Structures and Stabilities of Small DNA Dumbbells with Watson – Crick and Hoogsteen Base Pairs

Nuria Escaja,[b] Irene Gómez-Pinto,[a] Manuel Rico,[a] Enrique Pedroso,[b] and Carlos González*[a]

The structures and stabilities of cyclic DNA octamers of different sequences have been studied by NMR and CD spectroscopy and by restrained molecular dynamics. At low oligonucleotide concentrations, some of these molecules form stable monomeric structures consisting of a short stem of two base pairs connected by two mini-loops of two residues. To our knowledge, these dumbbell-like structures are the smallest observed to date. The relative stabilities of these cyclic dumbbells have been established by studying their melting transitions. Dumbbells made up purely of GC stems are more stable than those consisting purely of AT base pairs. The order of the base pairs closing the loops also has an important effect on the stabilities of these structures. The NMR data indicate that there are significant differences between the solution structures of dumbbells with G – C base pairs in the stem compared to those with A – T base pairs. In the case of dumbbells with G – C base pairs, the residues in the stem form a short segment of a B-DNA helix stabilized by two Watson – Crick base pairs. In contrast, in the case of d⟨pCATTCATT⟩, the stem is formed by two A – T base pairs with the glycosidic angles of the adenine bases in a syn *conformation, most probably forming Hoogsteen base pairs. Although the conformations of the loop residues are not very well defined, the thymine residues at the first position of the loop are observed to fold back into the minor groove of the stem.*

KEYWORDS:

DNA hairpin · DNA structures · Hoogsteen base pair · NMR spectroscopy · oligonucleotides

Introduction

In addition to the regular right-handed double helices, DNA may fold into a wide variety of structures, which include hairpins, cruciforms, triplexes, quadruplexes, and so on. Among these motifs, DNA hairpins have received considerable attention because of their possible implication in several biological processes. For this reason, the structures of many DNA hairpins have been determined in the last decade.[1, 2] Most of these studies have been carried out by means of NMR methods, and have focused on the most stable hairpins. In many cases, however, study by NMR methods is impeded by the flexibility inherent in short oligonucleotide fragments. In theses cases, cyclic oligonucleotides can be used as more suitable models than their linear analogues since the conformational constraints induced by cyclization of the chain increase the stabilities of the structures. Moreover, cyclization prevents the formation of dimeric duplex structures that may coexist with the hairpin forms. Although it has been shown that this assumption is not always valid,[3] cyclic oligonucleotides of moderate size are usually monomeric. Since the melting behavior of monomeric species can be analyzed more easily than that of dimeric duplexes, these cyclic molecules have been extensively used as convenient models for thermodynamic studies of sequence-dependent interactions in DNA.[4–10]

Cyclic oligonucleotides are also interesting because of their potential use in regulation of gene expression, as antisense or antigene compounds. These molecules display increased nuclease resistance and cellular uptake in relation to their linear analogues and are good candidates for diagnostic or chemotherapeutic agents.[11] In addition, it has been proposed that these molecules might serve as regulators of gene expression by acting as decoys for specific transcription factors.[12, 13]

DNA hairpins with mini-loops of only two residues have been studied since the early 1990s[14, 15] and have been classified into different types depending on the conformations of the unpaired bases in the loops.[2, 16] In addition, it has been shown that two-base-loop hairpins can occur in vivo,[17] and it has been suggested that these structures may play roles in gene regulation, recombination, or mutagenesis. Small DNA loops have also been studied in dumbbell-like cyclic oligonucleotides.[18] These molecules consist of two loops of unpaired nucleotides connected by a central stem of double-helical

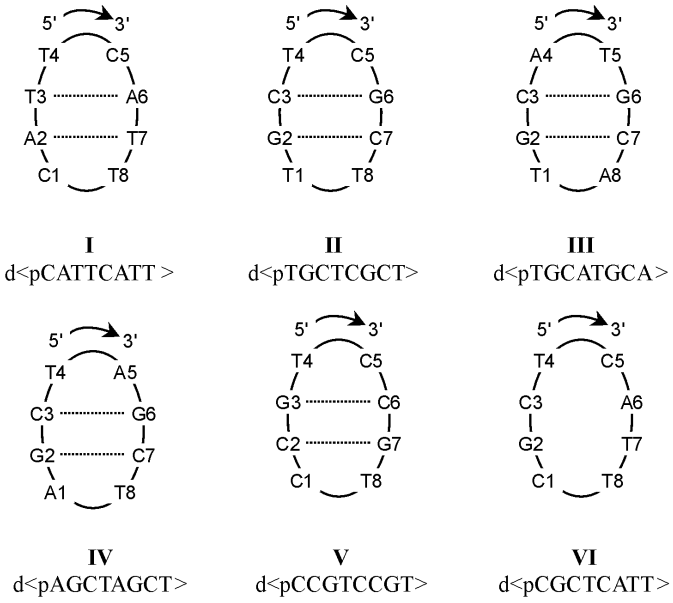
[a] *Dr. C. González, I. Gómez-Pinto, Prof. M. Rico*
Instituto de Química-Física
Rocasolano (C.S.I.C.)
Serrano, 119, 28006-Madrid (Spain)
Fax: (+34) 91564-2431
E-mail: cgonzalez@iqfr.csic.es

[b] *Dr. N. Escaja, Prof. E. Pedroso*
Departament de Química Orgànica
Universitat de Barcelona
C/. Martì I Franquès 1 – 11
08028-Barcelona (Spain)

 DOI: 10.1002/cbic.200300578

DNA. One of the most studied dumbbells is the decamer d⟨pCGCTTGCGTT⟩.[19] The structure and thermodynamics of this oligonucleotide have been extensively studied by various methods.[20–22] In solution, this molecule adopts two conformations. The structure of the first conformer consists of a central stem formed by three Watson – Crick G – C base pairs connected by two mini-loops made up of two thymine residues. In the other conformer, the closing GC pair in the 5′-GTTC-3′ sequence is partially disrupted and forms a G(*anti*) – C(*syn*) pair with only one hydrogen bond. The rest of the nucleotides in the stem form canonical Watson – Crick base pairs.

In a previous work we used small cyclic oligonucleotides as models to study four-stranded DNA structures in solution.[23] Cyclic octamers can form dimers, which are stabilized by intermolecular Watson – Crick base pairs. However, we found that under certain conditions, these dimers coexist in equilibrium with monomeric forms.[3] Surprisingly, some of these monomeric octamers, which can form only two intramolecular base pairs, are quite stable. Herein we report the structures and thermodynamic behavior of the monomeric forms of several cyclic octamers studied by NMR and CD spectroscopy. In four cases, the NMR data are consistent with dumbbell-like structures, each formed from a stem of two base pairs and two mini-loops of two residues. In the case of dumbbells formed by G – C pairs, the stems each consist of a short segment of B-form DNA. When the stem is made up of A – T pairs, however, the glycosidic angle of the adenine residues adopt a *syn* conformation, which suggests the formation of Hoogsteen base pairs. This structure resembles a short segment of the antiparallel Hoogsteen DNA helix recently found in the crystal structure of $d(ATATAT)_2$.[24]



Scheme 1. *Cyclic oligonucleotides studied in this paper.*

Results

Equilibrium between monomeric and dimeric forms

In a previous publication we showed that the ^{1}H NMR spectra of the cyclic octamers d⟨pCATTCATT⟩ (dumbbell **I** in this study) and d⟨pTGCTCGCT⟩ (dumbbell **II**) depend strongly on the oligonucleotide concentration.[3] This dependency results from an association equilibrium between the monomeric and dimeric forms of these molecules that is slow on the NMR time scale. Similar behavior was observed for all the cyclic octamers studied in this paper (see Scheme 1). Except in the case of **I**, the populations of dimeric forms are significant at the oligonucleotide concentrations usually employed in NMR studies. The populations of the dimers increase at high ionic strengths. Experiments were therefore carried out at low salt concentrations to reduce the populations of the dimeric forms.

