Dimeric Solution Structure of Two Cyclic Octamers: Four-Stranded DNA Structures Stabilized by A:T:A:T and G:C:G:C Tetrads

Núria Escaja,[‡] Enrique Pedroso,[‡] Manuel Rico,[†] and Carlos González*,[†]

Contribution from the Instituto de Estructura de la Materia, CSIC, C/, Serrano 119, 28006 Madrid, Spain, and Departament de Química Orgánica, Universitat de Barcelona, C/, Martí i Franquès 1-11, 08028 Barcelona, Spain

Received July 27, 2000. Revised Manuscript Received October 19, 2000

Abstract: The solution structure of two cyclic octamers of sequence d<pTGCTCGCT> and d<pCATTCATT> has been determined by two-dimensional NMR spectroscopy and restrained molecular dynamics. The two molecules dimerize at high oligonucleotide concentrations, forming a four-stranded symmetric structure. A complete relaxation matrix analysis of the NOE intensities was carried out, providing a set of around 400 accurate distance constraints. For each oligonucleotide, the calculation converges to very well-defined structures, consisting of two stacks, each with two intermolecular Watson−Crick base pairs. The central residues of each subunit are involved in the base pairs, forming two G:C:G:C or A:T:A:T tetrads, respectively. In both cases, the tetrads are formed by facing the minor groove side of the Watson−Crick base pairs. The tetrads are connected by short loops of two residues. While the residues in the second position of each loop are mainly disordered, nucleotides in the first position (thymines in both cases) are well-defined and form a cap at both ends of the stacks. This is the first time that oligonucleotide structures with minor groove aligned G:C:G:C or A:T:A:T tetrads are observed in solution. Although both oligonucleotides adopt a similar structure, the relative stability of the two molecules is different, with ΔG^0_{25} values for dimer formation of −32 and −13 kJ/mol, respectively.

Introduction

In recent years, a large number of different multistranded noncanonical DNA structures have been characterized, and their important role in biological processes has been put forward.[1−3] Among these structures, the guanine quadruplexes are probably the most extensively studied. G-rich sequences found in the telomers have been proposed to help the chromosomes to join together during cell division by forming guanine tetrads. Also, it has been suggested that quadruplex structures could be involved in processes such as genetic recombination, where homologous DNA recognition is required.[4] Since homologous recombination is a general process, involving purine and pyrimidine bases, observation of quadruplex structures with tetrads composed not exclusively by guanine bases is of primary importance. Some G-quadruplexes with one or two pyrimidine-containing tetrads,[5−7] and two tetrameric structures, consisting entirely of pyrimidine-containing tetrads, have been reported recently. Previous studies have demonstrated that the cyclic octamer d<pCATTCATT>[8] crystallizes forming dimers in the same way as the linear heptamer d(GCATGCT).[9] On the basis of the strong similarity between the structures of these two oligonucleotides of unrelated sequence, a novel DNA motif, named "bi-loop", was proposed.[8]

It is now accepted that crystal-packing forces may cause the DNA to adopt conformations which are unstable in aqueous solution. Thus, the biological relevance of the bi-loop motif would be strongly supported if these structures could be observed in solution. With this aim, we undertook the study of two cyclic octamers of sequence d<pTGCTCGCT> and d<pCATTCATT> by using NMR methods. In the case of d<pCATTCATT>, the three-dimensional structure has been obtained in the solid state.[8] Despite this, we think that valuable information can still be obtained from the studies in solution. More interesting is the case of d<pTGCTCGCT>, where no crystallographic structure is available. This sequence was chosen to facilitate the observation of the bi-loop in solution. Although the sequential requirements for bi-loop formation are unknown, it is reasonable that more stable dimers are formed if the intermolecular base pairs are G−C instead of A−T.

In a preliminary publication,[10] we established that both oligonucleotides can adopt two different conformations in solution. The two conformations observed are involved in a slow interconverting equilibrium that depends strongly on the oligonucleotide concentration. At low concentrations, both molecules are monomeric and adopt a dumbbell-like form, with four of the bases involved in two intramolecular Watson−Crick

* Corresponding author. Fax: +34-91-5642431. E-mail: carlos@malika.iem.csic.es.

