Journal of Biomolecular NMR, 6 (1995) 403–422
ESCOM

J-Bio NMR 313

Conformation of the circular dumbbell d<pCGC-TT-GCG-TT>: Structure determination and molecular dynamics

Johannes H. Ippel[a,*], Virginia Lanzotti[b], Aldo Galeone[c], Luciano Mayol[c], Jacquelien E. van den Boogaart[a], Jeroen A. Pikkemaat[a] and Cornelis Altona[a,**]

[a]*Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands*
[b]*Dipartimento STAAM, Università del Molise, Via Tiberio 21, I-86100 Campobasso, Italy*
[c]*Dipartimento di Chimica Sostanze Naturali, Università di Napoli Federico II, Via D. Montesano 49, I-80131 Naples, Italy*

Received 12 May 1995
Accepted 5 September 1995

Keywords: DNA; Conformational analysis; Mini-hairpin; Two-base loop

Summary

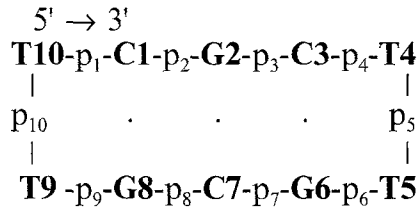
The circular DNA decamer 5'-d<pCGC-TT-GCG-TT>-3' was studied in solution by means of NMR spectroscopy and molecular dynamics in H_2O. At a temperature of 269 K, a 50/50 mixture of two dumbbell structures (denoted L2L2 and L2L4) is present. The L2L2 form contains three Watson–Crick C-G base pairs and two two-residue loops in opposite parts of the molecule. On raising the temperature from 269 K to 314 K, the L2L4 conformer becomes increasingly dominant (95% at 314 K). This conformer has a partially disrupted G(*anti*)-C(*syn*) closing base pair in the 5'-GTTC-3' loop with only one remaining (solvent-accessible) hydrogen bond between NH^{α} of the cytosine dC(1) and O6 of the guanine dG(8). The opposite 5'-CTTG-3' loop remains stable. The two conformers occur in slow equilibrium (rate constant 2–20 s^{-1}). Structure determination of the L2L2 and L2L4 forms was performed with the aid of a full relaxation matrix approach (IRMA) in combination with restrained MD. Torsional information was obtained from coupling constants. Coupling constant analysis ($^3J_{HH}$, $^3J_{HP}$, $^3J_{CP}$) gave detailed information about the local geometry around backbone torsion angles β, γ, δ, and ε, revealing a relatively high flexibility of the 5'-GTTC-3' loop. The values of the coupling constants are virtually temperature-independent. 'Weakly constrained' molecular dynamics in solvent was used to sample the conformational space of the dumbbell. The relaxation matrices from the MD simulation were averaged over $\langle r^{-3} \rangle$ to predict dynamic NOE volumes. In order to account for the 1:1 conformational mixture of L2L2 and L2L4 present at 271 K, we also included S^2 factors and $\langle r^{-6} \rangle$ averaging of the $\langle r^{-3} \rangle$-averaged relaxation matrices. On matrix averaging, the agreement of NOE volumes with experiment improved significantly for protons located in the thermodynamically less stable 5'-GTTC-3' loop. The difference in stability of the 5'-CTTG-3' and 5'-GTTC-3' loops is mainly caused by differences in the number of potential hydrogen bonds in the minor groove and differences in stacking overlap of the base pairs closing the minihairpin loops. The *syn* conformation for dC(1), favored at high temperature, is stabilized by solvation in the major groove. However, the conformational properties of the dC(1) base, as deduced from R-factor analysis and MD simulations, include a large flexibility about torsion angle χ.

Introduction

Recently it has been shown that DNA sequences known to form two-base loops in vitro also form tight loops in vivo (Davison and Leach, 1994). This ability may play important roles in gene regulation, recombination or mutagenesis.

To understand the role of base-sequence effects in loop folding in DNA minihairpins, we have studied the conformational aspects of the base pair reversal C-G to G-C in the closing base pair of the circular dumbbell d<pCGC-TT-GCG-TT> (see Scheme 1). In this dumbbell two types of loop are present, containing a 5'-CTTG-3' and a 5'-GTTC-3' sequence, respectively. According to the litera-

*Present address: Department of Biophysical Chemistry, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands.
**To whom correspondence should be addressed.



Scheme 1. Schematic structure of the circular dumbbell **I**.

ture (for an overview, see Hilbers et al., 1994), the thermodynamic stability of the hairpins with a G-C closing base pair is decreased relative to hairpins with the reversed combination C-G. Moreover, contradictory results have been reported on the loop size of hairpins with a closing G-C base pair (Ippel et al., 1995a).

We already have shown (Ippel et al., 1992,1995a,b) that the circular dumbbell d<pCGC-TT-GCG-TT> adopts two dominant conformers, which interconvert slowly (rate constant 2–20 s^{-1}) on the NMR chemical shift time scale. At a temperature of ≤269 K, the two conformers are present in a 50/50 mixture (denoted L2L2 and L2L4). The L2L2 form contains three Watson–Crick C-G base pairs and two two-residue loops (H2-family type) in opposite parts of the molecule. The conformer favored at high temperature (denoted L2L4) consists of two Watson–Crick base pairs, a loop of four residues at the 5'-GTTC-3' side and a loop of two residues at the 5'-CTTG-3' side (95% at 314 K). The 5'-GTTC-3' loop in this L2L4 form actually consists of a rather unstable G-C base pair, involving one hydrogen bond between the amino proton of dC(1) and O6 of dG(8), and a *syn* orientation of the dC(1) base. Although the exchange is slow, no doubling of resonances is found, except for the dG(8) imino proton, the two signals of which resonate at 12.3 ppm (L2L4) and 13.3 ppm (L2L2) at 269 K. Instead, line broadening at temperatures below 290 K is observed for some nonexchangeable protons of dC(1), T(9) and T(10), for the ^{31}P resonances of P(2) and P(10) and for the C1' ^{13}C resonance of dC(1). This broadening is indicative of local conformational changes in the 5'-GTTC-3' loop. Complete band-shape (CBS) calculations on signals of the dG(8) imino proton, together with chemical shift considerations, indicate multiconformational states in the *syn* domain of the dC(1) base. The rate constant of conversion between two major hypothetical sites in the L2L4 form is larger compared to that of the *anti–syn* conversion, and is estimated to be 600 s^{-1} at 269 K and 4300 s^{-1} at 295 K.

From a thermodynamic analysis (Ippel et al., 1995b) it was concluded that the thermal stability of the 5'-GTTC-3' loop in the L2L4 form is 4°–11° less than that of the 5'-CTTG-3' loop ($T_m = 352 \pm 2$ K). The transition from the L2L2 to the L2L4 form involves an enthalpy change of +25 to +30 kJ/mol, which value agrees with the calculated enthalpy difference at the T_m point in melting of the 5'-CTTG-3' loop ($\Delta H^0 = -89$ kJ/mol) relative to the melting of the thymidine residues in the 5'-GTTC-3' loop ($\Delta H^0 = -61$ kJ/mol).

In this paper, structures of the L2L2 and L2L4 forms are presented. Quantitative proton–proton distances obtained by means of the IRMA method and torsional information from coupling constants are used in the structure refinement. 'Weakly constrained' MD (wMD) with inclusion of solvent is used to probe the structural stability of the two forms. Moreover, the trajectory files of the wMD simulations allow us to recalculate the NOESY spectrum of a mixture of the L2L2 and L2L4 forms. Three factors are used to account for mobility in the molecule, i.e., S^2 factors, $\langle r^{-3} \rangle$ averaging and $\langle r^{-6} \rangle$ averaging of the relaxation matrices. Finally, a comparison is made between the geometries of the two types of loop, 5'-CTTG-3' and 5'-GTTC-3'.

Materials and Methods

NMR spectroscopy

The assignment of the ^1H, ^{13}C and ^{31}P resonances of the circular dumbbell (**I**) and sample conditions used during the experiments have been described by Ippel et al. (1992,1995a). Scheme 1 displays the schematic structure and numbering of the dumbbell.

Coupling constants

A 2D ^{13}C-^1H HSQC experiment (Bodenhausen and Ruben, 1980), recorded at 600 MHz at 298 K, was used to extract carbon–phosphorus and proton–phosphorus coupling constants, by using extensive folding in the ^{13}C domain (Schmieder et al., 1992). A total of 2173 experiments (averaged over 24 scans) of 2K data points were collected. The spectral widths in the ^1H (t_2) and ^{13}C (t_1) directions were 6250 and 3012 Hz, respectively. The offset in t_1 was placed at the center of the C1' carbon resonances. Carbon C5' resonances were folded once; C2' resonances were folded twice in the t_1 direction. The method described by Schmieder et al. (1991) was used to ensure pure absorption phase in t_1. Prior to Fourier transformation, apodization with a $\pi/2$-shifted sine-square bell in both dimensions was used, followed by zero-filling to 4K in the t_1 and t_2 domains.

Abbreviations: NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement and exchange spectroscopy; HMQC, heteronuclear multiple quantum coherence; HSQC, heteronuclear single quantum coherence; ZQ, zero quantum; IRMA, iterative relaxation matrix approach; rMM, restrained molecular mechanics; rMD, restrained molecular dynamics; wMD, weakly constrained molecular dynamics; vMD, weakly constrained molecular dynamics in vacuo; CBS, complete band-shape method; DSS, 3-trimethyl-silyl-1-propane sulfonic acid, sodium salt; T_m, melting midpoint; ppm, parts per million; ^{a}C, arabinofuranosyl-cytidine; mini-hairpin, hairpin with a two-residue loop.

Coupling constants involving H1', H2', H2", and H3' protons of **I** were obtained from 1D spectra recorded at 600 and 500 MHz, and simulated by means of a noniterative LAOCOON (Bothner-By and Castellano, 1964) spin-simulation program (maximum of eight nonequivalent spins) on an IBM mainframe computer. Coupling constants ^{3}J(H4'-H5") were obtained from slices in 2D HOHAHA spectra. Additionally, coupling constants involving H4', H5' and H5" to phosphorus were obtained by 2D simulation with the SPHINX/LINSHA program (Widmer and Wüthrich, 1986), and by means of 2D curve fitting of cross peaks in the ^{13}C-^{1}H HSQC experiment with the NMR2 program (New Methods Research Inc., Syracuse, NY). Least-squares fitting of the rotamer populations around torsion angles β and ε was done with the FORTRAN program TRISTAT (J. van Wijk, unpublished work). Pseudorotation parameters for sugar puckers were calculated by the PSEUROT 5.4 program (De Leeuw and Altona, 1983; Van Wijk et al., 1992).

Temperature conditions for NOESY buildup series

Exchangeable proton spectra of the circular dumbbell at different temperatures show that at >314 K an almost pure L2L4 form is present. At the lowest temperature reached (268 K), at best a 51/49 ratio of L2L2 and L2L4 is feasible. We have attempted to collect high-temperature NOESY spectra at 600 MHz with various mixing times to separate the contribution of the L2L4 conformer. Unfortunately, long-term temperature instability of the spectrometer, combined with the extremely small line width for the proton signals in the temperature range around 314 K, resulted in relatively strong t_1 noise, which becomes persistent at mixing times < 400 ms. Although peak volumes from NOESY spectra recorded at short mixing times are essential to the IRMA method, accurate determination of these peak volumes was therefore not possible at 314 K. Instead, a good-quality buildup series of NOESY spectra was collected at 271 K (4 mM DNA). At this temperature a mixture of about 49/51 L2L2/L2L4 is present. We already have shown that, except for the local conformation around the dC(1) base, chemical shifts and NOE intensities are very similar for both species (Ippel et al., 1995a). Thus, at first sight it appears plausible to derive geometries from this conformational mixture.

Buildup series of NOESY spectra

The buildup series of NOESY spectra was recorded at 271 K at 500 MHz. The mixing times were 50, 80, 130, 210 and 340 ms. ZQ-coherences were suppressed by random variation of the mixing times by 4 ms (Macura et al., 1981; Rance et al., 1985). For each spectrum, 388 experiments, averaged over 80 scans, were collected in 2K TPPI (Marion and Wüthrich, 1983) data points. The spectral width was 4800 Hz and the relaxation delay was 1.0 s. Spectra were identically processed on a Bruker X-32 workstation. The t_2 domain was zero-filled to 2K real points and multiplied with a π/2-shifted sine function; in the t_1 domain, zero-filling to 1K real points and multiplication with a π/3-shifted sine function was employed. After 2D Fourier transformation, the base line was corrected with a third-order polynomial in both the f_1 and f_2 directions. For the integration of near-diagonal peaks, the base line was specifically corrected over the entire spectral range by a second-order polynomial. Peak volumes were obtained by integration of nonoverlapping peaks. In case of overlap, relative integrals were derived from 2D curve fitting with the NMR2 program. The final list of integrated NOE volumes consists of 200 proton–proton distances, selected by the criterion that the integrals must be positive in the spectra recorded at the three highest mixing times.

Model building

The crude starting model of the circular dumbbell was constructed from two H2 family hairpins with the aid of the program MacroModel (Columbia University, New York, NY) on an Evans & Sutherland PS-350 graphics system, interfaced to a VAX 11/750.

Structure refinement by means of IRMA and rMD

To extract accurate proton–proton distances from 2D NOE spectra, we used an iterative relaxation matrix approach (IRMA), in which the effect of spin diffusion is corrected for (Boelens et al., 1988,1989). The IRMA procedure (v. 2.0) has been described in detail previously (Koning et al., 1990,1991a,b; Pieters et al., 1990). The main principle of IRMA is the backcalculation of a mixed NOE matrix. The mixed NOE matrix is a combination of experimental NOEs and calculated ones, derived from a starting structural model. Distances obtained from the mixed NOE matrix were used in refinement of the dumbbell structure by means of restrained molecular dynamics (rMD) calculations. For the IRMA calculation, the correlation time for the overall tumbling was taken to be 1.3 ns (1.9 ns with the inclusion of S^2 factors) and the diagonal leakage rate was 1.0 s^{-1}. For the methyl groups, fast rotation was taken into account by using an $\langle r^{-3} \rangle$ averaging for the three methyl proton positions. Furthermore, due to the fast rotation an extra diagonal leakage rate was added (Koning et al., 1991a,b). S^2 factors were introduced at an average value of 0.8. The S^2 factors for the intraresidue H6-H5 (0.92), H6-CH_3 (0.89) and H8/H6-H1' (0.90) proton pairs, and for the interresidue H8/H6(n)-H2'(n−1) (0.67) and H8/H6(n)-H2"(n−1) (0.60) pairs were taken from the literature (Koning et al., 1991a,b).