Circular dichroism and thermal denaturation

The thermal denaturation of the five cyclic oligonucleotides shown in Scheme 1 was monitored by UV spectroscopy and circular dichroism (CD) spectroscopy. As a result of the small sizes of these molecules, the hypochromicity effect caused by the dumbbell structure is low and the change in UV absorption upon dumbbell melting is small. The melting behavior of these molecules is better observed by following the changes in their CD spectra. For dumbbells **I** – **IV**, the main feature of each CD spectrum is the large positive band around 280 nm. As can be seen in Figure 1, this band diminishes drastically with increasing temperature. In the case of d⟨pCCGTCCGT⟩ (**V**; E in Figure 1), however, this band is much smaller than for **I** – **IV** and there is almost no change in the CD spectrum with temperature. The CD spectrum of this molecule is very similar to that of d⟨pCGCTCATT⟩ (**VI**), which cannot form a dumbbell-like structure and is used in this study as a control compound. For dumbbells **I** – **IV**, the changes in ellipticity at λ_{max} with temperature are plotted in Figure 2. These melting curves are concentration-independent and were analyzed by assuming a two-state equilibrium. The thermodynamic parameters obtained are shown in Table 1.

Spectral assignment

Sequential assignments of exchangeable and non-exchangeable protons were carried out by standard ^{1}H NMR methods. Many features of the non-exchangeable proton resonances are common to all the oligonucleotides. With the exception of dumbbell **II**, the sequences are repetitive and therefore only four spin systems are found in each spectrum (residue pairs 1/5, 2/6, 3/7, and 4/8 are degenerate). In the case of **II**, although the sequence is nonrepetitive, all non-exchangeable protons of G2 and G6 are degenerate. In addition, the proton chemical shifts in residues C3 and C7 are very similar but not all identical. Spin systems were identified in the TOCSY and COSY spectra in D_2O. The base protons of cytosine and thymine residues were identified by the H5-H6 or Met-H6 cross-peaks in the TOCSY spectra and connected to their sugar spin systems by the H6-H1′ cross-peak in the NOESY spectra. All intranucleotide H1′-base NOEs are medium or weak, which indicates that the glycosidic angles in all the nucleotides are those found in *anti* conforma-

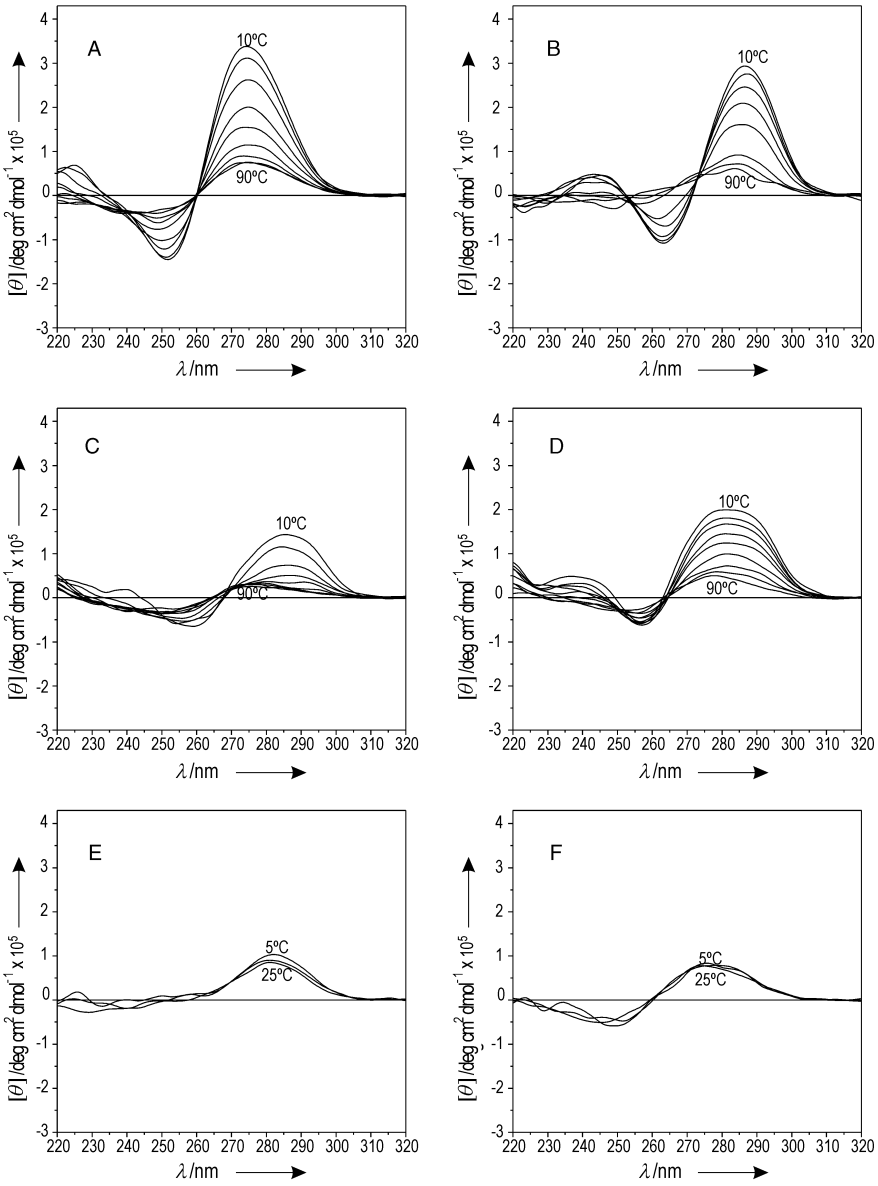


Figure 1. *CD spectra of the cyclic oligonucleotides: A) d⟨pCATTCATT⟩, B) d⟨pTGCTCGCT⟩, C) d⟨pAGCTAGCT⟩, D) d⟨pTGCATGCA⟩, E) d⟨pCCGTCCGT⟩, F) d⟨pCGCTCATT⟩.*

tions. The only exception is the case of dumbbell **I**, for which the intraresidual NOE between H1′ and H8 of the two equivalent adenine residues is particularly strong (its intensity at all mixing times is comparable to the NOE between H5 and H6 of C1). This strong NOE is clear evidence of a glycosidic angle of a *syn* conformation. Regions of the two-dimensional NOESY spectrum of dumbbell **I** illustrating these assignments are shown in Figure 3. The strong H1′-H8 NOE of A2 can be seen in the panel on the right of the figure.

Strong sequential sugar – base connections were observed between residues 2 and 3 (and the equivalent 6 and 7) in all the molecules, with the exception of dumbbell **V**. Sequential connections with other residues were weak or not observable; nevertheless complete sequential assignment could be carried out, except in the case of **V**. The adenine H2 protons were identified by their long spin-lattice relaxation times. Almost all the resonances were identified, including the assignment of some H5′/H5′′ protons. The complete assignment lists are shown in Table S1 in the Supporting Information and have been deposited at the BioMagResBank (BMRB).[25] In the case of dumbbell **V**, the poor chemical shift dispersion prevented the sequential assignment of protons H3′, H4′, H5′, and H5′′. Exchangeable protons were not detected for this oligonucleotide.