[†] Instituto de Estructura de la Materia.

[‡] Universitat de Barcelona.

(1) Gilbert, D. E.; Feigon, J. *Curr. Opin. Struct. Biol.* **1999**, *9*, 305−314.

(2) Patel, D. J.; Bouaziz, S.; Kettani, A.; Wang, Y. In *Oxford Handbook of Nucleic Acid Sructures*; Neidle, S., Ed.; New York: Oxford University Press, 1999; pp 389−453.

(3) Lebrun, A.; Lavery, R. *Curr. Opin. Struct. Biol.* **1997**, *7*, 348−354.

(4) McGavin, S. *J. Mol. Biol.* **1971**, *55*, 293−298.

(5) Kettani, A.; Kumar, R. A.; Patel, D. J. *J. Mol. Biol.* **1995**, *254*, 638−656.

(6) Kettani, A.; Bouaziz, S.; Gorin, A.; Zhao, H.; Jones, R.; Patel, D. J. *J. Mol. Biol.* **1998**, *282*, 619−636.

(7) Jing, N. J.; Hogan, M. E. *J. Biol. Chem.* **1998**, *273*, 34992−34999.

(8) Salisbury, S. A.; Wilson, S. E.; Powel, H. R.; Kennard, O.; Lubini, P.; Sheldrick, G. M.; Escaja, N.; Alazzouzi, E.; Grandas, A.; Pedroso, E. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 5515−5518.

(9) Leonard, G. A.; Zhang, S.; Peterson, M. R.; Harrop, S. J.; Helliwell, J. R.; Cruse, W. B. T.; d'Estaintot, B. L.; Kennard, O.; Brown, T.; Hunter, W. N. *Structure* **1995**, *3*, 335−340.

(10) González, C.; Escaja, N.; Rico, M.; Pedroso, E. *J. Am. Chem. Soc.* **1998**, *120*, 2176−2177.

10.1021/ja002778q CCC: $19.00
Published on Web 12/06/2000

Table 1. Free Energies at 25 ×bcC for Dimer Formation of d<pTGCTCGCT> and d<pCATTCATT>

	$\Delta G^0_{\text{dimer}}{}^a$ (kJ/mol)	$\Delta G^0_{\text{dumbbell}}{}^b$ (kJ/mol)	$\Delta G^0_{\text{dum-dimer}}{}^c$ (kJ/mol)
d<pTGCTCGCT>	−32	−6	−20
d<pCATTCATT>	−13	−3	−7

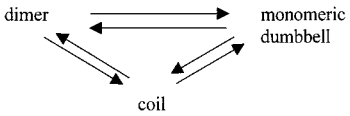
[a] $\Delta G^0_{\text{dimer}} + 2\,\Delta G^0_{\text{dumbbell}} + \Delta G^0_{\text{dum-dimer}} = 0$. [b] Estimated from DC melting curves of the monomeric forms (Escaja et al., manuscript in preparation). [c] Estimated from the van't Hoff analysis of the dumbbell-dimer equilibrium.

base pairs, and the other four forming two mini-hairpin loops of two bases each. At higher concentrations, the molecules are dimeric, forming a symmetric structure with four intermolecular Watson−Crick base pairs. In this paper, we report the detailed three-dimensional solution structure of the dimeric forms of these two octamers as obtained by restrained molecular dynamics calculations based on NMR derived experimental constraints. An estimation of the free energies of the dimer formation for the two molecules is also presented.

Methods

Experimental Details. The cyclic octamers were synthesized as reported by Alazzouzi et al.[11] Samples were suspended (in Na^+ salt form) in a phosphate buffer of either D_2O or 9:1 H_2O/D_2O (25 mM sodium phosphate buffer, pH = 7). All NMR spectra were acquired in a Bruker AMX spectrometer operating at 600 MHz, and processed with the UXNMR software. In the experiments in D_2O, presaturation was used to suppress the residual H_2O signal. A jump-and-return pulse sequence[12] was employed to observe the rapidly exchanging protons in 1D H_2O experiments. NOESY[13] spectra in D_2O were acquired with mixing times of 100, 200, and 300 ms. ROESY and TOCSY[14] spectra were recorded with the standard MLEV-17 spin-lock sequence, at 300 and 80 ms mixing time, respectively. In 2D experiments in H_2O, water suppression was achieved by including a WATERGATE[15] module in the pulse sequence prior to acquisition. The spectral analysis program XEASY[16] was used for semiautomatic assignment of the NOESY cross-peaks. Quantitative evaluation of the NOESY cross-peak intensities was carried out automatically with the integration routines included in the package. In the case of very weak or overlapping peaks, a manual integration was performed.