The estimated upper and lower bounds for the 200 distances were based on experimental error and were directly used in an rMD structure-refinement protocol. Distances involving the methyl groups were constrained

on the central methyl carbon atom, with a pseudo-atom correction of ±0.3 Å outside the upper and lower bounds calculated by IRMA (Koning et al., 1991a,b). In the rMD calculations the distances were entered via an extra harmonic pseudopotential, defined as described in the literature (Clore et al., 1985; Kaptein et al., 1985). The force constant K_{dc} for the distance constraints was set at 40 kJ/mol Å^2 in the first cycle and at 8 kJ/mol Å^2 in subsequent cycles. To keep the Watson–Crick base pairs intact, a pseudorestraint potential (1.7–2.0 Å) was added on the hydrogen bonds (8 kJ/mol Å^2). The dC(1)-dG(8) base pair in the L2L4 form was held together only by means of a weak constraint (<2.5 Å) on the NH^{β} dC(1)-O6 dG(8) hydrogen bond. The absence of a NOE between Me T(4) and Me T(9) at all mixing times was added to the constraint list as an anti-distance constraint of >5 Å. Observed NOEs from H2'/H2" T(4) and Me/H6 T(5) to the imino proton of dG(6) were put in as a constraint <5 Å.

Methodology of the rMD simulations

The rMD simulations were done in vacuo by means of the all-atom version of AMBER3.0a (Weiner et al., 1986; Seibel, 1989). The fully anionic phosphate charges were partially shielded by large (5 Å diameter) 'hydrated' Na^+ counterions, placed at a distance of 6 Å from the phosphorus atom at the bisectors of the PO_2^- groups (Singh et al., 1985). The rMD calculations were performed in the following way:

(1) 500 cycles of conjugate-gradient restrained energy minimization;

(2) heating stage from 0 to 300 K during 10 ps in 10 steps of 30 K intervals;

(3) equilibration during 30 ps at 300 K (except in the first cycle of the refinement: 20 ps at 300 K, 40 ps at 500 K and 20 ps at 300 K);

(4) data accumulation during 10 ps at 300 K; and

(5) coordinate averaging of the final 10 ps, followed by 500 cycles of conjugate-gradient restrained energy minimization.

Further computational details were as follows: a distance-dependent dielectric permittivity $1/\varepsilon r$ ($\varepsilon = 1$) was used; the nonbonded cutoff distance was 12 Å; updates for nonbonded interactions were carried out every 0.05 ps; overall rotational and translational motions about the center of mass were removed every 100 steps; a time step of 0.001 ps was used; and coordinates were saved every 0.1 ps. Initial velocities at the start of the warming up stage were calculated from the forces on the atoms. During the rMD runs, temperatures were maintained by coupling to a thermal bath, with a time constant $\tau = 0.6$ ps at the heating steps and 0.2 ps later on (Berendsen et al., 1984). The ensemble of structures in the range *anti* to high-*anti* and *syn* was generated by forcing, through torsional constraints on χ, the dC(1) base in L2L2 from an *anti* to a *syn* position and vice versa in L2L4 from *syn* to *anti*, during an in vacuo rMD run. The remainder of the structure was held rigid by means of the experimentally derived distance constraints obtained for both the L2L2 and L2L4 rMD models. All calculations were performed on a Silicon Graphics IRIS 4D35G workstation.

'Weakly constrained' molecular dynamics in water

'Weakly constrained' molecular dynamics simulations of the dumbbell in water (wMD) and in vacuo (vMD) were performed at 270 K, with an arbitrarily chosen force constant K_{dc} of 0.4 kJ/mol Å^2 on the distance constraints. Harmonic torsional constraints ($E_{tc} = 0.4$ kJ/mol) were set on β^+ torsion angles of residues dC(1), dC(3) and dG(6), and on γ^{tr} torsion angles of residues dC(1), dC(3), dG(6) and dG(8).

The molecule was immersed in a rectangular box of Monte Carlo TIP3P water molecules. Solvation was based on distance criteria; water molecules with an oxygen closer than 1.8 Å or a hydrogen closer than 1.5 Å to the solute were excluded from the system. The electronic charge on the water oxygen was taken as –0.834. Sodium counterions were placed at a distance of 2.5 Å from the negatively charged OA oxygen atom of the phosphates. Adapted force field parameters for the non-'hydrated' counterions were taken from the literature (Aqvist, 1990). The equilibrium distance between OA and Na^+ amounted to 2.2 Å. To restrict free diffusion through the solution, counterions were constrained between 1.0 Å and 5.0 Å distance of the phosphorus atom by means of a harmonic pseudopotential with a force constant of 8.0 kJ/mol Å^2. The system contained 3483 atoms for the L2L2 form and 3399 atoms for the L2L4 form. Box dimensions were approximately $37 \times 34 \times 29$ Å for both systems.

Constant volume and periodic boundary conditions with solute–solvent imaging were applied for all vMD and wMD simulations. Computational details were essentially the same in the rMD calculations, except for the following changes. Bonds involving hydrogen were constrained by the SHAKE algorithm (Van Gunsteren and Berendsen, 1977). The geometrical tolerance of the hydrogen-bond length was set to 0.0003 Å. The residue-based cutoff distance for nonbonded solute–solvent and solvent–solvent interactions was set to 8 Å. All intramolecular solute–solute nonbonded interactions were calculated, with the solute–solute imaging turned off. This was done to reduce electrostatic problems of highly charged DNA molecules in water. A constant dielectric permittivity, $\varepsilon = 1$, was used; the nonbonded pair list was updated every step. To ensure kinetic energy conservation, rotational and translational motions were not removed during the simulation. Before the start of the wMD simulations, 100 cycles of steepest descent followed by 400 cycles of conjugate-gradient restrained energy minimization were performed. Next, the system was heated during 10 ps to a temperature of 270 K in steps of 30 K. The lengths of

the simulation periods at 270 K were 99.6 ps for L2L2, 59.2 ps for L2L4 with χ of dC(1) unconstrained, and 60.0 ps for L2L4 with χ of dC(1) constrained. In the L2L4χ-constr wMD simulations, the *syn* orientation around torsion angle χ of the glycosydic bond was kept rigid by means of two harmonic torsional constraints (E_{tc}=420 kJ/mol) at torsion angles O4'-C1'-N1-C6 and O4'-C1'-N1-O2 of dC(1). Equilibrium of potential energy and hydrogen-bond energy was reached after approximately 20 ps (including the heating stage).

Analysis of structures

Analysis of torsion angles, bond angles and distances was performed by the TORSMON program, written in FORTRAN77 (Van Beuzekom, 1989). Twist and helix parameters were calculated with the 'lines and planes' routine in the QUANTA/CHARMm program (v. 3.0; Polygen Corp., Waltham, MA), with the program CURVES (Lavery and Sklenar, 1988) and with the NEWHELIX91 program (R.E. Dickerson, private communication). The quality of the fit between theoretically calculated, A_{ij}^{th}, and experimental, A_{ij}^{exp}, NOE intensities is expressed by the R-factor defined by Gonzalez et al. (1991).

Simulation of NOESY spectra

For the simulation of the NOESY spectra, intensities were taken from the theoretical NOE matrices calculated by IRMA (v. 2.0). The simulated spectra themselves were calculated with the aid of an earlier version (1.0) of IRMA and processed by the 2DNMR program (Utrecht University, v. 851204) on a VAX 11/750. Lines were assumed to have a pure Gaussian form. The line widths of the cross peaks in the simulated 2D spectra were taken according to experiment, and varied between 7 Hz (Me) and 20 Hz (H5'/H5").

The recommended IUPAC nomenclature for the torsion angles, helix parameters, and numbering of the residues is used throughout this work (IUPAC-IUB, 1983; EMBO, 1989).

Results

Coupling constant analysis

Coupling constants were derived from several NMR experiments in order to monitor the backbone geometry and sugar conformation of the residues in the circular dumbbell.

Conformation around δ (C5'-C4'-C3'-O3')

Proton–proton coupling constants of the sugar rings in the temperature range 309–314 K are reported in Table 1. ^{3}J(H1'H2') and ^{3}J(H1'H2") remain constant within 0.4 Hz between 290–313 K, and are consistent with the presence of predominantly S-type sugars in the dumbbell (75–94% at 313 K). Severe overlap in the H2'/H2" spectral region prevents determination of coupling constants ^{3}J(H2'H3') and ^{3}J(H2"H3') of most residues. For residues dG(2), T(5), dG(6), dC(7), dG(8) and T(9), additional coupling constants ^{3}J(H2'H3'), ^{3}J(H2"H3'), or ^{3}J(H3'H4') could be obtained with sufficient accuracy. This allows calculation of the phase angle (P_S) and the population of S-type sugar (%S) for these residues. The backbone torsion angle δ (C5'-C4'-C3'-O3') is directly related to the pseudorotation parameters P and Φ_M (Altona, 1982; Van Wijk et al., 1992). Results of the pseudorotation analysis are collected in Table 2.

The experimental δ values range from 140° to 147°, with the exception of δ dC(7). This sugar exhibits a larger flexibility (%S=75%), combined with a low P_S angle (117°). The values of δ for T(5) and dC(7) are in good agreement with earlier observations on a related H2-family hairpin (Pieters et al., 1990) and suggest specific base-sequence-dependent behavior.

Backbone torsion angles

Coupling constants involving the phosphate backbone in DNA are useful for detailed conformational analysis of the dumbbell. Of the five torsion angles in the backbone, i.e., α, β, γ, ε, and ζ, only torsion angles β, γ, and ε are

TABLE 1
^{1}H-^{1}H COUPLING CONSTANTS OF THE CIRCULAR DUMBBELL **I**

Residue	H1'H2'	H1'H2"	H2'H2"	H2'H3'	H2"H3'	H3'H4'	H4'H5'	H4'H5"	H5'H5"
dC(1)	9.3	5.9	–	–	–	(2)	(2)	7.3	–11.4
dG(2)	9.6	5.0	–14.3	5.2	–	(2)	–	–	–
dC(3)	10.3	4.6	–	–	–	1.5	(2)	6.9	–11.2
T(4)	9.6	5.0	–	–	–	(2)	(2)[a]	(2)[a]	–11.2[a]
T(5)	9.4	5.5	–13.7	5.4	1.5	(2)	(2)[a]	(2)[a]	–11.4
dG(6)	9.6	5.6	–14.5	–	1.6	1.6	2.2[a]	9.2	–11.2
dC(7)	7.9	6.2	–14.1	7.3	–	–	2.2[a]	4.3[a]	–11.2[a]
dG(8)	10.0	5.2	–13.7	–	1.5	(2)	–	–	–
T(9)	8.5	5.8	–13.8	–	2.3	(2)	(2)[a]	(2)[a]	–11.2[a]
T(10)	8.6	5.8	–	–	–	–	(2)[a]	(2)[a]	–11.2

Couplings (Hz) were obtained from a computer simulation of a 600 MHz 1D spectrum recorded at 313 K and from a 2D HOHAHA experiment recorded at 309 K. Estimated accuracy ±0.2 Hz. Values between brackets: accuracy ±1 Hz.

[a] Couplings obtained from a 2D simulation of cross peaks in the 2D ^{31}P-^{1}H correlated spectrum at 294 K. Estimated accuracy ±0.5 Hz.

TABLE 2
CONFORMATIONAL ANALYSIS OF TORSION ANGLES δ AND γ IN THE CIRCULAR DUMBBELL I

Residue	δ				γ	
	%S (Σ1')	%S (P5.4)	P_S	δ_S	$p\gamma^+$	$p\gamma^{tr}$
dC(1)	92	–	–	–	0.42	0.58
dG(2)	83	93[a]	161[a]	143	1.0[b]	0.0[b]
dC(3)	87	–	–	–	0.46	0.54
T(4)	83	–	–	–	0.8–1.0	0.0–0.2
T(5)	87	91	165	146	0.8–1.0	0.0–0.2
dG(6)	92	91	161	143	0.18	0.82
dC(7)	75	75	117	115	0.73	0.27
dG(8)	92	94	155	140	1.0[b]	0.0[b]
T(9)	78	83	169	147	0.8–1.0	0.0–0.2
T(10)	79	–	–	–	0.8–1.0	0.0–0.2

Angles were obtained from experimental ^1H-^1H coupling constants in the temperature range 309–313 K. Parameters: percentage S-type sugar %S (%), according to the sum rule Σ1' defined by Van Wijk et al. (1992) and calculated by the PSEUROT5.4 program (constrained values: $\Phi_N = 35°$, $\Phi_S = 35°$ and $P_N = 18°$); phase angle of S-type sugar P_S (degrees); torsion angle δ_S (degrees) and population fractions $p\gamma^+$ and $p\gamma^{tr}$ as determined by means of a sum rule (Altona, 1982).

[a] With $\Phi_S = 44°$ (corresponding to Φ_S dG(2) (Pieters et al., 1990)), %S is 84% and P_S is 160°.

[b] The presence of a large ^4J(H4'P) coupling for dG(2) and dG(8) is only possible for a practically pure β^{tr}/γ^+ torsion angle combination.

directly accessible by means of vicinal coupling constants (Altona, 1982). Measured coupling constants involving the backbone are given in Tables 1 and 3. No temperature dependence (271–314 K) in vicinal coupling constants along the backbone larger than 0.5 Hz was found, with the exception of ^3J(H3'P) of T(9), which yields 7.8 Hz at 294 K and 7.2 Hz at 314 K, and ^3J(H5"P) of dG(6), which yields 9.7 Hz at 290 K and 8.7 Hz at 314 K.