Exchangeable protons could be observed in the H_2O spectra of dumbbells **I** and **II**. In the case of **II**, the two imino signals of the guanine residues are not completely degenerate and their chemical shifts are characteristic of Watson – Crick base pairs. These signals were specifically assigned through their NOESY cross-peaks

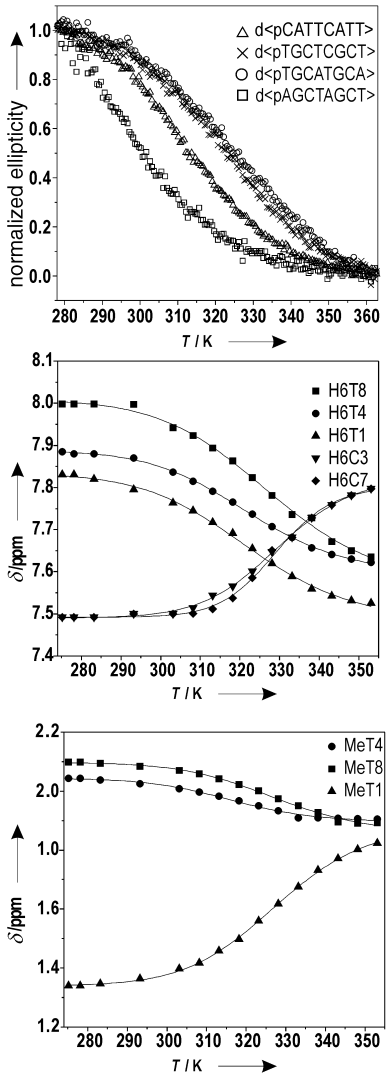


Figure 2. *Top: CD melting curves of the cyclic oligonucleotides d⟨pCATTCATT⟩, d⟨pTGCTCGCT⟩, d⟨pAGCTAGCT⟩, and d⟨pTGCATGCA⟩. Middle and bottom: NMR melting curves for the aromatic and methyl protons of d⟨pTGCTCGCT⟩.*

Table 1. *Thermodynamic parameters for dumbbell formation.*[a]

Sequence	T_m (NMR) [K]	ΔG^0_{298} (NMR) [kJ mol^{-1}]	ΔH^0 (NMR) [kJ mol^{-1}]	T_m (CD) [K]	ΔG^0_{298} (CD) [kJ mol^{-1}]	ΔH^0 (CD) [kJ mol^{-1}]
d⟨pCATTCATT⟩	311 ± 2	−3 ± 1	−77 ± 4	312 ± 1	−3 ± 1	−68
d⟨pTGCTCGCT⟩	324 ± 2	−6 ± 1	−79 ± 3	325 ± 1	−6 ± 1	−70
d⟨pTGCATGCA⟩	321 ± 4	−5 ± 1	−79 ± 5	327 ± 1	−5 ± 1	−69
d⟨pAGCTAGCT⟩	291 ± 3	1.3 ± 1	−76 ± 11	≈294	≈1.0	≈−68
d⟨pCCGTCCGT⟩			No structure			

[a] T_m, melting temperature; ΔG, free energy change; ΔH, entropy change.

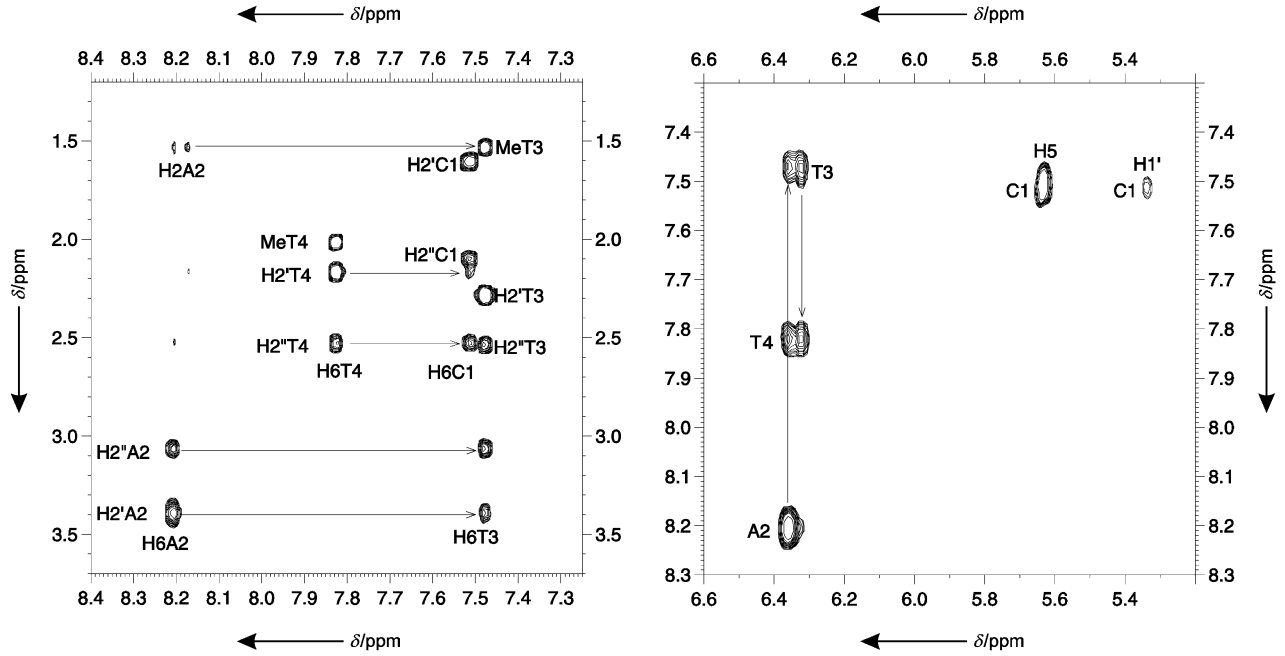


Figure 3. *Regions of the NOESY spectrum of d⟨pCATTCATT⟩ in D_2O. The aromatic-H2′/2″ region (left panel) and aromatic-H1′ region (right panel) are shown. Cross-peaks are labeled according to the numbering in Scheme 1. In the right-hand panel, only intraresidual aromatic-H1′ cross-peaks are labeled. Note that the residue pairs 1/5, 2/6, 3/7, and 4/8 are degenerate.*

with their base-paired cytosine moieties. The three thymine imino protons resonate at 10 – 12 ppm, which indicates that these bases are not paired. In the case of **I**, two imino protons were observed (13.3 and 10.9 ppm). The lower field imino moiety was assigned to the base-paired thymine residues, T3 and T7. Unfortunately, this signal is broad, and no cross-peak could be found either with H2 or with H8 of the base-paired adenine residue. Interestingly, the linewidth corresponding to these protons is much larger than those corresponding to the unpaired thymine moieties (see the Supporting Information). This result may indicate that another effect, not solvent exchange, is responsible for the observed line broadening. The behavior of the linewidths of these signals at different temperatures suggests a dynamic equilibrium involving the H3 proton of T3 (and the equivalent T7).

