characterized by extending the 5'-end of the duplex with five or six nucleotides arranged in a single helical A-type fashion; the remaining gap between the two ends is then easily bridged by one or two nucleotides (thereby forming a loop consisting of ca. 7 nucleotides). This is illustrated in Fig. 1 where interstrand distances between the phosphates in an A-type and B-type double helix are plotted. In contrast, the loop on top of a B-type duplex stem is characterized by propagating the base stacking pattern at the 3'-end of the double helical stem: extending the 3'-end of the stem by two or three nucleotides arranged in a single helical B-type fashion leaves a gap that can be closed by one or two nucleotides (thereby forming a loop comprising ca. 4 nucleotides). Thus, in our model the above mentioned experimental differences in RNA vs. DNA hairpin stability are a consequence of the structural differences in the double helical stems of the corresponding hairpins (A-type vs. B-type): it is the base stacking pattern that allows for shorter 'optimum' loops in DNA when compared to RNA hairpins.

In the case of RNA hairpin loops the validity of the above stated loop folding principle is demonstrated by the architecture of the anticodon loop in $tRNA^{Phe}$: crystallographic studies show that five bases stack in an A-type single helical way on top of the 5'-end of the anticodon stem, then the 'turn' is made and two nucleotides complete the seven-membered anticodon loop (cf. Chapter 1, Fig. 12).

DNA-hairpins are currently under investigation with aid of NMR spectroscopy. By far the most of our NMR data were obtained for d(ATCCTA-TTTT-TAGGAT) (Chapter 4).

MATERIAL AND METHODS

DNA models were constructed using the computer program Chem-X [12]. Energy minimizations were performed with the molecular mechanics program AMBER (version 2.0) [13,14]. The united atom version of the Force Field was used with hydrated counter ions added to the system.

RESULTS

According to the results of 2D-NMR spectroscopy obtained for d(ATCCTA-TTTT-TAGGAT), the B-type helical structure of the stem is propagated in the

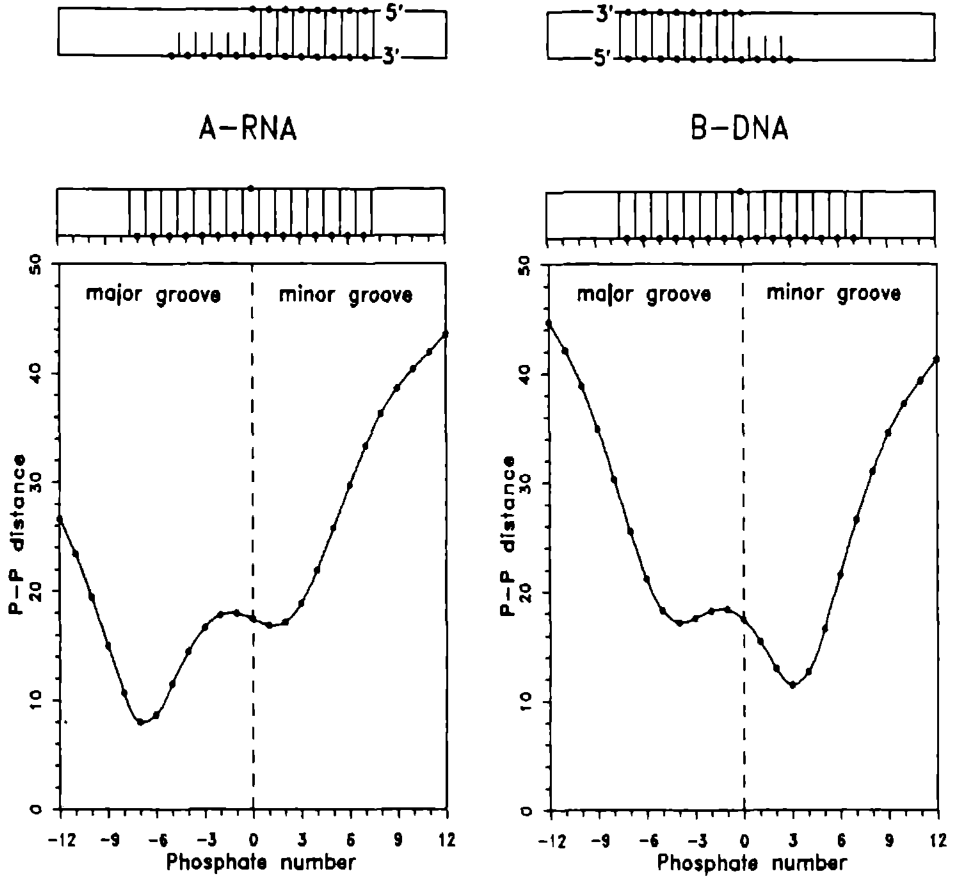


Fig. 1. Distances between a phosphate and the phosphate atoms in the opposite strand in 5'-3' and 3'-5' direction in a regular A-RNA and B-DNA double helix. A short interstrand phosphate distance is reached after five or six residues when in A-RNA the helix in the loop propagates the A-type helical folding over the major groove (i.e., in 3'-5' direction). In contrast a short interstrand distance is reached after two or three residues in B-DNA, when the helical pattern is propagated over de minor groove (in 5'-3'-direction). The numbering of the phosphates is indicated above the curves.

3'-direction into the loop region. This extension involves L1, L2 and more or less L3 as indicated by the inter-residual base-sugar crosspeaks [8]. The fourth loop thymine L4 stacks on top of T6 and favors the formation of a T-T wobble base pair. Using computer graphics methods, we generated a number of models which are in accordance with these NMR results; all structures were required to have torsion angles in the range normally found in deoxyribonucleotides. The models thus obtained were refined by molecular mechanics (MM) methods [9,10]. The lowest energy structure found is shown in Chapter 4 (Fig. 10); it is seen that this structure displays all characteristics prescribed by our loop folding principle (three bases stack in a B-type single helical fashion on top of the 3'-end of the stem; one nucleotide completes the loop). Using the lowest energy structure as a basis for 2D-NOE simulation [11] yielded a good (albeit not perfect) quantitative accordance between the simulated and experimental 2D-NOE spectrum of d(ATCCTA-TTTT-TAGGAT) [8]. Another conformational aspect of the lowest energy structure that finds experimental endorsement is the 'tightness' of the loop. Examination of the structure shows that all bases are more or less turned inward, i.e., towards the center of the loop. Moreover, the first and the last thymine base of the loop (L1 and L4, respectively) form a buckled wobble pair in which the N3-H of L4 is hydrogen-bonded to O4 of L4 and N3-H of L4 is hydrogen bonded to O2 of L1. As such, these features offer an explanation for the observed [1,3] retardation in the proton-exchange of the loop imino protons with the solvent; they also make the refusal [3] of the loop thymines to form a complex with d(AAAA) plausible. Taken together, we conclude that the data available at least prove the feasibility of the loop folding principle for DNA hairpins.

It should be realized that our loop folding principle is to be regarded as an abstraction, i.e., it only provides a framework that is to be used as a basis for understanding nucleic acid hairpin formation. As always, generalizations of this kind are in reality modulated by the specific 'needs' of the case under study; in other words, the bases in a specific hairpin will certainly try to optimize base stacking/pairing, evade highly localized strain situations, etc., within the conformational space set by our folding principle. Therefore, it is obvious that the details of the hairpin loop structure will depend on, e.g., base sequence in the loop and/or in the stem of the hairpin. Below this base sequence dependence is explored using molecular mechanics (MM) in order to map the degrees of freedom left in our framework for hairpin loop formation.

We performed a number of MM-calculations for DNA hairpins in which the base composition of the loop stretch was changed systematically. More specifically, so far we focussed primarily on changes affecting the 'wobble' pair formed by L1 and L4 in the lowest energy structure described above for the DNA hairpin d(ATCCTA-TTTT-TAGGAT). With aid of computer graphics we replaced the loop -TTTT- sequence by a -CTTG- sequence. The MM refined structure obtained for the hairpin after molecular mechanics calculations, using the united atom Force Field version, is presented in Fig. 2A. According to the calculations, a base pair is formed between the C and G residue of the loop sequence. The most conspicuous characteristic of this C-G base pair is the finding that it is not flat like a standard Watson-Crick base pair, but buckled, i.e., the base plane normals are not parallel but make an angle in the order of 30°. As will be outlined in the Discussion section this suggests that an interchange of the two bases will not bring the bases in the proper position to form a base pair and indeed when we replaced the sequence -CTTG- by -GTTC- we could not find a low energy structure in which some kind of base pairing between the G and C bases occurred (Fig. 2B). Thus the molecular mechanics calculations suggest that in the case of the -CTTG- loop sequence a base pair within the four membered loop is formed, albeit that this base pair appears to be buckled. In the case of the -GTTC- loop sequence no base pair is formed within the loop. Similar results were obtained from molecular mechanics calculations in which the -TTTT- loop sequence was changed to a -TTTA- base sequence. In the case of the -TTTA- sequence a base pair between T and A in positions L1 and L4 is formed. Close examination shows that the bases are somewhat twisted with respect to one another and, moreover, that the standard Watson-Crick N6-H...O4 hydrogen bond is not formed. Instead, the two adenine N6-H protons form hydrogen bonds to the O2-atoms of L2 and L3, respectively. The most conspicuous characteristic of this T-A pair, however, is the finding that again the base pair is not flat but buckled; moreover, neither one of the bases is perpendicular to the helix axis.

For the -ATTT- loop sequence the molecular mechanics calculations produce similar results as those obtained for the -GTTC- loop sequence. The -ATTT- sequence was further probed by building a structure in which the A

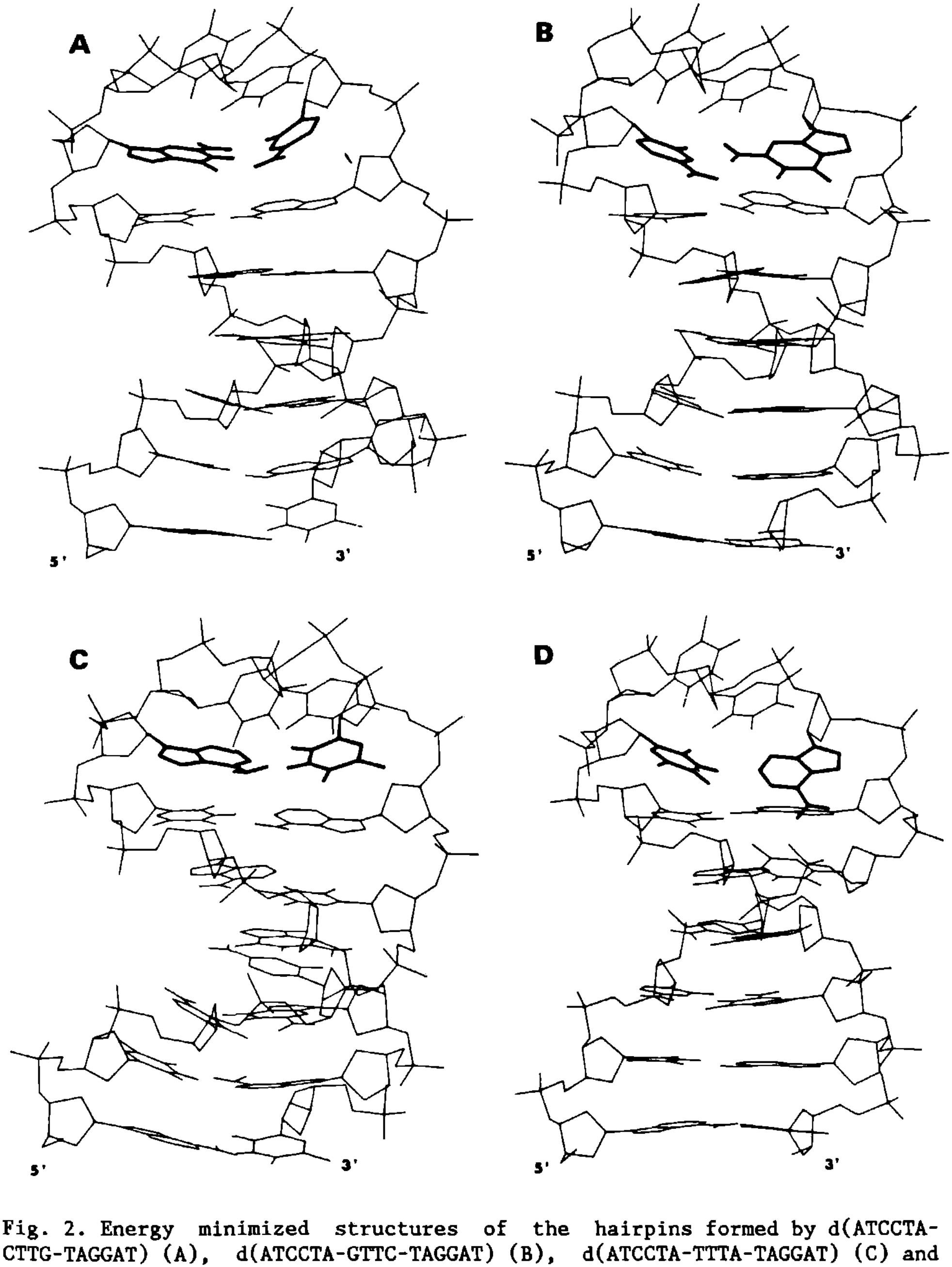


Fig. 2. Energy minimized structures of the hairpins formed by d(ATCCTA-CTTG-TAGGAT) (A), d(ATCCTA-GTTC-TAGGAT) (B), d(ATCCTA-TTTA-TAGGAT) (C) and d(ATCCTA-ATTT-TAGGAT) (D). Note that in case of the loop sequence -CTTG- and -TTTA- an additional base pair can be formed (A,C) whereas this does not occur for the -GTTC- and -ATTT- loop sequence (B,D).

and the T residues were forced to form a base pair while at the same time the sugar phosphate backbone was readjusted to accommodate the new situation. This turns out to yield a high energy structure; when the constraints are removed and the molecule is subjected to an energy minimization the A-T pair is disrupted. Thus the molecular mechanics calculations strongly suggest that in the case of a -PyTTPu-[1] loop sequence a base pair within the loop is formed, albeit that this base pair appears to be bent. In the case of a loop sequence -PuTTPy- no base pair is formed within the loop.

DISCUSSION

In the present chapter we extend our earlier studies [7-9] concerning the loop stacking pattern in DNA hairpins. To acquire more insight in the generality of the loop stacking pattern in DNA hairpins we have investigated the effects of the base composition of the loop on the structure of the hairpin. These studies were instigated by the observation that in the hairpin formed by d(ATCCTA-TTTT-TAGGAT) a T-T wobble pair can be formed [9] and by the results of Orbons et al. obtained for the mini hairpin formed by the octanucleotide d($C_mGC_mGTGC_mG$) [10] in which the cytidines are methylated at C5. Within the framework of our loop folding principle the loop formed in the latter hairpin may be considered as a four membered loop in which the fourth base in the loop G8 fills the gap left by the first three stacked loop bases and is then able to form a base pair with the first base, C_m3, of the loop sequence. Preliminary molecular mechanics calculations [8] as well as the refined molecular mechanics calculations presented in this chapter predict that such a base pair is only formed for a -PyTTPu- sequence and not for a -PuTTPy- loop sequence. These predictions are neatly confirmed by UV melting studies and ^{1}H-NMR spectroscopy [15] presented in Chapter 6, but these experiments do not allow us to conclude that the C-G base pair within the loop is not of a standard Watson-Crick type. However, if the base pair under consideration would be of the standard Watson-Crick type, it is hard to see why no base pair is formed when the purine and pyrimidine are interchanged. An explanation is suggested by the structural models derived from the molecular mechanics calculations. It was emphasized in the Results section that it follows from the molecular mechanics calculations that the base pair formed in the loop is buckled (Fig. 2). The reason for this can be gleaned from Fig. 3, where the hairpin formed by the hexadecanucleotide d(ATCCTA-CTTG-TAGGAT) has been plotted in such a way that the double helix is viewed along the helix axis. As for a standard B-type helix the sugar C1' atoms fall on a circle. The distance between the C1' atoms amounts 10.7 Å. This is not so for the C1' atoms of the base pair in the loop. These carbon atoms have approached each other to a distance of 9.7 Å,

---

[1] Py = pyrimidine (C,T); Pu = purine (G,A)

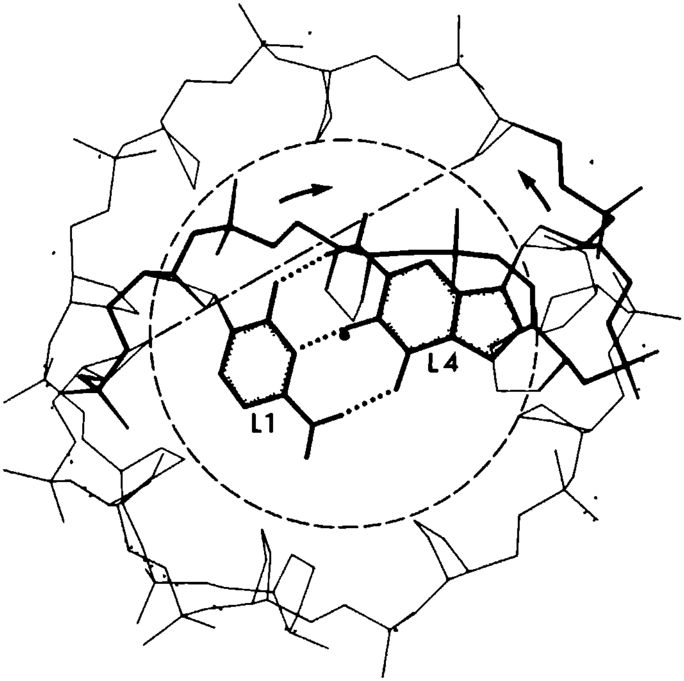


Fig. 3. View along the helix axis of the hairpin loop structure presented in Fig. 2A. For clarity all base pairs have been omitted except the C-G base pair which is formed in the loop and which is viewed from the side of the stem. The sugar phosphate backbone in the loop is accentuated by thick lines and the 5'-3' direction is indicated by arrows. The dotted circle is the projection of the cylinder on which an ideal helix the phosphorus atoms are located; the dashed circle is the projection of the cylinder on which in an ideal helix the C1' atoms are located. The C1' atoms of base pair A-T(6) which are at a normal distance, are connected by a dashed-dotted line. Note that the C1' of the first residue in the loop, i.e., cytidine (L1), has moved away from the C1'-circle into the direction of the helix axis, indicated by the central dot.

which impairs the formation of a 'flat' normal Watson-Crick base pair. If in this situation a base pair is formed it will necessarily be buckled. Since the bend in the base pair is displaced from the dyad axis (cf. Fig. 3), which can be defined with respect to the two C1' atoms, an asymmetry is introduced so that inversion of the base pair no longer allows formation of hydrogen bonds and therefore of the base pair. This is demonstrated in Fig. 4. It is of some interest to inquire why the C1' atoms of the buckled base pair have moved closer to each other than the corresponding atoms in standard B-type double helix. As has been pointed out earlier [6] the distance between the 3'- and 5'-terminal phosphate of a double helix is normally about 17.4 Å and, given the stacking pattern of the bases in the loop, it is not possible to span the distance by means of two nucleotides. Therefore, in order to accommodate the formation of a formal two-membered loop which connects these two terminal phosphates at least one of these phosphates should change its position and move towards the other. As has been mentioned above the C1' atom of L1 and the C1' atom of L4 have moved towards each other in particular the C1' atom of L1 has moved away from the C1' circle of the double helix (see Fig. 3) and is considerably closer to the helix axis, while the C1' atom of L4 remains at a normal position. The change in the position of the C1' atom of L1 is brought about by gradual alterations of the backbone torsion angles between the sugar of residue A6 and the sugar of residue L1. In turn the 3'-phosphate of L1 has moved rather close to the helix axis. In this way the distance between this phosphate and the 5'-phosphate of L4 has diminished too; the latter phosphate remains in its normal position found in double helical B-DNA (Fig. 3). Thus, this set of changes permits formation of a loop consisting of two nucleotides, which is necessarily closed by a buckled base pair. As we shall see in Chapter 6 this is not the only way in which a two-membered loop can be accommodated. At this point the results obtained can be summarized in the form of a more detailed loop folding principle as follows: the folding of a four-membered loop in a DNA hairpin with a B-type stem is achieved by extending the 3'-end of the duplex with three nucleotides arranged in a single helical stacking pattern. The remaining gap between the ends can be bridged by one nucleotide which turns inward and may form a base pair with the (complementary) base in the first position of the loop. This is possible for the loop sequence-CTTG- or -TTTA-, but not for the sequence -GTTC- or -ATTT-.

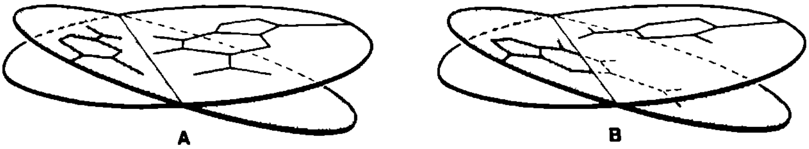


Fig. 4. Schematic representation of orientation of the first (left) and the fourth base (right) in the loop to bring out that for a -CTTG- sequence the bases are in a position to form a buckled base pair (A) whereas for the-GTTC- sequence this is not possible. The underlying reason is that the cross section of planes through the bases no longer forms a pseudo-dyad axis of symmetry.

It could be argued that in the first cases a loop comprising only two nucleotides is formed. We do not share this view since the pairing of the bases in these loops is not of the regular, non-buckled Watson-Crick type. Moreover, it appears that this kind of 'base pair' can only be formed in the case of a -PyTTPu- loop sequence. Therefore, we interpret the base pairing observed in the case of a -PyTTPu- loop sequence as a local optimization of the hairpin structure within the conformational space delineated by our loop folding principle.

It is of interest to see how general this rule of loop folding appears and how the present results relate to data in the literature. A detailed structural comparison can be made between the loop structure described for the hexadecanucleotides described in the present and our earlier papers and the mini-hairpin formed by $d(C_mGC_mGTGC_mG)$ which was studied in Altona's laboratory [10]. The structure of the latter molecule was deduced on the basis of J-coupling analysis and NOE measurements. There is a great similarity between the stacking pattern of this mini-hairpin loop structure and those deduced for our hexadecanucleotides, i.e, the stacking is propagated from the 3'-end of the stem into the loop. In the mini-hairpin closure of the loop is achieved by a sharp turn in the backbone which in our terminology allows the fourth base of the loop, i.e., G6 to turn inward and to form a base pair with C_m3. A sharp turn in the backbone is brought about by the unusual values of torsion angles β and γ of G6 which adopt a gauche$^+$ and trans conformation, respectively. At the corresponding position in the hexadecanucleotide d(ATCCTA-TTTT-TAGGAT), γ adopts a trans conformation as well but β does not have a gauche$^+$ conformation, but rather adopts a trans conformation as follows from the ^{31}P-^1H couplings (Chapter 4). This leads to an interesting difference between the hairpin loop structures. Model building and energy minimization shows that in the mini-hairpin the C_m-G base pair in the loop is a regular Watson-Crick base pair, whereas in the model of the hairpin loop discussed in this chapter the corresponding base pair is significantly buckled. This led to the conclusion that in the -CTTG- loop a base pair can be formed, but not in the -GTTC- loop. Conversely, it may be possible that in the mini-hairpin an interchange of the bases of the base pair in the loop does not lead to an disruption of this base pair. The underlying reason for the occurrence of a flat base pair in this loop may be found in the stacking pattern of the very small stem consisting of C-G pairs with methylated cytidines which could be different from that in the stem of the hexadecanucleotide hairpins. In the latter hairpins the loop is closed by an A-T base pair. Closure of the loop by a G-C pair may thus lead to differences. Indeed this is suggested by a study of Kearns and collaborators which was concerned with the hairpin formed by d(CGCGCG-TTTT-CGCGCG) which also carries a loop of four thymidines [16]. It was found that the exchange of the thymidine imino protons in this hairpin is faster than for the hexadecanucleotide studied in our laboratory under the same conditions [3]. This suggests that the loop structure/stability may be influenced by the terminal base pair adjacent to the loop.

The strongly related hairpin formed by d(CGCG-TTTT-CGCG) has been studied by Hare & Reid by means of 2D-NMR and Distance Geometry calculations [17]. Indeed the folding pattern in the loop differs (somewhat) from that in our hairpins, again suggesting that the stem composition may influence the loop structure, but to what extent remains to be established. Although the combined approach of 2D-NMR and distance geometry calculations is very promising for the elucidation of the detailed structure of nucleic acids, it is still in a developing stage. It is well known that some of the helix structures derived by means of the 2D-NMR distance geometry approach are underwound compared to those deduced by means of single crystal X-ray diffraction techniques [18]. In particular, for the hairpin at hand the 2D-

NMR distance geometry method leads to structural parameters some of which are most unlikely from an energetic point of view and thus physically non-realistic. However, as we have stated earlier the details of the hairpin loop structure are likely to depend on the composition of the base pair(s) adjacent to the loop. Thus the above formulation folding principle may need additional refinement to account for this phenomenon.

REFERENCES

1 Haasnoot, C.A.G., den Hartog, J.H.J., de Rooij, J.F.M., van Boom, J.H. and Altona, C. (1980) Nucl. Acids Res. **8**, 169-181.

2 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. and Berendsen, R.G. (1983) Cold Spring Harbor Symp. Quant. Biol. **47**, 403-409.

3 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. and van Boom, J.H. (1983) J. Biomol. Struct. Dyns. **1**, 115-129.

4 Hilbers, C.W., Haasnoot, C.A.G., de Bruin, S.H., Joordens, J.J.M., van der Marel, G.A. and van Boom, J.H. (1985) Biochimie **67**, 685-695.

5 Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biology **246**, 40.

6 Haasnoot, C.A.G., de Bruin, S.H., Hilbers, C.W., van der Marel, G.A. and van Boom, J.H. (1985) Proc. Int. Symp. Biomol. Struct. Interactions, Suppl. J. Biosci. **8**, 767-780.

7 Haasnoot, C.A.G., Hilbers, C.W., van de Marel, G.A., van Boom, J.H., Singh, U.C., Pattabiraman, N. and Kollman, P.A. (1986) J. Biomol. Struct. Dyns. **3**, 843-857.

8 Blommers, M.J.J., Haasnoot, G.A.G. and Hilbers, C.W. (1987) Structure and dynamics of Biopolymers, NATO ASIS Series E: Applied Sciences **133**, 78-91.

9 Haasnoot, C.A.G., Blommers, M.J.J., Hilbers, C.W. (1987) Springer Series in Biophysics **1**: Structure and dynamics and function of biomolecules (Eds. A. Ehrenberg, R. Rigler, A. Gräslund, L. Nilssen) 212-216.

10 Orbons, L.P.M., van Beuzekom, A.A. & Altona, C. (1987) J. Biomol. Struct. Dyns. **4**, 965-987.

11 Keepers, J.W. and James, T.L. (1984) J. Magn. Res. **57**, 404-426.

12 Davies, E.K. (1986) Chem-X. Chemical Chrystallography Laboratory. Oxford University. Chemical Design Ltd. Oxford.

13 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C. Alagona, G., Profeta, S. and Weiner, P. (1984) J. Amer. Chem. Soc. **106**, 765-784.

14 Singh, U.C., Weiner, S.J. and Kollman, P.A. (1985) Proc. Natl. Acad. Sci. USA **82**, 755-759.

15 Blommers, M.J.J., Walters, J.A.L.I., Haasnoot, C.A.G., van Aelen, J.M.A., van der Marel, G.A., J.H. van Boom & Hilbers, C.W. (1989) Biochemistry **28**, 7491-7498.