Conformation around β (P-O5'-C5'-C4')

Torsion angle β (P-O5'-C5'-C4') is monitored by the vicinal coupling constants ^3J(H5'P), ^3J(H5"P), and ^3J(C4'Pβ) (Altona, 1982). Coupling constants between proton and phosphorus are obtained by 1D simulation of proton spectra and, in cases where heavy overlap exists in the proton spectra, by 2D simulation (SPHINX/LINSHA) of 2D correlated ^{31}P-^1H spectra (Ippel et al., 1995a). To derive simultaneously carbon–phosphorus and proton–phosphorus coupling constants in crowded spectra, we have used a high-resolution 2D ^{13}C-^1H HSQC experiment, as described by Schmieder et al. (1992). When only two couplings are available, the rotamer populations can be calculated by the following modified sum rules ($\Phi\beta^+ = 55°$; $\Phi\beta^{tr} = 175°$) (Lankhorst et al., 1984,1985; Altona, 1982):

$$p\beta^{tr} = (24.6 - ({}^3J(H5'P) + {}^3J(H5"P))) / 19.7 \quad (1)$$

TABLE 3
^{13}C-^{31}P COUPLING CONSTANTS OF THE CIRCULAR DUMBBELL I

Residue	C3'P	C5'P	C4'Pβ	C4'Pε	C2'P	H2'P	H3'P	H4'P	H5'P	H5"P
dC(1)	–5.5	–6.3	7.4	11.3	<3[c]	–	(5)	<2	2.2	7.0
dG(2)	–4.8	–5.6	<18.9>[b]		<3[c]	–	6.0	3.9	<5.3>[d]	
dC(3)	[a]	–5.5	4.7	12.5	<3[c]	–	4.3	<2	3.2	10.9
T(4)	–5.4	–5.6	<17.3>[b]		(3)	–	8.4	(3)	5.0	0.9
T(5)	–7.2	(–5)	8.5	<3	5.5	2.4	8.3	2.3	0.8	7.5
dG(6)	[a]	–6.9	5.6	(12.5)	<3[c]	–	3.5	<2	2.0	9.4
dC(7)	[a]	(–5)	<17.4>[b]		<3[c]	–	(4)	2.7	3.2	6.3
dG(8)	[a]	–5.8	<19.1>[b]		<3[c]	–	(4)	3.2	<3.0>[d]	
T(9)	(–5)	–5.3	11.5	6.5	(4)	–	7.5	(3)	4.5	2.3
T(10)	–4.8	–5.5	<13.6>[b]		4.5	–	6.5	2.6	1.6	4.0

Couplings (Hz) were obtained by means of 2D curve fitting of a 2D ^{13}C-^1H HSQC experiment recorded at 298 K, together with averaged values of ^{31}P-^1H coupling constants (Hz), obtained by 1D simulation of a 500 MHz spectrum at 294 K and a 600 MHz spectrum at 313 K, by 2D curve fitting of a 2D HSQC experiment at 298 K and by 2D simulation of a 2D ^{31}P-^1H correlated spectrum at 294 K. Estimated accuracy for ^{31}P-^1H and ^{13}C-^{31}P coupling constants: ±0.3 Hz and ±0.5 Hz, respectively. Values between brackets: accuracy ±1 Hz.

[a] Resonance of H3' under HDO; ^2J(C3'P) could not be determined.

[b] Sum of ^3J(C4'Pβ) + ^3J(C4'Pε).

[c] Upper limit from the line broadening in the ^{13}C dimension of the HSQC spectrum.

[d] Isochronous H5' and H5" resonances; only an approximate value of the largest coupling between ^3J(H5'P) and ^3J(H5"P) can be estimated, assuming that one coupling constant is about 2.5–3 times larger than the other. When couplings are of similar size, the measured value is approximately equal to $\Sigma({}^3J(H5'P) + {}^3J(H5"P))/2$.

$$p\beta^{tr} = ({}^3J(C4'P) - 1.02)/9.92 \quad (2)$$

$$p\beta^{+} = ({}^3J(H5''P) - 1.8)/21.1 \quad (3)$$

A least-squares fit of the coupling constants yields the experimentally derived rotamer populations and torsion angles Φβ collected in Table 4. Residues dC(1), dC(3), dG(6), and dC(7) display significant amounts of β^+ rotamers ($p\beta^+ = 27\%$ and 40% for dC(1) and dG(6), respectively). Residue dC(3) shows the largest population of β^+ rotamer ($p\beta^+$ 48%), which yields a conformational benefit that will be discussed later.

The remaining residues of the dumbbell display pure β^{tr} rotamers, with a $\Phi\beta^{tr}$ angle around 175° for the dG(2) and dG(8) residues, as expected for B-DNA type duplexes (Altona, 1982; Kennard and Hunter, 1989). The thymidine residues in the two loops behave extraordinarily in this respect. The unresolved $^3J(C4'P\beta)$ coupling for residues T(4) and T(10) can be estimated from a comparison of the $^3J(C4'P\varepsilon)$ and proton–phosphorus couplings. We calculate $\Phi\beta^{tr}$ values of 163°, 209°, 170°, and 193°–198° for residues T(4), T(5), T(9), and T(10), respectively. The estimated uncertainty limits of the derived $\Phi\beta^{tr}$ values are larger for T(9) and T(10) (±5°–10°) than those obtained for the thymidines in the opposite loop (±4°), suggesting more flexibility within the β^{tr} range for T(9) and T(10) of the 5'-GTTC-3' loop relative to the β^{tr} rotamers of T(4) and T(5) in the 5'-CTTG-3' loop.

Conformation around γ (O5'-C5'-C4'-C3')

Backbone torsion angle γ (O5'-C5'-C4'-C3') is monitored by the proton–proton coupling constants $^3J(H4'H5')$ and $^3J(H4'H5'')$ (Altona, 1982). The conformational analysis of this torsion angle is summarized in Table 2. At the sharp turn at the 5' end of dC(1) and dG(6) a γ^{tr} rotamer is preferred; the populations are 58% and 82%, respectively. The ratio of populations of γ rotamers $p\gamma^+/p\gamma^{tr}$ is 73/27 in dC(7) and 46/54 in dC(3), which suggests that the backbone at this point is more flexible than for the dG(2) and dG(8) residues. The latter residues display a practically pure β^{tr}/γ^+ torsion angle combination, as shown by the large four-bond couplings J(H4'P) of 3.9 and 3.2 Hz, respectively, which arise from a favorable planar W-path between H4'(n) and P(n) (Sarma et al., 1973; Altona, 1982). The absence of this type of $^4J(H4'P)$ coupling for residues dC(1), dC(3) and dG(6), as deduced from the 2D ^{13}C-^1H HSQC spectrum and the 2D ^{31}P-^1H correlated spectrum, confirms the relatively high populations of β^+/γ^{tr} geometry for these residues. The $^3J(H4'H5')$ and $^3J(H4'H5'')$ coupling constants of the thymidine residues are less than 3 Hz, and consequently the population of γ^+ is certainly larger than 80% in these residues.

TABLE 4
CONFORMATIONAL ANALYSIS OF TORSION ANGLE β IN THE CIRCULAR DUMBBELL I

Residue	Experimental			Simulated		
	$p\beta^{tr}$	$\Phi(\beta^{tr})$	$\Phi(\beta^+)$	$p\beta^{tr}$	ΦrMD	ΦwMD
dC(1)	0.73±0.07	176	55[a]	0.34	49	63
dG(2)	0.98[b]	175[a]	55[a]	0.99	173	174
dC(3)	0.52±0.11	175[a]	72	0.00	64	68
T(4)	1.00[c]	163	–	1.00	166	167
T(5)	1.00[c]	209	–	1.00	204	190
dG(6)	0.60±0.11	175[a]	51	0.01	69	83
dC(7)	0.77±0.01	175[a]	55[a]	1.00	181	162
dG(8)	1.00[b]	175[a]	55[a]	0.98	175	166
T(9)	1.00[c]	169–170	–	0.99	163	176
T(10)	1.00[c]	193–198	–	1.00	212	194

Parameters: population fraction $p\beta^{tr}$, torsion angle of most abundant rotamer Φ (degrees) from experiment and averaged from the rMD and wMD simulations of L2L4χconstr and L2L4χunconstr.

[a] Fixed values of Φβ taken from Altona (1982) and Kennard and Hunter (1989).

[b] Calculated population by means of Eqs. 1 and 2.

[c] Fixed value in the calculation in order to limit the number of undetermined variables relative to the number of experimental parameters. Combinations with low rms fits are chosen.

Conformation around ε (C4'-C3'-O3'-P)

The conformation around ε can be monitored by the three vicinal couplings $^3J(H3'P)$, $^3J(C2'P)$ and $^3J(C4'P\varepsilon)$ (Lankhorst et al., 1984,1985). There are two allowed regions for the phosphate group, ε^{tr} and ε^-; the ε^+ region is sterically 'forbidden' and can be safely excluded from the analysis (Altona et al., 1974). A least-squares fit of the coupling constants resulted in the experimentally derived rotamer populations and torsion angles Φε collected in Table 5.

The epsilon torsion angle of the duplex residues adopts exclusively pure ε^{tr} rotamers, with $\Phi\varepsilon^{tr}$ between 200°–202° for dG(2) and 186° for dG(6), which corresponds to the ε^{tr} rotamer range found in X-ray studies of B-DNA type duplexes (Kennard and Hunter, 1989). In contrast, all four thymidine residues show mixtures of $\varepsilon^{tr}/\varepsilon^-$ rotamers, with a predominance in the ε^- domain. T(5) shows the highest conformational purity (80–85% ε^-), which concurs with the observation of a relatively large $^4J(H2''P)$ of 2.4 Hz, originating from the favorable W-path in an ε^-/S-type sugar geometry (Altona, 1982; Blommers et al., 1994). It is not seen for the other, less conformationally pure, thymidine residues (45–59% ε^-). The experimentally determined values of $\Phi\varepsilon^-$ for the thymidine residues show a loop-sequence-specific pattern, with normal values for the second loop residues (259°–260° and 266° for T(5) and T(10), respectively), and decreased values for the first loop residues (233°–238° and 252°–257° for T(4) and T(9), respectively).

^{31}P NMR spectroscopy

^{31}P chemical shifts are frequently used to obtain information about the conformational preference of the phos-

TABLE 5
CONFORMATIONAL ANALYSIS OF TORSION ANGLE ε IN THE CIRCULAR DUMBBELL I

Residue	Experimental			Simulated		
	$p\varepsilon^{tr}$	$\Phi(\varepsilon^{tr})$	$\Phi(\varepsilon^{-})$	$p\varepsilon^{tr}$	ΦrMD	ΦwMD
dC(1)	1.00	194	266[a]	0.98[b]	184	200[b]
dG(2)	0.95–1.00	200–202	266[a]	1.00	202	198
dC(3)	1.00	192	266[a]	1.00	196	191
T(4)	0.41–0.43	192[a]	233–238	0.58[c]	223	237[c]
T(5)	0.15–0.20	192[a]	259–260	0.00	280	290
dG(6)	1.00	186	266[a]	1.00	170	179
dC(7)	0.7–1.0	192[a]	266[a]	0.84	189	189
dG(8)	0.91–1.00	188–192	266[a]	0.99	187	180
T(9)	0.47–0.48	192[a]	252–257	0.37	204	259
T(10)	0.45–0.54	192[a]	266	0.04	228	270

Parameters: population fraction $p\varepsilon^{tr}$, torsion angle of most abundant rotamer Φ (degrees) from experiment, together with average values derived from the rMD and wMD simulations of the L2L4χconstr and L2L4χunconstr forms.

[a] Fixed values of Φε taken from Altona (1982) and Lankhorst et al. (1984,1985).

[b] Data obtained from the wMD simulation of L2L4 with unconstrained χ dC(1).

[c] Poorly defined minima of ε^{tr} and ε^{-} rotamers in wMD simulations; the overall average of ΦwMD is given.

phodiester torsion angles α and ζ (Gorenstein, 1984; Powers et al., 1990). Phosphorus resonances of the dumbbell were assigned previously by means of 2D ^{31}P-^{1}H correlated spectra (Ippel et al., 1995a). Scheme 1 shows the numbering of the phosphates. Eight out of the 10 signals resonate in a narrow region, coinciding with observations on other B-DNA type structures (Chen and Cohen, 1984; Roontga et al., 1990). P(5) is shifted upfield by 1.8 ppm, whereas P(4) shifts downfield by 1.1 ppm, relative to the remaining bulk. A downfield shift of a phosphorus signal (1.5–1.6 ppm) is usually associated with conformational changes around the phosphodiester torsion angles α and/or ζ, i.e., a gauche/gauche combination changes into a gauche/trans or trans/gauche combination (Roontga et al., 1990). An upfield shift of a phosphorus signal is often found in hairpin loops (Orbons et al., 1986,1987a,b; Williamson and Boxer, 1989a,b; Pieters et al., 1990; Blommers et al., 1991).

In contrast to P(4) and P(5), the corresponding phosphates P(9) and P(10) in the opposite loop show no particular deviations and resonate in the normal B-DNA region (Chen and Cohen, 1984; Roontga et al., 1990). When the temperature is raised from 270 K to 303 K (at 0.8 mM DNA), the chemical shifts of the phosphorus resonances in the stem show downfield shifts for P(3) (+0.10 ppm), P(6) (+0.07 ppm), and P(7) (+0.09 ppm) relative to 85% H_3PO_4. Downfield shifts of this magnitude are normally found in melting duplexes. However, the phosphorus signals in the loop behave differently. The P(4), P(5), P(9) and P(10) resonances shift +0.02, −0.26, +0.06, and −0.26 ppm, respectively, on going from 270 K to 303 K. The equivalent values for the differential shifts of the phosphorus resonances in the two loops (P(4) versus P(9) and P(5) versus P(10)) clearly show that changes in phosphate backbone conformation of the loops during the transition from L2L2 to L2L4 are minor. Therefore, the chemical shifts of the loop phosphates P(4), P(5), P(9) and P(10) are intrinsically related to base sequence, and hardly influenced by the shift of the equilibrium between L2L4 and L2L4. Furthermore, the shifts of P1 (+0.02 ppm), P2 (−0.03 ppm), and P8 (0.00 ppm) on going from 270 to 303 K indicate anomalous conformational behavior when compared to the duplex phosphates P(3), P(6) and P(7).

IRMA refinement of the L2L2 form

The buildup curves for the reference H6-H5 NOEs (distance 2.45 Å) of the residues dC(1), dC(3) and dC(7) at 271 K show virtually similar curvature with increasing mixing time, with mutual deviations of less than 5% in the linear part of the buildup curve. This shows that the molecule has an isotropic rotational-correlation time τ_c. The best fit of the reference distances with inclusion of local mobility led to $\tau_c = 1.9$ ns as the optimal correlation time. The H6-H5 NOE in the conformationally nonrigid dC(1) base was not used as a calibration peak, in order to preclude problems with the scaling of the experimentally derived distances. In total, 200 NOEs between nonexchangeable protons were used in the IRMA calculation, 60 of which were interresidual. It must be emphasized that during the refinement no explicit quantitative information on torsion angles was used. The validity of the model obtained by distance information was afterwards checked against the experimentally derived torsion angles.