Thermal denaturation monitored by NMR spectroscopy

Thermal denaturation was monitored by measurement of the chemical shift variation with temperature for different protons. Examples of these chemical shift profiles are shown in Figure 2. As can be observed in Figure 2 (and in more detail in Table S2), the melting temperatures for different protons are similar, which indicates that the transition between the dumbbell and a close random coil structure is mainly cooperative. In all cases, the average melting temperatures obtained by NMR spectroscopy are almost coincident with those estimated from CD data. Since these experiments were carried out at very different oligonucleotide concentrations, the agreement between melting temperatures indicates that the transition corresponds to a monomolecular process. Under these circumstances, the thermodynamic parameters can be directly estimated from the melting profiles.

The ΔG values obtained from CD and NMR data indicate that dumbbells **II** and **III** are the most stable, followed by dumbbell **I**. The stability of dumbbell **IV** is only marginal, with positive ΔG values at room temperature.

Structure calculation

The solution structures of dumbbells **I** and **II** were calculated from experimental distance constraints derived from NOE

intensities, and torsion angle constraints derived from *J* coupling constants. The structures of dumbbells **III** and **IV** could not be calculated because the number of experimental constraints that could be extracted from the spectra was very small. In the first case, the low stability of the dumbbell impeded the acquisition of a good-quality NOESY spectrum. In the second case, the dumbbell is more stable, but the presence of dimeric structures at a relatively low oligonucleotide concentration prevented the acquisition of good spectra of the monomeric form.

For dumbbells **I** and **II**, 186 and 70 distance constraints, respectively, could be derived from NOESY experiments (see Table 2). Most of the sequential constraints are mainly located in the step 2 →3 (and the equivalent step 6 →7), which indicates that these bases are stacked. In the case of **I**, no constraints involving exchangeable protons could be obtained because of the large linewidth of the imino protons of the base-paired thymine residues.

Table 2. *NMR restraints and structural statistics.*

Experimental distance constraints	Dumbbell **I** d⟨pCATTCATT⟩	Dumbbell **II** d⟨pTGCTCGCT⟩
Total number	186	70
intraresidue	120	36
interresidue	66	34
RMSD[a]		
bases central stem[b] [Å]	0.4 ± 0.1	0.5 ± 0.2
all heavy atoms central stem[b] [Å]	0.7 ± 0.2	0.8 ± 0.2
all heavy atoms [Å]	1.8 ± 0.7	2.5 ± 0.9
Residual violations		
sum of violations [Å]	13.7	1.0
max. violation [Å]	0.7	0.4
Average NOE energy [Kcal mol^{-1}]	51	6
range of NOE energies [Kcal mol^{-1}]	45 – 58	2 – 7

[a] Root mean square deviation. [b] Residues 2, 3, 6, and 7.

Coupling constants, *J*, for deoxyribose moieties were estimated from DQF-COSY experiments, and in all cases the *J* couplings are consistent with sugar puckers predominantly in the S domain (see Table S3). No stereospecific assignments for the H5′/H5″ protons could be performed, and consequently the γ angle could not be determined. However, some dihedral angle constraints could be obtained from analysis of the $^{31}P-^{1}H$ correlation experiments. Since all the ^{31}P resonances lie between −4.25 and −3.70 ppm, we can conclude that no α or ζ angle is that found in a *trans* conformation.[26] In addition, the small heteronuclear ^{3}J couplings between the phosphorus atom and the two H5′/5″ protons of A2 and T4 indicates that the β angles of these (and the equivalent residues A6 and T8) are those of a *trans* conformation.

These experimental constraints were used to calculate the structures by restrained molecular dynamics. Ten structures were calculated with the aid of the program DYANA 1.5,[27] and these structures were refined by use of the molecular dynamics package AMBER 5.1,[28] as explained in the Materials and Methods section. In the AMBER refinement, ten structures were calculated in vacuo by a high-temperature annealing procedure. The resulting structures were further refined by means of 220-ps NMR-restrained molecular dynamics trajectories including explicit solvent terms and a better evaluation of the electrostatic term. It has been shown that these calculations improve the quality of the final NMR-derived structures.[29–31] The final structures converged to a well-defined region of the conformational space, with mutual RMS deviations lower than 1.0 Å for the central stems of both molecules (see Table 2). Residual distance constraint violations are small, which demonstrates that the structures satisfy the experimental constraints. The higher number of distance constraints means that the residual violations in dumbbell **I** are larger than those in dumbbell **II**. Most of these violations involve thymine residues 4 and 8 in the loops.

Solution structures of d⟨pCATTCATT⟩ and d⟨pTGCTCGCT⟩

The ten structures resulting from the final refinement in water are displayed in Figure 4 and Figure 5. As expected, the

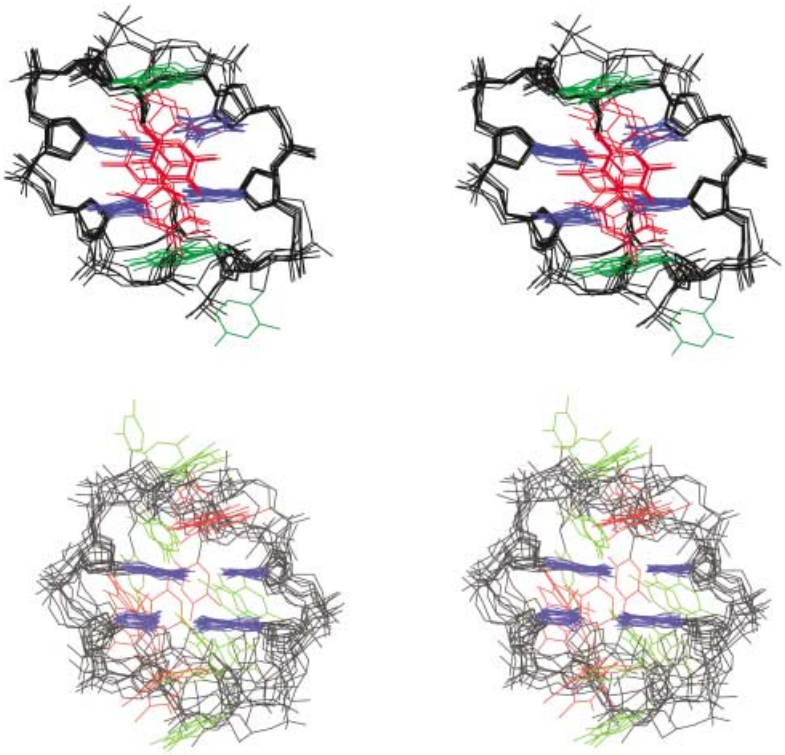


Figure 4. *Stereoscopic views of the superpositions of ten structures of the monomeric forms of (top) d⟨pCATTCATT⟩ (**I**) and (bottom) d⟨pTGCTCGCT⟩ (**II**). The sugar – phosphate backbones are shown in black. Nucleobases of the base-paired residues are in blue, and those in the two mini-loops are shown in green (1, 5) and red (4, 8).*

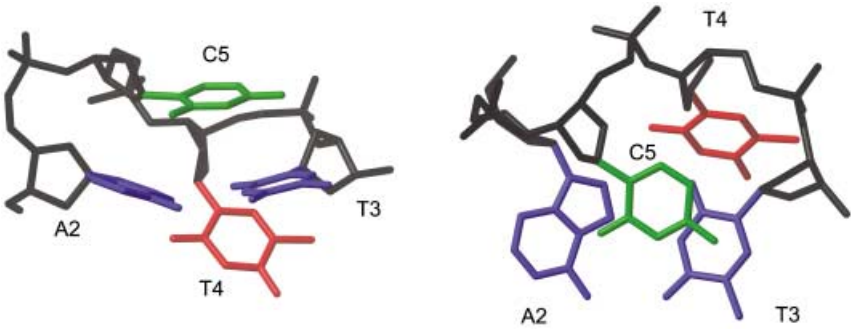


Figure 5. *Detail of the loops of d⟨pCATTCATT⟩.*

structures of **I** and **II** are made up of a short double-helical region of two base pairs connected by two-base loops. In the case of **II**, the base pairs are Watson – Crick, all the glycosidic angles are those of *anti* conformations, and the pseudorotation phase angles of the deoxyribose units are in the general S domain (see Table 3). The residues in the loop are disordered because of the low number of experimental constraints in these regions.