16 Ikuta, S., Chattopadhyana, K., Ito, H., Dickerson, R.E. & Kearns, D.R. (1986) Biochemistry **25**, 4840-4849.

17 Hare, D.R. & Reid, B.R. (1986) Biochemistry **25**, 5341-5350.

18 Patel, D.J., Shapiro, L. & Hare, D.R. (1987) Ann. Rev. Biophys. Chem. **16**, 423-454.

EFFECTS OF THE BASE SEQUENCE ON THE LOOP FOLDING IN DNA HAIRPINS

A NMR AND UV-MELTING STUDY

ABSTRACT

High-resolution NMR and UV-melting experiments have been used to study the hairpin formation of partly self-complementary DNA fragments in an attempt to derive rules that describe the folding in these molecules. Earlier experiments on the hexadecanucleotide d(ATCCTA-TTTT-TAGGAT) had indicated that within the loop of four thymidines a wobble T-T pair is formed (Chapter 4). In the present chapter it is shown that if the first and the last thymines of the intervening sequence are replaced by complementary bases sometimes base pairs can be formed. Thus, for the intervening sequences -CTTG- and -TTTA- with the pyrimidine in the 5'- and the purine in the 3'-position a base pair is formed leading to a loop consisting of two residues. For the intervening sequences -GTTC- and -ATTT- with the purine in the 5'- and the pyrimidine in the 3'-position this turns out not to be the case. It was found that it made no difference when the four-membered sequence was closed by a G-C base pair or an A-T base pair. Replacement of the two central thymidine residues by the more bulky adenine residues limits the hairpin to a four-membered loop scheme.

Very surprisingly, it was found from 2D NOE experiments that the T-A base pair, formed in the loop consisting of the -TTTA- sequence, is a Hoogsteen pair. It is argued that the pairing of the bases in this scheme may facilitate the formation of a loop of two residues, since the distance of the C1' atoms in this base pair is 8.6 Å instead of 10.4 Å found in the canonical Watson-Crick base pair. Combination of the data obtained for the series of DNA fragments studied shows that the results can be explained by a simple, earlier proposed, loop folding principle which assumes that the folding of the four-membered loop is dictated by the stacking of the double-helical stem of the hairpin.

INTRODUCTION

Extensive structural studies performed during the past decade [1,2] have demonstrated that the conformations which can be adopted by nucleic acids are much more varied, complicated, and interesting than the regular double helix originally envisioned. It is often believed that the conformational space available to nucleic acids will somehow be important in cellular processes, and therefore attempts have been made to explain the

---

Blommers, M.J.J., Walters, J.A.L.I., Haasnoot, C.A.G., van der Marel, G.A., van Boom, J.H. & Hilbers, C.W. (1988) Biochemistry **28**, 7491-7498.

conformational behavior on the basis of nucleotide sequences. Among the polymorphic conformations, hairpin loop structures play a special role, particularly in RNA molecules although it has been found for DNA that sequences capable of hairpin formation are frequently occurring near regulation and promotion sites.

We have found that the homologous partly self-complementary DNA fragments d(ATCCTA-Tn-TAGGAT) (in which n varies from 0 through 7) form hairpins of which the stability is at a maximum when the loop of the hairpin comprises four or five nucleotides [3,4]. This was in contrast to earlier studies of RNA hairpins for which optimal stability was found when there are six to seven nucleotides in the loop [5,6]. This difference in RNA vs. DNA behavior was rationalized [7,8] in terms of structural models for the architecture of RNA and DNA hairpin loops. These models yield [8] a simple framework in which the folding (i.e., structure) of the loop is dictated by the stacking of the bases in the double-helical stem of the hairpin: it is the base stacking pattern that allows for shorter 'optimum' loops in B-DNA when compared to A-RNA.

For DNA hairpins our loop folding principle states [7,8] a continued stacking of nucleotides in the loop section on top of the 3'-end of the B-type double-helical stem. The model predicts an optimum loop length of ca. four nucleotides. Notwithstanding its success in explaining the data available at that time, it was predicted [8,9], however, that the details of the hairpin loop folding will depend on ,e.g., base sequence in the loop and perhaps even on the base sequence in the stem. Since then, Orbons et al. [10,11] discovered that the octanucleotide d($C_mGC_mGTGC_mG$), in which the cytidines are methylated at C5, is able to fold into a 'mini-hairpin' structure in which the central -GT- part forms a loop consisting of only two nucleotides. At the same time it was assessed in our laboratory [12,13] that the two terminal thymidines of the central four-membered T-stretch in the hairpin formed by d(ATCCTA-TTTT-TAGGAT) are involved in T-T wobble pairing. The base stacking pattern in the loops of these two hairpins is very similar and fulfills the basic criterion in the loop folding model (i.e., propagation of the stack from the 3'-end of the stem into the loop). Therefore, these findings can be interpreted [9] in terms of a local optimization of the hairpin structure within the conformational space delineated by our loop folding principle.

In the present chapter we extend our investigations to obtain more insight regarding the loop architecture of DNA hairpins. In particular, we put to the test the prediction made in Chapter 5 on the basis of molecular mechanics calculations [9] that -PyTTPu- loop sequences can enhance the hairpin stability by base pairing within the loop whereas -PuTTPy- loop sequences cannot.

MATERIALS AND METHODS

The DNA fragments employed in the present study were synthesized according to the phosphate triester method [14]. Samples for ultraviolet melting studies and ^{1}H NMR spectroscopy were prepared by dissolving the oligonucleotides in a 5 mM sodium phosphate buffer with 1 mM sodium cacodylate, 0.1 mM EDTA, pH = 7.0; NaCl was added to attain the desired sodium ion concentration. For samples used for NMR spectroscopy the pH was adjusted to 6.0 in order to diminish proton exchange between the oligonucleotide and water.

UV-melting experiments were carried out at 260 nm on a Cary 118C spectrophotometer, using a constant temperature cell (Hellma). The temperature was controlled by circulating water from a cryothermostat (Mettler, WKS) first through a thermostatted cell holder and then through the cell. The heating rate was 1 °C/min. In our earlier studies [7,15] transition enthalpies were calculated from the differential melting curves, according to Gralla and Crothers [16]. In the present study we fitted the absorbance vs. temperature curves with a nonlinear least-squares program, using ΔH, ΔS, $E_{H,0}$, $E_{C,0}$, dE_H/dT and dE_C/dT as adjustable parameters. ΔH denotes the transition enthalpy and ΔS is the transition entropy. $E_{H,0}$ stands for the absorption of the helix structure at 0 °C , $E_{C,0}$ is the absorption of the coil structure at 0 °C, dE_H/dT and dE_C/dT denotes the linear temperature dependency of the absorption of the helix structure and the coil structure, respectively. In all cases excellent fits to the experimental data were obtained. The thermodynamic data derived in this manner are more accurate and more reproducible than those derived by the earlier method. The concentration of the samples used in the NMR as well as in the UV-melting experiments is given in the legends to the figures.

500 MHz and 600 MHz 1^H-NMR 1D- and 2D-spectra were recorded on Bruker AM-500 and AM-600 spectrometers, respectively, which were interfaced to an

Aspect 3000 computer. Water suppression techniques, applied to enable observation of imino proton resonances, were as described earlier [17]. The relative intensities of the imino proton resonances were determined by integration; the estimated accuracy is 20 % for the individual resonances. Only for the spectrum of d(ATCCTA-ATTT-TAGGAT) the accuracy was somewhat less, due to overlapping of some resonances.

Phase-sensitive 600 MHz NOESY spectra of the D_2O samples were recorded with a spectral width of 6000 Hz, using time proportional phase incrementation [18]. Two different spectra were acquired, with 4K points in the t_2- and 512 points, zero-filled to 1K points in the t_1-direction, with mixing times of 0.3 and 0.075 s, respectively. Phase-sensitive 600 MHz NOESY spectra of the H_2O-samples were recorded using a time-shared long pulse as observation pulse [17]. The spectral width was 23800 Hz. The carrier was placed at the low field end of the spectrum. The spectrum was acquired, with 4K points in the t_2- and 512 points zero-filled to 1K points in the t_1-direction, with a mixing time of 0.3 s. Before Fourier transformation each free induction decay was subjected to data shift accumulation treatment.

RESULTS

The choice of the oligonucleotides, studied in the UV-melting and ^1H NMR experiments presented below, was strongly guided by the outcome of earlier molecular mechanics calculations presented in Chapter 5 [9]. These modelling studies suggested that in the hairpin with -PyTTPu- loop sequences the terminal Py/Pu bases might be involved in base pairing. In contrast, in a hairpin loop with a -PuTTPy- sequence no low energy structure displaying the same feature could be found. Such a proposition can be checked by studying the imino proton spectrum of the hairpins considered [13]. The imino proton NMR spectra of d(ATCCTA-CTTG-TAGGAT) and d(ATCCTA-GTTC-TAGGAT) recorded in the presence of 5 mM NaCl are shown in Figs. 1A and 1B, respectively. The assignments of the resonances given in this figure are based on a comparison of the spectra with that of d(ATCCTA-TTTT-TAGGAT), [3] (cf. Fig. 3A) supplemented by one-dimensional Overhauser experiments and NMR-melting studies. In contrast to other chapters the numbering of the residues is counted from 1 to 16, as has been indicated in Fig. 5. Apart from the resonances around 11.0 ppm, which originate from the unpaired

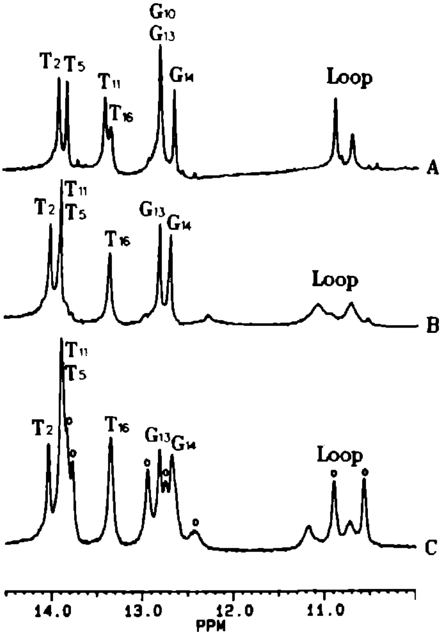


Fig. 1. 500 MHz imino proton spectra of the hairpins d(ATCCTA-CTTG-TAGGAT) (A) and d(ATCCTA-GTTC-TAGGAT) (B) at 5 mM NaCl. The imino proton resonances between 15 and 12 ppm belong to base pairs, those upfield from 11 ppm to unpaired residues in the loop region. The spectra were recorded at 274 K;the total strand concentration amounted to 1 mM DNA. The numbering of the resonances corresponds with the base numbering given in Fig. 5. After adjusting the salt concentration to 200 mM NaCl, spectrum (B) changes in a mixed hairpin/dimer spectrum (C). The resonances originating from dimer species are indicated with an open dot.

residues in the loop, the integrated peak intensities in spectrum A account for seven resonances , whereas in the spectrum B six resonances are found (Fig. 1). In agreement with UV-melting experiments (at NMR concentrations and at 0.2 M NaCl; to be discussed below) we observe for d(ATCCTA-GTTC-TAGGAT), the less stable hairpin, a mixture of hairpin and dimeric molecules when the ionic strength is raised (Fig. 1C). The extra resonances in that figure, which arise from the dimeric structure, are indicated with open dots. No such salt dependent or concentration dependent changes are observed for the more stable hairpin d(ATCCTA-CTTG-TAGGAT). Our experiments show that seven base pairs are present in the hairpin formed by d(ATCCTA-CTTG-TAGGAT), i.e., a loop consisting of only two nucleotides is apparent. On the other hand, the results also indicate that for the sequence d(ATCCTA-GTTC-TAGGAT) a loop of four residues is formed or that the stability or lifetime of the conceivable seventh G-C base pair is too small or too short, respectively, so that the corresponding imino proton resonance is not observed in the NMR spectrum. Of course, extra base pairing in the hairpin loop formed by d(ATCCTA-CTTG-TAGGAT) is expected to contribute to the overall stability of the folded form. UV-melting experiments were performed to investigate this aspect.

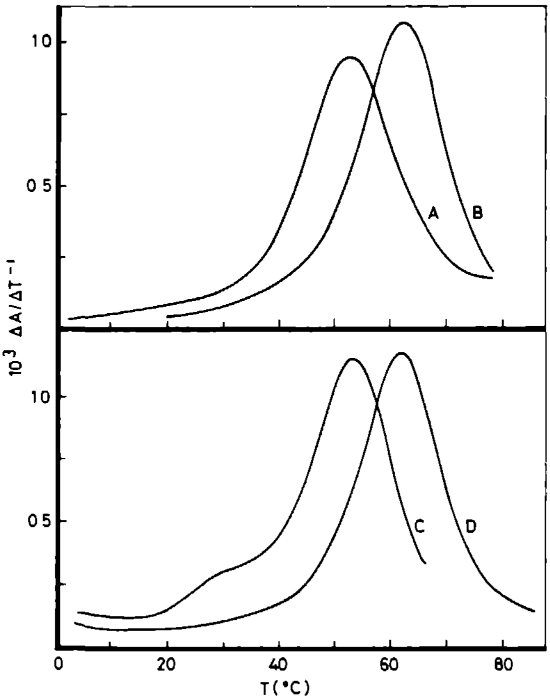


Fig. 2. Differential UV-melting profiles measured for the hexadecanucleotides d(ATCCTA-CTTG-TAGGAT) and d(ATCCTA-GTTC-TAGGAT) (200 mM NaCl). In the upper drawing the melting of d(ATCCTA-CTTG-TAGGAT) (curve B) is compared with the melting of d(ATCCTA-GTTC-TAGGAT) (curve A) at a total strand concentration of 3 μM. In the lower drawing the same comparison is made for 1 mM total strand concentration (NMR concentration). At both conditions the hairpins formed by d(ATCCTA-CTTG-TAGGAT) (curves B and D) are more stable than the hairpins formed by d(ATCCTA-GTTC-TAGGAT) (curves A and C). Note that for the latter fragment an extra transition, which is attributed to the melting of a dimeric species, appears at high concentration (curve C).

In Fig. 2 differential melting curves are presented for d(ATCCTA-CTTG-TAGGAT) (curve B) and for d(ATCCTA-GTTC-TAGGAT) (curve A) in the presence of 200 mM NaCl. The enhanced stability of d(ATCCTA-CTTG-TAGGAT) in comparison with that of d(ATCCTA-GTTC-TAGGAT) is obvious from the observed melting temperatures being 62 °C and 53 °C, respectively. These melting temperatures are independent of the oligomer concentration (vide infra) demonstrating that only mono-molecular transitions are involved. The corresponding ΔH values are 46 and 38 kcal/mol, respectively. The difference of 8 kcal/mol between these two values is in reasonable agreement with the value expected for the stacking enthalpy of an extra C-G base pair on top of an A-T base pair [19]. Since the concentration of the oligonucleotides, used in the UV-melting experiments, is much lower than we used in NMR experiments (3 μM vs. 1 mM) we also investigated the UV-melting behavior at a much higher oligonucleotide concentration. The results are given in Fig. 2, where differential melting curves are presented for d(ATCCTA-CTTG-TAGGAT) (curve D) and d(ATCCTA-GTTC-TAGGAT) (curve C) at 1 mM concentration. It is clear that for d(ATCCTA-CTTG-TAGGAT) the melting temperature is the same as at the low concentration (curve B) showing that also under 'NMR conditions' hairpin molecules are present. For d(ATCCTA-GTTC-TAGGAT) two transitions are observed of which one corresponds to that of curve A, indicating that this is the hairpin-to-coil transition. The transition occurring around 28 °C is concentration dependent, which shows that in this transition a duplex with an internal loop is melting out. These results correlate perfectly with the number of resonances in the NMR spectra obtained at similar conditions (cf. Fig. 1C).

For the analogous molecules d(ATCCTA-TTTA-TAGGAT) and d(ATCCTA-ATTT--TAGGAT) similar experiments were performed. Again for the former molecule seven imino proton resonances were found (see Fig. 3), where the imino proton spectrum is compared with that recorded for d(ATCCTA-TTTT-TAGGAT) showing six imino proton resonances at low field. The optical melting experiments were conducted for solutions with 3 μM DNA and 200 mM NaCl. In analogy with the results discussed above, we find a higher melting temperature for the d(ATCCTA-TTTA-TAGGAT) sequence, i.e., 57 °C, than for the d(ATCCTA-ATTT-TAGGAT) sequence , i.e., 50 °C. The corresponding ΔH values are 45 and 39 kcal/mol, respectively. Again in analogy with the first two hexadecanucleotides, we find that only for d(ATCCTA-ATTT-TAGGAT) a dimeric species is formed at high oligomer and/or high salt concentrations. The results are summarized in Table I.

We have also considered the influence of the base pair sequence in the common stem region of the molecules. To this end three hexadecanucleotides were synthesized with the sequence d(ATTACG-XTTY-CGTAAT), in which the sequence of the intervening residues was -TTTT-, -CTTG- and -GTTC- respectively. Again the same trends were observed as for the series of partly self-complementary molecules discussed above. This is demonstrated in Fig. 4 where the imino proton spectra of the hairpins formed by d(ATTACG-XTTY--CGTAAT) are shown. For the intervening sequence -CTTG- an extra imino proton resonance is observed at 12.9 ppm, demonstrating the formation of a

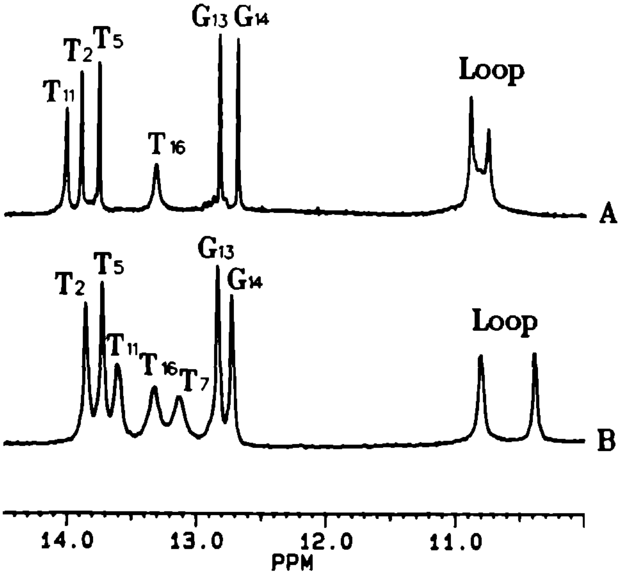


Fig. 3. 500 MHz imino proton spectra of the hairpins formed by d(ATCCTA--TTTT-TAGGAT) (A) and d(ATCCTA-TTTA-TAGGAT) (B) at 200 mM NaCl. The imino proton resonances between 15 and 12 ppm belong to base pairs, those upfield from 11 ppm to unpaired thymidines in the loop region. The spectra were recorded at 283 and 298 K, respectively; the total strand concentration amounts to 1 mM DNA. The numbering of the resonances corresponds with that in Fig. 5.

seventh base pair in the stem. This is not found for the intervening sequence -GTTC- but it is noted that a broad resonance is visible at 11.5 ppm for the unpaired (or partially paired) G residue. Also the difference in stability of these hairpins was as expected, i.e., the stability of the hairpin with the -CTTG- sequence was significantly higher than that of the other two hairpins (see Table I).

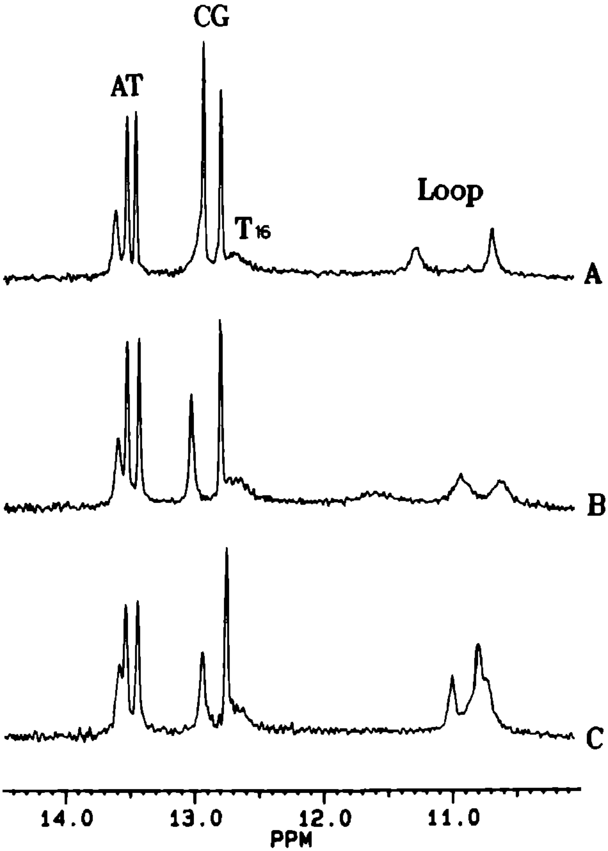


Fig. 4. 500 MHz imino proton spectra of the hairpins formed by d(ATTACG--CTTG-CGTAAT) (A), d(ATTACG-GTTC-CGTAAT) (B) and d(ATTACG-TTTT-CGTAAT) (C) at 5 mM NaCl. The imino proton resonances between 15 and 12 ppm belong to base pairs, those upfield from 11 ppm to unpaired thymidines in the loop region. The spectra were recorded at 298 K; the total strand concentration amounts 1 mM DNA.

Table I. Melting temperatures and enthalpies of the hairpin-to-coil transition of the examined DNA hairpins. All data were obtained by means of a least square fit of the experimental UV-absorbance - temperature curve (See Materials & Methods). a) This value is somewhat less reliable than the other melting enthalpies.

Oligonucleotide	T_m (°C)	ΔH (kcal/mol)
d(ATCCTA-CTTG-TAGGAT)	62	46
d(ATCCTA-GTTC-TAGGAT)	53	38
d(ATCCTA-TTTA-TAGGAT)	57	45
d(ATCCTA-ATTT-TAGGAT)	50	39
d(ATTACG-CTTG-CGTAAT)	67	40[a]
d(ATTACG-GTTC-CGTAAT)	53	38
d(ATTACG-TTTT-CGTAAT)	52	37
d(ATCCTA-TAAA-TAGGAT)	49	36
d(ATCCTA-TTTT-TAGGAT)	53	41
d(ATCCTA-AAAA-TAGGAT)	43	27
d(ATCCTA-AAAC-TAGGAT)	47	35
d(ATCCTA-AACA-TAGGAT)	48	34
d(ATCCT -AAAT -AGGAT)	47	30
d(ATCCTA-TTTTTT-TAGGAT)	47	41
d(ATCCTA-AAAAAA-TAGGAT)	38	24

Since earlier NMR experiments [3,12] and molecular mechanics calculations [8,9] indicated that the loop folding in the hairpins considered might be rather tight we investigated the influence of the introduction of more bulky bases in the loop sequence. Hairpins where adenines are introduced in the loop region, i.e., -AAAA-, -AAAC-, -AACA-, and -AAAAAA-, show a lowered stability with respect to hairpins where T residues are located at homologous positions (cf. Table I). A study of the hairpin formation by d(ATCCTA--TAAA-TAGGAT) shows that the imino proton spectrum of this molecule corresponds with the formation of a hairpin with six base pairs in contrast to d(ATCCTA-TTTA-TAGGAT) (vide supra); the imino proton resonance of T7 is observed in d(ATCCTA-TAAA-TAGGAT) as a broad line around 11 ppm (not shown). In line with this finding is the observation that the melting temperature drops from 57 °C for d(ATCCTA-TTTA-TAGGAT) to 49 °C for d(ATCCTA-TAAA-TAGGAT).

The hairpin formed by d(ATCCTA-TTTA-TAGGAT) was further examined by means of 2D-NMR. Part of its NOESY spectrum is presented in Fig. 5; in this region the H1' and H8/H6 cross peaks are found. This part of the spectrum was used to assign the ring proton resonances and the sugar H1' resonances in the standard manner [20-22]. It can be seen that when one starts at the cross peak generated by the 5'-terminal residue (A1), the sequential assignment can be continued to the H1' and H6 resonance of residue T8. On the other hand, when one starts at the 3'-end (T16) the assignment can be continued to T9. On the basis of these assignments the J-coupling patterns

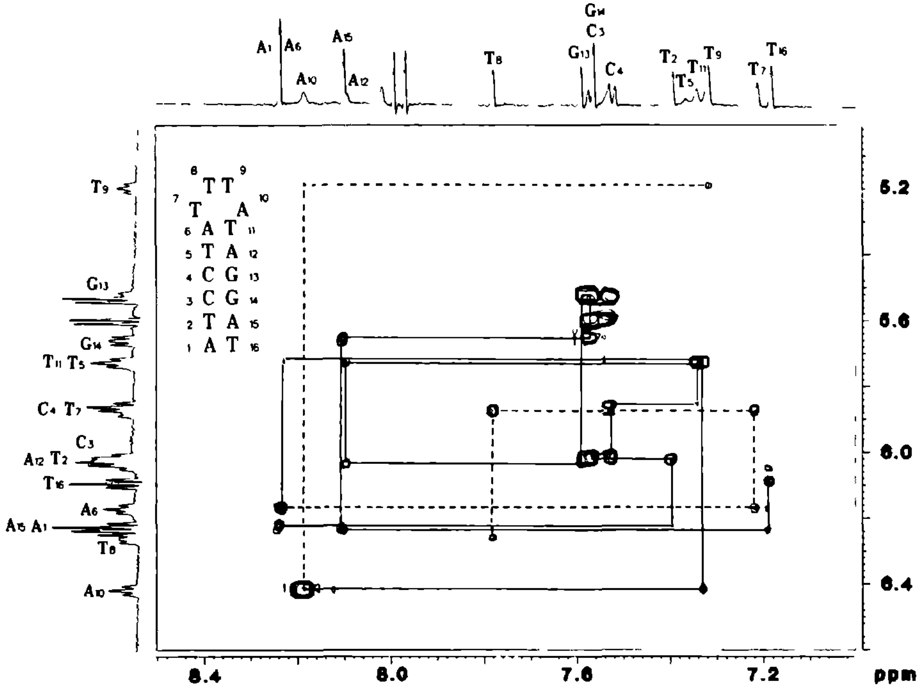


Fig. 5. Part of the 600 MHz phase sensitive NOESY spectrum of d(ATCCTA-TTTA--TAGGAT), dissolved in D_2O, exhibiting cross peaks between aromatic H6/H8 ring protons and sugar H1' protons. The spectrum was obtained for a 1 mM DNA sample at 1 mM NaCl, pH 7 and at a temperature of 298 K. The mixing time was 0.3 s. The sequential analysis of the cross peaks is indicated. The drawn lines represent the connectivity diagram for residues in the stem; the dashed lines represent the connectivity diagram of the residues in the loop. The assignment of the base H6/H8 protons is indicated along the horizontal axis, the assignment of the H1' sugar protons along the vertical axis.

of the H1'-protons were analyzed. It was found for all of the sugars that the sum $J_{1'2'} + J_{1'2''} > 13.3$ Hz which indicates that they predominantly adopt an S-type conformation [23]. Scrutiny of the H8/H6-H1' cross peak intensities shows that the intra residue cross peak intensity of residue A10 is abnormally high. This can be seen more clearly by considering the cross sections presented in Fig. 6 showing the Overhauser effects of three different ring protons, namely of H8 of A10 and of H6 of T7 and C4, to the H1' and H2'/H2" resonances. It is obvious that the intensity of the H8-H1' cross peak of adenine A10 is much higher than those of thymidine T7 and

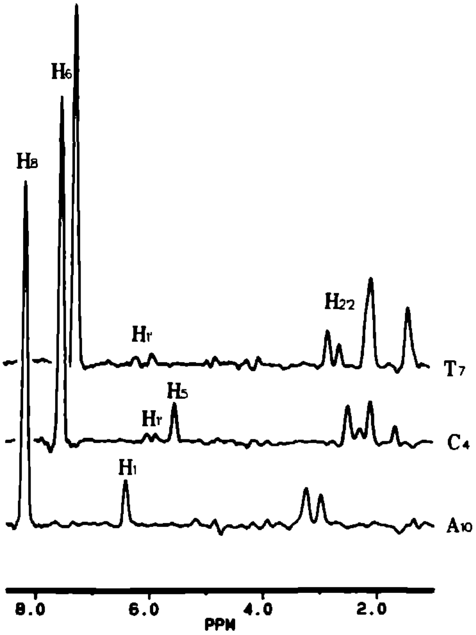


Fig. 6. Cross sections through the NOESY spectrum of Fig. 5 taken along ω_1 for three different ω_2 values so that cross peaks from the H6/H8 ring proton resonances to other proton resonances are visible for residues T7, C4 and A10. The assignments of various cross peaks are indicated.

cytidine C4, but comparable to the H5-H6 cross peak of cytidine C4. It follows from a NOESY experiment performed with a mixing time of 75 ms that this effect is not caused by spin diffusion. Then, in the spectral region considered, only the H8-H1' cross peak of A10 and the H5-H6 cross peaks of C4 are retained with about the same intensity, indicating that the distance between the adenosine H8 and H1' spins is near 2.5 Å. This strongly suggests that the adenine base is in a syn-conformation. However, the UV-melting experiments and the imino proton spectra indicate that this base (A10) is involved in base pairing (vide supra). This is only possible when the base pair is of the Hoogsteen type [24] (cf. Fig. 7). Confirmation of the occurrence of such a base pair can be obtained by consideration of intra base pair proton-proton distances. In a Hoogsteen pair the distance between the thymidine imino proton and H8 of the adenine residue is 2.6 Å. Thus, one expects to observe a cross peak between the resonances of these two protons in a NOESY spectrum recorded for a H_2O solution, as indeed is borne out by experiment (cf. Fig. 7). In addition, a spin diffusion induced cross peak between the imino proton resonance and the adenine H1' sugar resonance is observed. To avoid any possible confusion with cross peaks between imino proton resonances and adenine H2 resonances, the hairpin was deuterated at the purine C8 position. It followed that all resonances at and around 8.2 ppm (the position of the adenine H8 resonance) were reduced to 10 % of their original intensity. In the NOESY spectrum no connectivity between the imino proton resonance of thymine and the ring proton spectral region nor to the H1' proton spectral region was found. All these results demonstrate that T7 and A10 in d(ATCCTA-TTTA-TAGGAT) form a T-A Hoogsteen base pair with the adenine base in the syn-conformation.