Not surprisingly, after one cycle of IRMA it became obvious that the set of NOEs involving H6 and H5 of residue dC(1) could not be fitted to a single model, as the structure showed a highly buckled (60°) dC(1)-dG(8) base pair. Apparently, this buckle arose from contradictory constraints on the dC(1) H6-H1', H6-H2', H6-H2" and H6-H3' intraresidue proton pairs. In subsequent rMD cycles, all constraints involving the dC(1) H6 and H5 protons were removed, which resulted in an *anti* orientation of the dC(1) base, corresponding to the L2L2 model. The dC(1)-dG(8) base pair then became flattened, with a moderate buckle of less than 20°. Convergence of IRMA was rapidly achieved and after five cycles of refinement no further improvement was obtained. The total constraint energy dropped from 942 kJ/mol to 257 kJ/mol. The overall R-factor, averaged over five mixing times, decreased from 0.58 for the starting structure, cycle 0, to 0.40 after cycle 5 (Table 6). The averaged R-factor, calculated for the subset of the 60 interresidue NOEs, had a value of 0.96 at the start of the refinement, but did not improve significantly during the refinement (0.87 after

cycle 5), in contrast to the intraresidue R-factor, which dropped from 0.52 to 0.33. This is partly caused by the generally lower signal/noise ratio of the interresidue peaks compared to intraresidue peaks. When only the longest mixing time was considered, the value of the R-factor after cycle 5 of IRMA converged to 0.57 (0.69 at cycle 0). A final cycle with inclusion of S^2 factors (Lipari and Szabo, 1982a,b) resulted in a decrease of the R-factor (averaged over five mixing times) of the interresidue NOE subset from 0.87 to 0.71 (0.50 at $\tau_m = 340$ ms). In contrast, the intraresidue R-factor increased with inclusion of S^2 factors from 0.33 to a value of 0.35. The R-factor involving only the eight intraresidue NOEs of the H6 proton of residue dC(1) is significantly higher (R = 0.58) than the intraresidue R-factor of the subset of 140 intraresidue NOEs (R = 0.35), and indicates a poor fit for the NOEs of the dC(1) base.

IRMA refinement of the L2L4 form

The next step was to refine the structure of L2L4 in a similar way as for L2L2. The refined model of L2L2 was used as the starting model for L2L4, except that the dC(1) base in the model was rotated into the *syn* domain. The *syn* orientation of dC(1) was kept rigid during the rMD runs by means of two constraints ($E_{tc} = 420$ kJ/mol) around torsion angles O1'-C1'-N1-C6 and O1'-C1'-N1-O2 of the dC(1) residue. This was necessary to ensure that the rotamer stayed in the *syn* domain. Removal of these constraints resulted in a return to the *anti* orientation of the dC(1) base. Convergence of the L2L4 form during the IRMA procedure was rapidly achieved and resulted in an overall R-factor of 0.39 after two cycles (interresidue R-factor 0.69) (Table 6). However, again the subset of eight H6 dC(1) NOEs did not produce a good fit with experiment, as shown by the R-factor of 0.56.

TABLE 6
R-FACTORS (ALL, INTRARESIDUE AND INTERRESIDUE) AVERAGED OVER FIVE MIXING TIMES IN THE COURSE OF THE STRUCTURE REFINEMENT OF THE L2L2 AND L2L4 FORMS OF THE CIRCULAR DUMBBELL

Cycle no.	All	Intraresidue	Interresidue ($\tau_m = 340$ ms)[a]
L2L2			
Cycle 0	0.58	0.52	0.97 (0.69)
Cycle 1[b]	0.44	0.40	0.69 (0.57)
Cycle 2	0.41	0.34	0.85 (0.59)
Cycle 3	0.39	0.31	0.84 (0.63)
Cycle 4	0.40	0.33	0.80 (0.62)
Cycle 5	0.40	0.33	0.87 (0.57)
Including S^2	0.40	0.35	0.71 (0.50)
L2L4			
Cycle 1	0.40	0.33	0.83 (0.55)
Including S^2	0.39	0.34	0.69 (0.48)
1:1[c] L2L2/L2L4			
Including S^2	0.40	0.35	0.72 (0.48)

[a] Between brackets: calculated R-factors of a NOESY spectrum recorded at $\tau_m = 340$ ms.
[b] Force constant K_{dc} in cycle 1 was set to 40 kJ/mol Å^2; in all subsequent cycles K_{dc} was 8 kJ/mol Å^2.
[c] 1:1 Averaging of NOE matrices of L2L2 and L2L4.

A list of upper and lower bound distances obtained by IRMA, for both the L2L2 and L2L4 forms, is available from the authors as supplementary material (Table S1). In summary, the list shows a good agreement of experimental distances derived from IRMA and corresponding calculated distances from the model. At this point, it must be remembered that the IRMA refinements of the L2L2 and L2L4 forms were based on the same NOE data, hence it is not surprising to find that almost all calculated distances are identical within 0.1 Å for the two refined models. A comparison of the structures of L2L2 and L2L4, refined by means of distance constraints, results in a root-mean-square deviation (rmsd) of 0.41 Å, when the pC(1) nucleotide is excluded from the comparison.

Discussion

'Weakly constrained' MD in water

Although the major differences between L2L2 and L2L4 are clearly recognized, the subtle differences in the detailed structures are veiled by the use of 'mixed conformer' data. In order to study the geometries of both the L2L2 and the L2L4 forms in more detail, 'weakly constrained' MD calculations in water (wMD) and in vacuo (vMD) were performed. Normally, large forces on the NOE constraints are imposed in MD, resulting in a rigid structure which precludes dynamic processes, therefore 'free' MD runs are most informative (Pearlman and Kollman, 1991). However, 'free' MD with current force fields on flexible DNA structures tends to diverge from the limited conformational space deduced from experiment. Incorporation of ensemble averaging in the MD simulation can be used to retain proper sampling of the conformation space (Torda et al., 1989,1990; Schmitz et al., 1992).

A more simple approach is the reduction of NOE-restraint potential energies with a factor of 1–10 below the thermal energy 1/2kT of 1.1 kJ/mol at 270 K during the MD simulation. For the dumbbell, the force constant of the NOE constraints in the MD runs was reduced to an arbitrarily chosen value of 0.4 kJ/mol. The energy corresponds to a random excess distance of 2.4 Å outside the distance boundaries at 270 K and allows for restricted mobility in conformational space. The large number of tertiary 'long-range' NOEs between protons in the dumbbell makes this procedure well suited to account for structural differences in the L2L2 and L2L4 forms.

wMD simulations of the circular dumbbell

Five 'weakly constrained' MD calculations were performed on the circular dumbbell: L2L2 in vacuo (100 ps)

and in water (99.6 ps), L2L4(χunconstr) in vacuo (100 ps) and in water (59.2 ps), and L2L4(χconstr) in water (60.0 ps). In the simulation marked χconstr, torsion angle χ of dC(1) was constrained in the *syn* domain; χunconstr means that rotation around χ of dC(1) is allowed. The main conclusion of the wMD simulations is that the overall loop folding of the two H2-family type hairpin loops remains intact, even without explicit inclusion of solvent. We confine the results of the analysis to the 'weakly constrained' MD simulations in water (wMD).

Dynamic processes in the L2L2 and L2L4 forms

During the simulation period, torsion angles were monitored every 0.1 ps. The torsion angle analysis is summarized in Table S4 of the supplementary material. The backbone torsion angles β and ε in the wMD (and rMD) model of L2L4 are closely similar to experimentally determined values, obtained by means of coupling constant analysis at temperatures above 290 K. The results are shown in Tables 4 and 5. The β torsion angles determined for T(4) and T(5) (163° and 209°, respectively) are well reproduced in the L2L4 models obtained from the rMD calculations (166°–168° and 204°–210°). According to the static rMD models, the agreement in ε torsion angles (Table 5) seems less obvious. At this stage the introduction of an internal rotation model becomes important; this correctly predicts the occurrence of conformational mixtures of ε^{tr} and ε^{-}. In the supplementary material, conformational transitions are discussed in more detail.

Matrix averaging of the L2L2 and L2L4 forms (static R-factors)

Mixing the relaxation matrices of the minimized L2L2 and L2L4 models by 1:1 $\langle r^{-6}\rangle$ averaging, corresponding to a slow-exchange 1:1 equilibrium at 271 K between the two forms, does not result in a further improvement of the overall NOE fit (R = 0.40) (Table 6). In order to screen experimentally the differences between the two forms, if any, an R-factor calculation was performed in which individual NOE contributions per residue were separated. Table 7 displays the results for the 1:1 L2L2/L2L4 (rMD) conformational mixture. With the exception of residues dC(1) and dG(2), the number of integrated NOEs per residue varies from 9 to 21 intraresidue, and from 10 to 16 interresidue NOEs. In order to suppress noise in the analysis of the small number of NOEs available per residue, R-factors were weighted over mixing time by $1/\tau_c$.

The data show that intraresidue R-factors are relatively constant, with the exception of dC(1) (R = 0.46). In contrast, the individual interresidue R-factors are nonuniformly spread, ranging between 0.3–0.4 at positions between dC(1) and T(5) to a gradually increased maximum value of 1.22 at residue T(9). Similarly high R-factors in the dC(7)-T(10) part of the dumbbell are found for the separate L2L2 and L2L4 models. The R-factor for the subset of the eight dC(1) H6 NOEs drops from 0.58 to 0.52 in the 1:1 L2L2/L2L4 model. Although the R-factor remains high (especially for the H6-H2'/H2" NOEs; vide infra), this observation indicates that the time-averaged

TABLE 7
R-FACTORS (ALL, INTRARESIDUE, INTERRESIDUE) CALCULATED PER RESIDUE AND WEIGHTED OVER MIXING TIME

Residue	rMD (static) 1:1 L2L2/L2L4[a]			wMD (dynamic) 1:1 L2L2/L2L4[b]			wMD (dynamic) ensemble χ dC(1)[c]		
	All	Intra	Inter	All	Intra	Inter	All	Intra	Inter
dC(1)	0.45	0.46	0.27[d]	0.56	0.56[e]	0.52[d]	0.42	0.43[e]	0.33[d]
dG(2)	0.32	0.26[d]	0.41	0.47	0.31[d]	0.68	0.28	0.21[d]	0.36
dC(3)	0.36	0.33	0.42	0.41	0.43	0.37			
T(4)	0.34	0.33	0.36	0.44	0.40	0.52			
T(5)	0.33	0.38	0.23	0.48	0.48	0.46			
dG(6)	0.42	0.39	0.62	0.43	0.43	0.43			
dC(7)	0.37	0.26	0.73	0.44	0.42	0.50			
dG(8)	0.59	0.32	1.15	0.46	0.43	0.54			
T(9)	0.59	0.27	1.22	0.40	0.39	0.44			
T(10)	0.44	0.28	0.91	0.39	0.37	0.47			
Overall	0.40	0.35	0.72	0.44	0.43	0.49			

rMD (static): 1:1 conformational mixture of L2L2 and L2L4 after rMD ($\langle r^{-6}\rangle$ matrix averaging); wMD (dynamic): after wMD ($\langle r^{-3}\rangle$ matrix averaging for each dynamics run, followed by $\langle r^{-6}\rangle$ averaging). The overall R-factors are not weighted. All calculations were performed with $\tau_c = 1.9$ ns and inclusion of S^2 factors.

[a] R-factors after static averaging of IRMA structures of L2L2 and L2L4χ59 (ratio 1:1).

[b] R-factors after dynamic averaging of the L2L2 and L2L4χunconstr wMD trajectories. Overall dynamic R-factors for individual wMD trajectory files of the L2L2 and L2L4χunconstr forms: 0.46 (all), 0.45 (intra) and 0.50 (inter); 0.44 (all), 0.42 (intra) and 0.51 (inter), respectively.

[c] R-factors after $\langle r^{-6}\rangle$ averaging of an ensemble (20 structures) monitoring the *syn* to *anti* conversion of dC(1) and vice versa.

[d] Statistically not relevant, because of the small number of integrated NOEs, i.e., three interresidue NOEs for dC(1) and four intraresidue NOEs for dG(2).

[e] R-factors only for the eight dC(1) H6 intraresidue NOEs: 0.57, 0.52 and 0.50 for 1:1 rMD, 1:1 wMD and ensemble χ dC(1), respectively.

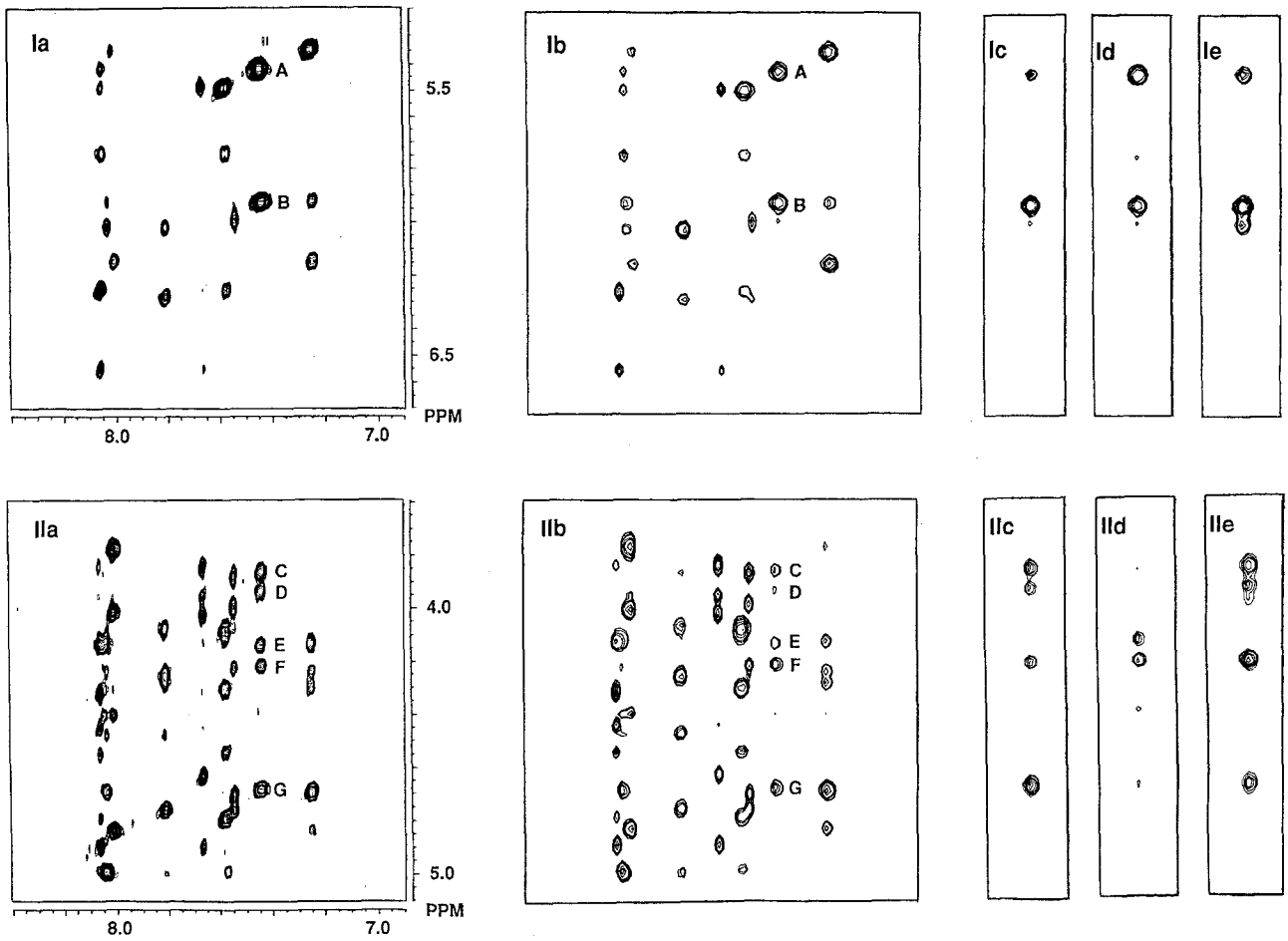


Fig. 1. Comparison between parts of the experimental and simulated NOESY spectra of **I**. I = base-H1'/H5 region; II = base-H3'/H4'/H5'/5" region. (a) Experimental spectrum ($\tau_m = 340$ ms); (b) simulated spectrum of $<r^{-6}>$-averaged 1:1 mixture L2L2/L2L4 (static); and simulated spectra of dC(1) H6 NOEs of (c) L2L2 (static); (d) L2L4 (static); and (e) $<r^{-3}>$-averaged 1:1 mixture L2L2/L2L4χunconstr (dynamic). Cross peaks indicated in the spectra: (A) H6-H1' dC(1); (B) H6-H5 dC(1); (C) H6-H5" dC(1); (D) H6-H5' dC(1); (E) H6 dC(1)-H5'/H5" dG(2) (overlap); (F) H6-H4' dC(1); and (G) H6-H3' dC(1).

structure of the dC(1) base in the circular dumbbell is better described by a conformational mixture of L2L2 and L2L4 rather than by a single conformer.