Melting Behavior. The dissociation of the dimeric forms of both molecules can be described with the following scheme:



Evidence of a monomeric dumbbell-like structure arises from NMR experiments[10] and CD melting curves carried out at low oligonucleotide concentration. Thermodynamic parameters for the melting of the monomeric forms were obtained by fitting the curves of elipticity at $\lambda_{\max}$ versus temperature with a nonlinear least-squares program. A similar procedure was used to fit the change on the chemical shifts of different protons with the temperature. Both methods give very similar results. Free energies for dumbbell formation of d<TGCTCGCT> and d<CATTCATT> are given in Table 1. More details on the stability of a number of octameric dumbbells of different sequences will be given in a separate publication (Escaja et al., manuscript in preparation).

The presence of these species makes the interpretation of spectroscopic melting curves of the dimeric forms difficult. Fortunately, the equilibrium between the dimeric and monomeric forms is slow on the NMR time scale, so that resonances from both species can be observed simultaneously under the appropriate conditions. The equilibrium constants can be determined from the ratio of the areas of equivalent peaks. Thermodynamic parameters for the dimer-dumbbell equilibrium are estimated from a van't Hoff analysis of the equilibrium constants at different temperatures (see Figure S2). Only temperatures quite below the T_m of the monomers are considered. Thermodynamic parameters for dimer formation can be obtained from the following equation:

$$\Delta G^0_{\text{dimer}} + 2\,\Delta G^0_{\text{dumbbell}} + \Delta G^0_{\text{dumbbell-dimer}} = 0$$

Experimental Constraints. Direct assignment of cross-peaks to specific distance constraints was impeded by the intrinsic ambiguity between inter- and intramolecular cross-peaks. Many cross-peaks could be directly assigned on the basis of a rough model of the molecule, such as intraresidual NOEs or some cross-peaks between base-paired residues. Trial assignments were made for the remaining cross-peaks, and the trial constraints were used in preliminary DYANA calculations. When a distance constraint was consistently violated in all the resulting structures, an alternative assignment was considered. After several cycles of assignment and structure calculation, a consistent set of constraints was obtained. All of these calculations were carried out with qualitative distance constraints (classified as 3, 4, or 5 Å). At this stage, some cross-peaks were still ambiguous since two or more assignments were consistent with the three-dimensional structure. These cross-peaks were not considered in the subsequent calculations. Despite these difficulties, the existence of only four different spin systems (five in the case of d<pTGCTCGCT>) reduces the signal overlap, facilitating the extraction of distance constraints in regions that are normally too crowded for oligonucleotides of this size.

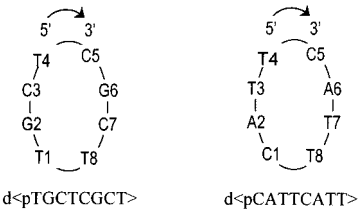
Refined structures were calculated with more accurate distance constraints. These distances were obtained from NOE cross-peak intensities by using a complete relaxation matrix analysis with the program MARDIGRAS.[17,18] No solvent exchange effects were taken into account in the analysis of NOE intensities in H_2O. Therefore, only upper limits were used in the distance constraints involving labile protons. Error bounds in the interprotonic distances were estimated by carrying out different MARDIGRAS calculations with different initial models, mixing times, and correlation times. Three initial models were chosen from the structures resulting from the "low-resolution" DYANA calculation (vide infra). Correlation times of 1.0, 2.0, and 4.0 ns were employed, assuming, in all cases, a single correlation time for the whole molecule (isotropic motion). Experimental intensities were recorded at three different mixing times (100, 200, and 300 ms) for nonexchangeable protons, and at a single mixing time (200 ms) for labile protons. Final distances were obtained by averaging a total of 27 MARDIGRAS outputs (three mixing times × three starting structures × three correlation times). Final distance errors were determined by averaging upper and lower bounds in all 27 individual runs. A lower limit of 1.8 Å was set in those distances where no quantitative analysis could be carried out, such as very weak intensities or cross-peaks involving labile protons. In addition to these experimentally derived constraints, Watson−Crick hydrogen bond restraints were used. Target values for distances and angles related to hydrogen bonds were set as described from crystallographic data.