In the case of **I**, the resulting structures each consist of a two-residue stem with two Hoogsteen T(*anti*) – A(*syn*) base pairs. The structures of the loops are better defined than in the case of **II**. The residues in the second position (in the 5′ to 3′ direction), C1 and C5, stack on top of the Hoogsteen base pairs, while those in the first position, T4 and T8, turn towards the minor groove of the stem (see Figure 5). These latter two residues are more disordered than C1 and C5, and adopt two families of conformations, depending on which thymine residue (T4 or T8) enters more deeply into the minor groove (see Figure 4). Thymines 4 and 8 exhibit the largest distance constraint violations (the sum of residual violations for them is approximately 10 Å), which indicates that none of the structures is totally compatible with the experimental constraints. The observation of two families of conformations for these residues may be a consequence of the short length of the double-helical stem, which is too small to accommodate two thymine bases.

In each of the two structures the phosphate backbone is more disordered than the bases, as is usually observed in NMR-derived oligonucleotide structures. The average values for the backbone dihedral angles are shown in Table 3. Most of the well-defined angles (those with order parameters larger than 0.85) are in the stem of the dumbbells and adopt values common in right-handed double-helical DNA.

Discussion

The CD and NMR data show that octameric cyclic oligonucleotides can form stable, dumbbell-like structures in solution. The structures and stabilities of these small DNA dumbbells depend strongly on the nucleotide sequence. A more detailed structural analysis of two of these molecules, d⟨pCATTCATT⟩ (**I**) and d⟨pTGCTCGCT⟩ (**II**), shows that they adopt very different conformations. Whereas the stem in dumbbell **II** is a very short segment of a B-DNA helix, in the case of **I** these residues form a short antiparallel helix with Hoogsteen base pairs. In the case of **II** and of other dumbbells with G – C pairs in their stems, direct evidence of Watson – Crick association is provided by the NOE cross-peaks between the guanine imino protons and the cytosine amino groups. In contrast, the imino resonance corresponding to the base-paired thymine residues in dumbbell **I** is broad and no cross-peaks with the adenine H8 could be

Table 3. *Geometric parameters.*[a]

	Dumbbell **I** d⟨pCATTCATT⟩														
	α		β		γ		δ		ε		ζ		χ		P_S
	Av	O.P	Av	O.P.	Av	O.P.	Av	O.P.	Av	O.P.	Av	O.P.	Av	O.P.	
1C	−68	0.7	169	0.3	127	0.5	147	1.0	−92	0.8	−70	0.8	−155	1.0	165 ± 3
2A	56	0.5	−163	0.5	−168	1.0	118	1.0	−128	1.0	87	0.6	54	1.0	117 ± 2
3T	−101	1.0	51	0.9	175	1.0	149	1.0	−155	1.0	−65	1.0	−123	1.0	172 ± 2
4T	127	0.2	177	0.9	91	0.7	154	1.0	−125	0.8	−112	1.0	−135	1.0	175 ± 3
5C	−79	0.9	168	0.2	167	0.7	147	1.0	−101	0.7	−67	0.8	−154	1.0	172 ± 2
6A	40	0.4	−150	0.3	−168	1.0	112	1.0	−146	0.8	100	0.3	49	1.0	163 ± 2
7T	112	0.8	75	0.8	−166	1.0	144	1.0	−156	1.0	−111	0.8	−120	1.0	113 ± 2
8T	98	0.1	179	0.9	99	0.6	149	1.0	−133	0.8	−116	1.0	−139	1.0	169 ± 4
	Dumbbell **II** d⟨pTGCTCGCT⟩														
	α		β		γ		δ		ε		ζ		χ		P_S
	Av	O.P.	Av	O.P.	Av	O.P.	Av	O.P.	Av	O.P.	Av	O.P.	Av	O.P.	
1T	−102	0.8	170	0.8	60	0.8	141	1.0	−89	0.7	−80	0.7	−85	0.9	153 ± 4
2G	102	0.5	−159	0.3	−164	1.0	127	1.0	−131	0.7	103	0.4	−76	1.0	137 ± 9
3C	−50	0.3	177	0.7	95	0.4	123	1.0	−175	0.7	−164	0.5	−117	1.0	132 ± 5
4T	136	0.4	179	1.0	123	0.4	134	1.0	−174	0.7	−108	0.6	−126	1.0	148 ± 5
5C	−167	0.4	177	1.0	144	0.5	138	1.0	−133	0.7	−89	0.8	−121	0.9	163 ± 8
6G	−168	0.4	180	0.9	162	0.5	119	1.0	−178	0.5	96	0.1	−74	1.0	146 ± 6
7C	−73	0.1	177	0.6	139	0.4	131	1.0	−130	0.6	−92	0.3	−112	0.9	146 ± 13
8T	30	0.4	−135	0.7	−170	0.8	146	1.0	−130	0.8	−143	0.2	−121	0.9	144 ± 5

[a] O.P., angular order parameter

found. However, the NOE data involving non-exchangeable protons clearly demonstrate that the adenine residues adopt *syn* conformations. This conformation is not compatible with the formation of two A – T Watson – Crick base pairs and, consequently, hydrogen bond restrictions corresponding to Hoogsteen base pairs were included in the calculation. Since the residual constraint violations in the final structures are small, and there are no significant distortions, we can conclude that the experimental information is compatible with the formation of two Hoogsteen A – T base pairs. The stem of the dumbbell structure of d⟨pCATTCATT⟩ resembles the antiparallel Hoogsteen helix recently observed in the crystal structure of $d(ATATAT)_2$.[24] Other Hoogsteen T(*anti*) – A(*syn*) base pairs have been observed in chemically modified oligonucleotides,[32] in some protein – DNA complexes,[33] or in the closing base pairs of some DNA hairpins, as in the structure of d(ATCCTA**TTTA**TAGGAT).[15] Hoogsteen A – T base pairs have also been found in DNA complexes with some intercalating drugs.[34]

When Hoogsteen base pairs are observed by NMR methods, they are usually found in dynamic equilibrium with different conformations. This is probably the case in the structure of dumbbell **I**, where the linewidth of the base-paired imino signal at different temperatures is characteristic of a dynamic equilibrium on an intermediate timescale. Similar behavior has been observed in some DNA mini-hairpins, in which this effect has been associated with a conformational exchange between *syn* and *anti* conformations of the glycosidic angle of the purines.[20, 22, 35] In the case of **I**, all the NOEs between non-exchangeable protons are consistent with the adenine glycosidic angles in a *syn* conformation, but a minor population of an *anti* conformer cannot be disregarded.