Matrix averaging of the L2L2 and L2L4 forms (time-averaged R-factors)

By introducing mobility in the molecule by means of the wMD calculation, a better agreement with experiment is obtained. The mobility of the molecule is introduced by two methods. Internal rotations on the picosecond time scale are treated by $<r^{-3}>$ averaging of the relaxation matrix computed for each separate structure in the wMD trajectories of L2L2 and L2L4. The slow exchange between the two forms was treated by $<r^{-6}>$ averaging of the two $<r^{-3}>$-averaged matrices. The corresponding time-averaged R-factors, calculated per residue, are listed in Table 7. In Fig. 1, parts of the experimental spectrum ($\tau_m = 340$ ms), containing the dC(1) H6 NOEs, are compared with the simulated spectra for the L2L2 and L2L4 and the static and dynamic 1:1 $<r^{-6}>$-averaged models.

On dynamic averaging, the overall R-factor of NOE intensities increases from 0.40 to 0.44 and this increase is due in particular to an increase of the intraresidue R-factor from 0.35 to 0.43. The interresidue R-factor drops significantly and goes from 0.72 to 0.49. The decrease is due in particular to the improvement of the calculated NOEs in the dG(6) to T(10) steps, as shown in Table 7. The R-factor is now reasonably constant throughout the molecule, with the exception of the NOEs between dG(2) and dC(1) ($R_{inter} = 0.68$). The decrease in R-factors of the strand between dG(6) and T(10) is caused by consideration of the rotational mobility of this region in both the L2L2 and L2L4 forms. In Fig. 2 the rmsd, expressed as amplitude of motion, is plotted per residue for both the L2L2 and L2L4χunconstr forms during the wMD runs, showing that the 5'-GTTC-3' loop, especially in the L2L4 form, displays a larger mobility compared to the opposite loop. In fact, the gradual increase of the rmsd, with a maximum at residue T(9), correlates very well with the deviations in the residue-based static R-factors obtained by rMD (Table 7).

In conclusion, the poor fit of NOEs in the 5'-GTTC-3' loop probably arises from the translation of time-aver-

aged NOE data into single-conformer distances. This means that the structural difference between the L2L2 and L2L4 forms lies not entirely in the orientation of the dC(1) base, but is also reflected in the detailed geometry of the 5'-GTTC-3' loop.

The observed line broadening of several resonances (the carbon signal of C1' dC(1), phosphorus resonances P(1), P(2) and P(10), and proton signals H1' dC(1), Me T(10), H1' T(10), H6 T(9) and H2" T(9)) in the 5'-GTTC-3' loop, due to conformational exchange between L2L2 and L2L4 at low temperatures (Ippel et al., 1995a), corresponds well to the regions of mobility deduced from the NOE data and the wMD simulations. However, chemical shifts do not vary significantly (with the exception of the imino proton of dG(8)) between 270 and 314 K; therefore no strong line broadening is expected for most phosphorus and proton signals based on the time scale (2–20 s^{-1}) of interconversion between the L2L2 and L2L4 forms. An alternative explanation for the specific deviation in NOE intensities and specific line broadening is that conformational heterogeneity exists only in one of the conformers, with the L2L4 form as the most probable candidate.

R-factor analysis of residue dC(1)

In contrast to residues dG(6) to T(10), incorporation into the model calculations of the relatively large dynamic mobility of residue dC(1) found in the L2L4χunconstr wMD run does not improve the NOE fit, as shown by the dynamic overall R-factor of 0.56. The simulated time-averaged NOEs involving the dC(1) H6 proton do not agree with the observed ones. For example, the predicted peak from H6 dC(1) to H1' T(10) in the simulated spectrum (not shown) is not observed experimentally. This is ascribed to deficiencies in the force field used. A high R-factor is also obtained for the L2L4χconstr wMD simulation, although the dC(1) base is constrained in the *syn* domain. Apparently, the wMD simulations of the L2L4 forms are unable to predict correctly the conformational behavior around the glycosyl bond of residue dC(1). An attempt to trace the origin of this situation was carried out as follows.

Figure 3 shows the experimental buildup curves for the base to H2'/H2" NOEs of residue dC(1). It can be seen that the pattern of the buildup curve of the H6-H2" NOE is uncommon when compared to the normal *anti* situation in dC(7). It reflects contributions of high-*anti* conformers at 271 K, where short distances are expected between H6 and H2' and H2" (Saenger, 1984). Unfortunately, above 300 K the H6-H2'/H2" cross peaks of residue dC(1) start to overlap in the experimental spectrum, thus preventing a check on the individual NOE contributions of the H6-H2' and H6-H2" pairs in the pure L2L4 form. However, in the temperature range between 268 and 290 K the relative intensities of the H6-H2' and H6-H2" NOEs of dC(1) do not change much with temperature; +117% and +111%, respectively. In contrast, the intensities of the intraresidue H6-H1' and H6-H3' NOEs of dC(1) do change on going from 268 to 314 K, i.e., decreasing for H6-H3' (−320%) and increasing for H6-H1' (+186%). The temperature-dependent NOEs of H6-H1' and H6-H3' point to a motion of the dC(1) base centered in the high-*anti* domain.

An ensemble of 20 structures, uniformly distributed in the 180° conformational circle between *syn*, high-*anti* and *anti*, reduces the intraresidue R-factor for the eight H6 NOEs of residue dC(1) (R = 0.50 versus 0.52, according to a static two-state '*syn–anti*' model), as shown in Table 7 (footnote e). A correct prediction of the approximately 1:1 intensity ratio (at $\tau_m = 340$ ms) of H6–H2' relative to H6-H2" is obtained only in the case of the ensemble containing the *anti*/high-*anti* and *syn* conformers. By this method, the fit of the interresidue NOEs of the adjacent dG(2) residue also improves compared to the time-averaged R-factor calculated from the two-state assumption (R = 0.36 versus 0.41).

In summary, the R-factor analysis shows that, except for the intraresidue NOEs of the sugar rings and the NOEs of dC(1), the 'weak constraint' method adequately reproduces the experimental NOESY spectrum in comparison with static structures and provides a valuable tool to study the less stable 5'-GTTC-3' loop in the slow-exchange limit of the dumbbell conformers. The orientation of the dC(1) base is, however, better described (although not perfect) by a multistate conformer population, with χ of dC(1) distributed between the *syn* state (χ = 59°), through the high-*anti* state (χ = −30°) and the *anti* state (χ = −120°).

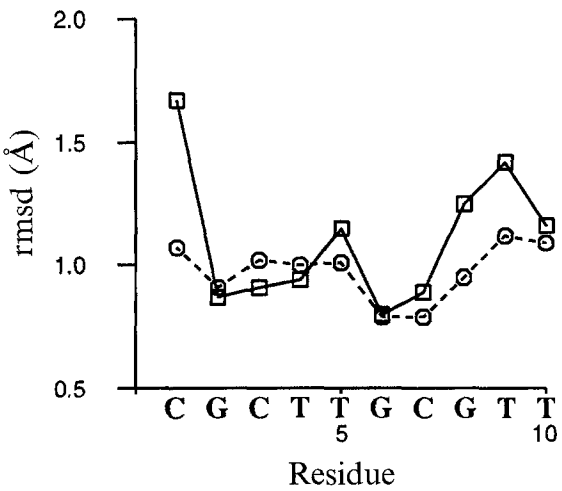


Fig. 2. Rmsd distance (Å) of amplitudes of motion, calculated per residue for the L2L2 form (○) and the L2L4χunconstr form (□) from the complete trajectory of the wMD simulations. Rotation and translation of the molecule were first removed by least-squares fitting of all structures. The residue-based rmsd values are expressed as Σ_i $rmsd_{atom}$ / N_i, i.e., the sum of individual atomic rmsd values in one residue normalized to N_i, the number of atoms involved. Phosphate atoms and methyl protons were excluded from the calculation.

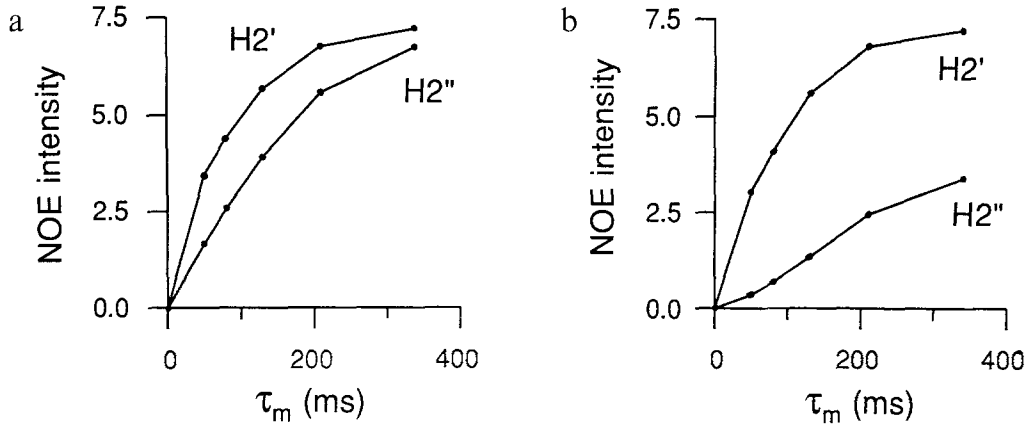


Fig. 3. NOE buildup curves at 271 K as a function of mixing time for intraresidue NOEs between the H6 proton and the H2' and H2" protons of (a) residue dC(1); and (b) residue dC(7).

Refined structure of the L2L2 form

A stereo plot of the model of the L2L2 form, refined by IRMA and rMD, is depicted in Fig. 4. The complete list of backbone torsion angles in the rMD models of L2L2 and L2L4 is given in Table 8. Figure 4 shows that the refined structure of L2L2 consists of a duplex stem with three C-G base pairs and a two-residue loop on both sides. Both loops contain a sharp turn at the 3'-5' loop–stem junction (β^+/γ^{tr}). As a consequence, the cross-chain phosphate–phosphate distances, P(9)-P(1) and P(4)-P(6), decrease by 5 Å to about 13 Å. However, analysis of coupling constants shows that the conformational purity of the γ^{tr} conformation of residue dC(1) at the potential L4 side is less pronounced compared to γ^{tr} of dG(6) at the other side (the populations are 58% and 82%, respectively). The same trend is seen for the β torsion angle, where the β^+ rotamer of dC(1) is less populated relative to dG(6) (27% versus 40%). The base pairs closing the loops, i.e. dC(3)-dG(6) and dC(1)-dG(8), are strained and buckle towards the loop by 12° and −23°, respectively (Table S2 of the supplementary material). The buckle relieves steric strain, otherwise present in closure of the intrastrand gap by two residues.

Both the T(4) and T(9) bases are located inside the minor groove, according to the H2-family loop-folding principle (Pieters et al., 1990). At first sight, this seems a sterically unfavorable situation. Closer inspection reveals that in the models the two bases do not interfere strongly with each other. The planes of the bases are placed at a mutual angle of 58° for both the L2L2 and L2L4 conformers; the distance between Me T(4) and Me T(9) in the L2L2 model is 6.20 Å (wMD 7.93 Å) and for the L2L4 model it is 5.44 Å (wMD L2L4χunconstr 5.13 Å; wMD L2L4χconstr 6.22 Å), thus steric hindrance between the two methyl groups is largely avoided. Nevertheless, the total interaction between T(4) and T(9) results in an opening of the minor groove of L2L2 with a total positive roll angle of 18° (calculated) between the bases T(5) and T(10).

Geometrical differences between the 5'-CTTG-3' and 5'-GTTC-3' loop folding

The most striking observation concerning the L2L2 structure is that the dC(1) base is positioned more or less outside the helix, in contrast to the corresponding base of dG(6) in the opposite loop. The difference in conforma-

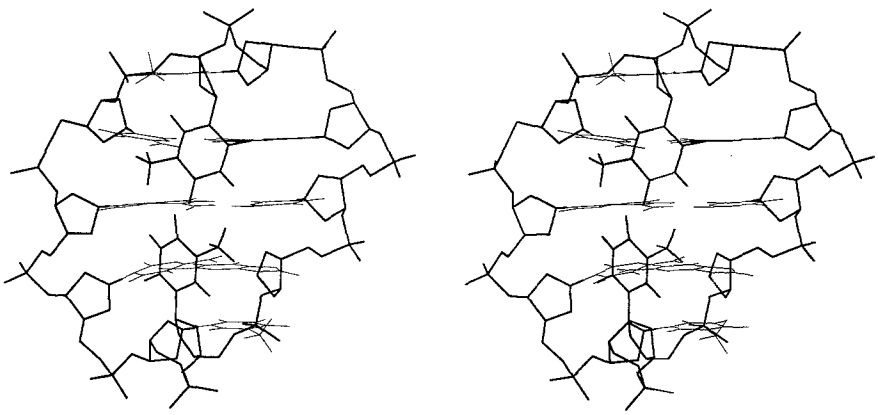


Fig. 4. Stereoview towards the minor groove of the minimized structure of L2L2 obtained by means of distance constraints. The 5'-CTTG-3' loop is placed at the top and the 5'-GTTC-3' loop at the bottom of the picture. For clarity, the bases involved in stacking are plotted by thin lines.