Torsion angle constraints for the sugar moieties were derived form the analysis of J-coupling data obtained from DQF−COSY experiments. Since only the sums of coupling constants were estimated, rather loose values were set for the dihedral angle of the deoxyriboses (δ angle between 110° and 170°, ν_1 between 5° and 65°, and ν_2 between −65 and −50°). Additional constraints for the γ angles of the backbone were used in those cases where stereoespecific assignments of H5′/H5" resonances could be made.

Structure Determination. Structures were calculated with the program DYANA 1.4[19] and further refined with the SANDER module

(11) Alazzouzi, E.; Escaja, N.; Grandas, A.; Pedroso, E. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 1506−1508.
(12) Plateau, P.; Güeron, M. *J. Am. Chem. Soc.* **1982**, *104*, 7310−7311.
(13) Kumar, A.; Ernst, R.R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1−6.
(14) Bax A.; Davies, D. G. *J. Magn. Reson.* **1985**, *65*, 355−360.
(15) Piotto, M.; Saudek, V.; Sklenar, V. *J. Biomol. NMR* **1992**, *2*, 661−665.
(16) Bartels, C.; Xia, T.; Billeter, M.; Güntert, P.; Wüthrich, K. *J. Biomol. NMR* **1995**, *6*, 1−10.
(17) Borgias, B. A.; James, T. L. *J. Magn. Res.* **1988**, *79*, 493−512.
(18) Borgias, B. A.; James, T. L. *J. Magn. Res.* **1990**, *87*, 475−487.

Scheme 1



of the molecular dynamics package AMBER 4.1.[20] Initial DYANA calculations were carried out on the basis of qualitative distance constraints. The resulting structures were used as initial models in the complete relaxation matrix calculations to obtain accurate distance constraints, as described in the previous paragraph. These structures were taken as starting points for the AMBER refinement. All AMBER calculations were performed in vacuo, with hexahydrated Na^+ counterions. The electrostatic term was calculated using a distance-dependent dielectric constant, and the cutoff value for nonbounded interactions was 10 Å. Usual force field parameters of the AMBER potential[21] were used. The temperature and the relative weights of the experimental constraints were varied during the simulations according to the following annealing strategy. Each structure was first restrained energy minimized and then submitted to 50 ps of molecular dynamics. During the first 5 ps the temperature was raised linearly from 300 to 1500 K, maintained at this value for the next 10 ps, and decreased to 300 K over 10 ps. This final value was maintained for the remaining 15 ps. The force constants for the distance restraints was varied simultaneously with the temperature. K_{NOE} was raised from 10 kcal/(mol·Å^2) up to 100 kcal/(mol·Å^2), kept at this value during the high-temperature period, and reduced to 20 kcal/(mol·Å^2). This value was maintained during the rest of the simulation and the subsequent restrained energy minimization. The last 5 ps of the trajectory were used for averaging, and the resulting structure was subjected to 1000 steps of restrained energy minimization using the steepest descent method.

Analysis of Structures and Trajectories. Analysis of the average structures as well as the MD trajectories was carried out with the programs Curves,[22] and MOLMOL.[23] Reliability factors (*R*-factors)[24,25] were calculated with the program CORMA.[17]

Results and Discussion

NMR Spectra. Sequential assignments of exchangeable and nonexchangeable protons were carried out following standard ^{1}H NMR methods. Almost all resonances were identified, including the stereoespecific assignment of some H5′/H5" protons. The complete assignment lists have been deposited at the BMRB[26] (accession numbers 4687 and 4694). Both exchangeable and nonexchangeable proton spectra indicate that the dimeric species are symmetric, since all of the equivalent resonances of residues 1−4 and 5−8 in d<pCATTCATT> (and between 2 and 4 and 6 to 8 in d<pTGCTCGCT>) are degenerate (see Scheme 1 for nucleotide numbering). Equivalent residues are indicated between brackets in the following discussion.