Although the loop region in the structure of dumbbell **I** is not as well defined as the stem, the conformation of the bases in the loops could be determined. According to the folding topology classification for two- and four-membered loops developed by Hilbers and co-workers,[16] the loops of this dumbbell belong to Type II. In this group the base in the first position of the loop turns into the minor groove and the second base lies over the base pair. In dumbbell **I**, cytosines 1 and 5 stack on top of the A – T base pairs, and thymines 4 and 8 approach the minor groove of the stem. Interestingly, the interaction of the loop thymine residues with the A – T stem resembles the contacts of the extrahelical thymine residues in the antiparallel Hoogsteen helix structure adopted by the oligonucleotide $d(ATATAT)_2$.[24] In this crystallographic structure, the terminal thymine residues of the oligonucleotide are not base paired, but enter into the minor groove of a symmetry-related duplex. This interaction may be important to stabilize the Hoogsteen conformation, since the minor groove is more hydrophobic in an antiparallel Hoogsteen helix than in a canonical Watson – Crick duplex.

In the case of dumbbell **II**, the structure of the molecule could not be obtained in enough detail to determine the conformation of the bases forming the loops. The poor definition of these residues could reflect an inherent flexibility in the loop regions of these molecules, but it could also be a consequence of the small number of structural constraints obtained as a result of the low concentration of the sample. It is interesting to compare the structure of dumbbell **II** with that of the decamer d⟨pCGCTTGCGTT⟩.[22] This cyclic oligonucleotide adopts a dumbbell structure formed by three G – C base pairs in which one of the closing loops has the same sequence as the loops of dumbbell **II**, 5′-C-NN-G-3′, while the other one is 5′-G-NN-C-3′. The structures of the stems in the two molecules are very similar, but in d⟨pCGCTTGCGTT⟩ the loop 5′-C-TT-G-3′ adopts a well-defined structure similar to that found in dumbbell **I**. Very similar loop conformations have been observed in the linear octamers d(CGCTCGCG) and d(CGCCTGCG).[36] In dumbbell **II** the loop conformations are not well defined, although they have almost identical sequences. This might be a consequence of the short helical stem of the octameric dumbbells. Since the two loops are very close to each other in dumbbell **II**, two nucleobases (one from each loop) cannot fit into the minor groove simultaneously. This effect is also observed in dumbbell **I**, but in this case the thymine residues in the loops still approach the minor groove because of its higher hydrophobicity in **I** than in **II**.

Thermodynamic data for this decamer and for the related dumbbell d⟨pGGCTTGCCTT⟩ have been reported.[21] The T_m values are 352 K and 365 K, respectively. Comparison of the latter value with the thermodynamic data obtained for dumbbell **II**, which has similar loops, indicates that the addition of one base pair to the stem increases the melting point by around 40 °C. In the case of d⟨pCGCTTGCGTT⟩, the presence of the less favorable 5′-G-NN-C-3′ loop decreases the T_m value by 13 °C, but this molecule is still 30 °C more stable than dumbbell **II**.

As shown in Table 1, the stabilities of the cyclic octamers studied here depend strongly on the nucleotide sequence. The difference in stability between d⟨pTGCTCGCT⟩ (**II**) and d⟨pCCGTCCGT⟩ (**V**) is particularly dramatic. Although these two octamers might form dumbbells of two G – C base pairs closed by two pyrimidine loops, **II** is very stable ($\Delta G^0_{298} = -6$ kJ mol^{-1}), whereas no structure can be observed for **V**. The molecules differ mainly in the order of the base-paired residues. In **II** the mini-hairpin loops have the sequence 5′-C-NN-G-3′, while in **V** the sequence is 5′-G-NN-C-3′. The difference in stability between **II** and **V** reflects the tendency, observed in other DNA hairpins with larger stems[14, 37] and also in DNA dumbbells,[22] for the sequence 5′-C-NN-G-3′ to adopt two-base loops, whereas the sequence 5′-G-NN-C-3′ has a propensity to form loops of four residues. Since the oligonucleotides studied here are all octamers, dumbbell structures with four-residue loops cannot be formed.

The interpretation of the different stabilities of dumbbells **III** and **IV** is not straightforward. Both molecules are similar to dumbbell **II**, but they have a purine base in each of the loops. Although no structure calculations could be carried out, the available NMR data indicate that these two oligonucleotides adopt dumbbell structures similar to that of dumbbell **II**. Several authors have reported that DNA hairpins with purines in the loops are less stable than hairpins with purely pyrimidine-containing loops,[37, 38] but this effect also depends on other factors, such as the number of residues in the loop, the sequence of the stem, and so on. Here, we observe that dumbbell **III**, with the purine base in the first position of the loop, has a ΔG value very similar to that of dumbbell **II**, with purely pyrimidine-

containing loops. Dumbbell **IV**, however, with the purine in the second position of the loop, is clearly less stable. Similar relative stabilities have been observed in linear mini-hairpins with analogous loop sequences,[36] but the differences we observed are more pronounced. This result is probably due to the small number of base-paired residues in the stem, which makes the effect of the loop residues more dramatic.

Conclusion

The results reported herein show that octameric cyclic oligonucleotides can form monomeric, dumbbell-like structures with two-residue mini-loops. The NMR data and the structural calculations carried out for d⟨pCATTCATT⟩ and d⟨pTGCTCGCT⟩ indicate that the base-paired residues are well defined in both cases, but that they adopt very different conformations. Whereas the stem in d⟨pTGCTCGCT⟩ is a very short segment of a B-DNA helix, in the case of d⟨pCATTCATT⟩ these residues form a short antiparallel helix with Hoogsteen T(*anti*) – A(*syn*) base pairs. In the latter case, the cytosine residues of the loops stack on top of the A – T base pairs while the unpaired thymine residues approach the minor groove of the stem. This interaction may be important for stabilization of the Hoogsteen conformation.

The thermal stabilities of these small dumbbells depend strongly on the nucleotide sequence, which confirms sequence-dependent effects previously observed in DNA hairpins, such as the influence of the closing base pair, or the presence of unpaired purines in the loops. However, the small sizes of these molecules make the impact of these effects on the stability of the dumbbell structures more pronounced.

Methods

Sample preparation: Cyclic oligonucleotides were synthesized by using previously described procedures.[39] Oligonucleotides for NMR samples were suspended either in D_2O or in H_2O/D_2O 9:1 (25 mM phosphate, pH 7.0). The oligonucleotide concentration was adjusted to avoid the presence of dimeric forms in the samples. In the case of **I**, the concentration was around 2 mM, but in other cases NMR spectra were recorded at oligonucleotide concentrations below 0.5 mM.

NMR spectroscopy: NMR spectra were acquired in a Bruker DMX spectrometer operating at 600 Hz, and processed with XWIN-NMR software. DQF-COSY, ^{31}P-decoupled DQF-COSY, TOCSY, and NOESY experiments were recorded in D_2O. NOESY experiments were also acquired in 90% H_2O/10% D_2O. NOESY[40] spectra of samples in D_2O were acquired with mixing times of 100 and 250 ms, and those of samples in H_2O with a mixing time of 200 ms. TOCSY[41] spectra were recorded with a standard MLEV-17 spin-lock sequence and 80-ms mixing time. ^{31}P resonances were assigned from proton-detected heteronuclear correlation spectra.[42] In the experiments in D_2O, presaturation was used to suppress the residual H_2O signal. A jump-and-return pulse sequence[43] was employed to observe the rapidly exchanging protons in 1 D H_2O experiments. Water suppression was achieved in 2 D experiments in H_2O by the inclusion of a WATERGATE[44] module in the pulse sequence prior to acquisition.