TABLE 8
BACKBONE TORSION ANGLES (DEGREES) OF THE DISTANCE-CONSTRAINED ENERGY-MINIMIZED MODELS OF L2L2 AND L2L4χ59[a]

Residue	α	β	γ	δ	ε	ζ	χ	P	Φ
L2L2									
dC(1)	275	53	163	113	181	269	−116	117	44
dG(2)	294	172	52	131	202	284	−96	138	45
dC(3)	248	67	166	147	195	269	−138	161	42
T(4)	190	170	54	147	214	274	−127	169	37
T(5)	300	208	51	151	290	111	−148	168	34
dG(6)	185	84	144	158	169	262	−88	208	37
dC(7)	290	179	57	126	186	249	−119	134	36
dG(8)	286	176	61	141	183	270	−109	168	34
T(9)	194	162	60	142	206	273	−141	163	34
T(10)	306	202	58	149	226	193	−154	177	32
L2L4χ59									
dC(1)	268	49	162	136	184	256	+59[a]	148	36
dG(2)	292	173	55	132	202	280	−91	140	45
dC(3)	252	64	168	149	196	274	−138	166	40
T(4)	186	166	55	148	223	270	−129	168	39
T(5)	304	204	49	152	280	125	−142	169	34
dG(6)	198	69	148	157	170	262	−89	203	36
dC(7)	294	181	54	135	189	237	−112	148	40
dG(8)	286	175	58	145	187	267	−116	170	36
T(9)	195	163	58	144	204	280	−139	162	36
T(10)	302	212	58	150	228[a]	188[a]	−147	176	34

[a] Torsion angles obtained after final energy minimization with extra experimentally derived torsional restraints around ε T(10) and β dC(1) to reduce divergence of the structure due to the unfavorable steric clash between the dG(8) and dC(1) bases. Using only distance constraints, two combinations of ε and ζ of T(10) are found; one is listed in the table and the other has ε/ζ = 268°/151° and χ dC(1) = 63°.

tional behavior of the two loops is caused by the different base sequence of the base pairs closing the loops. The closing base pair dC(3)-dG(6) of the 5'-CTTG-3' loop shows a better stacking overlap with the adjacent base pair dG(2)-dC(7) compared to the overlap at the dC(1)-dG(8) to dG(2)-dC(7) base-pair step, as shown by the individual base-pair steps in the rMD structure of L2L2 (Fig. 5).

The base–base stacking even diminishes further during the wMD simulation of L2L2, as shown by the value of average shift, Dx = −0.15 Å (rMD: Dx = 0.63 Å). The overlap of the dG(2)-dC(7)/dC(3)-dG(6) base-pair doublet in the duplex stem is directly controlled by the favorable geometry of nonbonded interactions between the bases of the 5'-GC-3'/3'-CG-5' base–base step. Normally, the 5'-GC-3'/3'-CG-5' base–base step in DNA has an energetically more favorable nearest-neighbor stacking overlap than the reversed 5'-CG-3'/3'-GC-5' step ($\Delta\Delta G_{MN} = -4.8$ kJ/mol; Delcourt and Blake, 1991). According to theoretical calculations, nonbonded interactions are by far the most important factor in DNA base stacking. Electrostatic contributions are relatively negligible in base stacking, at least in idealized B-DNA (Friedman and Honig, 1992). The next-nearest neighbor stacking may therefore largely be improved by adaptation of the local helix twist angle Ω between the base-pair doublet (Klug et al., 1979; Yuan et al., 1992). The relation of the two quantities is confirmed by a survey of high-resolution X-ray structures of B-DNA, where high correlation coefficients are found between local Dx/Dz displacements and local twist angles Ω (Bhattacharyya and Bansal, 1990). Comparison with other DNA structures in solution also reveals correlations between helix twist and base sequence. The 5'-CG-3'/3'-GC-5' steps in the duplex structure of $d(CGTACG)_2$, obtained by NMR, show large Ω values for the CG steps (47° for dC(1) to dG(2) and 43° for dC(5) to dG(6)) (Lane, 1990). Furthermore, similar trends for Ω and the nature of the base–base step 5'-CG-3'/3'-GC-5' versus 5'-GC-3'/3'-CG-5' ($\Omega = 38 \pm 4°$ and $27 \pm 3°$, respectively) are found in the nonselfcomplementary duplex 5'-d(CGTCA-CGCGC)-3', refined by means of rMD and iterative back-calculation of NOESY spectra (Metzler et al., 1990).

In the case of the L2L2 model, the local Ω angle at the base-pair step dC(1)-dG(8) to dG(2)-dC(7) is significantly higher than that found at the dG(2)-dC(7) to dC(3)-dG(6) step (30°–31° versus 22°–25°; see Table S3 of the supplementary material). The difference in local helix twist between the bases becomes even more pronounced (33°–42° versus 25°–27°) in the case of the L2L4 form. The small local helical twist of less than 27° between the dG(2)-dC(7) and dC(3)-dG(6) base pairs appears to result partly from the adaptation of an $\alpha^-/\beta^+/\gamma^{tr}$ backbone geometry in residue dC(3), a combination that swivels the P(3) phosphate and the sugar ring of dC(3) out of the minor

groove. The backbone reorientation creates more room for the dangling T(4) base, and additionally gives better stacking overlap of the dC(3) base onto the dG(2) base. Consequently, the dC(3) base is placed over the dG(2) base in an almost parallel fashion. In contrast, residue dG(8) shows the $\alpha^-/\beta^{tr}/\gamma^+$ torsion angle combination normally found in B-DNA duplexes. As a consequence, the minor groove at the 5'-GTTC-3' loop side is more narrow, and results intrinsically in more steric clash with the T(9) base.

Figure 6 displays plots of the 5'-CTTG-3' and 5'-GTTC-3' loops obtained from the wMD-averaged structure of L2L2, and superimposed by means of a least-squares fit of the thymidine residues in both loops. In this fit the bases of residues dC(3) and dG(8) are positioned on top of each other, with the N atom of the glycosyl linkage as the center. The geometries of the two -pTpTp- steps on each side of the dumbbell display virtually identical loop folding (rmsd = 0.59 Å; 0.35 Å in the rMD model). This means that the structure elements of the 5'-CTTG-3' and 5'-GTTC-3' loops, seen from the respective bases of residues dC(3) and dG(8), are conserved. However, the difference in base pairing of dG(8)-dC(1) relative to dC(3)-dG(6), together with the different local helix twist (30° and 22°, respectively) necessitates a horizontal displacement (4.2 Å) of the sugar of residue dG(6) relative to that of dC(1) at the 3'-5' loop–stem junction. The local displacement of 4.2 Å in the backbone of the stem sequence dG(6) to dG(8) relative to the sequence dC(1) to dC(3) is absorbed in the 5' to 3' direction, especially around torsion angles α and ζ in the sharp turns; a ζ^{tr}/α^- torsion combination occurs for the T(10)-dC(1) junction, in contrast to the ζ^+/α^{tr} torsion combination found for the T(5)-dG(6) junction (Table 8).

Furthermore, the two loops in the rMD structure of L2L2 differ in the hydrogen-bonding pattern of the thymine base (T(4) versus T(9)) located in the minor groove. The position of T(9) involves a hydrogen bond between O2 T(9)–NH^{β} dG(8) (average distance in the wMD simulation: 2.07 ± 0.26 Å) and, in addition, a weak hydrogen bond (2.8 Å) between H3 T(9) and O4 T(4) may occur. In contrast, three hydrogen bonds are formed between T(4) and atoms in the minor groove, i.e., O2 T(4)–NH^{β} dG(6), H3 T(4)–O2 dC(7) and O4 T(4)–NH^{β} dG(2) (average distances in the wMD simulation: 2.17 ± 0.29, 2.36 ± 0.35 and 2.14 ± 0.25 Å, respectively) and, in addition, the already noticed weak hydrogen bond O4 T(4)–H3 T(9), the first three of which have also been described for the hairpin structure studied by Pieters et al. (1990). The experimental observation of two separate amino proton signals of dG(6) shows that rotation around the C-NH_2 bond of dG(6) is hindered (Ippel et al., 1992). Similar evidence for slow rotation around this C-NH_2 bond at the corresponding position is observed in the H2-family hairpin d(CG^{a}C-TA-GCG). We suggest that the hydrogen bond between O2 T(4) and NH^{β} dG(6) is responsible for this slow rotation. Additionally, the wMD simulations correctly predict this particular hydrogen bond in the minor groove to be dynamically stable in both the L2L2 and L2L4 forms. In contrast, hydrogen bonds between O4 T(4) and NH^{β} dG(2), H3 T(4) and O2 dC(7), and O2 T(9) and NH^{β} dG(8) are predicted to be less stable in the

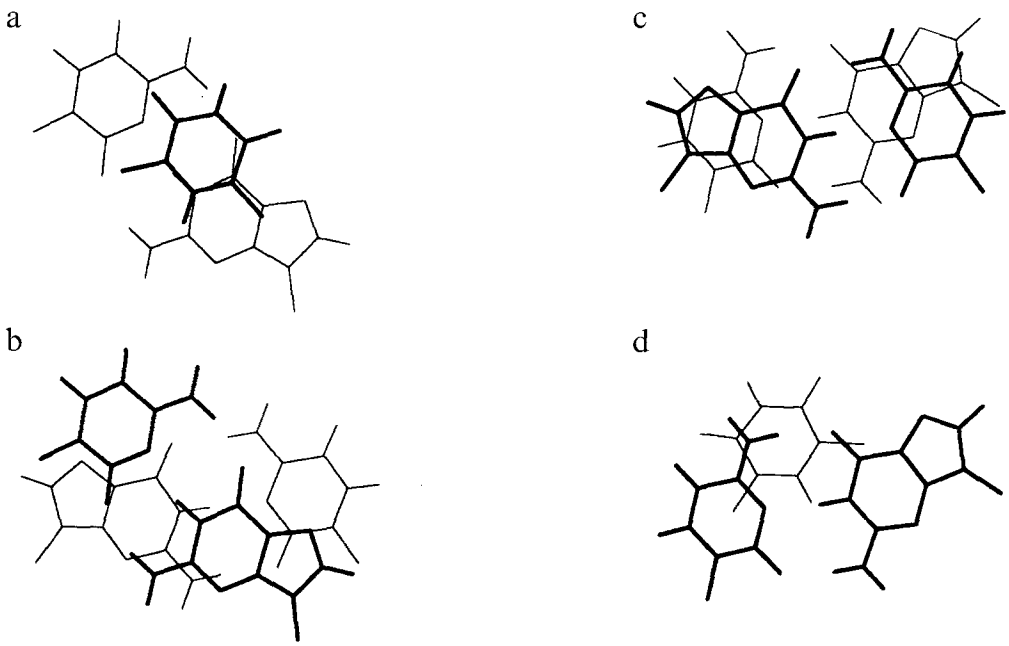


Fig. 5. Stacking overlap of bases in the L2L2 model, obtained by distance constraints. The bases are viewed parallel to the helix axis. (a) T10 (thick) relative to dC1-dG8 (thin); (b) dC1-dG8 (thick) relative to dG2-dC7 (thin); (c) dG2-dC7 (thick) relative to dC3-dG6 (thin); and (d) dC3-dG6 (thick) relative to T5 (thin).

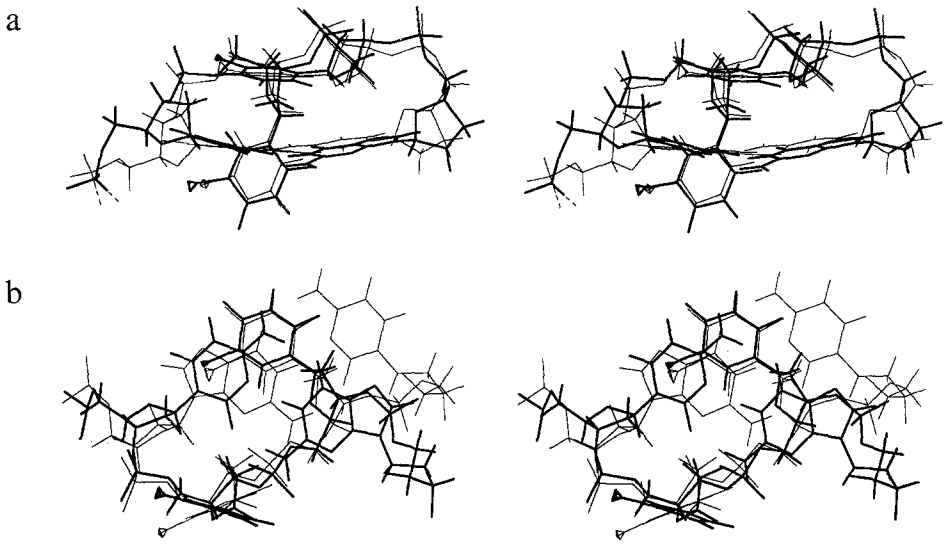


Fig. 6. Comparison of the 5'-CTTG-3' and 5'-GTTC-3' loops according to the wMD-averaged structure of the L2L2 model, viewed in a stereoplot. The two loops are superimposed on each other; the 5'-CTTG-3' loop is plotted by thick lines, the 5'-GTTC-3' loop by thin lines. The criterion for superposition was a least-squares fit of the pair of thymidine residues in both loops (rmsd = 0.59 Å). The angle of view in (a) is towards the minor groove. In (b) the point of view is parallel to the helix axis, looking from inside the minor groove towards the end of the molecule. In both situations the 5'-terminus starts at the left. The triangles at position C7 of the thymine bases represent the methyl protons, averaged due to fast rotation of the methyl groups in the wMD run.

L2L4 form, with average distances of 2.71, 3.47 and 3.46 Å, respectively.