(a) Nonexchangeable Protons. Many features of the nonexchangeable proton resonances are common in both oligonucleotides. Fragments of the two-dimensional NOESY spectrum of d<pTGCTCGCT> in D_2O are shown in Figure 1. All intra-nucleotide H1′-base NOEs are medium or weak, indicating that the glycosidic angle in all the nucleotides is in an *anti* conformation. Only intra-residual NOEs were observed for the bases of residues 1 and 5. However, strong sugar-base sequential connections were observed between residues 2 → 3 (and the corresponding 6 → 7). Many unusual NOEs were detected between residues 2(6) and 3(7) and residue 4(8). Some of them were assigned as interstrand constraints during the structure calculation (see below). Interestingly, unusual values are observed for the chemical shifts of the H4′ and H5′/5" protons of residue 4(8), which are shifted upfield (see Figure 1). These shifts are due to the proximity of the purine base of residue 2(6) in the three-dimensional structure.

(b) Exchangeable Protons. A very narrow imino signal is observed in the two molecules, at 14.46 ppm in d<pCATTCATT> and at 13.45 ppm in d<pTGCTCGCT>. In the 2D-NOESY spectra in H_2O of d<pTGCTCGCT>, intense cross-peaks are observed between these signals and the amino protons of C3(7) (see Figure 2). Similar cross-peaks are observed in the case of d<pCATTCATT> between the imino proton of the thymine 3(7) and the H2 and the amino protons of the adenine (see Figure 1 in the Supporting Information). In both cases, these NOE patterns are characteristic of Watson−Crick base pairs. In d<pTGCTCGCT>, only one of the amino protons of the cytosines resonates at low field (8.77 ppm), indicating that the other amino (6.50 ppm) is not hydrogen bonded. These cytosine amino protons show a different behavior than that observed in other G:C:G:C tetrads, where both cytosine amino protons are involved in hydrogen bonds and resonate at low field.[5,6] Interestingly, the exchangeable protons involved in base-pair hydrogen bonding are very well-protected against solvent exchange, since the signals are narrow in a large range of temperatures (see Figure 3 in ref 10).

Melting Behavior. The melting behavior of the dimeric forms of both molecules is peculiar since the equilibrium between both species is slow on the NMR time scale over the entire range of temperatures where the two forms coexist. As can be seen in Figure 3, the signals of the dimeric form (the major ones at low temperature) do not change with the temperature. Thermodynamic parameters for the monomer−dimer equilibria were calculated from a van't Hoff analysis of the equilibrium constants, estimated from the relative intensities of the NMR signals between equivalent protons in the two forms (see Fig S2 in Supporting Information). This analysis was carried out at temperatures rather below the T_m of the monomers, were the two-state approximation is valid. Since the melting of the monomers can be studied independently at lower concentrations by NMR and CD melting experiments (T_m of 52 °C for d<pTGCTCGCT> and 40 °C for d<pCATTCATT>, Escaja et al., manuscript in preparation), the parameters for dimer formation can be derived from simple thermodynamical considerations. The results are shown in Table 1. Both dimers are rather stable, with ΔG values of −32 kJ/mol and −13 kJ/mol at 25 °C for d<pTGCTCGCT> and d<pCATTCATT>, respectively (the melting temperature at 5 mM is estimated to be 56 °C for d<pTGCTCGCT> and 25 °C for d<pCATTCATT>). These results have been obtained at low salt