The spectral analysis program XEASY[45] was used for the semi-automatic assignment of the NOESY cross-peaks. Quantitative evaluation of the NOESY cross-peak intensities was carried out automatically with the aid of the integration routines included in the package. In cases of overlapping or very weak peaks, the integration region was selected manually. When the corresponding peaks on both sides of the diagonal were equally reliable, the average intensity was considered.

Thermal denaturation followed by CD: Circular dichroism spectra were collected on a Jasco J-720 spectropolarimeter fitted with a thermostatted cell holder and interfaced with a Neslab RP-100 water bath. CD melting experiments were recorded at 280 nm, with a heating rate of 20 °C h^{-1} from 5 to 90 °C. The spectra were normalized to facilitate comparisons.

Thermal unfolding monitored by NMR spectroscopy: NMR melting curves were measured by monitoring the changes in chemical shifts for well-resolved resonances in 1 D spectra recorded at different temperatures (5 – 80 °C). Only those protons experiencing a chemical shift variation larger than 0.1 ppm over this range of temperatures were used in the fitting. Both NMR and CD melting curves were fitted by using Microcal Origin 5.0 software and assuming a two-state equilibrium between a closed dumbbell form with two-base pairs and a random coil closed circle.[46] All evaluations of thermodynamic parameters were made on the assumption that the melting reactions take place with no appreciable net heat capacity changes. The transition enthalpy and free energies were calculated from the expressions derived by Breslauer et al.[46]

Experimental constraints: Qualitative distance constraints were obtained from NOESY experiments. NOE cross-peaks were classified into strong, medium, or weak categories according to their intensities. The corresponding distance constraints for each group were 2.5 Å, 3.5 Å, and 5 Å, respectively. Refined structures were calculated with more accurate distance constraints. These distances were obtained from NOE cross-peak intensities by use of a complete relaxation matrix analysis with the aid of the program MARDIGRAS.[47] No solvent exchange effects were taken into account in the analysis of NOE intensities in H_2O, and therefore only upper limits were used in the distance constraints involving labile protons. Error bounds in the interproton distances were estimated by carrying out several MARDIGRAS calculations with different initial models, mixing times, and correlation times. Three initial models were chosen from the structures resulting from preliminary "low-resolution" DYANA calculations.[27] Correlation times of 1.0, 2.0, and 4.0 ns were employed, with a single correlation time assumed for the whole molecule in all cases (isotropic motion). Experimental intensities were recorded at two different mixing times (100 and 250 ms). Final constraints were obtained by averaging the upper and lower distance bounds in all the MARDIGRAS runs.

Sums of *J* coupling constants involving H1′, H2′, and H2″ protons were estimated from DQF-COSY cross-peaks. The population of major S conformers were calculated from the sums of *J* coupling constants, Σ2″ or Σ1′, according to the expression:

$$f_s = \frac{31.5 - \sum 2''}{10.9} \quad (1)$$

or

$$f_s = \frac{\sum 1' - 9.8}{5.9} \quad (2)$$

where f_s is the population of S-type puckering.[48] When both sums were available, the average estimation was used. In most cases the population of S-type conformer is greater than 70%, and torsion angle constraints were therefore included. Since only the sums of

coupling constants were estimated, loose values were set for the dihedral angles of the deoxyribose moieties (δ angle between 110° and 170°, ν_1 between 5° and 65°, and ν_2 between −65 and −50°). Additional constraints for the ε angles of the backbone were used to avoid *gauche+* conformations.

In addition to these experimentally derived constraints, Hoogsteen or Watson–Crick hydrogen bond restraints were used for dumbbells **I** and **II**, respectively. Target values for distances and angles relating to hydrogen bonds were set as those described by crystallographic data.[49] Distance constraints with their corresponding error bounds were incorporated into the AMBER potential energy by definition of a flat-well potential term.

Structure determination: Structures were calculated with the aid of the program DYANA 1.5[27] and further refined by use of the SANDER module of the molecular dynamics package AMBER 5.1.[28] The ten best DYANA structures that resulted were taken as starting points for the AMBER refinement. Structures were minimized in the gas phase for 5000 cycles, with hydrated sodium ions introduced as counterions. Random velocities were assigned at an initial temperature of 300 K. For equilibration, the ten starting models were then subjected to 5 ps MD simulation ($T = 300$ K). The temperature was then raised gradually from 300 to 1000 K in 10 ps, and the trajectory was followed for 20 ps longer at 1000 K. These heated structures were then cooled down to 300 K over 20 ps, and further equilibrated at 300 K for 20 ps. Finally, the ten structures obtained at the end of the simulated annealing process were optimized for 5000 cycles. All these simulations were carried out with use of SHAKE[50] to constrain chemical bonds. A time step of 1 fs was used for integration.

The ten structures obtained after the annealing process were further refined by use of MD simulations including explicit water molecules and eight neutralizing ions, and the introduction of long-range electrostatic effects by the particle mesh ewald (PME) method.[51] This lengthy procedure allowed us to relax a few local distortions in the structure of the DNA that appear in the structures derived from the simulated annealing process. The structures were first minimized by standard equilibration protocols and were then subjected to 220 ps of restrained molecular dynamics simulations at constant pressure (1 atm) and temperature (300 K). Periodic boundary conditions and the PME technique have been used to account for long-range effects. The SHAKE program[50] was used, which allowed us to use a 2-fs time step for integration. The AMBER-95 force field[52] was used for the DNA, and the TIP3P model for the water molecules.

Analysis of structures and trajectories: Structural analyses of trajectories in solution were performed by use of analysis modules in the AMBER package. The average structures were determined from the last 10 ps of every trajectory in solution. Helical analysis of MD-averaged structures was carried out with the aid of the programs Curves[53] and MOLMOL.[54]

Supporting information available: Figure S1: NMR spectra of d⟨pCATTCATT⟩ in H_2O at different temperatures. Table S1: assignment tables. Table S2: thermodynamic parameters for NMR melting transitions. Table S3: *J* coupling constants and populations of the major deoxyribose conformers.

We gratefully acknowledge Dr. D. Laurents and Prof. J. A. Subirana for their very useful comments. This work was supported by the Spanish Ministry of Science and Technology (Grant no. BQU 2001-3693), the Generalitat de Catalunya (Grant nos. 2000SGR18, 2001SGR49), and the Centre de Referència de Biotecnologia.