Refined structure of the L2L4 form

Compared to the model of L2L2, the rMD model of L2L4 contains a similar overall backbone conformation (Table 8). The major difference lies in the *syn* conformation around χ dC(1). In order to reduce the steric clash between the H6/H5 protons of dC(1) and the dG(8) base, the dC(1)-dG(8) base pair is buckled. Two orientations of χ dC(1) in the *syn* domain appear possible, χ = 49° or 63°, denoting buckling towards dG(2) ($\kappa_{dC(1)\text{-}dG(8)} = 36°$) and towards T(10) ($\kappa_{dC(1)\text{-}dG(8)} = -6°$), respectively (see Table S2 of the supplementary material). The introduction of extra experimentally derived torsional constraints around ε T(10) and β dC(1) during energy minimization shows that the latter model, with χ dC(1) = 59°–63°, is in better agreement with experiment (Table 8).

The rMD models of L2L4 imply that no hydrogen bond is formed between the imino proton of dG(8) and N3 of dC(1), whereas the NH^{α} proton of the dC(1) base in a *syn* conformation can form a hydrogen bond with O6 of dG(8) (distances of 1.85 and 1.98 Å in L2L4χ49 and L2L4χ59, respectively), even when no constraint on the latter hydrogen bond is present. Figure 7 displays the rMD model of L2L4χ59. The C(*syn*)-G(*anti*) base pair is highlighted in the plot. The weak hydrogen bond between O4 T(4)–H3 T(9), which was present in the L2L2 model, is broken in the L2L4 model (> 3 Å); this is indirectly caused by the rearrangement of the T(9) base position.

None of the wMD simulations of L2L4 predicts the conformation of χ dC(1) correctly. The conformation of the dC(1) base, originally started in a *syn* conformation, gradually moves to an *anti* orientation in the L2L4χunconstr simulation. At the same time, the stacking overlap of dC(1) with the base of dG(2) decreases. Finally, in the overall structure, the dC(1) base is placed in an *anti* position with no base pairing to the dG(8) base. The wMD L2L4χconstr simulation with χ dC(1) constrained in the *syn* domains leads in the averaged structure to complete destacking of dC(1) and no base pairing to dG(8). Nevertheless, the wMD simulations of the L2L4 form show interesting rearrangements of the base of T(9), moving in- and outwards of the minor groove. The H6 T(9) to H1' dG(8) and the Me T(9) to H4' dG(8) NOEs are useful markers for this conformational motion of T(9). The calculated $\langle r^{-3} \rangle$ time-averaged distance for the H6 T(9)–H1' dG(8) pair, based on the three wMD simulations, is larger than the corresponding distance for the H6 T(4)–H1' dC(3) pair (3.45 versus 2.81 Å, respectively), in better agreement with the experimentally derived distance (3.6–3.8 Å for T(9)) than the distances obtained from the static models (2.6 Å for T(9)). The movement of T(9) out of the minor groove in the L2L4 form also results in a decrease of total roll angle in the duplex stem of L2L4 and allows the T(4) base to dip deeper into the minor groove compared to the L2L2 model.

Geometrical differences between L2L2 and L2L4

A comparison of the wMD-averaged structures reveals an overall rmsd value of 1.24 Å between L2L2 and L2L4χunconstr and an rmsd value of 1.05 Å between

L2L2 and L2L4χconstr, i.e., values that are indeed larger than the rmsd value of 0.41 Å from the static models (the p1C(1) nucleotide was excluded in all comparisons). In Fig. 8, the superimposed L2L2 and L2L4χunconstr forms are shown. The difference between L2L2 and L2L4 occurs predominantly in the 5'-GTTC-3' loop (rmsd of residues dC(7) to dG(2) (minus p1C(1)): 1.42 Å). The geometry of the 5'-CTTG-3' region is in this respect well preserved in both forms (rmsd of residues dG(2) to dC(7): 0.49 Å).

In conclusion, the smaller number of hydrogen-bonding interactions to residue T(9), two in total compared to four for residue T(4) in L2L2, and the less efficient stacking overlap of the dC(1) base obviously make the 5'-GTTC-3' loop less stable compared to the opposite loop with the reversed base sequence 5'-CTTG-3'. The loss of hydrogen bonds involving the bases located in the minor groove of L2L4 causes the 5'-GTTC-3' loop to be even more destabilized than in the L2L2 model. Conformational analysis based on coupling constants also shows differences in backbone flexibility between the two types of loop, i.e., a resonably well-defined backbone geometry in the 5'-CTTG-3' loop (L2) compared to the more flexible 5'-GTTC-3' loop (L4).

Why syn?

MD and NMR experiments have shown that the 5'-CTTG-3' loop is stable over a large temperature range. In contrast, the dC(1)-dG(8) base pair and, indirectly, the 5'-GTTC-3' loop appear to be unstable and flexible, especially in the L2L4 form in which a *syn* orientation of dC(1) is found. The presence of this *syn* conformation remains

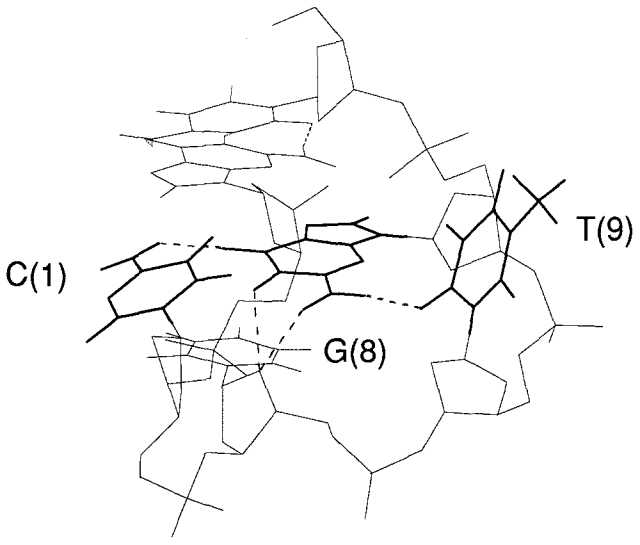


Fig. 7. View of the 5'-GTTC-3' loop in the minimized structure of L2L4, obtained by means of distance constraints and additional constraints on experimentally derived torsion angles. The dC(1)(*syn*)-dG(8)(*anti*) base pair and the T(9) base are highlighted. The hydrogen-bond network is displayed in broken lines and involves hydrogen bonds between NH^{β} dC(1)-O6 dG(8), NH dG(8)-O4' T(10), NH^{α} dG(8)-O4' T(10) and NH^{β} dG(8)-O2 T(9).

difficult to understand. On geometrical grounds, no advantage can be rationalized as to why the molecule wants to adopt a *syn* position in the cytosine dC(1). Firstly, stacking overlap of the dC(1) base in the duplex becomes less pronounced in a *syn* conformer. Secondly, base pairing from dC(1) to dG(8) involves a maximum of one hydrogen bond between O6 dG(8) and NH^{α} dC(1), compared to three H-bonds for Watson–Crick base pairing. Thirdly, upon formation of L2L4 the energy-favorable hydrogen bond between O2 T(9) and NH^{β} dG(8) in the minor groove is partly lost.

According to the wMD calculations, hydration of the O2 and NH_2 atoms of dC(1) in the major groove seems to be a likely mechanism for the stabilization of the *syn* form. In the wMD L2L4χconstr run, two specific water molecules in the major groove solvate the NH_2 dC(1) protons during a period of 60 ps, whereas the O2 atom is solvated by one water molecule. The slow exchange rate between L2L2 and L2L4 is then probably related to the high activation energy needed for the disruption of base stacking of dG(2) and T(10), together with breakage of solvation bridges during rotation around the glycosyl bond of the dC(1) base.

In this respect, the analogy with frayed C-G base pairs at the termini of DNA duplexes is striking. From NOESY spectra, we find that the 5'-terminal cytosine in 5'-CGC... B-DNA duplex stems involves large intraresidue H6-H1' NOE intensities, typical for large populations of *syn* rotamer at room temperature (calculated 30–40% in a two-state model). This stands in contrast to the reversed sequences of the type 5'-GCG..., in which the 5'-terminal guanine residue displays pure *anti* rotamers, as evidenced by the small intraresidual NOE between H8 and H1', about a factor of 10 smaller than the H6-H1' NOE in the 5'-terminal cytosine (J.H. Ippel, unpublished results). Apparently, the guanine base exhibits less flexibility around the glycosidic bond compared to the cytosine residue. This observation is in agreement with the high stability of the '5'-terminus' dG(6) base in the 5'-CTTG-3' loop, relative to the '5'-terminus' dC(1) base in the 5'-GTTC-3' loop.

Interestingly, a *syn* conformation is also found in residues at the 3'-5' loop–stem junction of the DNA hairpins 5'-d(ATCCTA-TTTA-TAGGAT)-3' (Blommers et al., 1991) and 5'-d(ATCCTA-GTTA-TAGGAT)-3' (Hilbers et al., 1991), and in the RNA minihairpins 5'-r(UGAGC-UUCG-GCUC)-3', 5'-r(UGAGC-UUCI-GCUC)-3' and 5'-r(GGAC-UUCG-GUCC)-3' (Sakata et al., 1990; Cheong et al., 1990; Varani et al., 1991; Varani and Tinoco, 1991). The proposed base pairs are of the Hoogsteen T*(anti)*-A*(syn)* and wobble G*(anti)*-A*(syn)* type in the DNA hairpins and the reversed wobble U*(anti)*-G*(syn)*/I*(syn)* type in the RNA hairpins. Structures of the T*(anti)*-A*(syn)* DNA hairpin and the U*(anti)*-G*(syn)* RNA hairpin have been generated by means of distance geom-

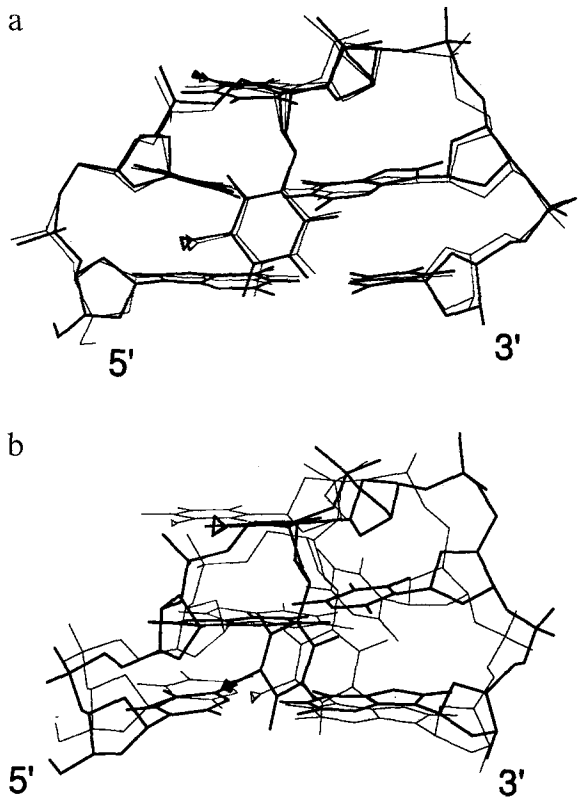


Fig. 8. Views of the wMD-averaged structures of L2L2 (thick line) and L2L4χunconstr (thin line). (a) Superposition of the 5'-CTTG-3' loop; (b) superposition of the 5'-GTTC-3' loop. The rms distance for the two structures is 1.24 Å. Structures superimposed by least-squares fitting of the 5'-GCTTGC-3' loop fragment show an rmsd value for this region of 0.49 Å. The rmsd in a best fit of the opposite 5'-CGTTCG-3' fragment is worse, i.e., 1.42 Å. The pC(1) nucleotide is excluded in all comparisons.

etry and restrained MM and were found to exhibit severe buckle (30°–50°) of the base pair closing the hairpin loop. In both cases, the strained geometry remains in strong disagreement with the high thermodynamic stability observed for these hairpins.

The present study of a dumbbell shows similar buckling of the closing base pair, caused by translating time-averaged NOE data directly into distances with neglect of equilibria between *anti* and *syn* forms. These observations raise doubts about the presence of pure *syn* rotamers instead of *syn*/high-*anti*/*anti* equilibria (at low temperatures) for residue A (equivalent to Hoogsteen ⇄ Watson–Crick base pairing) and residue G (equivalent to reversed wobble ⇄ wobble pairing) in the DNA and RNA hairpin sequences, respectively. Temperature-dependent exchange phenomena on a slow or intermediate time scale have also been found for T*(anti)*-A*(syn)* base pairs in the hairpin 5'-d(CGCGT-ATAT-ACGCG)-3' (Kallick and Wemmer, 1991) and in echinomycin–DNA complexes (Gilbert et al., 1989; Gilbert and Feigon, 1992). Moreover, temperature-dependent line broadening was found specifically for the H1' signal of dG(10) in an RNA hairpin 5'-r(GGGCGC-UGGG-ACGCCCGUC)-3', containing the CUGGGA loop of the HIV-1 TAR element (Colvin et al., 1993). Furthermore, chemical exchange was found for single G*(anti)*-G*(anti)* (Borden et al., 1992), tandem G*(anti)*-A*(anti)* (Lane et al., 1992) and G*(anti)*-εA*(syn)* (De los Santos et al., 1991) base-pair mismatches in DNA duplexes.

Conclusions

From quantitative measurements on distances and coupling constants and molecular dynamics simulations in H_2O, detailed structures of the L2L2 and L2L4 forms of the circular decamer d<CGC-TT-GCG-TT> have been obtained. The L2L2 and L2L4 forms are in slow exchange (rate constant 2–20 s^{-1}), but appear very similar, as indicated by the presence of one set of nonexchangeable proton signals and coupling constants for sugars and backbone, which are invariant to temperature. In the two forms, both hairpin loops are characterized by a sharp turn (β^+/γ^{tr}) in the 3'-5' loop–stem junction and a thymidine located in the minor groove. Three Watson–Crick C*(anti)*-G*(anti)* base pairs are formed in the L2L2 form, but in the L2L4 form the 5'-GTTC-3' loop contains a single hydrogen-bonded G*(anti)*-C*(syn)* base pair.

In order to refine both structures, a NOESY buildup series at 271 K was obtained for a 1:1 conformational mixture of L2L2 and L2L4. Quantitative distances were extracted by the IRMA method, in combination with restrained MD. 'Weakly' restrained molecular dynamics simulations in H_2O at 270 K were subsequently used to sample the internal rotation space. The trajectories of the wMD simulation were used to predict torsional parameters and NOE volumes. Three different types of inclusion of mobility were used in the NOE relaxation calculations: S^2 factors, $\langle r^{-3}\rangle$ matrix averaging and $\langle r^{-6}\rangle$ matrix averaging. Averaging over $\langle r^{-3}\rangle$ accounts for the presence of conformational microstates on the picosecond time scale, whereas averaging over $\langle r^{-6}\rangle$ (one to one) of the two $\langle r^{-3}\rangle$-averaged relaxation matrices of the L2L2 and L2L4 forms accounts for the slow exchange between the two forms at 271 K.