Similar experiments were performed for the hexadecanucleotide d(ATCCTA--CTTG-TAGGAT) which also forms a hairpin with an additional base pair (vide supra). Here we did not find any indications that the seventh base pair, the C7-G10, forms a Hoogsteen pair. Instead, the intensity pattern of the NOE cross peaks is in accordance with the presence of a Watson-Crick type base pair.

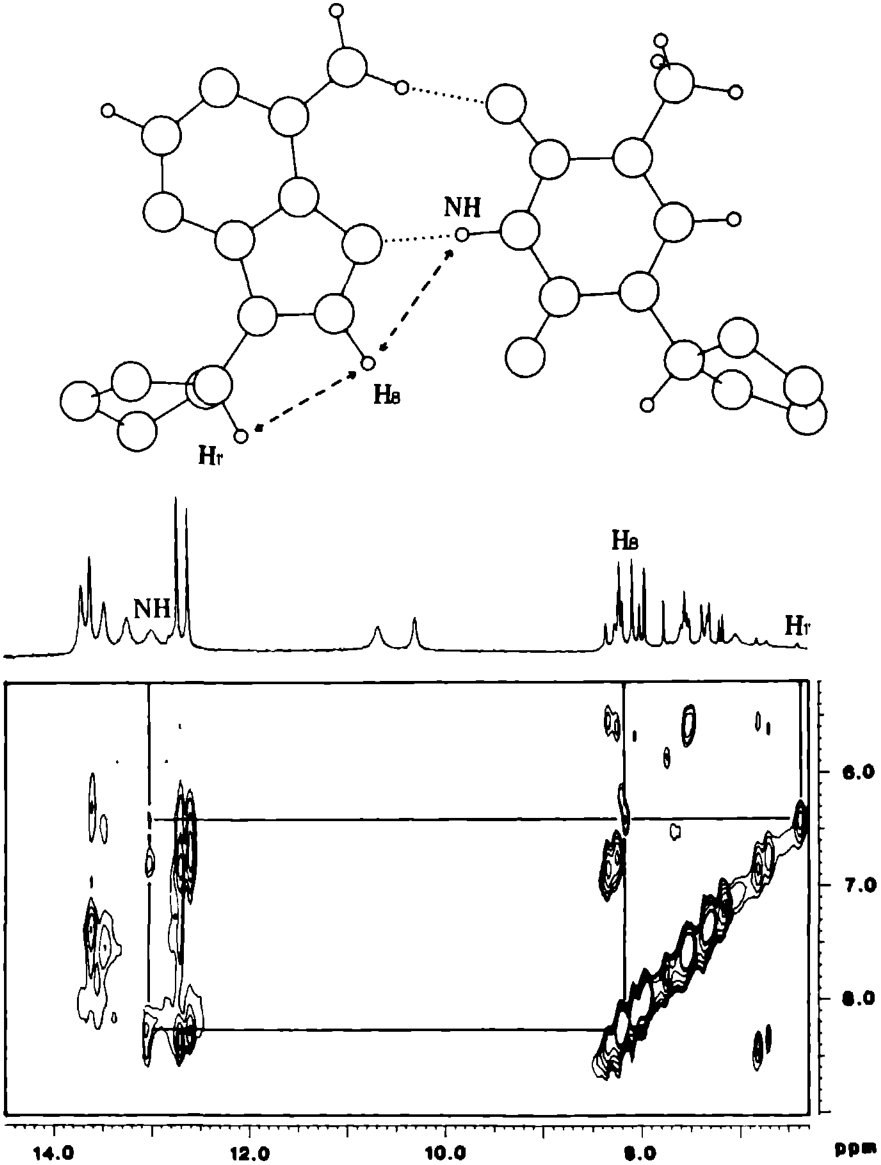


Fig. 7. Part of the 600 MHz NOESY spectrum of d(ATCCTA-TTTA-TAGGAT) dissolved in H_2O. The spectrum was obtained for a 1 mM DNA sample containing 1 mM NaCl, pH 6, at a temperature of 298 K. The mixing time was 0.3 s. The cross peaks between the imino proton resonance of residue T7 and the H8 and H1' resonances of residue A10 are indicated. The geometry of the Hoogsteen base pair displayed on top corresponds to the observed intensities of the cross peaks.

DISCUSSION

The experiments performed in the present study were inspired by two earlier observations. In our laboratory, it was found that the hexadecanucleotide d(ATCCTA-TTTT-TAGGAT) folds into a hairpin loop in which in the expected loop region a wobble T-T pair is formed [12,13] so that apparently a loop of two residues is present. For the octanucleotide d($C_mGC_mGTGC_mG$) Orbons et al. observed that a mini-hairpin can be generated also leading to a loop of two bases [10,11]. In Chapter 4, we concluded to the existence of a wobble T-T pair on the basis of sequential NOE measurements and NOE measurements between imino protons. It followed from these experiments that the first and the last thymine base of the loop is stacked upon the stem of the hairpin and that their imino protons are arranged in such a way that these bases can form a pair. That actual hydrogen bond formation occurs follows from NMR melting experiments which show that the mentioned imino protons are bleached out from the spectrum at the same temperature as the majority of the imino protons in the stem. This indicates that the T-T pair has a lifetime similar to that of the other base pairs in the molecule. The resonance position of the imino protons in the T-T pair is somewhat higher than is normally found for uridine or thymine imino protons involved in N-H...O hydrogen bonds. This we do not completely understand at this moment. In contrast, for the hairpins studied in this chapter the existence of a seventh base pair is only accepted when an extra resonance is observed between 12.5 and 14 ppm and this needs some explanation. Contrary to the observations described above we found that for molecules with -PuTTPy- loop sequences the exchange of the imino protons in the loop region is enhanced compared by those of the stem region of the hairpin and also compared to the loop region of molecules with -TTTT- and -PyTTPu- loop sequences (cf. Figs. 1,3 and 4). The imino proton resonances of the -PuTTPy- loop sequences are strongly broadened. In the case of d(ATTACG-GTTC-CGTAAT) we were fortunate to be able to see the resonance of G7, as a very broad line but separated from the resonances at 11 ppm. For the other molecules with -PuTTPy- loop sequences, the corresponding resonances are bleached out completely or hidden beneath the broadened resonances near 11 ppm. This can only mean that for the -PuTTPy- sequences the base pair is not at all or transiently formed, in contrast to the -TTTT- loop sequence. This conclusion is further corroborated by the occurrence of a seventh resonance in 12.5-14 ppm region for the molecules with -PyTTPu- loop sequences, which is diagnostic for the existence of a seventh base pair, and by the comparative UV-melting experiments.

The observation of the two-membered loops came thus originally as a surprise because in a regular B-type (or A-type) helix the distance between the two terminal phosphates at one end of the stem is about 17.5 Å and bridging this distance by two nucleotides was thought to be sterically impossible [25]. For the aforementioned hexadecanucleotide d(ATCCTA-TTTT-TAGGAT) and octanucleotide d($C_mGC_mGTGC_mG$) it was observed that the torsion angle γ of the residue directly following the two-membered loop adopts a trans conformation instead of the standard gauche$^+$ conformation found in normal B-DNA [11,12]. This adjustment of the backbone conformation decreases the distance between the 5'-end phosphate of the stem and the second loop residue, thus facilitating the formation of a two-membered loop.
Another interesting feature of this type of loop closure is displayed by modelling studies [9,26] (Chapter 5) which show that the C-G base pair closing such a two-membered loop is not flat but buckled and hence points to a strained conformation.

The results of the present chapter show that there is an additional mechanism by which the distance between the terminal phosphates at one end of the stem can be diminished and this is the totally unexpected formation of a Hoogsteen base pair. The distance between the C1' sugar carbons, which is 10.4 Å in a standard Watson-Crick base pair, is diminished to 8.6 Å in a Hoogsteen pair (cf. Chapter 1; Fig. 10). This will concomitantly decrease the distance between the phosphates at the end of the stem where the Hoogsteen pair is situated and will additionally facilitate the formation of a two-membered loop. The occurrence of Hoogsteen base pairs as part of the end of double-helical regions in nucleic acids is relatively rare. In yeast-tRNA^{Phe} two reversed Hoogsteen base pairs, U8-A14 and T54-m^{1}A58, are observed [27]. They facilitate the occurrence of sharp bends in the molecule and may in this respect resemble the role of the T-A Hoogsteen base pair in our hairpin. Hoogsteen T-A base pairs have also been observed in the complex of the bis-intercalator triostin-A and the double helix formed by d(CGTACG)$_2$ [28]. Here, the reduced distance between the C1' carbon atoms of the A-T pairs serves to stabilize the complex through a close packing of the oligonucleotide near the end of the triostin-A molecule. These observations

illustrate that nucleic acids quite easily adjust their structure in order to accommodate local energetic needs so as to avoid highly localized strain situations and to optimize local interactions [9]. Our experiments do show that the formation of loops consisting of residues in partly self-complementary molecules depends on the sequence of the intervening residues. With a pyrimidine at the 5'- and a purine at the 3'-end of this intervening sequence two-membered loops are easily formed in the molecules considered, while inversion of this sequence makes the formation of such a loop much more difficult if not impossible. It is noted that replacement of the two central thymidines by more bulky bases appears to invalidate this simple rule in the sense that in that case a four-membered loop is retained.
We also studied the salt and concentration dependence of the UV-melting profiles for the oligonucleotides. At higher strand and salt concentrations the d(ATCCTA-PuTTPy-TAGGAT) and not the d(ATCCTA-PyTTPu-TAGGAT) sequences give rise to dimer formation. The first type of fragments does not form the two-membered loop, but in the duplex form the Py and Pu are involved in base pair formation which explains the relative preference for the dimer structure compared to the second fragment where Py and Pu can pair in the hairpin form. Perusal of the literature shows that these considerations can be extended to some self-complementary sequences as well. For instance a comparison of the melting behavior of the self-complementary sequences d(CGCG-AATT-CGCG) and d(CGCG-TATA-CGCG) shows that it is easier for the latter molecule to form hairpin structures. Thus at 0.07 mM strand concentration d(CGCG-AATT-CGCG) does not form a hairpin in 0.01 M NaCl [29] while at 0.38 mM strand concentration d(CGCG-TATA-CGCG) still is able to form hairpins in 0.1 M NaCl [30].

At first sight the above pieces of information seem rather unrelated: in some cases a two-membered loop closed by a distorted base pair (buckled or Hoogsteen-type) is apparent, in other cases sometimes only subtly changed sequences give rise to four-membered loops. Although it may seem a matter of semantics we are inclined to think of the above mentioned two-membered loops as four-membered loops in which local optimization (e.g., base pairing) occurs. Such a view has its completely accepted analogies, e.g., the extra base pair T54-m^1A58 in $tRNA^{Phe}$ does not make the TΨC-loop a three membered one nor is the m_2G26-A44 base pair in the same molecule normally reckoned to be part of the anticodon stem.

Seen in this light all the above hairpins fall within the scope of a general loop folding model for which under specific conditions further architectural details can be stated. For example, the present results demonstrate that for intervening loop sequences -PyXXPu- (Py = pyrimidine, X = thymidine, Pu = purine complementary to Py) the terminal bases are involved in some kind of base pairing scheme, while reversal of this sequence, i.e., -PuXXPy-, makes base pairing much more difficult if not impossible. Bulkiness of the two central bases in the loop is another limiting factor in local optimization: when the first X base is a purine (cf. d($C_mGC_mGTGC_mG$)) an extra base pair may be formed, but replacing both X-bases by purines (cf. d(ATCCTA-TAAA-TAGGAT)) limits the hairpin to the basic four-membered loop scheme.

Although the observed effects can thus be formulated in terms of a simple rule, the available dataset is still too small to declare the model generally applicable to all DNA hairpins. A case in point is formed by the results reported for d(CGCG-TTTT-CGCG) [31] and d(CGCGCGCG-TTTT-CGCGCGCG) [32]. The authors claim a loop folding for the four central T's different from our loop folding model in the sense that stack propagation from the 3'-end of the stem into the loop is limited to one thymidines, the second T is supposed to be rotated away from the center of the loop. Although we have some doubts about the ability of the used NOESY-distance geometry methods to deduce such conformational details in nucleic acids in an 'ab initio' way [23], we cannot offer an explanation for this aberration apart from the observation that these loops have in common that they are folded on top of an all C-G stem. This is also true for the results obtained by Xodo et al. [34]. These authors found that neither the melting temperatures nor the melting enthalpies measured for d(CGCGCG-TTTTT-CGCGCG) and d(CGCGCG-AA-AAA-CGCGCG) differed within experimental accuracy. This is in contrast to our results; upon substitution of all intervening thymidines by adenines the melting temperature as well as the transition enthalpy is significantly lower (cf. Table I). Very recently, similar results were obtained by Breslauer and collaborators [35]. These authors demonstrate that the stability of the related hairpins formed by d(CGAACG-XXXX-CGTTCG) decrease in the order X = T, C, G and A, respectively. It is noted in passing that the NOESY cross peak patterns of the loop regions of d(CGAACG-TTTT-CGTTCG) and d(CGAACG-AAAA-CGTTCG) agree with the inter-residue connectivity pattern observed for d(ATCCTA-TTTT-TAGGAT) [12]: connectivities are observed between the loop residues X7 to X9, while no connectivity is observed between the loop residues X9 and X10. These results are in perfect agreement with our loop folding principle, i.e., the base stacking pattern of the loop bases is propagated on top of the 3'-end of a B-type double-helical stem, while after the third loop base the sugar-phosphate backbone changes its direction [12].

The thesis put forward by Xodo et al. [34] that the self-complementary sequences d(CGCGC-GATC-GCGCG) and d(CGCGC-GTAC-GCGCG) form hairpins in which the loops consist of only the two central (A/T) bases and G6-C9 base pairs close the loop is obviously also deviating from our results, i.e., -PuXXPy-sequences forming four-membered loops only. However, we note that their thesis is based on UV-melting data recorded for a very limited set of hairpins. Their main argument is that the enthalpy changes on denaturation of the hairpins is ca. 3 kcal higher than that of the 'unconstrained' hexamer duplex $d(CGCGCG)_2$. That this finding in itself is not sufficient for drawing conclusions regarding the loop size is exemplified by our previously reported results for d(ATCCT-ATTTT-AGGAT). In the latter molecule the hairpin denaturation enthalpy is also slightly higher (ca. 3 kcal/mol) than that of the corresponding helix formed by the 'unconstrained' hexamer duplex d(ATCCTA)*d(TAGGAT) [4,8], but NMR experiments indicate [7,8] that the stem of the hairpin formed by this molecule involves only five base pairs (no imino proton resonance was found for the conceivable pair A6-T10). In other words, the enthalpy contribution of the loop to the overall hairpin stability may be equal or exceed that of one base pair [4,7,8].

In summary, the present results demonstrate that nucleotide sequence and composition have a marked influence on loop folding and hairpin stability. Notwithstandingly, the results can be explained by a rather simple model in which under certain conditions local optimization of the hairpin structure within the conformational space delineated by our loop folding principle may occur.

REFERENCES

1 Saenger, W. (1984) Principles of Nucleic Acid Structure. 1th edition, Springer-Verlag. New York.

2 Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A. & Weiner, A.M. (1987) Molecular Biology of the Gene. 4th ed. The Benjamin/Cummings Publishing Company Inc. Menlo Park, California.

3 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) J. Biomol. Struct. Dyns. **1**, 115-129.

4 Hilbers, C.W., Haasnoot, C.A.G., de Bruin, S.H., Joordens, J.J.M., van der Marel, G.A. & van Boom, J.H. (1985) Biochimie **67**, 685-695.

5 Uhlenbeck, O.C., Borer, P.N., Dengler, B & Tinoco, I. (1973) J. Mol. Biol. **73**, 483-496.

6 Gralla, J. & Crothers, D.M. (1973) J. Mol. Biol. **73**, 497-511.

7 Haasnoot, C.A.G., de Bruin, S.H., Hilbers, C.W., van der Marel, G.A., van Boom, J.H. (1985) Proc. Int. Symp. Biomol. Struct. Interactions, Suppl. J. Biosci. **8**, 767-780.

8 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) J. Biomol. Struct. Dyns. **3**, 843-857.

9 Haasnoot, C.A.G., Blommers, M.J.J., Hilbers, C.W. (1987) Springer Series in Biophysics **1**: Structure, Dynamics and Function of Biomolecules (Eds. A. Ehrenberg, R. Rigler, A. Gräslund, L. Nilsson), 212-216.

10 Orbons, L.P.M., van der Marel, G.A., van Boom, J.H. & Altona, C. (1986) Nucl. Acids Res. **14**, 4187-4196.

11 Orbons L.P.M., van Beuzekom, A.A. & Altona, C. (1987) J. Biomol. Struct. Dyns. **4**, 965-987.

12 Blommers, M.J.J., Haasnoot, C.A.G., Hilbers, C.W, van Boom, J.H., van der Marel, G.A. (1987) Structure and Dynamics of Biopolymers, NATO ASIS Series E:Applied Sciences **133**, 78-91.

13 Hilbers, C.W., Blommers, M.J.J., Haasnoot, C.A.G., van der Marel, G.A.& van Boom, J.H. (1987) Fresenius Z. Anal. Chem. **327**, 70-71.

14 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) Cold Spring Harbor Symp. Quant. Biol. **47**, 403-409.

15 Tibanyenda, N., de Bruin, S.H., Haasnoot, C.A.G., van der Marel, G.A., van Boom, J.H., Hilbers, C.W. (1984) Eur. J. Biochem. **139** , 19-27.

16 Gralla, J. & Crothers, D.M. (1973) J. Mol. Biol. **78**, 301-319.

17 Haasnoot, C.W. & Hilbers, C.W. (1983) Biopolymers **22**, 1259-1266.

18 Marion, D. & Wüthrich, K. (1983) Biochem. Biophys. Res. Com. **113**, 967-974.

19 Breslauer, K.J., Frank, R., Blocker, H. & Marky, L.A. (1986) Proc. Natl. Acad. Sci. USA **83**, 3746-3750.

20 Haasnoot, C.A.G., Westerink, H.P., van der Marel, G.A. & van Boom, J.H. (1983) Biomol. Struct. Dyns. **1**, 131-149.

21 Scheek, R.M., Russo, N., Boelens, R. Kaptein, R. & van Boom, J.H. (1983) J. Am. Chem. Soc. **105**, 2914-2916.

22 Hare, D.R., Wemmer, D.E., Chou, S.-H., Drobny, G. & Reid, B.R. (1983) J. Mol. Biol. **171**, 319-336.

23 Rinkel, L.J. & Altona, C. (1987) J. Biomol. Struct. Dyns. **4**, 621-649.

24 Hoogsteen, K. (1963) Acta Crystallogr. **16**, 907-916.

25 Tinoco, I., Uhlenbeck, O.C. & Levine, M.D. (1971) Nature (London) **230**,

362-370.
26 van Beuzekom, A.A. (1989) Ph.D. Thesis, State University Leiden
27 Rich, A. & RajBhandary, U.L. (1976) Ann. Rev. Biochem. **45**, 805-860.
28 Wang, A.H.-J., Ughetto, G., Quigley, G.J., Hakoshima, T., van der Marel, G.A., van Boom, J.H. & Rich, A. (1984) Science **225**, 1115-1121.
29 Marky, L.A., Blumenfield, K.S., Kozlowski, S. & Breslauer, K.J. (1983) Biopolymers **22**, 1247.
30 Wemmer, D.E., Chou, S.H., Hare, D.R. & Reid, B.R. (1985) Nucl. Acids Res. **13**, 3755.
31 Hare, D.R. & Reid, B.R. (1986) Biochemistry (1986) **25**, 5341-5350.
32 Wolk, S.K., Hardin, C.C., Germann, M.W., van de Sande, J.H. & Tinoco,I. (1988) Biochemistry **27**, 6960-6967.
33 van de Ven, F.J.M. & Hilbers, C.W. (1988) Eur. J. Biochem. 178, 1-38.
34 Xodo, L.E., Manzini, G., Quadrifoglio, F., van der Marel, G.A. & van Boom, J.H. (1988) Biochemistry **27**, 6321-6326.
35 Senior, M.M., Jones, R.A. & Breslauer, K.J. (1988) Proc. Natl. Acad. Sci. USA **85**, 6242-6246.

THREE-DIMENSIONAL STRUCTURE OF A DNA HAIRPIN IN SOLUTION.
2-DIMENSIONAL NMR STUDIES AND MULTI-CONFORMATIONAL ANALYSIS OF d(ATCCTA-TTTA-TAGGAT).

ABSTRACT

The B-type hairpin formed by d(ATCCTA-TTTA-TAGGAT) was studied by means of 2-dimensional NMR spectroscopy and multi-conformational analysis. The ^{1}H-resonances of the stem region were assigned almost completely and the ^{1}H- and ^{31}P-spectrum of the loop region was interpreted completely; this includes the stereo specific assignments of the H5' and H5" resonances. The conformational details of this hairpin loop were examined by means of multi-conformational analysis. In this procedure the conformational space spanned by the glycosidic torsion angle and pseudorotational parameters of each mononucleotide unit were sampled systematically on the basis of the available experimental J-coupling constants and NOE-intensities. Inter-residue distances between protons were calculated by means of a procedure in which the simulated and experimental NOEs were fitted, using a relaxation matrix approach, but without the introduction of a molecular model. The backbone torsion angles β, γ and ε were deduced from homo- and hetero-coupling constants. The structure of the loop region was then examined with the aid of multi-conformational analysis. The obtained conformation was subsequently docked on a B-type stem. Restrained molecular dynamics, restrained as well as free molecular mechanics calculations strongly corroborate the results obtained by means of the multi-conformational analysis.

The structural appearance of the hairpin loop is a compactly folded conformation: the first base of the central -TTTA- region forms a Hoogsteen T-A pair with the fourth base. This Hoogsteen base pair is stacked upon the sixth base pair of the B-type double helical stem. The second base of the loop is folded into the minor groove, whereas the third loop base is partly stacked on the first and fourth base. The phosphate backbone exhibits a sharp turn between the third and fourth loop-nucleotide. The peculiar structure of this hairpin loop is discussed in relation to loopfolding in DNA and RNA hairpins and supports a general model for loopfolding.

INTRODUCTION

The studies conducted in the previous chapters on the folding patterns in hairpins have shown that the structure formed in loop regions depends on the type of double helix, which forms the stem of the hairpin, probably on

---

Blommers, M.J.J., van de Ven, F.J.M., van der Marel, G.A., van Boom, J.H. & Hilbers, C.W. Eur. J. Biochem. (submitted)

the base pair sequence of the stem and on the base sequence of the loop region. For instance, for the DNA hairpins with four membered loops, studied in this thesis, it was found that the introduction of complementary bases in the loop region sometimes results in the formation of a base pair [1] (Chapter 6). This turned out to be the case for, for instance, the hairpin formed by d(ATCCTA-TTTA-TAGGAT). In the loop formed by the central residues -TTTA- a Hoogsteen base pair is formed between its first and fourth base. As has been indicated in Chapter 6 the formation of such a base pair on top of a regular double helix may facilitate the closure by two nucleotides of the gap formed by the 3'- and 5'-terminal phosphates of this Hoogsteen pair. However, additional determination of the backbone conformation (with respect to a normal B-type backbone) remains necessary. Therefore, in the present chapter the details of the loop structure of the hairpin formed by the title compound is studied in detail; since this molecule gives rise to excellent NMR spectra it is well suited to this purpose. In the course of the investigations new methods were introduced to derive the loop structure from the NMR data. In particular, it is mentioned that the application of multi-conformational analysis was investigated to obtain an unbiased and reliable sampling of the conformational space of the loop region.

MATERIALS AND METHODS

Experimental

The oligonucleotide d(ATCCTA-TTTA-TAGGAT) was synthesized via the phosphotriester method [2]. Samples were prepared by dissolving the oligonucleotide to a concentration of 3 mM in a solution containing 25 mM sodium phosphate, 1 mM sodium cacodylate and 0.1 mM EDTA at pH 7.0. NaCl was added to 200 mM sodium concentration. The NMR samples were prepared by three-fold lyophilization from D_2O (99.8 %).

400 MHz and 600 MHz ^{1}H-NMR experiments were performed on Bruker AM-400 and AM-600 spectrometers both interfaced to an Aspect 3000 computer. Phase sensitive 2D- spectra were recorded using time proportional phase incrementation [3]. The 400 MHz NOESY spectra were recorded with a spectral width of 4000 MHz, using a mixing time of 75, 125, 200, 300, 400 and 500 ms respectively. The 600 MHz NOESY spectrum was recorded with a spectral width of 6000 Hz, using a mixing time of 200 ms. The spectra were acquired, with 2K points in the t2- and 512 points in the t_1-direction. The time domain

data were zero-filled in both directions and were multiplied with a shifted sine function before Fourier transformation. The 400 MHz DQF-COSY spectrum was recorded with a spectral width of 4000 Hz, and was acquired with 1K points in the t1- and 4K points in the t2-direction. The ^{1}H-detected ^{31}P-^{1}H hetero-nuclear multiple quantum coherence (HMQC) spectrum [4] was obtained with the pulse sequence and phase list depicted in Fig. 1. This experiment was performed on a Bruker AM400 spectrometer equipped with a QNP probe. 256 experiments were recorded, each 352 scans, with a spectral width of 4000 Hz in t2 and 1500 Hz in t1.

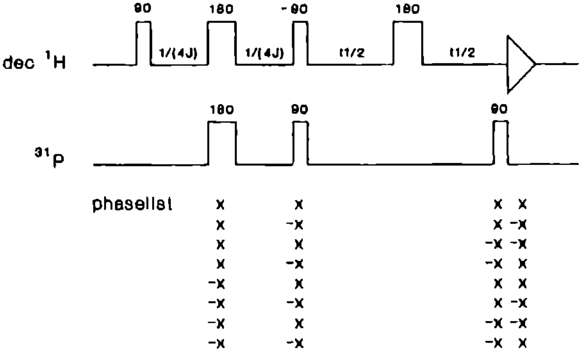


Fig. 1. Pulse scheme for the acquisition of the ^{1}H-detected hetero multiple quantum coherence (HMQC) ^{31}P-^{1}H spectrum. The phase list applied in this experiment is indicated.

Conformational analysis

The conformational analysis of the hairpin was performed according the following procedure. In the first step the conformation of each sugar residue was determined from the J_{HH}-coupling constants and/or sums of couplings determined for the sugar spin systems. Then, the orientation of each base relative to its own sugar was derived with the aid of intra-residue NOE intensities. Subsequently, distances between residues were evaluated from inter-residue NOEs. In addition, the domains of backbone torsion angles were derived from J-coupling information. These results were then combined with a multi-conformation generation step to obtain dinucleotide structures which are consistent with the available NMR constraints. This approach was extended to longer fragments, i.e., the -TTTA-fragment. The procedure is outlined in Fig. 2; a more detailed description following the scheme indicated in this figure is presented below.