[1] G. Varani, *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 379–404.
[2] C. Hilbers, H. Heus, M. van Dongen, S. Wijmenga in *Nucleic Acids and Molecular Biology* (Eds.: F. Eckstein, D. M. J. Lilley), Springer, **1994**, *Vol. 8*, pp. 56–104.
[3] C. González, N. Escaja, M. Rico, E. Pedroso, *J. Am. Chem. Soc.* **1998**, *120*, 2176–2177.
[4] D. E. Wemmer, A. S. Benight, *Nucleic Acids Res.* **1985**, *13*, 8611–8621.
[5] T. M. Paner, P. V. Riccelli, R. Owczarzy, A. S. Benight, *Biopolymers* **1996**, *39*, 779–793.
[6] T. M. Paner, M. Amaratunga, A. S. Benight *Biopolymers* **1992**, *32*, 881–892.
[7] R. Owczarzy, P. M. Vallone, R. F. Goldstein, A. S. Benight, *Biopolymers* **1999**, *52*, 29–56.
[8] M. J. Doktycz, T. M. Paner, A. S. Benight, *Biopolymers* **1993**, *33*, 1765–1777.
[9] M. J. Doktycz, R. F. Goldstein, T. M. Paner, F. J. Gallo, A. S. Benight, *Biopolymers* **1992**, *32*, 849–864.
[10] M. Amaratunga, E. Snowden-Ifft, D. E. Wemmer, A. S. Benight, *Biopolymers* **1992**, *32*, 865–879.
[11] E. T. Kool, *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 1–28.
[12] C. S. Lim, N. Jabrane-Ferrat, J. D. Fontes, H. Okamoto, M. R. Garovoy, B. M. Peterlin, C. A. Hunt, *Nucleic Acids Res.* **1997**, *25*, 575–581.
[13] T. Hosoya, H. Takeuchi, Y. Kanesaka, H. Yamakawa, N. Miyano-Kurosaki, K. Takai, N. Yamamoto, H. Takaku, *FEBS Lett.* **1999**, *461*, 136–140.
[14] M. J. Blommers, J. A. Walters, C. A. Haasnoot, J. M. Aelen, G. A. van der Marel, J. H. van Boom, C. W. Hilbers, *Biochemistry* **1989**, *28*, 7491–7498.
[15] M. J. Blommers, F. J. van de Ven, G. A. van der Marel, J. H. van Boom, C. W. Hilbers, *Eur. J. Biochem.* **1991**, *201*, 33–51.
[16] M. J. van Dongen, S. S. Wijmenga, G. A. van der Marel, J. H. van Boom, C. W. Hilbers, *J. Mol. Biol.* **1996**, *263*, 715–729.
[17] A. Davison, D. R. Leach, *Nucleic Acids Res.* **1994**, *22*, 4361–4363.
[18] L. J. Rinkel, I. Tinoco Jr., *Nucleic Acids Res.* **1991**, *19*, 3695–3700.
[19] J. H. Ippel, V. Lanzotti, A. Galeone, L. Mayol, J. E. van den Boogaart, J. A. Pikkemaat, C. Altona, *J. Biomol. Struct. Dyn.* **1992**, *9*, 821–836.
[20] J. H. Ippel, V. Lanzotti, A. Galeone, L. Mayol, J. E. Van den Boogaart, J. A. Pikkemaat, C. Altona, *Biopolymers* **1995**, *36*, 681–694.
[21] J. H. Ippel, V. Lanzotti, A. Galeone, L. Mayol, J. E. Van den Boogaart, J. A. Pikkemaat, C. Altona *Biopolymers* **1995**, *36*, 701–710.
[22] J. H. Ippel, V. Lanzotti, A. Galeone, L. Mayol, J. E. van den Boogaart, J. A. Pikkemaat, C. Altona, *J. Biomol. NMR* **1995**, *6*, 403–422.
[23] N. Escaja, E. Pedroso, M. Rico, C. González, *J. Am. Chem. Soc.* **2000**, *122*, 12732–12742.
[24] N. G. Abrescia, A. Thompson, T. Huynh-Dinh, J. A. Subirana, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2806–2811.
[25] B. R. Seavey, E. A. Farr, W. M. Westler, J. L. Markley, *J. Biomol. NMR* **1991**, *1*, 217–236.
[26] S. S. Wijmenga, B. N. M. van Buuren, *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 287–387.
[27] P. Guntert, C. Mumenthaler, K. Wüthrich, *J. Mol. Biol.* **1997**, *273*, 283–298.
[28] D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, D. M. Ferguson, R. J. Radmer, G. L. Seibel, U. C. Singh, P. K. Weiner, P. A. Kollman, 5 ed.; University of California, San Francisco, **1997**.
[29] D. E. Konerding, T. E. Cheatham III, P. A. Kollman, T. L. James, *J. Biomol. NMR*. **1999**, *13*, 119–131.
[30] R. Soliva, V. Monaco, I. Gómez-Pinto, N. J. Meeuwenoord, G. A. Marel, J. H. Boom, C. González, M. Orozco, *Nucleic Acids Res.* **2001**, *29*, 2973–2985.
[31] I. Gómez-Pinto, V. Marchán, F. Gago, A. Grandas, C. González, *ChemBioChem* **2003**, *4*, 40–49.
[32] J. Isaksson, E. Zamaratski, T. V. Maltseva, P. Agback, A. Kumar, J. Chattopadhyaya, *J. Biomol. Struct. Dyn.* **2001**, *18*, 783–806.
[33] J. Aishima, R. K. Gitti, J. E. Noah, H. H. Gan, T. Schlick, C. Wolberger, *Nucleic Acids Res.* **2002**, *30*, 5244–5252.
[34] D. E. Gilbert, G. A. van der Marel, J. H. van Boom, J. Feigon, *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 3006–3010.
[35] L. J. Rinkel, G. A. van der Marel, J. H. van Boom, C. Altona, *Eur. J. Biochem.* **1987**, *163*, 287–296.
[36] J. H. Ippel, H. van den Elst, G. A. van der Marel, J. H. van Boom, C. Altona, *Biopolymers* **1998**, *46*, 375–393.
[37] V. P. Antao, I. Tinoco Jr., *Nucleic Acids Res.* **1992**, *20*, 819–824.
[38] D. Z. Avizonis, D. R. Kearns, *Biopolymers* **1995**, *35*, 187–200.

[39] E. Alazzouzi, N. Escaja, A. Grandas, E. Pedroso *Angew Chem*. **1997**, *109*, 1564 – 1567; *Angew Chem*. *Int*. *Ed*. *Engl*. **1997**, *36*, 1506 – 1508.
[40] A. Kumar, R. R. Ernst, K. Wüthrich, *Biochem*. *Biophys*. *Res*. *Commun*. **1980**, *95*, 1 – 6.
[41] A. Bax, D. J. Davies, *J. Magn*. *Reson*. **1985**, *65*, 355 – 360.
[42] V. Sklenar, H. Miyashiro, G. Zon, H. T. Miles, A. Bax, *FEBS Lett*. **1986**, *208*, 94 – 98.
[43] P. Plateau, M. Güeron, *J. Am*. *Chem*. *Soc*. **1982**, *104*, 7310 – 7311.
[44] M. Piotto, V. Saudek, V. Sklenar, *J. Biomol*. *NMR* **1992**, *2*, 661 – 665.
[45] C. Bartels, T. Xia, M. Billeter, P. Güntert, K. Wüthrich, *J. Biomol*. *NMR* **1995**, *6*, 1 – 10.
[46] K. J. Breslauer *Methods Enzymol*. **1995**, *259*, 221 – 242.
[47] B. A. Borgias, T. L. James, *J. Magn*. *Reson*. **1990**, *87*, 475 – 487.
[48] L. J. Rinkel, C. Altona, *J. Biomol*. *Struct*. *Dyn*. **1987**, *4*, 621 – 649.
[49] W. Saenger, *Principles of Nucleic Acid Structure*, Springer, **1984**.
[50] J. P. Rickaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput*. *Chem*. **1977**, *23*, 327 – 341.
[51] T. E. Darden, D. York, L. Pedersen *J. Chem*. *Phys*. **1993**, *98*, 10089 – 10092.
[52] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am*. *Chem*. *Soc*. **1995**, *117*, 5179 – 5197.
[53] R. Lavery, H. Sklenar, 3.0 ed., Laboratory of Theoretical Biochemistry CNRS, Paris, **1990**.
[54] R. Koradi, M. Billeter, K. Wüthrich, *J. Mol*. *Graphics* **1996**, *14*, 29 – 32.

Received: February 14, 2003 [F 578]