The main conclusion from the wMD dynamics trajectories is that the overall folding of the L2L2 and L2L4 forms remains stable within the time span of the simulation. In the L2L2 form, the 5'-GTTC-3' loop is more flexible compared to the 5'-CTTG-3' loop. The breakage of the dC(1)-dG(8) base pair in the L2L4 form amplifies this flexible conformation of the 5'-GTTC-3' loop. The *syn* orientation of the dC(1) base in the L2L4χconstr and L2L4χunconstr wMD simulations leads to divergent structures that are not encountered experimentally. However, the dynamic R-factor calculated for a 1:1 mixture of L2L2 and L2L4χunconstr from wMD shows that in general a good fit per residue is obtained for both the 5'-

CTTG-3' and 5'-GTTC-3' loops (Table 7). Especially the movement of the T(9) base in- and outwards of the minor groove is adequately described as a dynamical process. The interresidue R-factors for residues dG(8) and T(9) drop from 1.15 and 1.22 in the static models to 0.54 and 0.44 in the dynamic models, respectively. The reorientations of the dC(1) base are not predicted very well by the wMD simulations, as shown by the relatively large intraresidue R-factor for dC(1) (R = 0.56) as well as the large interresidue R-factor of dG(2) (R = 0.68) after averaging the NOE matrices of the L2L2 and L2L4 forms. From the presence of a hydrogen-bonded NH^{α} amino proton of dC(1) at higher temperatures, it is likely that in the L2L4 form the dC(1) base is less destacked from the duplex stem than predicted by the various wMD runs. The absence of NOEs between dC(1) and T(10), and the appearance of NOEs between H6 dC(1) and H5'/H5" dG(2) suggest that the dC(1) base in the *syn* domain is partly buckled toward dG(2), with a possibility of adopting high-*anti* conformers. This fact is confirmed by the better fit of simulated NOEs of H6-H2'/H2" dC(1) after inclusion of a multiconformational state around the χ torsion angle of dC(1) from *syn* to high-*anti* and *anti*.

The detailed structures of the two loops in the L2L2 form differ predominantly by the geometry in the 3'-5' stem–loop junction, dC(3) and dG(8). In dC(3) the torsion angle combination $\alpha^{-}/\beta^{+}/\gamma^{tr}$ occurs, while in residue dG(8) the normal $\alpha^{-}/\beta^{tr}/\gamma^{+}$ geometry is formed. The advantage of the $\alpha^{-}/\beta^{+}/\gamma^{tr}$ combination in dC(3) is the widening of the minor groove at the position of the dangling T(4) base. Furthermore, the poor stacking overlap between the dC(1)-dG(8) and dG(2)-dC(7) base pairs, with respect to the better stacking overlap of the dC(3)-dG(6) closing base pair in the opposite loop, results in partial destacking of the dC(1) base. Additionally, three hydrogen bonds are formed between the T(4) base of the 5'-CTTG-3' loop and the minor groove. According to the wMD simulations, the 5'-GTTC-3' loop contains only one hydrogen bond between the dangling T(9) base and the minor groove, and this bond is lost upon formation of the L2L4 form.

The structural results explain the lower thermal stability of hairpins containing a G-C closing base pair compared to hairpins with the reversed C-G base pair (Ippel et al., 1995a).

Acknowledgements

This research was financially supported by the Italian Council of National Research and MURST (60 and 40%, respectively) and by the Netherlands Foundation of Chemical Research (SON). Part of this work was carried out under the auspices of the Leiden Research School BIOMAC (Structure and Function of Biomacromolecules). An EMBO fellowship to V.L. is gratefully acknowledged. We thank Dr. C.J.M. Huige for assistance with the computational methods.

References

Altona, C., Van Boom, J.H., De Jager, J.R., Koeners, H.J. and Van Binst, G. (1974) *Nature*, **247**, 558–561.

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

Aqvist, J. (1990) *J. Phys. Chem.*, **94**, 8021–8024.

Berendsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) *J. Chem. Phys.*, **81**, 3684–3690.

Bhattacharyya, D. and Bansal, M. (1990) *J. Biomol. Struct. Dyn.*, **8**, 539–572.

Blommers, M.J.J., Van de Ven, F.J.M., Van der Marel, G.A., Van Boom, J.H. and Hilbers, C.W. (1991) *Eur. J. Biochem.*, **201**, 33–51.

Blommers, M.J.J., Nanz, D. and Zerbe, O. (1994) *J. Biomol. NMR*, **4**, 595–601.

Bodenhausen, G. and Ruben, D.J. (1980) *Chem. Phys. Lett.*, **69**, 185–189.

Boelens, R., Koning, T.M.G. and Kaptein, R. (1988) *J. Mol. Struct.*, **173**, 299–311.

Boelens, R., Koning, T.M.G., Van der Marel, G.A., Van Boom, J.H. and Kaptein, R. (1989) *J. Magn. Reson.*, **82**, 290–308.

Borden, K.L.B., Jenkins, T.C., Skelly, J.V., Brown, T. and Lane, A.N. (1992) *Biochemistry*, **31**, 5411–5422.

Bothner-By, A.A. and Castellano, S. (1964) *J. Chem. Phys.*, **41**, 3863–3869.

Chen, C.-W. and Cohen, J.S. (1984) In *Phosphorus-31 NMR: Principles and Applications* (Ed., Gorenstein, D.G.), Academic Press, New York, NY, pp. 233–263.

Cheong, C., Varani, G. and Tinoco Jr., I. (1990) *Nature*, **346**, 680–682.

Clore, G.M., Gronenborn, A.M., Brünger, A.T. and Karplus, M. (1985) *J. Mol. Biol.*, **186**, 435–455.

Colvin, R.A., White, S.W., Garcia-Blanco, M.A. and Hoffman, D.W. (1993) *Biochemistry*, **32**, 1105–1112.

Davison, A. and Leach, D.R.F. (1994) *Nucleic Acids Res.*, **22**, 4361–4363.

De los Santos, C., Kouchakdjian, M., Yarema, K., Basu, A., Essigmann, J. and Patel, D.J. (1991) *Biochemistry*, **30**, 1828–1835.

De Leeuw, F.A.A.M. and Altona, C. (1983) *J. Comput. Chem.*, **4**, 428–437.

Delcourt, S.G. and Blake, R.D. (1991) *J. Biol. Chem.*, **266**, 15160–15169.

EMBO Workshop on DNA Curvature and Bending (1989) *EMBO J.*, **8**, 1–4.

Friedman, R.A. and Honig, B. (1992) *Biopolymers*, **32**, 145–159.

Gilbert, D.E., Van der Marel, G.A., Van Boom, J.H. and Feigon, J. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 3006–3010.

Gilbert, D.E. and Feigon, J. (1992) *Nucleic Acids Res.*, **20**, 2411–2420.

Gonzalez, C., Rullmann, J.A.C., Bonvin, A.M.J.J., Boelens, R. and Kaptein, R. (1991) *J. Magn. Reson.*, **91**, 659–664.

Gorenstein, D.G. (1984) In *Phosphorus-31 NMR: Principles and Applications* (Ed., Gorenstein, D.G.), Academic Press, New York, NY, pp. 7–36.

Hilbers, C.W., Blommers, M.J.J., Van de Ven, F.J.M., Van Boom, J.H. and Van der Marel, G.A. (1991) *Nucleos. Nucleot.*, **10**, 61–80.

Hilbers, C.W., Heus, H.A., Van Dongen, M.J.P. and Wijmenga, S.S. (1994) In *Nucleic Acids and Molecular Biology*, Vol. 8 (Eds, Eckstein, F. and Lilley, D.M.J.), Springer, Berlin, pp. 56–104.

Ippel, J.H., Lanzotti, V., Galeone, A., Mayol, L., Van den Boogaart, J.E., Pikkemaat, J.A. and Altona, C. (1992) *J. Biomol. Struct. Dyn.*, **9**, 821–836.

Ippel, J.H., Lanzotti, V., Galeone, A., Mayol, L., Van den Boogaart, J.E., Pikkemaat, J.A. and Altona, C. (1995a) *Biopolymers*, **36**, 681–694.

Ippel, J.H., Lanzotti, V., Galeone, A., Mayol, L., Van den Boogaart, J.E., Pikkemaat, J.A. and Altona, C. (1995b) *Biopolymers*, **36**, 701–710.

IUPAC-IUB Joint Commission on Biochemical Nomenclature (1983) *Eur. J. Biochem.*, **131**, 9–15.

Kallick, D.A. and Wemmer, D.E. (1991) *Nucleic Acids Res.*, **19**, 6041–6046.

Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. and Van Gunsteren, W.F. (1985) *J. Mol. Biol.*, **182**, 179–182.

Kennard, O. and Hunter, W.N. (1989) *Nucleic Acids Vol. 1a, Crystallographic and Structural Data I*, New Series, Group VII, Springer, Berlin, pp. 255–360.

Klug, A., Jack, A., Vismamitra, M.A., Kennard, O., Shakked, Z. and Steitz, T.A. (1979) *J. Mol. Biol.*, **131**, 669–680.

Koning, T.M.G., Boelens, R. and Kaptein, R. (1990) *J. Magn. Reson.*, **90**, 111–123.

Koning, T.M.G., Boelens, R., Van der Marel, G.A., Van Boom, J.H. and Kaptein, R. (1991a) *Biochemistry*, **30**, 3787–3797.

Koning, T.M.G., Van Soest, J.J.G. and Kaptein, R. (1991b) *Eur. J. Biochem.*, **195**, 29–40.

Lane, A.N. (1990) *Biochim. Biophys. Acta*, **1049**, 205–212.

Lane, A., Martin, S.R., Ebel, S. and Brown, T. (1992) *Biochemistry*, **31**, 12087–12095.

Lankhorst, P.P., Haasnoot, C.A.G., Erkelens, C. and Altona, C. (1984) *J. Biomol. Struct. Dyn.*, **1**, 1387–1405.

Lankhorst, P.P., Haasnoot, C.A.G., Erkelens, C., Westerink, H.P., Van der Marel, G.A., Van Boom, J.H. and Altona, C. (1985) *Nucleic Acids Res.*, **13**, 927–942.

Lavery, R. and Sklenar, H. (1988) *J. Biomol. Struct. Dyn.*, **6**, 63–91.

Lipari, G. and Szabo, A. (1982a) *J. Am. Chem. Soc.*, **104**, 4546–4559.

Lipari, G. and Szabo, A. (1982b) *J. Am. Chem. Soc.*, **104**, 4559–4570.

Macura, S., Huang, Y., Suter, D. and Ernst, R.R. (1981) *J. Magn. Reson.*, **43**, 259–281.

Marion, D. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967–974.

Metzler, W.J., Wang, C., Kitchen, D.B., Levy, R.M. and Pardi, A. (1990) *J. Mol. Biol.*, **214**, 711–736.

Orbons, L.P.M., Van der Marel, G.A., Van Boom, J.H. and Altona, C. (1986) *Nucleic Acids Res.*, **14**, 4187–4196.

Orbons, L.P.M., Van der Marel, G.A., Van Boom, J.H. and Altona, C. (1987a) *J. Biomol. Struct. Dyn.*, **4**, 939–963.

Orbons, L.P.M., Van Beuzekom, A.A. and Altona, C. (1987b) *J. Biomol. Struct. Dyn.*, **4**, 965–987.

Pearlman, D.A. and Kollman, P.A. (1991) *J. Mol. Biol.*, **220**, 457–459.

Pieters, J.M.L., De Vroom, E., Van der Marel, G.A., Van Boom, J.H., Koning, T.M.G., Kaptein, R. and Altona, C. (1990) *Biochemistry*, **29**, 788–799.

Powers, R., Jones, C.R. and Gorenstein, D.G. (1990) *J. Biomol. Struct. Dyn.*, **8**, 253–294.

Rance, M., Bodenhausen, G., Wagner, G., Wüthrich, K. and Ernst, R.R. (1985) *J. Magn. Reson.*, **62**, 497–510.

Roontga, V.A., Jones, C.R. and Gorenstein, D.G. (1990) *Biochemistry*, **29**, 5245–5258.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer, New York, NY.

Sakata, T., Hiroaki, H., Oda, Y., Tanaka, T., Ikehara, M. and Uesugi, S. (1990) *Nucleic Acids Res.*, **18**, 3831–3839.

Sarma, R.H., Mynott, R.J., Wood, D.J. and Hruska, F.E. (1973) *J. Am. Chem. Soc.*, **95**, 6457–6459.

Schmieder, P., Zimmer, S. and Kessler, H. (1991) *Magn. Reson. Chem.*, **29**, 375–380.

Schmieder, P., Ippel, J.H., Van den Elst, H., Van der Marel, G.A., Van Boom, J.H., Altona, C. and Kessler, H. (1992) *Nucleic Acids Res.*, **20**, 4747–4751.

Schmitz, U., Sethson, I., Egan, W.M. and James, T.L. (1992) *J. Mol. Biol.*, **227**, 510–531.

Seibel, G. (1989) *AMBER 3.0 Revision A Program Manual*, University of California at San Francisco, San Francisco, CA.

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

Torda, A.E., Scheek, R.M. and Van Gunsteren, W.F. (1989) *Chem. Phys. Lett.*, **157**, 289–294.

Torda, A.E., Scheek, R.M. and Van Gunsteren, W.F. (1990) *J. Mol. Biol.*, **214**, 223–235.

Van Beuzekom, A.A. (1989) Ph.D. Thesis, Leiden University, Leiden.

Van Gunsteren, W.F. and Berendsen, H.J.C. (1977) *Mol. Phys.*, **34**, 1311–1327.

Van Wijk, J., Huckriede, B.D., Ippel, J.H. and Altona, C. (1992) Methods Enzymol., **211**, 286–306.

Varani, G., Cheong, C. and Tinoco Jr., I. (1991) *Biochemistry*, **30**, 3280–3289.

Varani, G. and Tinoco Jr., I. (1991) *J. Am. Chem. Soc.*, **113**, 9349–9354.

Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A. (1986) *J. Comput. Chem.*, **7**, 230–252.

Widmer, H. and Wüthrich, K. (1986) *J. Magn. Reson.*, **70**, 270–279.

Williamson, J.R. and Boxer, S.G. (1989a) *Biochemistry*, **28**, 2819–2831.

Williamson, J.R. and Boxer, S.G. (1989b) *Biochemistry*, **28**, 2831–2836.

Yuan, H., Quintana, J. and Dickerson, R.E. (1992) *Biochemistry*, **31**, 8009–8021.