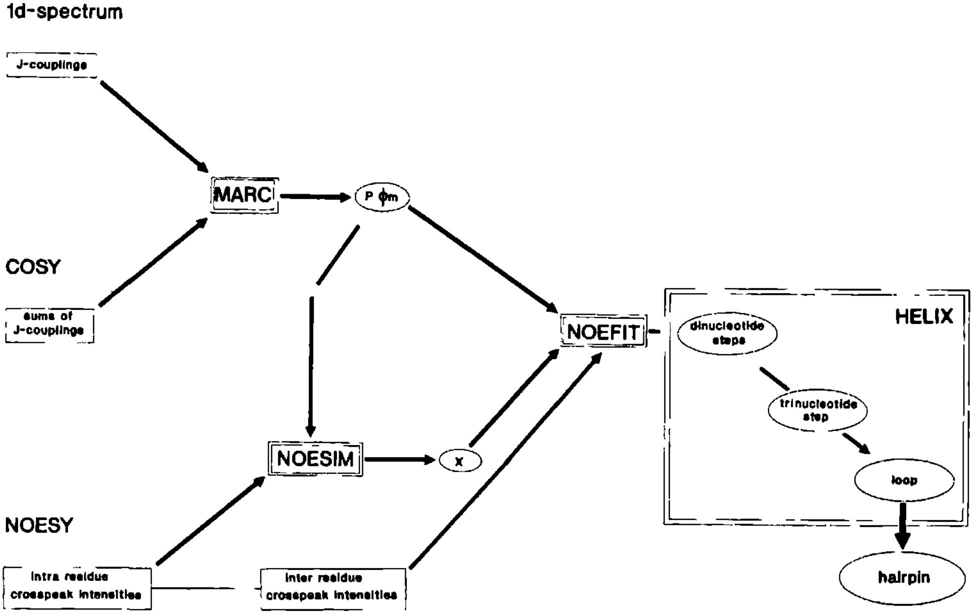

1d-spectrum
J-couplings
MARC
P ϕm
COSY
sums of J-couplings
NOEFIT
HELIX
dinucleotide steps
trinucleotide step
loop
hairpin
NOESIM
x
NOESY
Intra residue crosspeak intensities
Inter residue crosspeak intensities

Fig. 2. Flow chart of the procedure which is followed in the conformational analysis of the DNA hairpin. J-couplings and NOE cross peak intensities are used in the subsequent stages indicated in this figure: MARC (multi-conformational analysis of deoxy-ribose conformation), NOESIM (NOE simulation), NOEFIT (NOE fit) and HELIX (multi-conformational analysis of di-, tri- and tetra-nucleotides).

Stage 1: Multi-conformational analysis of the deoxyribose conformation

For the deoxyribose sugars the H-C-C-H torsion angles, ϕ_{ij}, are related to the pseudorotation parameters [5], which characterize the sugar conformation, by the following equations [6]:

$$
\begin{aligned}
\phi_{1'2'} &= 121.4 + 1.03\ \phi_m \cos(P - 144)\\
\phi_{1'2''} &= 0.9 + 1.03\ \phi_m \cos(P - 144)\\
\phi_{2'3'} &= 2.4 + 1.03\ \phi_m \cos(P)\\
\phi_{2''3'} &= 122.9 + 1.03\ \phi_m \cos(P)\\
\phi_{3'4'} &= -124.0 + 1.03\ \phi_m \cos(P + 144)
\end{aligned}
\tag{1}
$$

where P is the phase angle of pseudorotation and ϕ_m the pucker amplitude of the sugar. On the basis of these equations, sets of torsion angles were generated by systematically varying P and ϕ_m in steps of 5° or any other chosen step size, if so desired. These torsion angles served to evaluate the corresponding values of the J-couplings between the sugar proton spins. To this end the EOS-Karplus equation:

$$
\begin{aligned}
{}^3J_{HH} = {} & 13.22 \cos^2 \phi - 0.99 \cos \phi + \\
& \Sigma[\ 0.87 - 2.46 \cos^2 (\ \xi_i\phi + 19.9\ |\Delta\chi_i|\ )\ ]\ \ \Delta\chi_i
\end{aligned}
\tag{2a}
$$

was used for the calculation of $J_{1'2'}$, $J_{1'2''}$, $J_{2'3'}$ and $J_{2''3'}$, whereas for $J_{3'4'}$ the equations

$$
\begin{aligned}
{}^3J_{HH} = {} & 13.24 \cos^2 \phi - 0.91 \cos \phi + \\
& \Sigma[\ 0.53 - 2.41 \cos^2 (\ \xi_i\phi + 15.5\ |\Delta\chi_i|\ )\ ]\ \ \Delta\chi_i
\end{aligned}
\tag{2b}
$$

$$
\Delta\chi_i = \Delta\chi_{i,\alpha} - 0.19\ \Sigma\ \Delta\chi_{i,\beta}
\tag{2c}
$$

were applied [7]. In these equations ϕ represents the torsion angle defined by the two vicinal protons; $\Delta\chi_i$ stands for the difference in electronegativity between the substituents on the H-C-C-H fragment and hydrogen. In Eq. 2b the value of $\Delta\chi_i$ needs correction for the electronegativity of β-substituents, as indicated in Eq. 2c. ξ_i is +1 or -1 depending on the orientation of the substituent [7]. For certain values of P, the J1'2" and

J2'3' are sensitive to the so-called Barfield transmission effect. Corrections for this effect were introduced on the basis of the following expressions [8]:

$$\begin{aligned} \Delta J_{1'2''} &= -2.0 \cos^2 (P-234) \qquad 144 < P < 324 \\ \Delta J_{2'3'} &= -0.5 \cos^2 (P-270) \qquad 180 < P < 360 \end{aligned} \qquad (3)$$

Thus, at this point we have obtained a set of five coupling constants for each generated P, ϕ_m-pair, i.e., for each corresponding sugar conformation. In practice deoxyribose sugars are not rigid but they may adopt N-type or S-type conformations, which are in rapid equilibrium [5,9]. Therefore a dataset of J-couplings was generated, for different ratios of N- and S-conformations, which is composed of the weighted average of the J-couplings of each conformation:

$$J_{av} = (1 - x_S)\, J_N + x_S\, J_S \qquad (4)$$

To this end the fraction of sugars in the S-conformation, x_S, was systematically varied between 0 and 1 in steps of 0.05. The five coupling constants thus obtained for a large number of values P_N, ϕ_{mN}, P_S, ϕ_{mS} and x_S were written to a database. The size of the database depends on the user's choice of the boundaries of the conformational space and on the step size. Once the database is created it can be used in conjunction with a program which tests each line for the occurrence of J-couplings which satisfy the experimental constraints. E.g., when the $J_{1'2'}$ = 10 ± 0.5 Hz and Σ2' (= $J_{1'2'} + |J_{2'2''}| + J_{2'3'}$) = 24 ± 1 Hz are used as input, all values P_N, ϕ_{mN}, P_S, ϕ_{mS} and x_S which correspond with these constraints are written to a file. The computer program which is developed to perform the analysis is called MARC (multi-conformational analysis of deoxy-ribose conformation).

Stage 2: Determination of the glycosidic torsion angle χ

The torsion angle χ can now be derived from intra-residue cross peak intensities. The procedure used was as follows. For the mononucleotides the distances between H8/H6 and the sugar protons were calculated as a function of the phase angle of pseudorotation, P, and the glycosidic torsion angle χ. To this end P and χ were systematically varied between 0 and 360°. The results were used to determine the values of the elements of the relaxation matrix **R**, which is part of the Solomon-Bloch equations [10,11] that describe the time-dependence of the magnetization, **M**:

$$\frac{d}{dt}\ \mathbf{M}(t) = \mathbf{R}\ \mathbf{M}(t) \tag{5}$$

The elements of the relaxation matrix are:

$$R_{ii} = \sum_j (W_0^{ij} + 2\ W_1^{ij} + W_2^{ij}) \tag{6a}$$

$$R_{ij} = W_2^{ij} - W_0^{ij} \tag{6b}$$

where i and j stand for spins i and j respectively and the W's are the well known transition probabilities [10]:

$$W_0^{ij} = \frac{1}{10}\ \frac{\gamma^4\ \hbar^2}{r_{ij}^6}\ \tau_c \tag{7a}$$

$$W_1^{ij} = \frac{3}{10}\ \frac{\gamma^4\ \hbar^2}{r_{ij}^6}\ \frac{\tau_c}{1 + \omega^2\tau_c^2} \tag{7b}$$

$$W_2^{ij} = \frac{3}{5}\ \frac{\gamma^4\ \hbar^2}{r_{ij}^6}\ \frac{\tau_c}{1 + 4\omega^2\tau_c^2} \tag{7c}$$

r_{ij} is the distance between the spins i and j and τ_c is the rotation correlation time. The other symbols have their usual meaning. $\gamma^4\hbar^2/10 = 5.70\ 10^{10}\ s^{-2}Å^6$ for protons. The use of these equations implies that isotropic tumbling of the molecule is assumed. The value of the rotational correlation time can be estimated with the aid of the Stokes-Einstein equation [12]:

$$\tau_c = V_h\ \eta\ /\ kT \tag{8}$$

V_h is the hydrated volume, η is the viscosity of the solvent and k is Boltzman constant. For a spherical molecule dissolved in H_2O the value of τ_c at room temperature is then about $4.1\ 10^{-13}$ times the molecular weight. A formal solution to the differential matrix equation (Eq. 5) is:

$$\mathbf{M}(t) = e^{-t\mathbf{R}}\,\mathbf{M}(0) \tag{9}$$

where $\mathbf{M}(0)$ is the magnetization at times $t=0$. This equation can be solved by calculating the truncated Taylor expansion of the exponent (see Chapter 1), but it is better to solve this equation numerically in order to account for spin diffusion. Therefore the matrix $\mathbf{R}$ is diagonalized to λ [13]. If

$$\mathbf{X}^{-1}\mathbf{R}\,\mathbf{X} = \lambda \tag{10}$$

where $\mathbf{X}$ denotes the matrix of eigen-vectors and λ is the matrix of eigen-values of $\mathbf{R}$, then

$$e^{-t\mathbf{R}} = \mathbf{X}\,e^{-t\lambda}\,\mathbf{X}^{-1} \tag{11}$$

and substitution in (1) yields

$$\mathbf{M}(t) = \mathbf{X}\,e^{-t\lambda}\,\mathbf{X}^{-1}\,\mathbf{M}(0) \tag{12}$$

The intensities of the auto- and cross-peaks observed in a NOESY spectrum contained in the matrix $\mathbf{a}$:

$$\mathbf{a} = \mathbf{X}\,e^{-\lambda\tau m}\,\mathbf{X}^{-1} \tag{13}$$

which is the matrix of NOE intensities, τ_m is the mixing time. Using the relaxation matrix approach the intensities of the NOE cross peaks between the resonances of H6/H8 on one hand and the resonances of H1', H2', H2" and H3' on the other hand were determined as a function of P and χ (e.g., Fig. 3; Chapter 1). The program NOESIM [14] (see also Chapter 4) calculates the NOE cross peak intensities from a set of inter-proton distances using Eqs. 5-13. In the calculations also the influence of the other sugar protons including the H5' and H5" was included; the influence of the protons of the 5'-neighboring nucleotide was accounted for by positioning three protons at a distance of 2.3, 3.2 and 4.0 Å from H6/H8. These distances were taken from regular B-DNA. They only serve to describe the relaxation of the aromatic proton more realisticly. The results are plotted in Fig. 3. For this

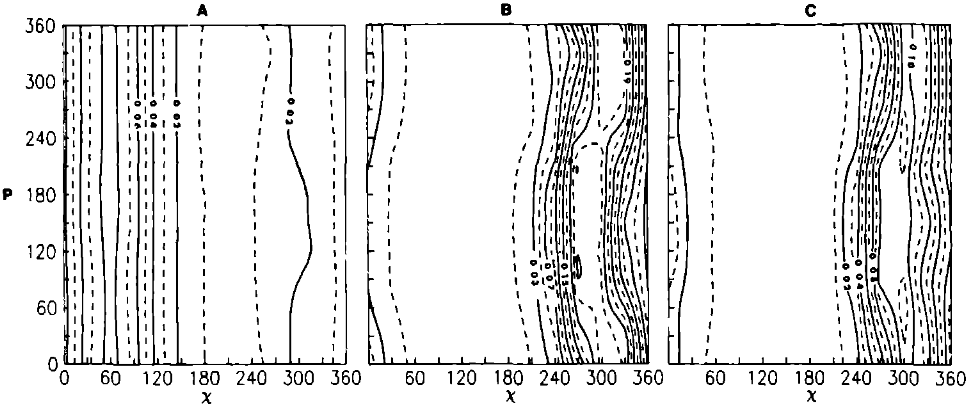


Fig. 3. Contour plots of NOE cross peak intensities as a function of the pseudorotational parameter P and the glycosidic torsion angle χ. The NOE intensities, presented for cross peaks between the aromatic proton H6 or H8 and the H1',H2' and H2" sugar protons. The calculation was carried out, using the program NOESIM, which accounts for spin diffusion (see text). In the calculations the spectrometer frequency was set to 600 MHz and a rotational correlation time of 2 ns was assumed. The calculated NOE intensity of the H5-H6 cross peak in cytidine was 0.08.

particular case the calculations were performed for a correlation time τ_c = 2.0 ns, spectral frequency = 600 MHz and a mixing time τ_m = 0.2 s. It appeared that the results were not very sensitive for the value of τ_c around 2 ns. Plots like Fig. 3 can be used to provide a direct estimate of χ if P is known. To this end the experimental cross peak intensities are combined with the contour plot of Fig. 3 to obtain, if possible, an unique solution.

The peak intensities (volumes of NOE cross peaks) were calculated using the PEAKFIND algorithm developed in our laboratory by T.J.L. Bosman and J.P.M. van Duynhoven. In this procedure the successive rows and columns of the 2D-data matrix are smoothed and then differentiated twice. With this method maxima in rows and columns can be identified easily. Then by means of cluster analysis in the f_1 and f_2 dimension the maxima of the peak positions are determined. Peak-volumes are determined by means of a curve fitting procedure applied to each peak. These are scaled with the same factor which is needed to scale the H5-H6 NOE of the cytidine residues to the calculated H5-H6 NOE intensity, using the NOESIM program. In this procedure an absolute experimental error for the NOE cross peak intensity of 0.01 was assumed. In practice this leads to an accurate estimation of χ. We note that according to our experience the four experimental base-sugar NOE intensities form a consistent set. This means that these intensities yield, within experimental error, conformations which fit into the conformational space available for a mononucleotide.

Stage 3: Estimation of inter-residue distances

Inter-residue distances were determined by means of a computer program NOEFIT which makes use of the relaxation matrix approach introduced in the preceding section. This program uses the conformational parameters, P, ϕ_m, and χ, determined in the stages 1 and 2 of the structure determination as input for the calculation of the intra-residue proton distances. Furthermore a list of the measured NOE cross peak intensities together with the spectrometer frequency and the rotational correlation time is entered into NOEFIT. The calculations are started by setting all inter-residue distances to the irrealistic value of 20 Å. Then NOE cross peak intensities are simulated and compared with the experimental values. Subsequently, the simulated NOEs are fitted to the experimental values by changing the inter proton distances by means of a steepest descent procedure. The strongest NOE is fitted first, then the strongest but one NOE is considered, etc. In each step of this procedure the distances determined in the preceding steps were included in the fitting process. This process was continued until all NOEs could be fitted to 0.001 of their intensity. The total procedure was then repeated but now with the inter-residue distances, derived in the preceding step, as input parameters. This results in a set of distances which are within 0.1 Å from the input. The calculations were executed for experiments conducted at several different mixing times (75 ms- 500 ms) to make sure that for all of these mixing times the same distances are obtained. Distances less than 3 Å were reproducible within ± 0.2 Å.

Stage 4: Multi-conformational analysis

The distances estimated in the preceding section are used as constraints in a multi-conformational analysis in which the torsion angles of the phosphate backbone are changed systematically. Fortunately, for some torsion angles the available conformational space may be restricted to quite narrow domains if sufficient homo- and hetero- J-coupling information is available by which the torsion angles can be defined. Of course, this will lead to a more efficient analysis and, in the end, to better defined structures.

The multi-conformational analysis was carried out with the computer program HELIX (developed by F.J.M. van de Ven) which will be described in detail elsewhere. First, dinucleotides are analyzed; the conformational space of a dinucleotide is, provided bond distances and bond angles are constant, in principle 11-dimensional. Thus a proper sampling of the conformational space requires systematic variation of the 11 parameters, i.e., P, ϕ_m and χ of the 5'- and 3'-nucleotide and the backbone torsional angles ε, ζ, α, β and γ. Because we have already determined the conformation of the mono-nucleotide units, the conformational search can be restricted, in an ideal case, to five dimensions: namely to the variation of the backbone angles ε, ζ, α, β and γ. This leads to a large number of conformations for the dinucleotide units in the molecule, the total number depending on the number of steps in which the torsion angles are varied. Each conformation was tested against the available distance constraints and rejected if these were violated. All other distances in the conformation are confined to the interval from 1.8 to 30.0 Å, otherwise the conformation was rejected as well. Torsion angles of the structures which satisfy the distance constraints are stored. In the next step the procedure is extended from the dinucleotide analysis to the investigation of trinucleotide units. This is only possible when distance constraints are available between the first and the third nucleotide. Further extension of the analysis to tetra nucleotides and beyond can be carried out in principle, but again long range distance constraints are required in order to keep the number of generated conformations within reasonable bound, i.e., to keep the conformational space occupied by the generated conformations within such limits that an actual structure determination is performed.

Restrained molecular dynamics

The results obtained by means of the multi-conformational analysis were reproduced by means of a restrained molecular dynamics calculation [15] using the program QUANTA with the CHARMM force field [16]. A suitable loop structure, docked on a B-type stem, was used as starting structure. This structure was energy minimized and subjected to a restrained molecular dynamics calculation at 400 K. At these conditions the influence of the non-bonded interactions is negligible. The distance constraints (upper and lower bound distances) were introduced by means of pseudo square well potentials. The potential for the distances between upper and lower bounds is set to zero. In addition, the torsion angles β, γ, δ and ε were constrained at the standard gauche $^{+}$, trans or gauche^{-} rotamers as deduced from the NMR experiments. The procedure was as follows: the heating of the system from 0 to 400 K took place in 1 ps. Then the system was equilibrated at 400 K during 1 ps. Subsequently the dynamics run was performed during 10 ps at this temperature. The time step of the dynamic calculations was 1 fs. 100 structures were sampled, i.e., after each 0.2 ps.

Restrained and free molecular mechanics calculations.

The structure of the hairpin, which is obtained by means of the multi-conformational analysis was further examined by means of a restrained molecular mechanics calculation using the program DISCOVER [17] and its standard force field. Sodium counter ions were added to the negatively charged phosphates. The hairpin structure, obtained as described above, was placed in a shell of water molecules of 5 Å. The system thus comprised the hairpin, 15 sodium ions 425 water molecules. The cut-off radius for the non-bonded interactions was 8.5 Å. These were switched using a cubic switching function between 8.5 and 10 Å. The long range cut-off was 10 Å.

RESULTS

Spectrum interpretation

Part of the spectrum of the hairpin formed by d(ATCCTA-TTTA-TAGGAT) has already been interpreted in an earlier study [1] (Chapter 6). This concerns the imino proton spectrum, the H1'-sugar proton spectrum and the H8/H6 ring proton spectrum. The results obtained for the non-exchangeable protons were based on the sequential analysis of a NOESY spectrum, recorded with a mixing time of 300 ms at 298 K. In the present study, the experiments were extended to other mixing times, and temperatures. Despite these experimental differences, the spectra hardly differ from those obtained earlier and consequently the sequential analysis of the H6/H8-H1' cross peaks leads to results which are essentially the same to those mentioned above. This means that a NOE-walk can be made from the H8- resonance of residue A1 to the H1'-

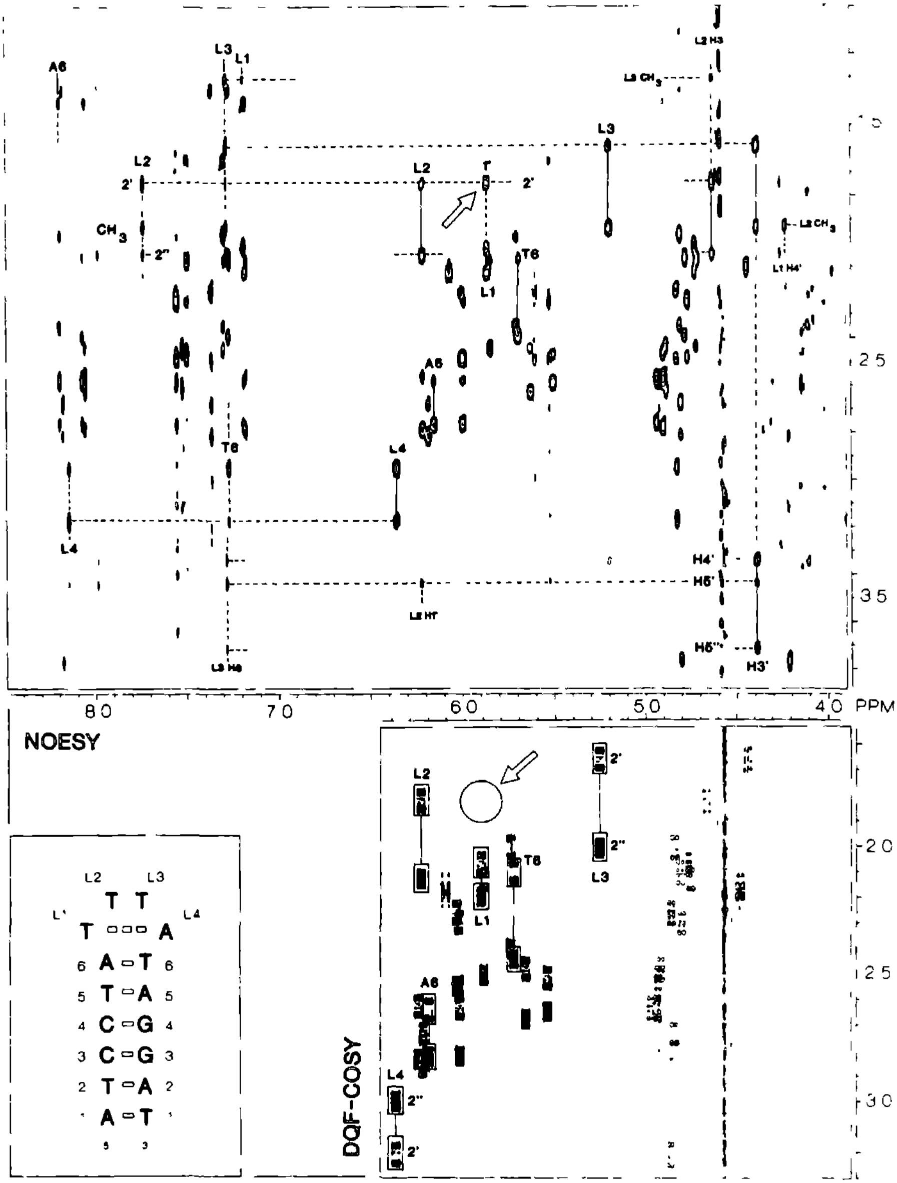


Fig. 4. Part of the 600 MHz NOESY spectrum of d(ATCCTA-TTTA-TAGGAT) (top) and the corresponding region of the 400 MHz DQF-COSY spectrum (bottom). The cross peaks which are crucial in the interpretation of this part of the spectrum are indicated (see text). The unusual NOE between the L1H1' and L2H2' resonances is indicated by an arrow. It is confirmed that at the corresponding place in the DQF-COSY spectrum no cross peak is present (arrow). The numbering of the resonances of the hairpin, which is used in this chapter, is indicated in the left bottom corner.

resonance of residue L2 and from the H6 resonance of residue T1 to the H8 resonance of residue L4. The residues in the stem are numbered 1 to 6 as indicated in Fig. 4 and the four residues which form the loop are labeled L1, L2, L3 and L4, respectively. Once these results are available the assignment procedure can be easily extended to the H1'-H2'/H2" and H3'-H2'/H2" cross peaks by combining suitable NOESY and DQF-COSY spectra. As followed already from our earlier J-coupling analysis of the $J_{1'2'}$- and $J_{1'2''}$ coupling constants [1] (Chapter 6), the sugar entities occur predominantly in a S-type conformation. This means that the $J_{2''3'}$-coupling constant is small and as a result the cross peaks in the DQF-COSY spectrum (Fig. 4) between the corresponding resonances are absent or very low in intensity. Moreover, the fine-structure of the H1'-H2' and H1'-H2" cross peaks differs significantly [18]. This can be inspected in Fig. 4, where these cross peaks for the residues L3 and L4 have been indicated. Note that the H2' resonance of L4 resonates downfield from the H2"-resonance, which is quite unusual. On the other hand, in the NOESY spectrum the H1'-H2" cross peaks then are more intense than the H1'-H2' cross peaks. Thus the individual H2' and H2" resonances can be easily identified.

It is well known that a sequential analysis can also be conducted using H8/H6-H2'/H2" connectivities. Indeed, via these cross peaks a NOE walk can be performed from residue A1 into the loop including residue L1. In contrast to the aforementioned H8/H6-H1'-walk no H6-H2'/H2" connectivities are observed between residues L1 and L2, while a weak cross peak is present between the L3H6- and L2H2'-resonance (cf. Fig. 4). It is possible that a cross peak is present between the L3H6- and L2H2"-resonance, but we cannot be entirely sure because of partial overlap with the T6H6-T6H2'-cross peak. No connectivities are found between the H8 and H2'/H2" resonances of residue L4 and L3, respectively. The absence of these cross peaks is also obvious at longer mixing times (e.g., at 500 ms). The normal sequential analysis can be continued, in the 3'-direction, from L4 to the terminal residue T1. The results of these sequential walks are summarized in Fig. 5; steps about which we are somewhat less certain because of overlap are indicated.

Also the sequential connectivities between the CH_3- and H5- resonances on one hand and H8- and H6-resonances on the other hand, which were obtained in a similar manner, are incorporated in Fig. 5; these endorse the interpretation obtained by means of the other NOE walks. Some of the assignments

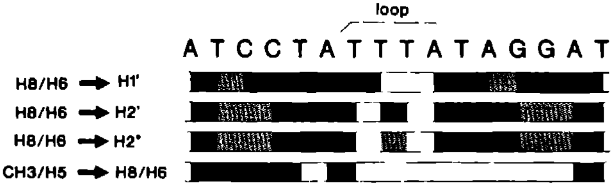


Fig. 5. Connectivity diagram in which the appearance of NOE cross peaks between neighboring residues is indicated. The dark region correspond to unambiguous cross peaks. The dashed regions correspond to ambiguities for the reason of peak overlap. The blank regions correspond to the absence of cross peaks. Note the absence of cross peaks between the third and fourth loop residue. The connectivities between the first and second residue also suggest a deviation from the B-type conformation at this place in the loop section.

which are crucial in the discussion of the loop structure are indicated in Fig. 4 and will be discussed below.