(19) Güntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol*. **1997**, *273*, 283−298.
(20) Pearlman, D. A.; Case, D. A.; Caldwell, J. C.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. AMBER 4.0, University of California, San Francisco, 1991.
(21) Weiner, P.; Kollman, P. A. AMBER4.0, University of California, San Francisco, 1991.
(22) Lavery, R.; Sklenar, H. CURVES. Curves 3.0, helical analysis of irregular nucleic acids, Laboratory of Theoretical Biochemistry CNRS, Paris, 1990.
(23) Koradi R.; Billeter, M.; Wüthrich, K. *Mol. J. Graphics* **1996**, *14*, 29−32.
(24) González, C.; Rullmann, J. A. C.; Bonvin, A. M. J. J.; Boelens, R.; Kaptein, R. *J. Magn. Reson*. **1991**, *91*, 659−664.
(25) Thomas, P. D.; Basus, V. J.; James, T. L. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 1237−1241.
(26) Seavey, B. R.; Farr, E. A.; Westler, W. M.; Markley, J. L. *J. Biomol. NMR* **1991**, *1*, 217−236.

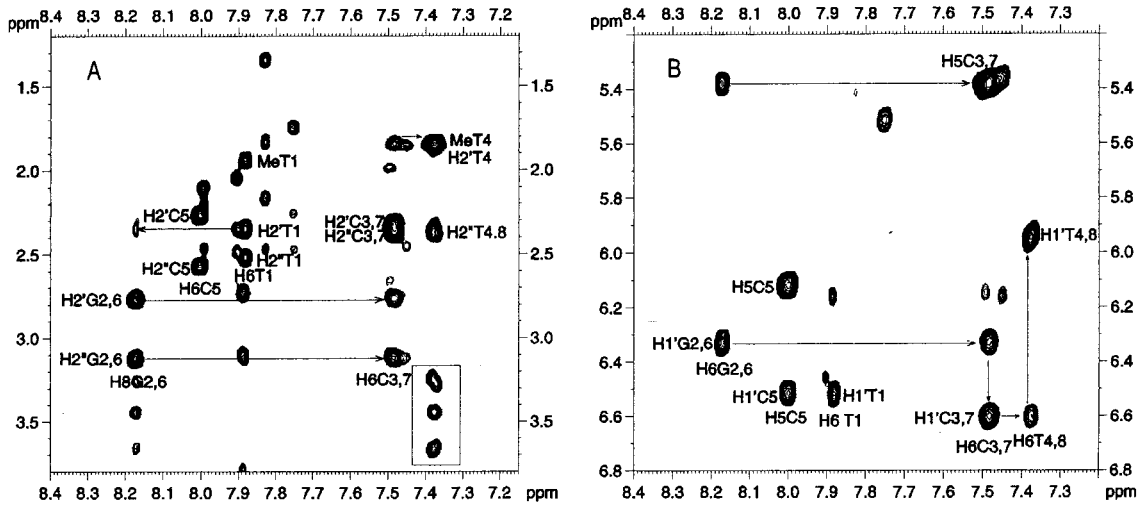


Figure 1. Regions of the NOESY spectrum (300 ms mixing time) of d<pTGCTCGCT > in D_2O (2.4 mM oligonucleotide concentration, $T = 5$ °C, pH = 7). Sequential assignment pathways are drawn in the H1′-aromatic region. Intraresidual NOEs between H4′/5′/5" with H6 of T4(8) are squared. At this concentration, some weak NOEs from the monomeric form can be observed.