We now return to the identification of the sugar proton resonances. Starting from the assigned H3'-resonances, the DQF-COSY and NOESY spectra were used to assign the H4'-resonances. Then in the NOESY spectrum cross peaks from H3'- to H4'-, H5'- and H5"-resonances were incorporated in the analysis in order to extend the assignment to the H5'- and H5"-signals. If available in the DQF-COSY spectrum H4'-H5'/H5" cross peaks were used to confirm the assignments (vide infra). Because of overlap not all of the H4', H5'- and H5"-resonances could be identified, but the most important ones, i.e., those from the loop region, were accessible to analysis (see Fig. 6). The results are listed in Table I. It is noted that we were able to make stereo specific assignments for the H5' and the H5" spins. The ability to do this is intimately connected with the determination of the value of the torsion angle γ, and will be discussed in one of the following sections. Here we make use of those results and just summarize the assignment procedure. It is established that if γ adopts a gauche$^+$ conformation the H5' and H5" can be assigned by making use of the fact that the NOE between the H3' and H5" is strong and the NOE between the H3' and H5' will be weak or even absent (cf. the corresponding distances in Fig. 7). In case it has been concluded that γ is in the trans domain the discrimination between H5' and H5" can be based on the difference of the $J_{4'5'}$ and $J_{4'5''}$ coupling constants [19] (Fig. 7). Then, in the DQF-COSY spectrum the corresponding cross peaks

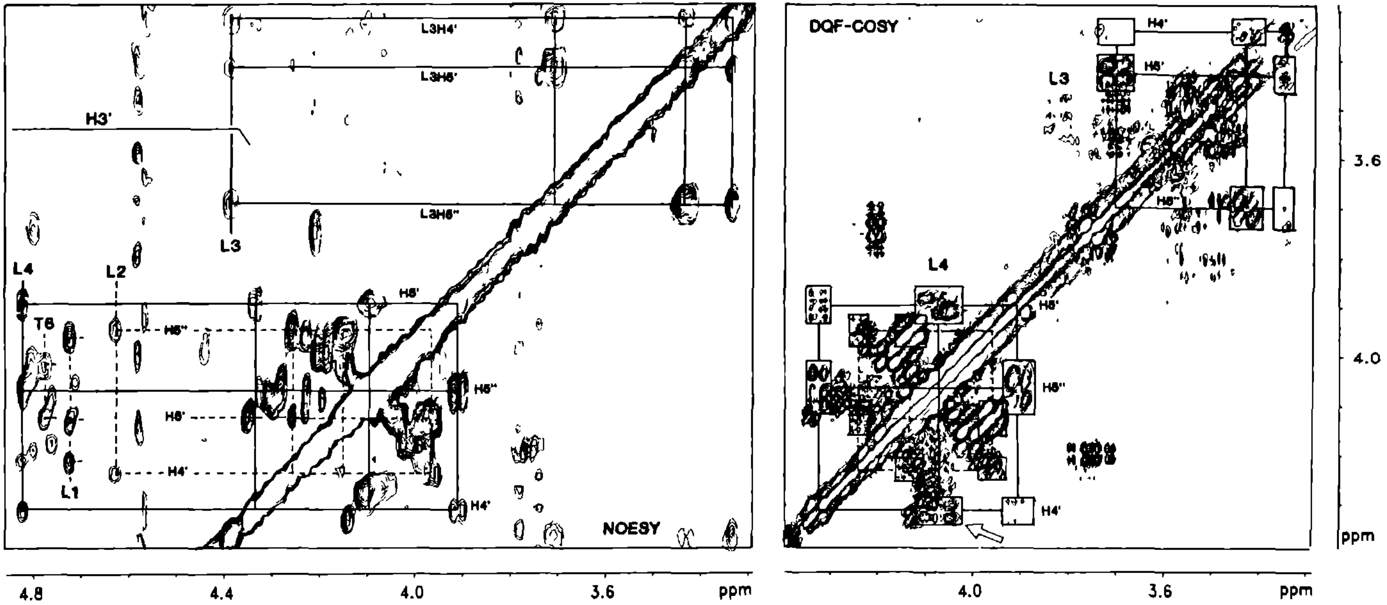


Fig. 6. Part of the 600 MHz NOESY spectrum (left) of d(ATCCTA-TTTA-TAGGAT) in which the sugar phosphate backbone protons are found. Some of the cross peaks which correspond to intra nucleotide NOEs of protons in the loop region have been indicated. The corresponding part of the 400 MHz DQF-COSY spectrum is shown (right). The NOE cross peak intensities and coupling patterns appear to be crucial in the assignment and interpretation (see text). The cross peak pattern of the L4H4'-H5" cross peak (arrow) corresponds to a γ^t-β^t conformation of the backbone.

Table I. Chemical shifts of the non-exchangeable proton resonances of d(ATCCTA-TTTA-TAGGAT) at 310 K.

Res	H6 H8	H5 CH_3 H2	H1'	H2'	H2"	H3'	H4'	H5'	H5"
A1	8.20		6.18	2.67	2.80	4.82			
T2	7.38	1.33	6.01	2.19	2.47	4.84			
C3	7.54	5.60	6.01	2.23	2.47	4.78	4.20		4.11
C4	7.51	5.52	5.86	2.43	2.43	4.73			
T5	7.31	1.63	5.72	1.96	2.36	4.82			
A6	8.21		6.17	2.57	2.76	4.91	4.28		4.08
L1	7.20	1.39	5.88	2.00	2.11	4.74	4.23	4.15	3.98
L2	7.75	1.92	6.23	1.73	2.04	4.64	4.26	4.14	3.96
L3	7.30	1.29	5.21	1.57	1.92	4.40	3.33	3.43	3.71
L4	8.16	8.00	6.36	3.16	2.94	4.84	4.33	3.91	4.08
T6	7.29	1.33	5.70	2.05	2.38	4.79	4.13		4.03
A5	8.09		6.01	2.57	2.75	4.95	4.28		4.08
G4	7.57	--	5.51	2.45	2.58	4.91			
G3	7.57	--	5.64	2.43	2.62	4.89			
A2	8.07		6.23	2.55	2.78	4.91			
T1	7.19	1.39	6.08	2.10	2.10	4.46	4.00		

are expected to be weak and strong respectively. When γ is in the gauche$^-$ region the situation is just reversed; $J_{4'5'}$ is large and $J_{4'5''}$ is small leading to corresponding changes in the DQF-COSY spectrum. In this case the conclusion may be endorsed by considering the NOE intensity of the H3'-H5' and H3'-H5" cross peaks. The first is expected to have a relatively large intensity, while the latter is weak or absent. The assignment of these resonances of residue L3 is indicated in Fig. 6. The NOE cross peaks with L3H3' are used to identify the H4'-, H5'- and H5"-resonances. The strong H5'-H5" cross peak in the NOESY as well as in the DQF-COSY spectrum establishes that the H4' resonates upfield from the H5'/H5" resonances. Because L3-γ adopts a gauche$^+$ conformation (note the weak H4'-H5' and H4'-H5" cross peaks in the DQF-COSY spectrum), the H5'-H3' NOE must be weaker than the H5"-H3' NOE. This allowed us to assign L3H5' and L3H5". The same approach has been used for the assignment of the resonances of the other sugar-backbone protons in the loop region. The assignment of these resonances of L4 is also indicated in Fig. 6. In this case we find a strong H4'-H5'

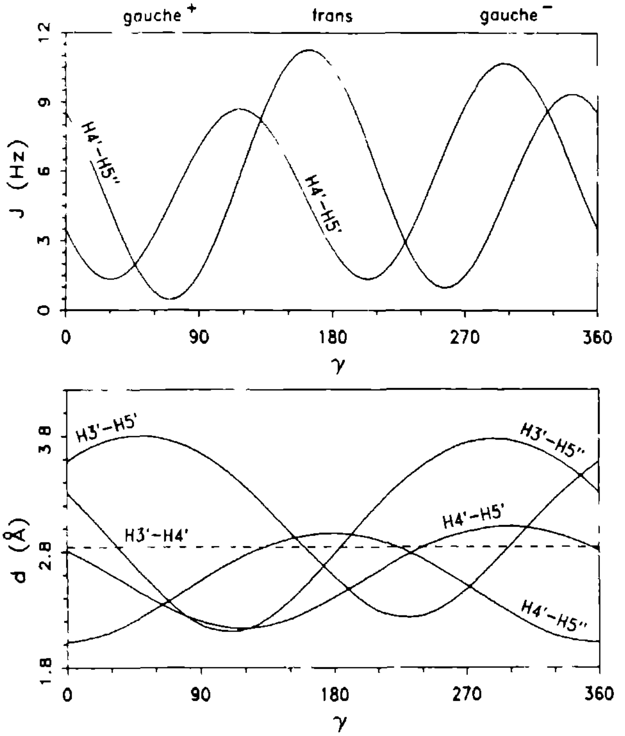


Fig. 7. ^{1}H-^{1}H coupling constants for $J_{4'5'}$ and $J_{4'5''}$ calculated as a function of the torsion angle γ [53], together with ^{1}H-^{1}H distances of the proton pairs H3'-H4', H3'-H5", H3'-H4', H4'-H5' and H4'-H5". These couplings and distances illustrate that a stereospecific assignment of the H5' and H5" resonances can be achieved when the cross peaks of NOESY and COSY type spectra are evaluated.

cross peak in the DQF-COSY spectrum, which is indicative for γ^{t}. This is corroborated by the observation of H3'-H5' and H3'-H5" NOEs of equal intensity.

The ^{31}P-spectrum of the title compound has been investigated by means of a ^{31}P-^{1}H-HMQC experiment (Fig. 8A). The bulk of resonances of the phosphorus atoms of the stem could not be identified specifically (they are clustered in the region ω_1 = -0.9 - 0.9 ppm and ω_2 = 4.0 - 4.5 ppm), but all P-resonances of the loop region, designated P6-P10 (see Fig. 8B) have been assigned, using the H3'(n)-P(n) and H5'/H5"(n)-P(n-1) cross peaks, as has been indicated. The resonance of P8 is shifted somewhat upfield relative to the others. We mention in passing that P9 exhibits a cross peak to L3H2'; we will return to this observation shortly.

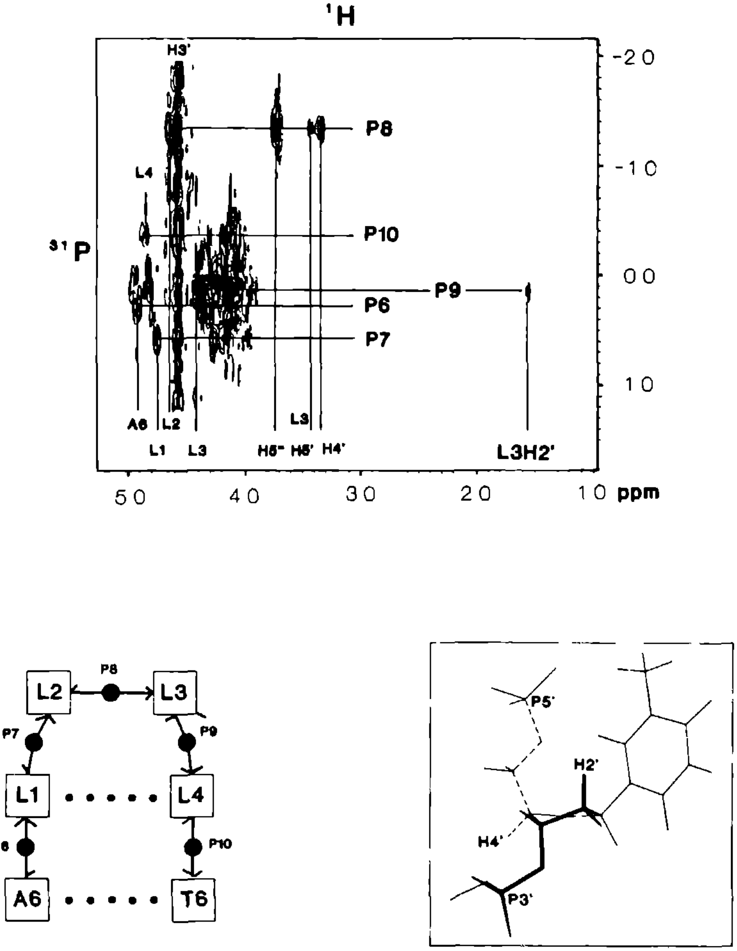


Fig. 8. ^{31}P-^{1}H HMQC spectrum of d(ATCCTA-TTTA-TAGGAT). The assignment of the phosphorus resonances of the loop (P6-P10) is indicated. The number of the phosphates is indicated in (B). The presence of a small cross peak between P9 and L3H2' is obvious (A). This illustrates that a long range $J_{H2'P}$ is present in this particular case. This coupling exists when four successive bonds in P3'-O3'-C3'-C2'-H2' form a flat W, as is depicted with heavy lines in the structure (C). This happens when the ε adopts a gauche^{-} conformation and when the sugar has a S-type pucker.

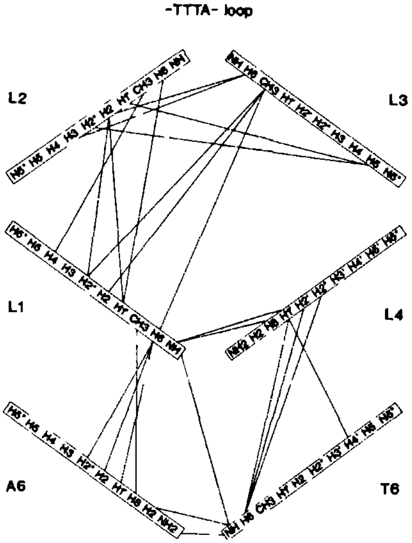


Fig. 9. Scheme of the loop region of the hairpin formed by d(ATCCTA-TTTA-TAGGAT). The characterization of the residues corresponds with that in Fig. 4. The connectivities observed between the protons of the loop section of the hairpin are indicated by lines connecting the protons in the different residues.

Examination of the chemical shifts presented in Table I leads to the observation of some interesting chemical shift values for the residues in the loop region. An upfield shift is observed for the methyl resonance of L2, which is not totally unexpected for a loop methyl resonance. The H1', H4' and H5' resonances of residue L3 show strong upfield shifts. Note that we assigned H4' of L3 upfield from its H5'/H5" resonances. Remarkable downfield shifts are found for L2H6 and the H2' and H2" resonances of residue L4.

Structure determination

After completion of the assignments we now examine which structural details can be derived from the spectra utilizing both NOE and J-coupling information.

The sugar conformation

We first consider the H1'-resonances. These are already well resolved in the 1D-spectrum, and the J1'2' and J1'2" coupling constants can be derived fairly accurately through simulation of the H1'-spectrum. This does not apply for the other sugar resonances, due to severe spectral overlap. Necessary additional J-coupling information was obtained from the fine structure of the cross peaks present in the DQF-COSY spectrum, either by simulation of the cross peaks or by taking cross sections through the cross peaks in the DQF-COSY spectrum along the ω_2-direction and directly reading of the J-coupling constants or sums of J-coupling constants. The accuracy of the latter method was checked by means of simulation of these cross sections. Although the coupling constants were obtained after correction for resonance overlap, they are generally somewhat less accurate than the data obtained from 1D-spectra due to the limited digital resolution. The results are listed in Table II.

By means of the multi-conformational analysis program MARC (see Materials and Methods), the J-coupling information was translated into the pseudorotation parameters, P and ϕ_m, and the fraction of the S-puckered sugar (cf. Table II). It follows from these calculations that the sugar rings occur almost exclusively in a S-type conformation as was already concluded from the preliminary analysis (Chapter 6). Moreover, it appears that for certain sugars, the set of J-couplings match a narrow domain of S-conformations; this applies for residues A1, T5, L3, G4 and A2. It is particularly interesting that, for the loop residue L3, P and ϕ_m are well-defined and the sugar occurs entirely as a S-pucker. The phase angle is somewhat larger than normally found in B-DNA possibly a result of the tight loop structure or because ε has an abnormal orientation (see later). Due to a lack of data, the sugar conformation of residues T1 and C4 is ill-defined, which is reflected in the range of values which can be adopted by P, ϕ_m, and x_S. In fact P and ϕ_m of these two sugars correspond with the domains used as input for the calculation (Table II).

Table II. J-coupling information obtained for the sugar rings and conformational parameters obtained for the mononucleotide units of the title compound. The $J_{1'2'}$ and $J_{1'2''}$ represent the coupling constants between the H1' and H2' and H1' and H2" respectively. The sums of the coupling constants: $\Sigma 2' = J_{1'2'} + J_{2'3'} + |J_{2'2''}|$ and $\Sigma 2'' = J_{1'2''} + J_{2''3'} + |J_{2'2''}|$. The values obtained for the pseudorotational parameters, P and ϕ_m, of the S-puckered sugar and the fraction of the S-type sugar were obtained by the MARC analysis. P(S) was sampled between 117° and 198° in steps of 4.5°. ϕ_m(S) was sampled between 32 and 44 in steps of 3. The molar fraction x(S) was sampled between 0.0 and 1.0 in steps of 0.05. The numbers in the column err indicate the constraints: (1) all constraints +/- 0.5 Hz; (2) as (1) but with constraints of the sum +/- 1.0 Hz; (3) as (2) but with constraints $J_{1'2'}$ and $J_{1'2''}$ +/- 0.75; (4) only $J_{1'2'} + J_{1'2''} = 14.0$ is used as constraint. The values for the glycosidic torsion angle χ were estimated from the combined use of NOE intensities and Fig. 3 (see text).

Res	$J_{1'2'}$	$J_{1'2''}$	Σ2'	Σ2"	err	P(S)	ϕ_m(S)	x(S)	χ
A1	8.7	5.3	27.4	21.2	1	189-198	41-44	0.95-1.0	135-215
T2	9.1	5.6	29.6	20.3	2	166-189	32-35	1.0	242-248
C3	9.3	5.0	29.6	20.3	2	130-180	38-41	0.9-1.0	160-300
C4	8.1	5.6	-	22.8	1	117-198	35-44	0.75-0.9	160-300
T5	8.4	5.3	29.3	20.7	3	180-189	35	0.95-1.0	228-242
A6	9.9	5.6	28.0	20.3	3	166-193	35-44	0.95-1.0	220-225
L1	9.2	5.2	28.6	21.5	1	139-189	38-44	0.80-1.0	225-232
L2	10.2	5.4	28.0	21.5	1	166-189	41-44	0.9-1.0	212-221
L3	8.4	5.6	28.0	20.3	2	198	41	1.0	165-210
L4	9.3	5.3	28.0	20.3	2	162-198	38-44	0.9-1.0	130-160
T6	9.5	5.6	29.8	20.9	1	139-175	32-41	0.9-1.0	239-241
A5	9.5	5.6	28.5	21.1	1	157-193	35-41	0.9-1.0	210-220
G4	9.8	5.3	27.9	19.6	2	171	41	1.0	150-320
G3	9.0	5.3	27.9	20.3	1	180-198	35-44	1.0	160-300
A2	8.7	5.3	29.1	21.5	1	135-144	38-44	0.8-0.85	180-218
T1	-	-	-	-	4	117-198	32-44	0.75-1.0	220-230

The glycosidic torsion angle

This torsion angle was determined by combining the results obtained for the phase angle, P, and the NOE intensities measured for the intra residue connectivities between the H8/H6- and the H1'-, H2'-, H2"- and H3'-resonances. The experimental cross peak intensities were calculated with the aid of the program PEAKFIND (cf. Materials and Methods) from the 600 MHz NOESY spectra; those acquired for a mixing time of 200 ms are collected in Table III. These intensities can be directly used to read of the glycosidic torsion angle in the different P-χ plots given in Fig. 3. The results are presented in Table II. Examination of these results shows that some of the torsion angles are well constrained, i.e., those of T2, T5, A6, L1, L2, T6, A5 and T1, while others are not. In particular, the glycosidic torsion angle of the C and G residues and the 5'-terminal residue are not well defined. This latter result is caused by the fact that some of the cross peak intensities could not be determined because of peak overlap. Apart from the C and G residues it can be concluded that the values of χ are in the anti-range, normally found for B-DNA with the exception of those of the loop-residues L3 and L4. The adenine of L4 is in a high-syn conformation which permits Hoogsteen base pairing with the thymine of the first loop residue. The glycosidic torsion angle of thymine of L3 differs a little from B-DNA. This orientation of the base may be a result of a local optimization of base-base stacking interactions.

The backbone torsion angles:

The torsion angle γ (O5'-C5'-C4'-C3')

It has been indicated (vide supra) that for the loop region of the hairpin the spectral interpretation can be extended to the 4', 5' and 5" sugar proton resonances, which involves stereo-specific assignment of the 5'- and 5"-signals. The latter result could be obtained in combination with the determination of the torsion angle γ. We will now elaborate on this feature.

In A- and B-type helical structures γ is found in the gauche$^+$ domain. In NOESY spectra deviations from this situation can be readily picked up by considering the cross peaks between the H6/H8 and the H5'/H5"-resonances. If γ is in the gauche$^+$-region no NOE cross peak will be observed between these proton pairs for sufficiently short mixing times (≈75 ms) for the size of the molecule considered here. On the other hand, for γ in the trans or gauche$^-$ domain, the distances between H6/H8 and H5' or H5" become rather short and the NOE cross peaks will be correspondingly more intense and observable in NOESY spectra. For a further characterization of this torsion angle we consider Fig. 7, where the values of the J-couplings, $J_{H4'H5'}$ and $J_{H4'H5''}$ [19] and the distances between the proton pairs H3'-H5', H3'-H5", H4'-H5', H4'-H5" and H3'-H4' have been plotted as a function of γ. We focus on the three staggered conformations and for the first one, gauche$^+$, we note

A6	H8-H1'	1127
	H8-H2'	7618
	H8-H2"	1038
	H8-H3'	500
	H1'-H2'	1960
	H1'-H2"	6061
	H1'-H4'	965
	H2"-H3'	7372

L1	H6-H1'	473
	H6-H2'	989
	H6-H3'	919
	H6-H5"	839
	H1'-H2'	4484
	H1'-H2"	10906
	H1'-H4'	1078
	H2'-H3'	5981
	H2"-H3'	14265
	H3'-H4'	2674
	H3'-H5'	3790
	H3'-H5"	5302
	H2"-H5"	1242

L2	H6-H1'	637
	H6-H2'	6380
	H6-H2"	1319
	H6-CH_3	8684
	H6-H5'	478
	H1'-H2'	3394
	H1'-H2"	9903
	H1'-H4'	500
	H2'-H3'	6306
	H2"-H3'	3255
	H3'-H4'	1586
	H3'-H5"	2480

L3	H6-H1'	780
	H6-H2'	2220
	H6-CH_3	8425
	H6-H3'	886
	H6-H4'	507
	H6-H5'	1413
	H6-H5"	649
	H1'-H2'	5026
	H1'-H2"	10767
	H1'-H4'	1461
	H2'-H3'	8788
	H2"-H3'	4012
	H2'-H2"	21102
	H3'-H4'	5043
	H3'-H5'	1451
	H3'-H5"	4377
	H4'-H5'	5383
	H4'-H5"	7518

T6	H6-H1'	604
	H6-H2'	12006
	H6-H2"	4018
	H6-H3'	705
	H1'-H2'	5028
	H1'-H4'	774
	H2'-H3'	7214
	H2"-H3'	4133
	H3'-H4'	8697
	H3'-H5"	1683

L4	H8-H1'	2820
	H8-H2'	4327
	H8-H2"	1451
	H1'-H2'	2130
	H1'-H2"	12487
	H1'-H4'	500
	H2'-H3'	7558
	H2"-H3'	5109
	H2'-H2"	21610
	H3'-H4'	2371
	H3'-H5'	7184
	H3'-H5"	5082
	H4'-H5'	3800
	H5'-H5"	19102

Table III. The experimental NOE cross peak volumes in arbitrary units of the 600 MHz NOESY spectrum, obtained by means of curve fitting using the program PEAKFIND. The mixing time was 200 ms. The volume of the H5-H6 cross peaks was 10064.

A6-L1	H1'-H6	500
	H8-CH_3	1803
L1-L2	H1'-H6	2506
	H1'-H2'	5951
	H2"-H2'	1227
	H4'-CH_3	3979
L1-L3	H2"-CH_3	500
	H2'-CH_3	500
	H6-CH_3	1006
L2-L3	H2'-H6	653
	H3'-H6	842
	H1'-H5'	710
	H2"-H5'	2022
L4-T6	H1'-H6	1215
	H2'-H6	865
	H2"-H6	3462
	H1'-H4'	819

that only relatively weak NOE cross peaks are expected between H3'- and H5'-resonances. On the other hand the distances between H4' and H5'/H5" are rather short (≈2.4 Å) so that intense cross peaks are expected for these proton pairs. Furthermore, it can be gleaned from Fig. 7 that the J-coupling constants $J_{4'5'}$ and $J_{4'5''}$ will be small so that in a DQF-COSY spectrum the cross peaks between the corresponding resonances will be of low intensity or even absent. By combining this set of parameters γ can be assigned unambiguously to the gauche$^+$ domain.

When γ adopts a trans-conformation strong NOEs of comparable intensity are expected between the H3' and H5'/H5" proton pairs. This is typical of this conformation and does not occur for either the gauche$^+$- or the gauche$^-$-conformation. In case γ occurs in the gauche$^-$ domain, the intensities of the H3'-H5'/H5" NOE cross peaks will differ substantially. To be able to distinguish this situation from γ being in the gauche$^+$-region, for which also this difference in intensity occurs (vide supra), it is noted that if γ falls in the gauche$^-$ domain, the $J_{4'5''}$ is large giving rise to an intense cross peak in the DQF-COSY spectrum. Using these spectral features γ can be characterized unambiguously. If, in addition, the J-couplings $J_{4'5'}$ and

$J_{4'5''}$ can be determined quantitatively, more precise estimates of γ can be obtained. To derive more accurate values of γ, however, a three-state equilibrium between the states γ^+, γ^t and γ^- should be assumed in a multiconformational analysis. Thus, very accurate couplings are indispensable to achieve such.

Analysis of the NOESY and DQF-COSY spectra of d(ATCCTA-TTTA-TAGGAT), along the lines indicated above shows that for most of its residues, γ adopts the gauche$^+$-conformation. Only for the ring protons A1H8 and L3H6 NOE connectivities to H5'/H5" resonances are detected, suggesting that in these cases γ may adopt other conformations. Since the rotation around the C4'-C5' bond of the 5'-terminal sugar is virtually unhindered, it is not surprising that in this case γ, in addition to the gauche$^+$-conformation adopts other conformations which lead to substantial NOEs between these spin pairs. For residue L3 the possible deviation from the gauche$^+$-conformation, as suggested by the aforementioned NOE cross peaks, is not confirmed by strong cross peaks between H4'-H5' or H4'-H5" resonances in the DQF COSY spectrum. This suggests that γ remains in the gauche$^+$ domain. Indeed, as will be discussed below, the cross peaks between L3H6 and the C5' proton resonances arise from the peculiar structure of the loop region: the NOEs can be explained assuming the spin diffusion path L3H5' - L2H1'/H2'/H2" - L3H6 (cf. Fig. 9). It has been mentioned before that the base of residue L4 is in a syn-conformation and therefore no NOE cross peaks between the H8 and H5'/H5" are to be expected. However, equally strong NOEs are observed between the H3' and H5'/H5" resonances. Moreover, in the DQF-COSY spectrum a strong cross peak is seen between the H4' and H5" resonances with a fine structure dominated by a large active coupling constant (≈10 Hz; cf. Fig. 6). This shows that the torsion angle γ of the last residue in the loop region adopts a trans conformation. The results are summarized in Table IV.

The torsion angle β (P-O5'-C5'-C4')

This torsion angle can be monitored by means of the hetero nuclear coupling constants $J_{H5'P}$ and $J_{H5''P}$ (cf. Chapter 1, Fig. 5). In B-type DNA β falls in the trans domain and the proton phosphorus coupling constants are small, i.e., 1-2 Hz [20]. In case β adopts a gauche$^+$- or gauche$^-$-conformation these coupling constants may increase to about 23 Hz (cf. Chapter 1, Fig 5). They will be manifest in the ^{1}H-DQF COSY spectrum as large passive couplings. For the H5'- and H5"-cross peaks of d(ATCCTA-TTTA-TAGGAT) which are sufficiently well resolved (and this is the case for those of the loop region, which interest us most) such large couplings are not observed. This indicates that β adopts a trans conformation, at least for the residues in the loop region.

The torsion angle ε (C4'-C3'-O3'-P)

The torsion angle ε is only found to occur in two staggered conformations, namely gauche$^-$ and trans; the gauche$^+$ conformation is 'forbidden' [20]. Although the $J_{H3'P}$ coupling constant might in principle serve to monitor this torsion angle, it is of limited practical value because for both conformers (ε^- and ε^t) it has roughly the same value. Therefore, it cannot be used to distinguish between these two possibilities. Long-range coupling constants may be observed when four bonds in succession lie in the same plane and form a W-shaped conformation [5]. We note in passing that this is the case for the atomic sequence H4'-C4'-C5'-O5'-P in B-DNA, where β and γ are in the trans and gauche$^+$ domain, respectively (see dashed lines in Fig. 8C). Indeed, in the ^{31}P-^{1}H HMQC-spectrum a cross peak is observed between L3H4' and P8, confirming our interpretation in the preceding sections that β and γ of residue L3 are in the trans and gauche$^+$ domain, respectively. Due to overlap such conclusions cannot be draw for the other residues. In this

Table IV. Torsion angle domains deduced for the hairpin formed by d(ATCCTA-TTTA-TAGGAT). The torsion angles are expressed in terms of the classical rotamers gauche$^+$ (g$^+$), trans (t) and gauche$^-$ (g$^-$). The torsion angles in the first three columns were derived from J-couplings, the torsion angle χ via NOEs by application of the program NOESIM and the torsion angles α and ζ by means of multi-conformational analysis (HELIX).

	β	γ	ε	χ	α	ζ
B-DNA type stem:	t	g$^+$	t	anti	g$^-$	g$^-$
A6	t	g$^+$	t	anti	g$^-$	g$^-$
L1	t	g$^+$	t	anti	g$^-$	g$^-$
L2	t	g$^+$	t	anti	t	g$^-$
L3	t	g$^+$	g$^-$	anti	g$^-$	g$^+$
L4	t	t	t	high-syn	g$^+$	g$^-$
T6	t	g$^+$	t	anti	g$^-$	g$^-$

HMQC spectrum a cross peak is also observed between the H2' resonance of residue L3 and its P3'-resonance (Fig. 8A). If the sugar adopts a S-type conformation and the torsion angle ε is gauche$^-$ the four bonds between the H2' and P3' atoms are in one plane and form a W (see bold lines in Fig. 8C) which accounts for the observation of a long range coupling through the mentioned cross peak. We therefore conclude that ε adopts a gauche$^-$-conformation in this particular case. Since only one H2'-P3' cross peak is observed, all other torsion angles ε are assumed to be in the trans conformation.

The values so far obtained for the torsion angles are summarized in Table IV. In cases where we had no indications to the contrary, e.g., for the stem region, the angles were assigned standard B-DNA values.

Structure of the loop region

An essential part of this structure has been established already in our earlier studies [1] (Chapter 6). There it was shown that the first and the last residue in the loop, i.e., thymidine L1 and adenine L4, form a Hoogsteen base pair. We now attempt to determine the total loop structure. The NOE cross peaks that form the underlying basis for this determination are indicated in Fig. 4. We first discuss some qualitative aspects of the structure determination which follow from a consideration of the available Overhauser effects. The sequential NOE walks used in the assignment procedure (vide supra) in conjunction with the formation of the Hoogsteen base pair between L1 and L4 indicated already that the first and last base of the loop are stacked upon the terminal base pair of the stem region, i.e., upon base pair A6-T6. The connectivities are schematized in Fig. 9. While standard B-type connectivities are found for the A6-L1 step, the connectivity pattern of the L4-T6 step deviates somewhat: the L4H8-T6CH$_3$ cross peak is absent, but this can be explained easily with the observation that the adenine base is in a syn- instead of an anti-orientation. The inter-residue H6-H2' cross peak is rather weak and a L4H1'-T6H4' NOE was observed. The conclusion that the two base pairs are stacked is supported by the observation of a NOE effect in 1-dimensional NOE experiments between the imino proton resonances of T6 and L1.

The connectivities observed between the resonances of the loop residues L1 and L2 indicate that in this step the nucleotide chain is strongly

Table V. Upper and lower bounds of the distances between protons of the residues of the loop and of the residues of the base pair closing the loop, which were used as constraints in the multi-conformational analysis (HELIX), restrained molecular dynamics and restrained molecular mechanics calculation.

A6-L1	H1'-H6	3.0-5.0	L1-L2	H1'-H6	2.8-4.1
	H2'-H6	3.1-4.3		H1'-H2'	2.5-3.0
	H2"-H6	2.5-3.0		H2"-H2'	2.8-3.4
	H8-CH_3	3.2-5.4		H4'-CH_3	2.8-4.9
				H2'-H6	4.0-30.0
				H2"-H6	4.0-30.0
				H6-CH_3	4.0-30.0
L1-L3	H2'-CH_3	3.5-5.0	L2-L3	H2'-H6	3.7-5.0
	H2"-CH_3	3.5-5.0		H2"-H6	1.8-30.0
	H6-CH_3	3.5-5.0		H3'-H6	3.6-5.0
	H1'-CH_3	4.0-30.0		H1'-H5'	3.4-7.0
	H1'-H6	4.0-30.0		H2"-H5'	3.1-7.0
	H2'-H6	4.0-30.0		H6-CH_3	4.0-30.0
	H2"-H6	4.0-30.0		H1'-H6	4.0-30.0
L3-L4			L4-T6	H1'-H6	3.4-4.7
	H1'-H8	4.0-30.0		H2'-H6	3.5-7.0
	H2'-H8	4.0-30.0		H2"-H6	2.7-3.2
	H3'-H8	4.0-30.0		H1'-H4'	3.6-5.0
				H1'-CH_3	4.0-30.0

deviating from the B-type conformation. Although an (unusually) strong Overhauser effect is observed between the L1H1' resonance and its 3'-neighbor L2H6 resonance, the cross peaks between the H2'/H2" resonances and of L1 and the ring proton resonance of L2 were absent; these were not even observed when the NOESY was recorded with a mixing time of 500 ms. Instead, a cross peak is observed between the L2H2' and L1H1' resonances. It is noted in passing, that we have not confused this cross peak (indicated with an arrow in Fig. 4) with an intra-residue connectivity because the corresponding cross peak then expected in the DQF-COSY spectrum is not found. Additional NOEs which indicate that the L1-L2 step deviates from the normal B-type, are those between the L1H2" and L2H2' and between the L1H4' and the L2CH_3 resonance (cf. Fig. 4).

In this respect the connectivities between the spins of L2 and L3 are interesting as well. A weak NOE is observed between the base proton resonance H6 of L3 and the H2', (probably) H2" and H3' resonances of L2, while no NOEs are observed between this base proton and the L2H1' resonance. Furthermore, we find NOEs between the L3H5' resonance and the L2H2" and L2H1' resonances respectively. Again deviations from the B-type chain conformation are indicated. In addition to these connectivities between nearest neighbors we find NOEs between the next nearest neighbors L1 and L3, i.e., the resonances of $L3CH_3$ on the one hand and L1H6, L1H2' and L1H2" on the other hand are connected by NOE cross peaks. These 'long-range' distance constraints are, of course, a prerequisite in the determination of a detailed structure of the L1-L2-L3 portion of the loop. No cross peaks could be observed between the resonances of the third and the fourth residue of the loop, implying that the structure determination of this dinucleotide step can only follow from constraints in the backbone torsion angles (Table IV). Fortunately, long range distance constraints are found between L1 and L4, which forms the Hoogsteen base pair, i.e., cross peaks are observed between L3NH and L4H8 and L4H1'. The NOE contacts between the loop residues and the NOEs between the loop residues and the terminal base pair A6-T6 in the stem are schematized in Fig. 9.

For all measured and assigned NOE cross peaks of the six central nucleotides of d(ATCCTA-TTTA-TAGGAT) the intensities, i.e., the peak volumes, were determined by the PEAKFIND algorithm (Materials & Methods) and collected in Table III. These intensities were used to establish interproton distances between the loop residues. This was performed by means of the NOEFIT program (Materials & Methods) and resulted in a set of distances for which upper and lower bounds were introduced by adding 25 % and -10 % of the measured distances respectively (Table V).

These distance constraints together with torsion angle domains (Tables II and IV) and sugar pseudorotation parameters (Table II) served as input for the multi-conformational analysis carried out with the program HELIX (Materials & Methods). In the analysis the torsion angles were varied between their permitted bounds in steps of 15° to scan the defined conformational space; torsion angles for which these bounds were not available, e.g., α and ζ, were first varied between 0° and 360°. Structures generated in this way which were not or only partly violating the constraints were stored and examined. In the next steps the step size was reduced to 10° and the conformational space was examined in the regions which resulted in the 'best' structures during the first phase. Finally, the structures with violations were discarded. This procedure was followed for all dinucleotide steps in the loop region and resulted in 132 conformations for the first, the L1-L2 step and 220 conformations for the L2-L3 step. Because of the absence of NOE data between the residues L3 and L4 the structure of this dinucleotide was ill-defined at this stage of the analysis.

Subsequently, the conformations obtained for the L1-L2 and L2-L3 steps were combined into L1-L2-L3 conformers, which were tested against the L1-L3 distances. This resulted in a family of 327 trinucleotides that satisfy the L1-L3 constraints. To provide an impression of the limited conformational space occupied by these conformers twelve of these trinucleotides were superimposed and presented in Fig. 10. The RMS difference between these conformations was 0.5 Å. It can be seen that this part of the loop is well defined; the second base is bulged out and the first and third base overlap

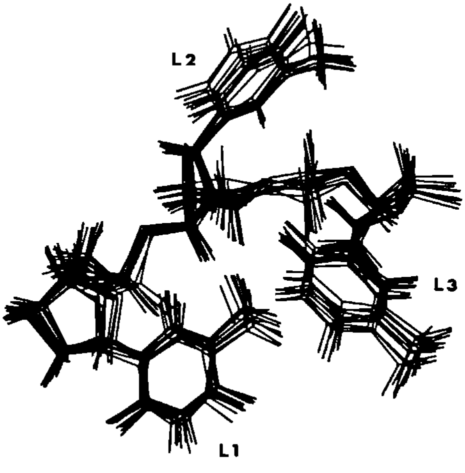


Fig. 10. Superposition of twelve representative structures found for the L1-L2-L3 section with the aid of multi-conformational analysis. The structural variation gives an impression of the structural variance. The root mean square deviation between the structures was 0.5 Å.

to some extent. The trinucleotide conformers were combined with several different sets of L3-L4 conformers to form the tetranucleotide L1-L2-L3-L4. The resulting conformations were tested for their ability to form a Hoogsteen base pair between L1 and L4. It followed that these constraints were only matched by those conformations for which the backbone angles α and ζ of L4 and L3 respectively fall in the gauche$^+$ domain. Subsequently, the complete multi-conformational analysis was repeated in such a way that the torsion angles of the phosphate backbone of the L1-L2-L3-L4 part were sampled in steps of 3° around the values given in Table VI. It appeared that only a minor part, i.e., 57 structures satisfy the constraints. The RMS difference between these conformations was 0.1 Å. This illustrates that the applied constraints define the structure of the loop. It is noted that in this last step we did not combine all possible L1-L2-L3 combinations with those for L3-L4. This is caused by practical reasons connected with cpu-time. The multi-conformational analysis of the dinucleotide steps A6-L1 and L4-T6 resulted in conformations, which correspond with the B-DNA domain. One of these conformations is included in Table VI.

The available NOEs and J-coupling constants appear to be consistent with a large number of different conformations in the double stranded region of the hairpin, i.e., the stem region. This confines a detailed conformational analysis to the loop region. The loop structure was docked on the stem region which had the standard B-helix conformation. The torsion angles of the loop structure are given in Table VI. Energy minimization of this

Table VI. Values for the torsion angles of the loop structure depicted in Fig. 11.

res	α	β	γ	δ	ε	ζ	χ	P	ϕ_m
A6	300	180	60	159	205	285	226	190	38
L1	290	195	55	133	180	250	229	140	41
L2	190	180	50	159	240	240	217	178	43
L3	310	200	70	164	295	55	210	198	41
L4	100	170	185	158	180	265	130	180	41
T6	310	190	60	137	180	300	240	150	38

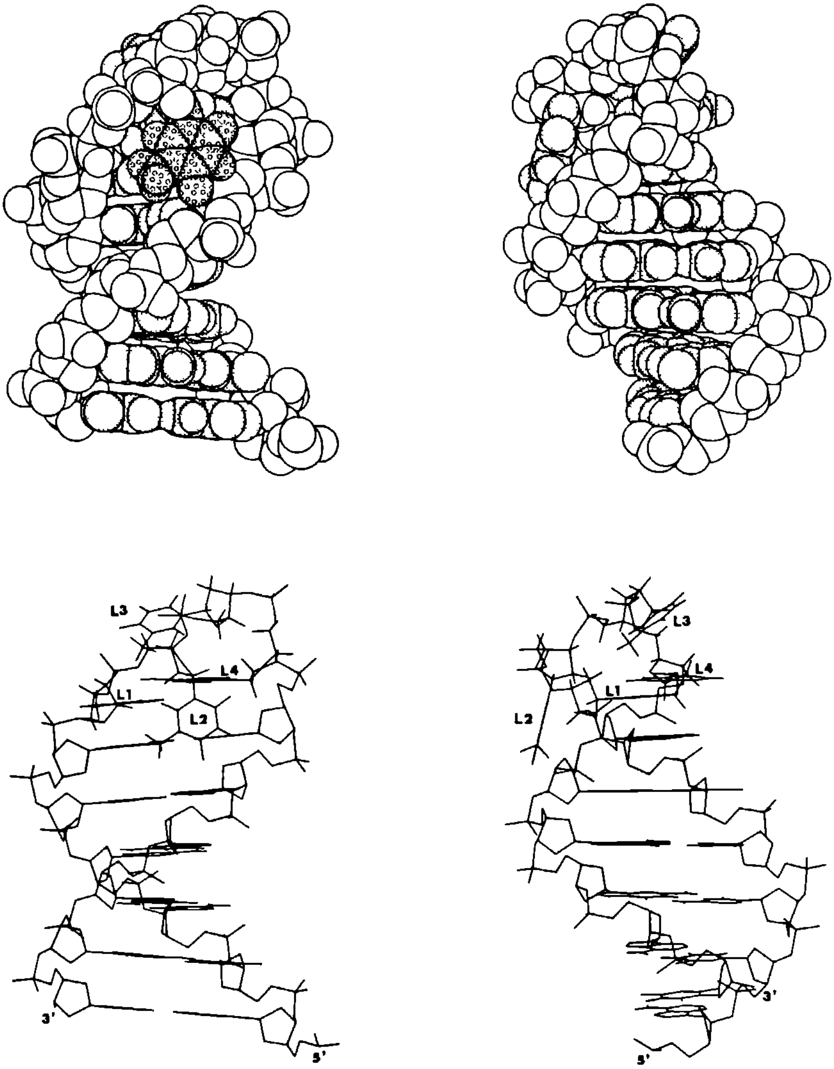


Fig. 11. Plots of the hairpin model in which the loop structure obtained with multi-conformational analysis was docked on a B-type stem. The base of L2 folds downwards into the minor groove. The base of L3 stacks in a non-regular fashion on the base of L4. The structure is displayed with stick models as well as solid models in two different orientations. The aromatic bases in the solid models are hatched.

structure was applied in order to optimize the conformation of the loop-stem junction. This did not alter the conformation of the hairpin loop, which indicates that no steric conflicts are present in this conformation. The resulting hairpin structure is presented in Fig. 11.

The hairpin structure was further studied with the aid of restrained molecular mechanics calculations (see Materials & Methods). In these calculations we included all torsion angle and distance information obtained for the stem and loop sections of the hairpin except the constraints necessary for defining the Hoogsteen base pair formation. The refined structure is presented in Fig. 12A. Note that the A and T which are involved

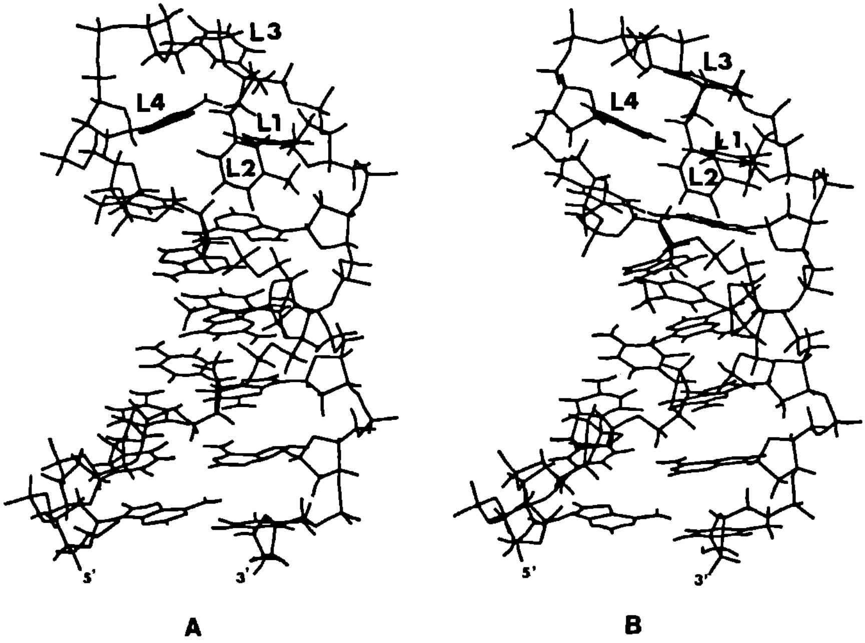


Fig. 12. Structure of the hairpin (A) obtained after a restrained MM calculation and (B) obtained after a unconstrained MM calculation using the structure depicted in (A) as starting structure. It can be viewed that the stacking between the loop bases is more efficient in the free MM structure. In addition, the NH proton of L2 forms a hydrogen bond with the O2 atom of T6.

in the Hoogsteen pairing come close to each other to permit hydrogen bonding, but that the adenine is somewhat tilted. However, after an energy minimization in which all constraints were removed, the Hoogsteen A-T pair became flat (see Fig. 12B). As can be viewed from this figure, the stacking of the various bases in the loop region is optimized: L3 stacks partly on L4 and partly on L1. The Hoogsteen pair stacks on top of the sixth base pair of the B-type stem. All conformational properties of the loop section have been conserved during the calculation. The only new aspect of the energy-refined structure is the formation of a hydrogen bond between the imino proton of residue L2, which is folded into the minor groove, and the O2 atom of residue T6.

All structural details deduced from our NMR data of this hairpin are incorporated in this model. The first loop-thymine forms a Hoogsteen pair with the fourth adenine, which is in high-syn conformation. This base pair is stacked on a B-type stem. The second loop base is bulged out and is folded into the minor groove. This goes together with the unusual torsion angle α^t. The third base is partly stacked on L1 and L4. The turn in direction of propagation of the phosphate backbone goes together with the unusual torsion angles α^+ and ζ^+ between the third and fourth loop residue.

DISCUSSION

Structure determination

In this chapter several new approaches have been introduced to derive the structure of a nucleic acid molecule, in this case the hexadecanucleotide d(ATCCTA-TTTA-TAGGAT), from the raw NMR data.

In contrast to the procedure where the sugar conformation is determined from the J-coupling information with the program PSEUROT [21] here this conformation is obtained by means of a multi-conformational analysis (cf. materials & methods). In PSEUROT the values of the P(N), ϕ_m(N), P(S), ϕ_m(S) and x(S) are obtained with the aid of a least square fit procedure: the RMS difference between the experimental and calculated coupling constants is minimized. This procedure can only be used when detailed coupling data are available. When not all five H-H coupling constants can be measured, assumptions must be made for some of the parameters (e.g., P(N) and ϕ_m(N)). Therefore, with the current data available for the hexadecanucleotide a PSEUROT analysis would result in a strongly biased output. Using the MARC procedure the relation between the NMR parameters and the conformational parameters is made explicit.

Furthermore, a convenient method was introduced to estimate the glycosidic torsion angle, χ, by combining the experimental cross peak intensities and NOE contour plots (cf. Fig. 3). In this method spin diffusion is accounted for and it is ensured that the resulting structure is within the conformational space which can possibly be spanned by a mononucleotide in contrast to some earlier work [22].

To obtain information about the backbone torsion angle γ it is necessary to have available a stereo specific assignment of the H5' and H5" sugar protons. So far the rule of Remin and Shugar was in common use, i.e., it was assumed that the H5'-resonance is found downfield from the H5"-resonance [23]. In the more irregular loop regions this approach may no longer hold and a procedure is introduced to obtain the stereo specific assignments for these protons. This was particularly important when establishing the correct domains for the γ- torsion angles of the L3 and L4 residues, i.e., Remin's and Shugar's rule does not hold for L3 and L4.

Inter-nucleotide distances were estimated directly (by means of the program NOEFIT, from NOE cross peaks without invoking a structural model while still accounting for the effects of spin diffusion. The availability of such a method is particularly important when in advance nothing is known about the structure which in our situation applied to the loop region.

The structural constraints obtained by means of the method described above were translated into a structural model by means of multi-conformational analysis using the program HELIX.

During the past years several computational strategies have been developed to obtain the structure of nucleic acids from NMR data. Three types of approaches or combinations thereof have been utilized: distance geometry calculations [24-26], iterative relaxation matrix approaches [28,29] and restrained molecular dynamics [15]. The methods have been continuously subjected to improvements or are in the process of being improved with respect to proper sampling of the conformational space and with respect to the question to what extent the resulting structure is determined by the NMR constraints [29]. As mentioned above here we investigate the application of a multi-conformational analysis approach. In principle this method permits an unbiased, systematic and complete sampling of the conformational space of a molecule and the relationship between the remaining structure or family of structures and the NMR data remains transparent. A major disadvantage of this approach is that it is very demanding in terms of computer time. In particular when the sampling step size is small, i.e., a fine grid search is conducted, millions of conformations have to be tested against the experimental constraints. This problem may of course be alleviated to some extent by a courser sampling but this has the obvious disadvantage that one could miss part of the conformational space which might match the experimental constraints. It is in fact our experience that when one starts looking at longer fragments, e.g., in our case the loop sequence (vide infra), where say the allowed conformations of the trinucleotide are combined with those of the following dinucleotide, tracing of a conformer of the resulting tetra nucleotide which fits the experimental constraints is the most time consuming part of the procedure. Once one has found such a conformer the 'surrounding' conformational domain fulfilling the experimental constraints can be explored relatively rapidly. The calculations presented here were conducted on a PC-AT and conversion of the program HELIX to more powerful computers will certainly help. Nevertheless improvements of the methodology are needed to make it applicable to larger systems. One approach that might improve matters is to combine the multi-conformational analysis with other approaches, e.g., restrained molecular dynamics or genetic algorithms [30,31].

In this study we have also compared the results of the multi-conformational analysis (MCA) with a restrained molecular dynamics calculation (RMD) of the hexadecanucleotide. This calculation resulted in a set of 100 structures. The results obtained by the two methods are compared in Fig. 13 where conformational wheels of the backbone torsion angles of the loop region are drawn. First we note that the range of torsion angles found with the MCA method are always within the bounds of those found by means of restrained molecular dynamics. Furthermore, some of the torsion angle variations are comparable for both methods, e.g., for the step L1-L2. Other torsion angles, however, adopt considerable wider ranges in RMD. This happens for instance for ε(L2), ζ(L2), α(L4) and β(L4). This is partly due to the fact that in the restrained molecular dynamics calculations sampling took place outside allowed regions, i.e., the ε^{-}-ζ^{t} combination in L2 and the α^{t}-β^{+} combination in L4. This clearly illustrates that J-coupling information is crucial for a detailed structural analysis of the loop region. From the above comparison of the multi-conformational analysis and restrained molecular dynamics it may not be concluded that for the L3-L4 the torsion angle variation are as small as indicated by the MCA results. The bounds may be somewhat wider because only a subset of the structure of the L1-L2-L3 fragment could be analyzed in the MCA study of the L1-L2-L3-L4 fragment otherwise the computer time needed to complete the calculations became prohibitively long. The actual torsion angle variations, however, are expected to be smaller than indicated by the RMD calculation.

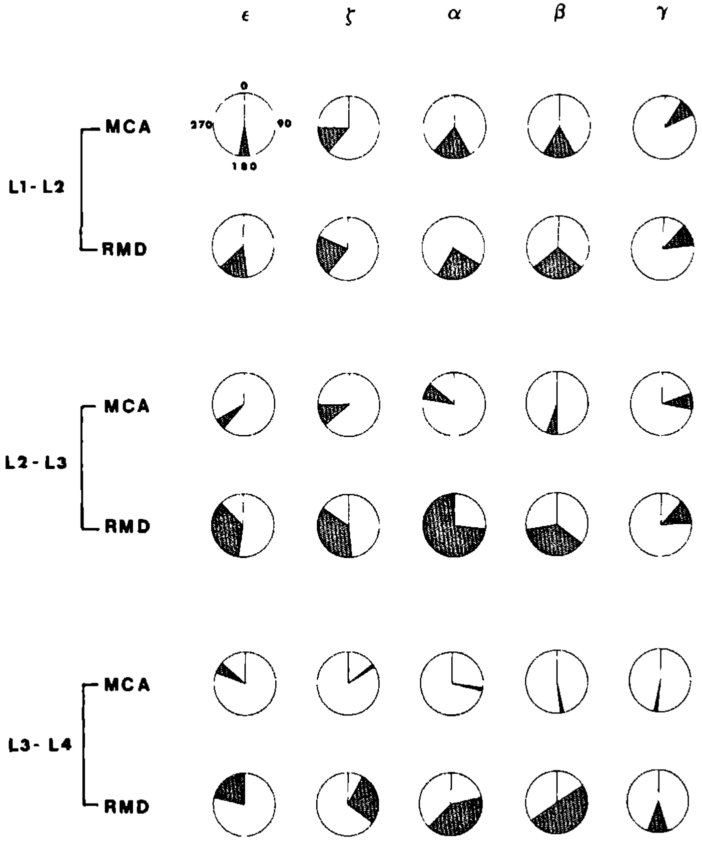


Fig. 13. Conformation wheels for the backbone torsion angles of the loop section. In these wheels the range of values which occur during the multi-conformational analysis (MCA) and during the restrained molecular dynamics calculation (RMD) are given for the L1-L2, L2-L3 and L3-L4 steps.

The structure presented in Fig. 12A satisfies all experimental constraints. It must be emphasized that the NOEs which have been observed with NMR are time-averaged. Because the DNA molecule is not rigid, the NOEs reflect a population of structures. The structure presented in Fig. 12A can therefore be interpreted as a structure which corresponds with these time-averaged NOEs. The structure presented in Fig. 12B, obtained after energy refinement, can be regarded as one of the preferred conformations.

Chemical shifts

The conformation of the hairpin formed by d(ATCCTA-TTTA-TAGGAT) was obtained by means of a multi-conformational analysis, using J-couplings and NOE intensities. Chemical shifts are related to the structure as well and at this place we discuss the most remarkable chemical shifts in the perspective of the obtained structure of the hairpin. When we compare the chemical shifts with those of the strongly related hairpin d(ATCCTA-TTTT-TAGGAT) [14] (see Chapter 4) we note some differences, especially concerning some of the resonances of the protons in the loop region. The most prominent shifts are summarized: upfield shifts are observed for the resonances of L1CH_3 (-0.29 ppm), L2CH_3 (-0.14 ppm), L3H1' (-0.67 ppm), L3H4' (-0.67 ppm) and L3H5' (-0.46 ppm). Downfield shifts are observed for L2H6 (0.27 ppm), L4H2' (0.87 ppm) and L4H2" (0.38 ppm). The chemical shifts of the resonances corresponding to protons in the stem hardly shift.

The sugar residue of L3 is partly stacked on the syn-oriented adenine in position L4. As a result, the H1' and H4' and H5' are positioned about 3.5 Å above the plane of the adenine. The adenine ring currents result in noticeable upfield shifts for these proton resonances. The high-syn orientation of the adenine brings the H1', H2' and H2" protons of the same nucleotide near the base plane and these protons experience downfield ring current induced shifts. Giessner-Prettre and Pullman have calculated ring current shifts of the sugar protons as a function of the glycosidic torsion angle χ [32]. The downfield shifts, experienced by the H1', H2' and H2" resonances of L4 find a good explanation based on these calculations.

The hairpin loop structure

The most surprising aspect of the loop structure of d(ATCCTA-TTTA-TAGGAT) is the occurrence of a Hoogsteen T-A pair in the loop region of the molecule and the fact that the second loop residue is turned into the minor groove. The formation of the Hoogsteen pair classifies the molecule as an example of a DNA hairpin which is able to form a base pair within a loop of four nucleotides. The conformational adjustments which can accommodate this base pair formation were discussed before (Chapter 6). On the basis of the results obtained in the present chapter we can give a more detailed discussion of the loopfolding. We first consider the conformation of the second nucleotide in the loop. Stabilization of the loop structure does not take place through base-base stacking of its base with its neighbors L1 and L3 as in d(ATCCTA-TTTT-TAGGAT), but through its interaction in the minor groove (cf. Figs. 11-12). One side of the aromatic ring is shielded from the solvent, which may contribute to the overall structural stability instead of the vertical base-base stacking [33]. When the structure obtained for this hairpin is subjected to energy minimization a hydrogen bond is formed between the amino proton of A6 and the O4 atom of L2. The formation of such a H-bond may contribute to the stability as well, however, it should be noted that the putative H-bond is only an outcome of the Force Field calculations.

Next we consider the conformation of the sugar-phosphate backbone of the loop region (cf. Tables VI-VII). In Fig. 14 the hairpin is viewed from the top (the loop region is directed towards the reader) along the helix axis. Upon entering the loop region from the 5'-end, the sugar phosphate backbone continues its course with some local deviations as in a B-type helix. Between the third and the fourth residues in the loop, residue L3 and L4, there is a sharp change (a 180° turn) in the direction of the backbone and it continues with a B-type conformation into the stem region. Consideration of the backbone angles between L3 and L4 (Tables VI-VII) shows that four torsion angles differ essentially from the A-or B-type backbone conformation [33], i.e., ε^{-}-ζ^{+}-α^{+}-β^{t}-γ^{t} instead of ε^{t}-ζ^{-}-α^{-}-β^{t}-γ^{+}.

The conversion of the torsion angles $\alpha(3')$ and $\zeta(5')$ in a dinucleotide unit from the gauche^{-}, gauche^{-} (present in A- or B-type helices) to the gauche^{+}, gauche^{+} domain, can induce a complete change (i.e., a 180° turn) in the direction of the phosphate backbone. This has been observed for the first time in the crystal structures of the ribonucleotides UpA [34] and A^+pApA [35] and in the crystal structure [36] and NMR structure [37] of the cyclic deoxyribonucleotide d<pApA>. In all of these examples the remaining part of the backbone in the dinucleotide unit pertains to an A-type backbone conformation. This includes the sugar moieties which are N-puckered even for the deoxyribose sugars in the cyclic dinucleotide d<pApA>. In the latter case the sugar(s) may switch to a S-type conformation, while retaining α(3') and ζ(5') in the gauche$^+$ domain and preserving the preservation of the 180°-turn in the backbone, but this requires adjustment of the other torsion angles otherwise severe steric hindrance of the H3'-atoms of the adjacent sugars will occur (Chapter 3). In d(ATCCTA-TTTA-TAGGAT) the sugar moieties of the L3 and L4 residues are S-puckered (cf. Tables II,VI,VII) and indeed

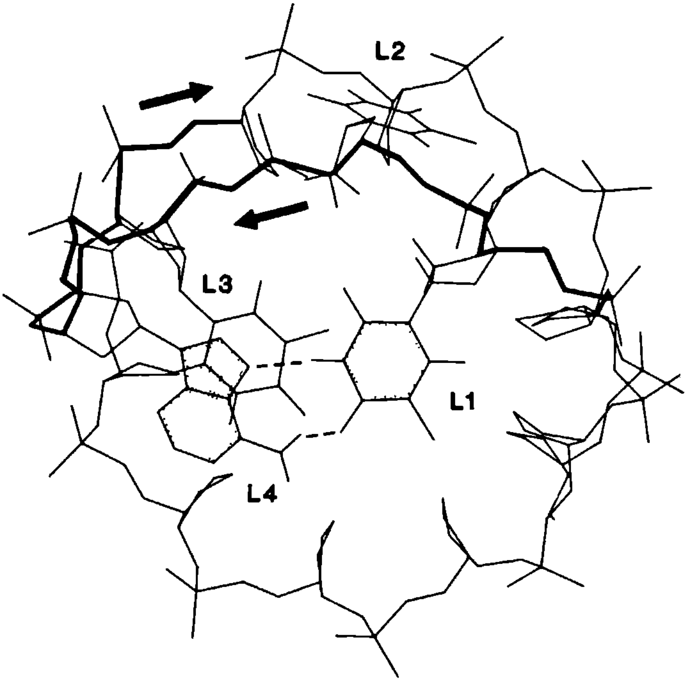


Fig. 14. Plot of the hairpin structure viewed from the loop region along the helix axis. The bases of the stem are deleted for reasons of clarity. The Hoogsteen T-A pair is hatched. The phosphate backbone in the loop region is emphasized with thick lines and it is indicated that the direction of the backbone is changed 180° between the third and fourth loop residue.

compared to the situation ε^t and γ^+ seen in the turns mentioned above here we have ε^- and γ^t. To obtain the 180°-turn it is not necessary, however, to switch ε to the gauche$^-$-domain as we have seen in the loop of the hairpin formed by d(ATCCTA-TTTT-TAGGAT). It is noted in passing that the 180° turn in this hairpin is also localized between the L3 and L4 residues. Its conformation was estimated using the same procedure followed for the conformational analysis of the title compound (see Chapter 4). The turn in this hairpin is described by the backbone torsion angles ε^t-ζ^+-α^+-β^t-γ^t. As a result in this molecule the bases L3 and L4 remain far enough apart from another to permit interaction with a third base. Indeed L3 is partly stacked upon L2 in this hairpin. Similar situations are encountered in the other molecules introduced above. In the hairpin formed by d(ATCCTA-TTTA-TAGGAT) residue L2 is turned into the minor groove which allows L3 to stack upon L4. This is accommodated by changing ε to the gauche$^-$-domain.

Despite the fact that the second thymidine (L2) in the loop region turns into the minor groove the phosphate backbone does not change direction. This orientation of the thymidine brings about a change in α(L2) which is in the trans orientation instead of the normal gauche$^-$ domain. The direction of the backbone is maintained by compensations in the values of

Table VII. Torsion angles comprising the 180° turn in four hairpins for which accurate data were obtained. The values of the backbone torsion angles are indicated with the classical rotamers gauche$^+$ (+), trans (t), gauche$^-$ (-) or - anti clinal (around -120°). The 180° turn is compared with that of the cyclic dinucleotide d<pApA> and with the torsion angles in the backbone of B-DNA.

	ε	ζ	α	β	γ
d(ATCCTA-TTTT-TAGGAT)	t	+	+	t	t
d(ATCCTA-TTTA-TAGGAT)	-	+	+	t	t
d($C_mGC_mGTGC_mG$)	t	-ac	-	+	t
d(CGC_aTAGCG)	-	+	t	+	t
d<pApA>	t	+	+	t	+
B-DNA	t	-	-	t	+

torsion angles in the 3'-direction. The most outstanding are those observed for ε(L2) and ζ(L2). These angles both fall in the anticlinal region instead of the trans and gauche$^-$-region, respectively. It is interesting to look at this problem in a somewhat different manner. Yathindra and Malathy have noted that the six bond backbone P-O5'-C5'-C4'-C3'-O3'-P of a nucleotide unit may, in a first approximation, be represented by two segments, i.e., P-O5'-C5'-C4' on the 5'-side and C4'-C3'-O3'-P on the 3'-side. The distance between the terminal atoms in these segments is almost constant, i.e., close to 3.9 Å in a wide variety of structures [38]. These two segments have been designated hemi-nucleotides and nucleotide conformations can be conveniently described by virtual bonds which join the terminal atoms of the aforementioned segments and by the 'torsion angles' which can be defined for these bonds, i.e., α_V for (P-C4') and ζ_V for (C4'-P). In normal B-DNA the values for these virtual torsion angles α_V and ζ_V are on the average 160° and 210° respectively [38]. Summation of α_V(L2), ζ_V(L2) and α_V(L3) yields 167°, while the corresponding value for B-DNA is 182°. Thus,the change in α(L2) is compensated by other angles so that the course of the backbone is maintained as in B-type DNA.

At this point it is interesting to compare the present results with those obtained for the 'mini-hairpins' formed by d($C_mGC_mGTGC_mG$) and d(CGC^aTAGCG) [39,40]. In these sequences C_m denotes a cytidine methylated at the C5-position and C^a represents a cytidine in which the deoxyribose sugar is replaced by an arabinose sugar. By a combination of NMR structural studies and molecular mechanics or molecular dynamics calculations similarly detailed data were obtained for the backbone torsion angles as presented in the present paper. The loop in the first hairpin formed by d($C_mGC_mGTGC_mG$) resembles that in the hairpin formed by d(ATCCTA-TTTT-TAGGAT) [14]. In this loop, formed by the sequence -C_mGTG-, the first and the last residue are base paired and the stacking of the bases is as in the -TTTT- loop. In the loop of the second hairpin, formed by the sequence -C^aTAG- of d(CGC^aTAGCG), again the first and the last nucleotide form a base pair, but in this case the second loop residue (the thymidine) is turned into the minor groove as in the d(ATCCTA-TTTA-TAGGAT) molecule. The presence of the arabinose sugar could be thought to play a predominant role in inducing this effect, but the d(ATCCTA-TTTA-TAGGAT) hairpin provides an example, that this may occur in an all deoxyribose phosphate backbone as well.

In both 'mini-hairpins' the change in direction in the sugar phosphate backbone is between the third and the fourth residue in the loop region (in our terminology) as is found in the hexadecanucleotide hairpins. Although not all of the torsion angles effecting the 180°-turn in the 'mini-hairpins' are the same as in the hexadecanucleotides some interesting correspondences between them can be noted (cf. Table VII).

First the sugar moieties of L3 and L4 all are S-puckered; also in all four hairpins the angle γ(L4) adopts a trans orientation. Furthermore, in d(CGC^{a}TAGCG) ε(L3) and ζ(L3) are gauche$^-$ and gauche$^+$ respectively as in the corresponding d(ATCCTA-TTTA-TAGGAT). On the other hand α(L4) and β(L4) in both mini-hairpins as well as ζ(L3) in d(C$_m$GC$_m$GTGC$_m$G) differ from the angles in the corresponding hexadecanucleotide hairpins (cf. Table VII). These results show that it is possible to achieve the 180° turn with different combinations of torsion angles. The data for the mini-hairpins were obtained, however, after energy minimization calculations and other combinations of angles, e.g., the ones found for the hexadecanucleotides, might yield lower energy conformations. This is already indicated by the experimental values obtained for β; this angle is not conformationally pure, apart from the gauche$^+$- the energetically more favorable trans-conformation contributes to the structure [41]. Furthermore, molecular dynamics simulations show that the ζ^+-α^--β^+ combination easily converts to the ζ^+-α^+-β^t conformation which is observed in the hexadecanucleotides.

In conclusion, the loop folding principle is operative in the DNA hairpins which are presented in Figs. 15 C-F. Other DNA hairpin structures havebeen published [42-44], but in these studies the analysis was restricted to NOEs. Therefore they are less well-defined and it is difficult to relate these to the folding principle. The hairpin formed by d(CGCG-TTTT-CGCG) for example was studied with NOESY and distance geometry. The backbone between the residues consisting of O3'-P-O5'-C5'-C4' is completely undefined [42] and exhibits large variation from structure to structure. The position of the bases was more reproducible: the third and fourth loop thymidine stack on C9, extending the stacking pattern of the stem. The first loop thymine stacks at the 3'-end of the first helical strand and the second loop thymine turns the corner and closes the loop. Whether this hairpin and the title compound have the similar folding cannot be answered here, but the stacking pattern of the bases in the loop is similar.

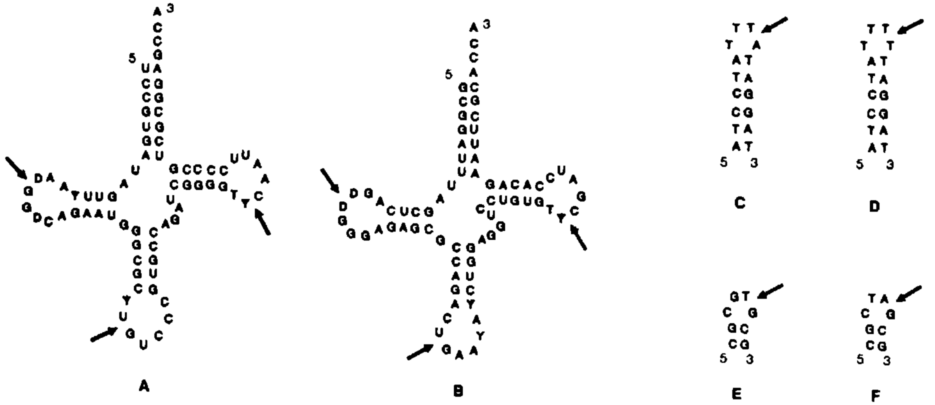


Fig. 15. Positions in tRNA loops (A-B), indicated by arrows, where sharp turns (180 -turns) occur in the direction of the sugar phosphate backbone. In tRNA^{Phe} (A) and tRNA^{Asp} (B) this turn is located after the 5th loop residue, counting from the 5'-end of the stem. In contrast, for B-DNA hairpins (C-F) it is found that the turn is located after the third residue, counting from the 3'-end of the stem. The octanucleotides are presented as four-membered loops according to our folding principle [1]. Within these loops the first and forth base form a regular C-G pair. The position of the turn corresponds with those of the hexadecanucleotides.

A class of molecules exhibiting sharp changes in their backbone chain are the tRNAs. Detailed data are available for yeast tRNA^{Phe} [45] and yeast tRNA^{Asp} [46]; 180° turns (called π-turns in the tRNA literature [33]) are observed at the 3'-side of the nucleotides 9, 17, 33, 46 and 55 (as indicated in Fig 15). Nucleotide 9 and 46 are involved in the junction between the D-arm and the amino acyl acceptor arm and between the anticodon arm and the T-arm respectively. The remaining turns are present in the three loop regions. The combination of torsion angles leading to the π-turns are different from those observed for the DNA hairpins (see Table IV in Chapter 3); extensive discussions have been presented in the literature of the combinations of torsion angles leading to turns in tRNA which will not be repeated here, except that we mention that the turn with ζ^- and α^+ or α^t is the most common. The π-turns in the loop region all occur between the fifth and the sixth residue, counting from the 3'-terminal loop residue (in the 5'-direction; see Fig 15). This is in contrast to the situation in the

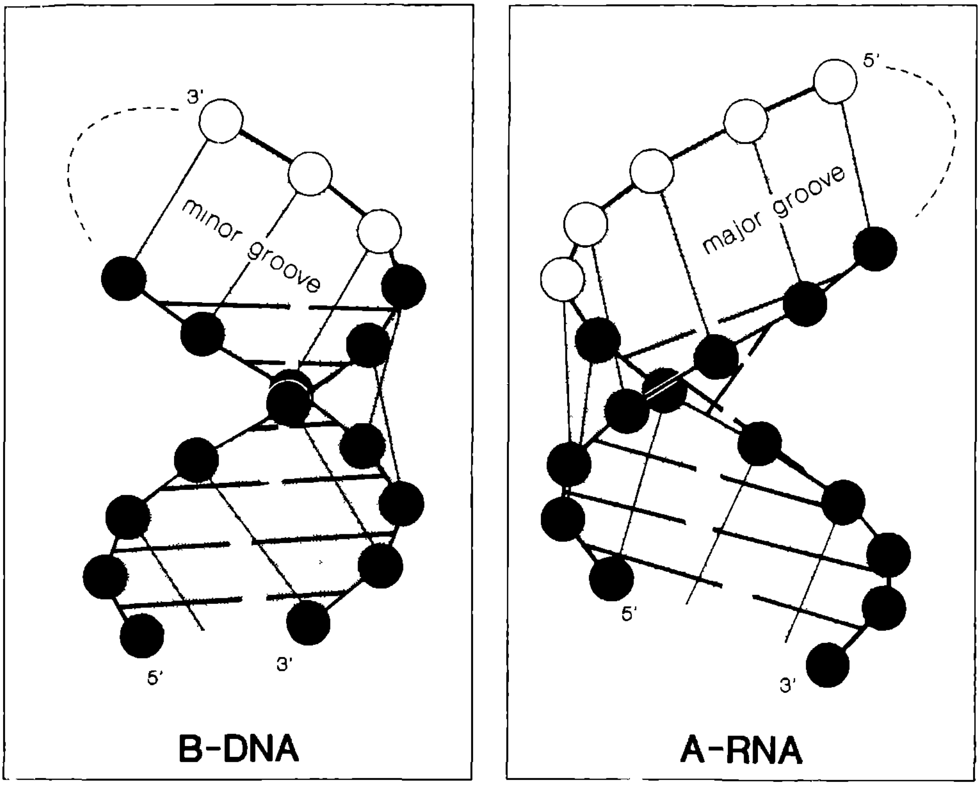


Fig. 16. Plots of the backbone phosphorus atoms in a B-type and A-type helical arrangement. When the A-type helix is extended with 5 residues (phosphates) over the major groove, the loop can be closed easily with two nucleotides. On the other hand, when the B-type helix is extended at the 3'-end (over the minor groove) then the distance over the minor groove can be bridged by one or two nucleotides. Note that the phosphate distances crossing the minor groove in B-DNA and the major groove in A-RNA are comparable in size.

aforementioned four DNA hairpins, where after three residues, counting from the 5'-end the sharp turn occurs. These observations are in accordance with the loop folding principle formulated for the formation of hairpin loops in DNA and RNA, except that at the time of the formulation this conclusion was based upon the assumption that the stacking present in the stem of the hairpin was extended into the loop region [47,48]. For the hairpins discussed here this does not occur in all cases. We do see, however, that the phosphate chain maintains the course of an A-type or B-type backbone as prescribed by the folding principle and as indicated in the earlier representation of this principle by means of a plot of the distances between the phosphates in the loop and the terminal phosphate of the stem region [47,49]. The base stacking, however, adepts to local energetic needs.

Fig. 16 shows the structure of an A- and B-type helix schematically. At the 3'-end of the B-type double helix the phosphate backbone is extended with three residues (phosphates). At the 5'-end of the A-type helix the backbone is elongated with five phosphates. The position of the phosphates in the loop are drawn with white spheres, the phosphates in the stem are indicated with black spheres. The distances between the phosphates crossing the minor groove after 3 nucleotides in B-DNA and the distances between the phosphates crossing the major groove after 5 nucleotides in A-RNA is almost identical (12 Å). As can be gleaned from Fig. 16 the geometry of hypothetical trees which connect the phosphates over the minor groove in B-DNA is quite similar to those drawn over the major groove in A-RNA. It is noted that the P-P distance of 12 Å is correlated with the water structure, which was observed with the aid of X-ray diffraction of the A-DNA structure formed by the iodinated tetramer $d(iCGCG)_2$ [50]. It was argued that the water structure stitches together the phosphates over the major groove. This observation suggests that the water structure in the grooves of hairpin loops have a contribution to the overall loop folding of hairpin loops. As can be inferred from Fig. 16, the lines between the phosphates, which may correspond to 'forces which connect the phosphates', keep intact till the turn is made.

RNA loops occurring in tRNA consist of 7-9 nucleotides. Examination of the secondary structure of 16S-RNA indicates that RNA hairpins may form also four-membered loops. Indeed melting experiments performed for the series of partly self-complementary oligonucleotides r(GGAUAC-N_x-GUAUCCA) in which N = A, C or U and x = 3,4,5,7 and 9 revealed that hairpins with a 4-5 membered loop are most stable [51]. A recent study of the hairpin formed by r(GCGAUU-UCU-GACCGCC) demonstrates that even three-membered loops in RNA are possible [52]. These results can be explained in terms of the folding principle in the following way: first we consider, e.g., a seven-membered RNA loop which has a folding in which the nucleotide chain is extended at the 5'-end with five nucleotides. The remaining gap is then closed by two nucleotides. The nucleotides in the loop will optimize the hairpin stability by adjusting their relative positions. In this situation the bases of the loop adjacent to the stem may form a pair and reduce the loop to five nucleotides. Such a base pair formation may also occur between the second and the sixth loop base, but this will probably induce strain in the molecule. This is indeed obvious in the NMR study of the hairpin formed by r(GCGAUU-UCU-GACCGCC): three ribose sugars have adopted a S-type conformation which is probably necessary to relax the strain in the molecule. A further probing of this hairpin in terms of the folding principle goes beyond the presently obtained structural data.

CONCLUSION

The ^{1}H-spectrum of the hairpin formed by d(ATCCTA-TTTA-TAGGAT) was assigned almost completely by means of a sequential analysis of the NOESY and DQF-COSY spectra. Stereospecific assignments of the H5' and H5" protons were achieved by means of a semi-quantitative evaluation of the NOE and COSY cross peak intensities. The assignment of the resonances of the ^{31}P atoms in the loop was achieved with the aid of a HMQC-experiment.

The orientation of the phosphate backbone torsion angle β and γ in the loop were estimated with the aid of an analysis of the NOESY and DQF-COSY spectra. The orientation of the backbone torsion angle ε follows from the HMQC spectrum, i.e., the long-range 31P-1H coupling between the P3' and H2'.

Experimental ^{1}H-^{1}H coupling constants and NOE cross peak intensities were used to obtain structural details of the hairpin structure, especially the structure of the loop nucleotides. The detailed conformation of the loop was established by means of a multi-conformational analysis in which models were generated with a systematic variation of the torsion angles of the phosphate backbone. These models were tested against the available distance constraints. This resulted in a series of structures which satisfy all the constraints. These structures appeared to be very similar and indicate that this part of the molecule is determined near atomic resolution.

The conformation of the loop was docked on a B-type stem. This structure was further refined with restrained molecular mechanics. The structure has a compactly folded loop. Within the four membered loop a Hoogsteen type T-A pair is formed, comprising the first and last base in the

-TTTA- sequence. This pair is stacked upon the sixth base pair of the stem. The second base is folded into the minor groove and the third base is partly stacked on the first and fourth loop residue. The turn in the phosphate backbone takes place between the third and fourth nucleotide in the loop. Comparison of this structure with those published for other DNA and RNA hairpins support a general model for the folding of hairpin loops.

REFERENCES

1 Blommers, M.J.J., Walters, J.A.L.I., Haasnoot, C.A.G.,, van der Marel, G.A., van Boom, J.H. & Hilbers, C.W. (1989) Biochemistry **28**, 7491-7498.

2 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) Cold Spring Harbor Symp. Quant. Biol. **47**, 403-409.

3 Marion, D. & Wühtrich, K. (1983) Biochem. Biophys. Res. Commun. **113**, 967-974.

4 Sklenar, & Bax, A.(1988) J. Am. CHem. Soc. USA **109**, 2221-2222.

5 Altona, C. (1982) Recl. Trav. Chim. Pays-Bas **101**, 413-433.

6 de Leeuw, H.P.M., Haasnoot, C.A.G. & Altona, C. (1980) Isr. J. Chem. **20**, 108-126.

7 Haasnoot, C.A.G., de Leeuw, F.A.A.M. & Altona, C. (1983) Tetrahedron **36**, 2783-2792.

8 de Leeuw, F.A.A.M., van Beuzekom, A.A. & Altona, C. (1983) J. Comp. Chem. **4**, 438-449.

9 van de Ven, F.J.M. & Hilbers, C.W. (1988) Eur. J. Biochem. **178**, 1-38.

10 Solomon, I. (1955) Phys. Rev. **99**, 559-565.

11 Macura, S. & Ernst, R.R. (1980) Mol. Phys. **41**, 95-117.

12 Cantor, R.R. & Schimmel, P.R. (1980) Biophysical Chemistry 1th ed. W.H. Freeman & Co. San Fransisco.

13 Perrin, C.C. (1970) Mathematics for chemists. 1th ed. J. Wiley & Sons Inc. New York.

14 Blommers, M.J.J., Haasnoot, C.A.G., Hilbers, C.W., van Boom, J.H. & van der Marel, G.A. (1987) Structure and Dynamics of Biopolymers, NATO ASIS Series E: Applied Sciences **133**, 78-91.

15 Nilges, M.G., Clore, M., Gronenborn, A.M., Brunger, A.T. & Karplus, M. (1987) Biochemistry **26**, 3718-3733.

16 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. & Karplus, M. (1983) J. Comp. Chem. **4**, 187.

17 Discover 2.5 (1989) Biosym technologies Inc. San Diego.

18 Chazin, W.J., Wühtrich, K., Hyberts, S., Rance, M., Denny, W.A. & Leupin, W. (1986) J. Mol. Biol. **190**, 439-453.

19 Haasnoot, C.A.G., de Leeuw, F.A.A.M., de Leeuw, H.P.M. & Altona, C. (1979) Recl. J. R. Neth. Chem. Soc. **98**, 576-577.

20 Lankhorst, P.P., Haasnoot, C.A.G., Erkelens, C. & Altona, C. (1984) J. Biomol. Struct. Dyns. **1**, 1387-1405.

21 de Leeuw, F.A.A.M. & Altona, C. (1983) QCEP Bull **3**, 69-70.

22 Gronenborn, A.M., Clore, G.M. & Kimber, B.J. (1984) Biochem. J. **221**, 723-736.

23 Remin, M. & Shugar, D. (1972) Biochem. Biophys. Res. Commun. **48**, 636-642.

24 Patel, D.J., Dhapirio, L. & Hare, D. (1987) Q. Rev. Biophys. **20**, 35-112.

25 Patel, D.J., Dhapirio, L. & Hare, D. (1987) Annu. Rev. Biophys. Biophys. Chem. **16**, 423-454.

26 Reid, B.R. (1987) Q. Rev. Biophys. **20**, 1-34.

27 Boelens, R., Koning, T.M.G., van der Marel, G.A., van Boom, J.H. & Kaptein, R. (1989). J. Mag. Res. **82**, 290-308.

28 Borgias, B.A. & James, T.L. (1988) J. Magn. Res. **79**, 493-512.

29 Metzler, W.J., Hare, D.R. & Pardi, A. (1989) Biochemistry **28**, 7045.

30 Lucasius, C.B. & Kateman, G (1989) International conference on genetic algorithms. **3** Editor Shaffer, J.D. San Mateo, CA., 170-176.

31 Lucasius, C.B., Blommers, M.J.J., Buydens, L.M.C. & Kateman, G. (1990) "A genetic algorithm for conformational analysis of DNA". Handbook of the genetic algorithm. (in press)

32 Giessner-Prettre, C. & Pullman, B. (1977) J. Theor. Biol. **65**, 189-201.

33 Saenger, W. (1984) Principles of Nucleic Acid Structure. 1th ed. Springer Verlag. New-York.

34 Sussman, J.L., Seeman, N.C., Kim, S.-H. & Bermann, H.M. (1972) J. Mol. Biol. **66**, 403-421.

35 Suck, D., Manor, P.C. & Saenger, W. (1976) Acta Crystallogr. Sect B: Struct. Crystallogr. Cryst. Chem. **B32**, 1727-1737.

36 Frederick, C.A., Coll, M., van der Marel, G.A., van Boom, J.H. & Wang, A.H.-J. (1988) Biochemistry **27**, 8350-8361.

37 Blommers, M.J.J., Haasnoot, C.A.G., Walters, J.A.L.I., van der Marel, G.A., van Boom, J.H. & Hilbers, C.W. (1988) Biochemistry **27**, 8361-8369.

38 Yathindra, N. & Malathi, R. (1983) Nucleic Acids: The vector of Life, 229-252.

39 Orbons, L.P.M., van Beuzekom, A.A. & Altona, C. (1987) J. Biomol. Struct. Dyns. **4**, 965-987.

40 Pieters, J.M.L. de Vroom, E., van der Marel, G.A., van Boom, J.H., Koning, T.M.G., Kaptein, R. & Atlona, C. (1989) Biochemistry **29**, 788-799.

41 Pieters, J.M.L. (1989) Thesis. State University, Leiden, The Netherlands.

42 Hare, D.R. & Reid, B.R. (1986) Biochemistry **25**, 5341-5350.

43 Gupta, G., Sarma, M.H. & Sarma, R.H. (1987) Biochemistry **26**, 7715-7723.

44 Williamson, J.R. & Boxer, S.G. (1989) Biochemistry **28**, 2819-2831.

45 Westhof, E. & Sundaralingam, M. (1986) Biochemistry **25**, 4868-4878.

46 Westhof, E., Dumas, P, & Moras, D. (1985) J. Mol. Biol. **184**, 119-145.

47 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) J. Biomol. Struct. Dyns. **3**, 843-857.

48 Haasnoot, C.A.G., Blommers, M.J.J. & Hilbers, C.W. (1987) Structure, Dynamics and Function of Biomolecules. Springer Ser. Biophys. **1**, 212-216.

49 Haasnoot, C.A.G., de Bruin, S.H., Hilbers, C.W., van der Marel, G.A. & van Boom, J.H. (1985) Proc. Int. Symp. Biomol. Struct. Interactions, Suppl. J. Biosci. **8**, 767-780.

50 Conner, B.N., Takano, T., Tanaka, S., Itakura, K. & Dickerson, R.E. (1982) Nature **295**, 294-299.

51 Groebe, D.R. & Uhlenbeck, O.C. (1988) Nucl. Acids Res. **16**, 11725-11735.

52 Puglisi, J.D., Wyatt, J.R. & Tinoco, I. (1990) Biochemistry **29**, 4215-4226.

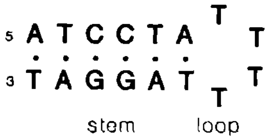
53 Haasnoot, C.A.G., de Leeuw, F.A.A.M., de Leeuw, H.P.M. & Altona, C. (1979) Receil, J. Roy. Netherl. Chem. Soc. **98**, 567-577.

SUMMARY

ASPECTS OF LOOP FOLDING IN DNA HAIRPINS

NUCLEAR MAGNETIC RESONANCE, MULTI-CONFORMATIONAL ANALYSIS, UV-MELTING AND MOLECULAR MECHANICS CALCULATIONS

Natural occurring RNA and DNA contain structural elements which may be called hairpins. Such hairpins consist of a double-helical stem of which the two strands are connected by a single stranded loop. An example is given in the scheme below:



The formation of hairpins, which occurs in cellular RNA and DNA has interesting and probably important biological implications. The investigations described in this thesis, were aimed to gain insight into the spatial structure and conformational properties of DNA hairpins. In order to achieve this, these hairpins were studied with nuclear magnetic resonance (NMR), multi-conformational analysis, UV-melting experiments and by means of molecular mechanics calculations.

Chapter 1 gives an overview of the 2-dimensional NMR methods, which are of importance for the structural investigations of nucleic acids. In addition, the nomenclature which is used for the characterization of nucleic acid structures is introduced. The relation between NMR parameters such as the nuclear Overhauser effect (NOE), the proton-proton distances and J-coupling constants on the one hand, and the conformation of the nucleotide units in the DNA molecule on the other hand, is visualized by way of a number of representative plots. Finally, the results of the structural investigation of DNA hairpin molecules, which preceded the studies presented in this dissertation, are discussed briefly.

Chapter 2 describes the analysis of a 3'-5' cyclic DNA molecule that consists of two nucleotides: the d<pApA>. The study of this molecule is

interesting in relation to the investigation of hairpin molecules, because the sugar-phosphate chain between the two nucleotides turns 180°. Such changes are also observed in the conformation of the sugar-phosphate chain of loops in hairpin molecules (Chapters 4 and 7) and they appear to be better interpretable as a result of the examination of the dinucleotide. Moreover, the d<pApA> molecule is related to the cyclic dinucleotide r<pGpG> which functions as the natural activator of the cellulose synthase in Acetobacter xylinum. With the aid of the determination of 1H-1H, 1H-31P and 13C-31P coupling constants, it was possible to derive the structure of d<pApA>. The conformation of the sugars in this cyclic dinucleotide is, for deoxyribose rings, of the unusual N-type. The 180° turn in the sugar-phosphate chain of the d<pApA> is caused by the torsion angles α and ζ, which both adopt a gauche$^+$ conformation. This conformation appears to be also a part of the 180° change in direction which occurs in the studied hairpin loops (Chapters 4 and 7). At lower temperatures, the d<pApA> in solution appears to form a dimer, in which a new type of base stacking plays a role. At elevated temperatures and/or in extreme dilutions the dimer transfers into two monomers which can change conformation.

In Chapter 3 the d<pApA> molecule is studied with the aid of molecular mechanics calculations. 'All possible' conformations were generated, by means of systematically sampling the conformational space, allowed by this molecule, by way of a multi-conformational analysis. From this compilation the low energy conformations (conformers) were obtained with the aid of energy minimization methods. It appeared that the lowest energy conformer gives a good description of the experimentally determined structure. Through a comparison of the structures of low energy conformers, a better understanding of the conformational properties of the molecule was obtained. Particularly, the occurrence of the N-type deoxyribose sugars in the lowest energy conformer could be understood.

Chapter 4 describes NMR studies of the hairpin structure formed by the d(ATCCTA-TTTT-TAGGAT) sequence. 2-dimensional NMR experiments which were performed for this hairpin (roughly 160 protons) allowed an almost complete assignment of the resonances of the 1H-spectrum. On the basis of a sequential analysis of the cross peaks in the NOESY spectrum (2D-NOE spectrum) and an analysis of coupling constants, it could be concluded that the stem of the hairpin can be characterized as a B-type double-stranded helix. The stacking of the bases in the stem is propagated into the loop: the B-type stacking is extended at the 3'-end of one of the chains of the stem into the loop with three nucleotides. The sugar-phosphate chain changes its direction between the third and fourth nucleotide in the loop. The fourth base turns towards the interior of the loop and forms a T-T pair with the first base in the loop. The conformation which is responsible for the 180° turn in the phosphate chain between the third and fourth residue in the loop, was derived with the aid of multi-conformational analysis. The structure was also examined with molecular mechanics calculations. Coordinates, which were obtained from these calculations of the lowest energy structure, were used for a simulation of the NOESY spectrum and the results of the calculation and experiment were quite similar.

In Chapter 5 the folding principle for hairpins, which was first proposed by Haasnoot et al. (J. Biomol. Struct. Dyns (1986), 3, 843-857) is further elaborated. This principle was formulated on the basis of the observed difference in stability of DNA and RNA hairpin molecules as a function of their loop size. On the basis of the geometric properties of A-RNA versus B-DNA it is 'advantageous' to extend the 5'-end of the A-type helix with about five nucleotides. The remaining gap, which has to be bridged to the 3'-end of the stem, is then easily closed by two nucleotides. For a B-type helix it is 'advantageous' to extend the 3'-end with two or three nucleotides. The remaining gap to the 5'-end of the stem is then easily closed with one or two nucleotides. This principle, whose validity was supported by the structure of the hexadecanucleotide d(ATCCTA-TTTT-TAGGAT) (Chapter 4) appears to require extension on the basis of the same structure. The formation of a wobble T-T pair within the loop structure, brought forth the question whether this pair could be substituted by a complementary C-G or A-T combination. For this purpose the performed molecular mechanics calculations in this chapter suggest that base pairing in a loop of four nucleotides is indeed possible for the loop sequences-CTTG- and -TTTA- but not for the loop sequences -GTTC- and -ATTT-.

These results formed a basis for the experiments described in Chapter 6. The hairpins with the aforementioned loop sequences were synthesized and subjected to UV-melting studies and NMR-experiments. It appeared that the hairpins with the loop sequences -CTTG- and -TTTA- are significantly more stable than those with the loop sequence -GTTC- and -ATTT-. With the aid of

NMR experiments of the hairpins with the -TTTA- and -CTTG- loop sequences an extra base pair was observed. For the -GTTC- and -ATTT- loops no such indication is found. In addition, the influence of the base pair sequence of the stem and the replacement of the two central thymines by adenines was studied. 2D-NMR experiments performed for the hairpin formed by d(ATCCTA-TTTA-TAGGAT) lead to the discovery that the additional T-A pair, formed within the loop of four nucleotides, has an unusual Hoogsteen configuration.

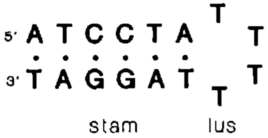
The conformational analysis of nucleic acids with the aid of NMR is complicated. In the first place the experimental data, e.g., NOE intensities and J-coupling constants must be translated into H-H distances and torsion angles and, subsequently, these must be translated into a structure. Chapter 7 describes a method in which the (loop) structure of the hairpin formed by d(ATCCTA-TTTA-TAGGAT) could be derived. In this procedure the sugar conformation and the glycosidic torsion angles were derived initially through multi-conformational calculations, in which the available J-couplings and NOE intensities were used. Intra- and inter-nucleotide proton-proton distances were estimated by means of a relaxation matrix calculation without the necessity of introducing a molecular model. The torsion angles β, γ and ε were estimated from homo- and hetero-coupling constants, when possible. The structure of the loop was subsequently derived with the aid of multi-conformational analysis. The hairpin formed by d(ATCCTA-TTTA-TAGGAT) appears to have a B-type stem, the first and the fourth base of the loop form a Hoogsteen pair, which is stacked upon the last base pair of the stem (Chapter 6). The second base of the loop folds back into the minor groove of the B-type double helix. The third is partly stacked on the first and partly stacked on the fourth loop base. The sugar-phosphate chain makes a 180° turn between the third and fourth residue of the loop. Furthermore, a comparison is made between the structures of the hairpins formed by d(ATCCTA-TTTT-TAGGAT) and d(ATCCTA-TTTA-TAGGAT). In spite of the differences in the stacking pattern of the bases in the loop, several structural resemblances are obvious: in both structures, the sugar-phosphate chain in the loop follows its course in the 3'-direction according to the pattern of a B-type helix, though with some local deviations. After the third residue, the phosphate backbone makes a sharp 180° turn and the remaining gap is closed by one nucleotide. The conformation of the region which exhibits the change in direction of the chain, is almost a copy of the base stacking pattern in the loop. The thus far obtained results suggest that the folding in these hairpins can be described with the aid of relatively simple rules.

SAMENVATTING

ASPECTEN VAN LUSVOUWING IN DNA-HAARSPELDEN

KERNMAGNETISCHE RESONANTIE, MULTI-CONFORMATIE ANALYSE,
UV-SMELT EN MOLECULAIRE MECHANICA BEREKENINGEN

Natuurlijk voorkomend RNA en DNA bevat structuurelementen welke men haarspelden zou kunnen noemen. Zulke haarspelden bestaan uit een dubbele helix, waarvan de twee strengen verbonden zijn via een enkelstrengs lus. Een voorbeeld wordt gegeven in het onderstaande schema:



De haarspeldvorming, die in cellulair RNA en DNA voorkomt heeft interessante en waarschijnlijk belangrijke biologische implicaties. Het in dit proefschrift beschreven onderzoek was er op gericht inzicht te verwerven in de structuur en de conformationele eigenschappen van DNA haarspelden. Daartoe werden deze bestudeerd met kernmagnetische resonantie (NMR), multi-conformatie analyse, UV-smeltmetingen en middels moleculaire mechanica berekeningen.

Hoofdstuk 1 geeft een overzicht van de 2-dimensionale NMR-methoden, welke belangrijk zijn voor het structuuronderzoek van nucleïnezuren. Tevens wordt de nomenclatuur geïntroduceerd welke gebruikt wordt ter karakterisering van nucleïnezuurstructuren. De relatie tussen NMR parameters zoals het nucleaire Overhauser effect (NOE), de proton-proton afstanden en J-koppelingsconstanten enerzijds, en de conformatie van de nucleotide eenheden in het DNA-molecuul anderzijds, wordt in beeld gebracht via een aantal representatieve voorstellingen. Tenslotte worden de resultaten van structuuronderzoek aan DNA-haarspeld moleculen, dat voorafging aan het in deze dissertatie beschreven onderzoek, kort besproken.

Hoofdstuk 2 beschrijft de analyse van een uit twee nucleotiden bestaande 3'-5'-cyclische DNA-molecuul, het d<pApA>. De studie van dit molecuul is interessant in het kader van het onderzoek aan haarspeld moleculen omdat de suiker-fosfaat-keten tussen de twee nucleotiden 180° van richting verandert. Dergelijke veranderingen worden ook waargenomen in de conformatie van de suiker-fosfaat-keten van de lussen in haarspeldmoleculen (hoofdstuk 4 en 7) en blijken door het onderzoek aan het dinucleotide beter interpreteerbaar. Verder is het d<pApA> molecuul verwant aan het cyclische dinucleotide r<pGpG> dat als de natuurlijke activator fungeert van het cellulose synthase in Acetobacter xylinum. Middels de bepaling van ^{1}H-^{1}H, ^{1}H-^{31}P en ^{1}H-^{13}C koppelingsconstanten kon de structuur van het d<pApA> worden afgeleid. De conformatie van de suikers in dit cyclische dinucleotide is van het, voor deoxyriboseringen, ongebruikelijke N-type. De 180°-richtingsverandering in de suiker-fosfaat-keten van het d<pApA> wordt veroorzaakt door de torsiehoeken α and ζ, die beiden een gauche$^+$ waarde aannemen. Deze conformatie blijkt ook onderdeel te zijn van de 180°-richtingsverandering die in de bestudeerde haarspeld lussen optreedt (hoofdstuk 4 en 7). Bij lagere temperatuur blijkt het d<pApA> in oplossing een dimeer te vormen, waarbij een nieuwe vorm van basestapeling een rol speelt. Bij hogere temperatuur en/of grote verdunningen gaat het dimeer over in twee monomeren die van conformatie kunnen veranderen.

In hoofdstuk 3 wordt het d<pApA> molecuul bestudeerd met behulp van moleculaire mechanica berekeningen. 'Alle mogelijke' conformaties werden gegenereerd door een systematische bemonstering van de voor dit molecuul toegestane conformatieruimte middels een multi-conformatie analyse. Uit deze verzameling werden de lage-energie-conformaties (conformeren) verkregen met behulp van energie-minimalisatie berekeningen. Het bleek dat de laagste-energie-conformeer een goede beschrijving is van de experimenteel bepaalde structuur. Door vergelijking van de structuren van de lage-energie-conformeren werd een beter inzicht verkregen in de conformatie-eigenschappen van het molecuul. Met name het voorkomen van de deoxyribosesuikers in de N-toestand in de laagste-energie-structuur kon hierdoor worden begrepen.

Hoofdstuk 4 beschrijft NMR studies aan de haarspeldstructuur gevormd door de d(ATCCTA-TTTT-TAGGAT) sequentie. 2-dimensionale NMR-experimenten die aan deze haarspeld (ca 160 protonen) werden uitgevoerd maakten het mogelijk de resonanties van het ^{1}H-spectrum vrijwel volledig toe te kennen. Op basis van een sequentiële analyse van de kruispieken in het NOESY-spectrum (2D-NOE spectrum) en een analyse van koppelingsconstanten kon geconcludeerd worden dat de stam van de haarspeldstructuur kan worden gekenmerkt als een B-type dubbelstrengshelix. De stapeling van de basen in de stam wordt voortgezet in de lus. De B-type stapeling wordt aan het 3'-einde van één van de ketens van de stam in de lus met drie nucleotiden verlengd. De suiker-fosfaat-keten verandert tussen het derde en vierde nucleotide in de lus van richting. De vierde base draait naar binnen en vormt een T-T paar met de eerste base van de lus. De conformatie die zorg draagt voor de 180°-richtingsverandering welke fosfaatketen ondergaat tussen het derde en vierde residu van de lus, werd met behulp van multi-conformatie analyse bepaald. De structuur werd ook onderzocht met moleculaire mechanica berekeningen. Coördinaten, welke via deze berekeningen werden verkregen voor de laagste-energie-structuur, werden gebruikt voor een simulatie van het NOESY-spectrum en dit resulteerde in een kwalitatieve overeenkomst tussen experiment en simulatie.

In hoofdstuk 5 wordt het vouwingsprincipe voor haarspeldlussen dat voor het eerst werd voorgesteld door Haasnoot e.a. (J. Biomol. Struct. Dyns. (1986), 3, 843-857) verder uitgewerkt. Voornoemd principe werd geformuleerd op basis van het waargenomen verschil in stabiliteit van DNA- en RNA-haarspeldmoleculen als functie van hun lus grootte. Op grond van de geometrische eigenschappen van A-RNA versus B-DNA is het 'voordeliger' om het 5'-einde van de A-type helix te verlengen met ongeveer vijf nucleotiden. De overgebleven ruimte die naar het 3'-einde van de stam overbrugd moet worden kan dan gemakkelijk gesloten worden met twee nucleotiden. Voor een B-type helix is het 'voordeliger' om het 3'-einde te verlengen met twee of drie nucleotiden. De ontstane ruimte naar het 5'-einde van de stam is dan eenvoudig te overbruggen met één of twee nucleotiden. Dit principe, waarvan de geldigheid kon worden ondersteund middels de voor het hexadecanucleotide d(ATCCTA-TTTT-TAGGAT) gevonden structuur (hoofdstuk 4), blijkt op de basis van deze zelfde structuur uitbreiding te behoeven. De vorming binnen de lus structuur van een T-T wobble paar deed de vraag rijzen of dit paar ook vervangen kon worden door een complementaire C-G of A-T combinatie. De daartoe in dit hoofdstuk uitgevoerde moleculaire mechanica berekeningen suggereren dat baseparing in een lus van vier nucleotiden inderdaad mogelijk is voor de lus sequenties -CTTG- en -TTTA- maar niet voor de lus sequenties -GTTC- en -ATTT-.

Deze resultaten legden de basis voor de experimenten beschreven in hoofdstuk 6. De haarspelden met bovengenoemde lussequenties werden gesynthetiseerd en onderworpen aan UV-smeltstudies en NMR-experimenten. Het bleek dat haarspelden met lussequenties -CTTG- en -TTTA- significant stabieler zijn dan die met de lussequentie -GTTC- en -ATTT-. Met behulp van NMR experimenten kon aangetoond worden dat voor de -CTTG-en -TTTA- lussequenties een extra basepaar gevormd wordt. Voor de -GTTC- en -ATTT- werden daarvoor geen aanwijzingen gevonden. Tevens werd de invloed van de basepaarvolgorde in de stam en vervanging van de twee centrale T's in de lus door A's bestudeerd. 2D-NMR experimenten uitgevoerd aan de haarspeld gevormd door d(ATCCTA-TTTA-TAGGAT) leidden tot de ontdekking dat het extra T-A base paar dat gevormd wordt binnen de lus van vier, de ongebruikelijke Hoogsteen-configuratie heeft.

De conformatie-analyse van nucleïnezuren met behulp van NMR is gecompliceerd. Eerst moeten de experimentele data, d.w.z. NOE-intensiteiten en J-koppelingsconstanten, vertaald worden naar H-H afstanden en torsiehoeken en vervolgens moeten deze gegevens vertaald worden naar een structuur. In hoofdstuk 7 wordt een methode beschreven waarmee de (lus)structuur van de haarspeld gevormd door d(ATCCTA-TTTA-TAGGAT) kon worden afgeleid. In deze procedure werden eerst de suikerconformaties en de glycosidische torsiehoeken afgeleid middels multi-conformatie berekeningen, waarmee gebruik gemaakt werd van beschikbare J-koppelingen en NOE-intensiteiten. Intra- en inter-nucleotide proton-proton afstanden werden bepaald door gebruik te maken van een relaxatie matrix berekening zonder dat daarbij een moleculair model behoefde te worden geïntroduceerd. De torsiehoeken β, γ and ε werden voor zover mogelijk bepaald uit homo- en hetero-nucleaire koppelingsconstanten. De structuur van de lus werd vervolgens afgeleid met behulp van multi-conformatie analyse. De haarspeld gevormd door d(ATCCTA-TTTA-TAGGAT) blijkt een B-type stam te hebben, de eerste en vierde base van de lus vormen een Hoogsteen-paar, dat gestapeld is op het laatste basepaar van de stam (hoofdstuk 6). De tweede base van de lus vouwt terug in de 'minor groove' van de B-type stam; de derde is deels gestapeld op de eerste en deels op de vierde base in de lus. De suiker-fosfaat-keten verandert tussen het derde en vierde residu van de lus 180° van richting. Vervolgens werd een vergelijking uitgevoerd van de structuur van de haarspelden gevormd door d(ATCCTA-TTTT-TAGGAT) en d(ATCCTA-TTTA-TAGGAT). Ondanks het feit dat het stapelingspatroon van de basen in de lus verschilt, hebben ze een aantal belangrijke structurele overeenkomsten. In beide haarspelden vervolgt de fosfaatketen in de lus zijn weg in de 3'-richting volgens het patroon van een B-type helix, zij het soms met locale afwijkingen. Na het derde residu maakt de fosfaatketen een scherpe 180°-draai en de resterende ruimte wordt gesloten door één nucleotide. De conformatie van het gebied waarin de richtingsverandering van de keten plaatsvindt is een weerspiegeling van het base-stapelingspatroon in de lus. De tot nu toe beschikbare resultaten suggereren dat de vouwing in deze DNA- haarspelden met behulp van relatief eenvoudige regels kan worden beschreven.

CURRICULUM VITAE

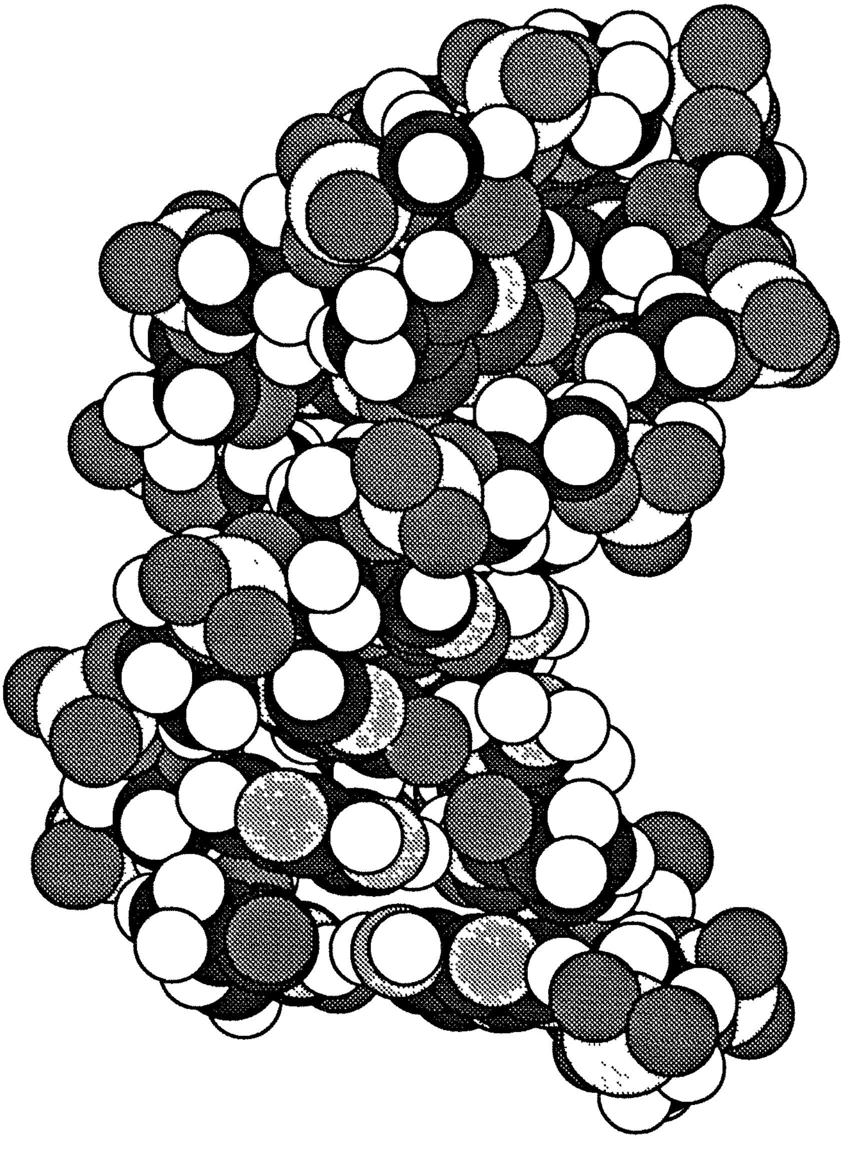
Marcel J.J. Blommers werd op 2 maart 1957 geboren te Amersfoort. Hij behaalde het diploma Atheneum-B in 1976 aan het Eemland college in Amersfoort en begon datzelfde jaar met de opleiding tot analist. Hij werkte tijdens het studiejaar 1978-1979 als stagiair bij Organon Int. te Oss en behaalde het HBO-B diploma Biochemie in 1979 aan de Hogere en Middelbare Laboratoriumschool in Oss. In 1979 begon hij met de studie scheikunde aan de Katholieke Universiteit te Nijmegen. De kandidaats studie werd gedurende 14 maanden onderbroken in verband met het vervullen van de militaire dienstplicht. Het kandidaats examen S3 (scheikunde met natuurkunde) werd in 1983 afgelegd. Zijn doctoraal studie (oude stijl) omvatte de hoofdvakken Biochemie (prof. H. Bloemendal en prof. H.J. Hoenders) en Biofysische Chemie (prof. C.W. Hilbers) aangevuld met capita in de Analytische Chemie (prof. G. Kateman). Het doctoraal diploma werd behaald in september 1985 met het predicaat "met veel genoegen". Aansluitend trad hij in dienst van de faculteit Wiskunde en Natuurwetenschappen te Nijmegen. Gedurende vier jaar was hij werkzaam als wetenschappelijk assistent op de afdeling Biofysische Chemie waar het in dit proefschrift beschreven onderzoek werd verricht. Hij gaf werkcolleges Fasenleer en Kinetiek, was betokken bij tweede jaars practica en de college-cyclus Computers in de Chemie, en hij begeleidde doctoraal studenten. In april 1988 bezocht hij de conferentie getiteld "Colloquium on computer simulations in protein engineering and drug design." in Amsterdam. In augustus 1988 bezocht hij, op kosten van Shell, de conferentie "XIII international conference on magnetic resonance in biological systems" in Madison, Wisconsin, U.S.A. Een deel van de in dit proefschrift beschreven onderzoek werd gepresenteerd in lezingen en posters in nationale en internationale congressen en in de vorm van wetenschappelijke publicaties. Hij is vanaf december 1989 werkzaam als toegevoegd onderzoeker bij de afdeling NMR spectroscopie, onderdeel van het Bijvoet centrum voor biomoleculaire research, van de Rijks Universiteit te Utrecht. Marcel Blommers trouwde in december 1982 met Mieke Friederichs en in juli 1988 werd hun dochter Marloes geboren.

NAWOORD

Op deze plaats wil ik iedereen bedanken die aan het tot stand komen van dit proefschrift heeft bijgedragen. Allereerst wil ik Jan Aelen bedanken voor zijn bijdrage in het uitvoeren van de smeltmetingen. De doctoraal studenten Harrie Marsman, Els Heijmen, Jos Tissen, Leon Overgaauw, Geert-Jan Willems, Rik Leenders, Niels van der Stappen, Frank Leemhuis en Erik Esveld dank ik voor hun inzet. Voor de synthese van de in dit proefschrift beschreven DNA moleculen verdienen Hans van den Elst, Gijs van der Marel en Jacques van Boom bijzondere vermelding. Jos Joordens en Sybren Wijmenga zorgden voor een optimaal functioneren van de 400, 500 en 600 MHz spectrometers. Hilbert Bruins Slot en Ad Thiers bedank ik voor de ondersteuning van het rekenwerk op de VAX van het CAOS-CAMM centrum. Van Theo Bosman ontving ik bereidwillige hulp bij het IBM mainframe gebeuren. Met Carlos Lucasius vond een aangename interactie plaats tussen chemometrie en NMR. Voor hun stimulerende wetenschappelijke ondersteuning noem ik Hans Walters, Kees Haasnoot en Frank van der Ven. Waardevol was de 'academische vorming' van mijn promotor. Babs Klink, Nelly Kersten en John Roef stonden altijd voor mij klaar voor het uitvoeren van haast-klussen. Goede herinneringen heb ik aan de prettige samenwerking met mijn collega's Evert de Jong, Henk Bulsink, Marcel Akkerman, John van Duynhoven, Margret Mooren en Paul Folkers. De illustratie op de omslag werd verzorgd door Joost van Opheusden. Verder wil ik mijn ouders en schoonouders bedanken voor hun materiële en morele steun. Mieke is een onmisbare steun geweest in de afgelopen jaren.

BRUKER

FT-NMR
FT-IR
FT-MS

Your partner in science