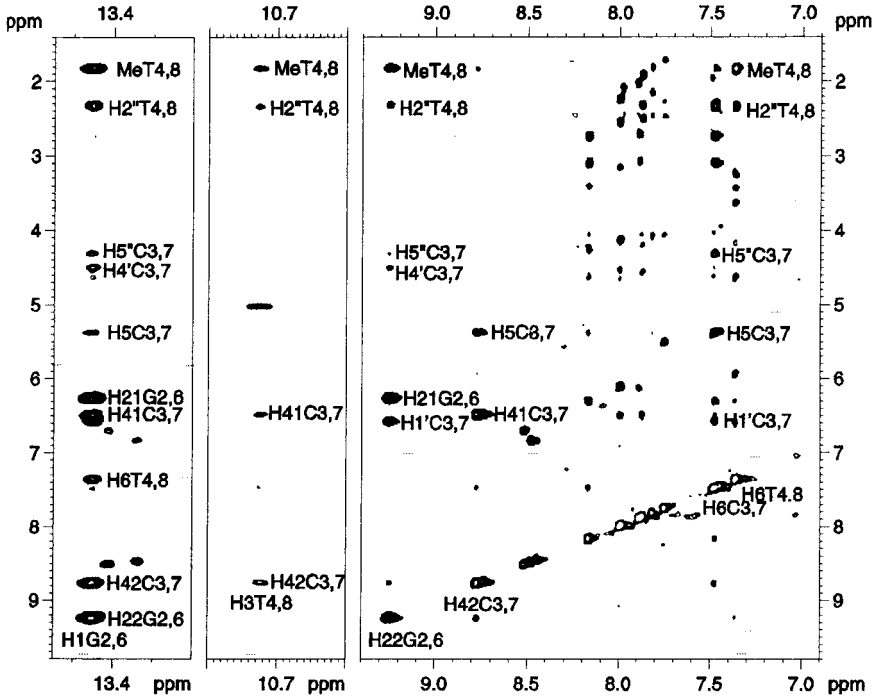


Figure 2. Regions of the NOESY spectrum (150 ms mixing time) of d<pTGCTCGCT> in H_2O (2.4 mM oligo concentration, $T = 5$ °C, pH = 7). Watson−Crick base pairing can be established from the H1G2(6)−H42C3(7), H1G2(6)−H41C3(7,) and H1G2(6)−H5C3(7) cross-peaks shown in the figure. The NOEs H3T4(8)−H42C3(7) and H3T4(8)−H41C3(7) define the orientation of the T4(8).

concentration, but preliminary data indicate that the dimeric forms are more stable at larger salt concentration.

Inter- versus Intramolecular Base Pairs. The dimers of d<pTGCTCGCT> and d<pCATTCATT> can be stabilized either by formation of intermolecular base pairs or by association of two dumbbell-like monomers with intramolecular base pairs. Taking into account the conformational restraints imposed by loops of only two residues, and that the glycosidic angles of all the nucleotides are *anti*, the association of two dumbbell-like molecules can only occur through the major groove side of the Watson−Crick pair. Under these circumstances, minor groove association is sterically impeded by the sugar−phosphate backbone.

Direct evidence of minor groove association arises from the chemical shifts of the amino protons and some unambiguous NOE contacts. As mentioned before, only one of the exchangeable cytosine amino protons is low-field shifted, indicating that only that amino proton is involved in a hydrogen bond. In contrast, NMR data on major groove aligned G−C pairs indicate that both cytosine amino protons are forming hydrogen bonds.[5,6] Moreover, the two guanine amino protons can be observed and present rather narrow signals, suggesting that neither of them are exposed to the solvent. A number of NOEs involving these amino protons, like the NOE with the H1′ of cytosines, clearly indicates that major grove association is not taking place. Also, some NOE contacts described in major groove G:C:G:C tetrads, like the strong NOE between H5C and H8G, are not observed.

In the case of d<pCATTCATT>, no NMR studies on major groove aligned A:T:A:T tetrads are available for comparison. However, all our data indicate that the structure is very similar to that of d<pTGCTCGCT>. Especially informative are the NOEs with H2A, like the strong contact between H2A and H1′T,

Table 2. NMR Restraints and Structural Statistics

	d<pCATTCATT>	d<pTGCTCGCT>
experimental distance constraints		
total number	392	486
intraresidue	220	252
sequential	102	144
range > 1.	70	90
intramolecular	340	430
intermolecular	52	56
RMSD		
all bases well-defined[b]	0.3 ± 0.1	0.3 ± 0.1
all heavy atoms well-defined[b]	0.9 ± 0.2	0.5 ± 0.1
backbone.	1.0 ± 0.2	0.7 ± 0.1
all bases.	1.4 ± 0.3	2.7 ± 0.8
all heavy atoms.	1.3 ± 0.3	2.0 ± 1.2
sum of violations	15.5 Å	12.7 Å
max. violation	0.48 Å	0.39 Å
R-factor (×100)[a]	6.93	7.24
av NOE energy	64 kcal/mol	91 kcal/mol
range of NOE energies	44 to 85 kcal/mol	77 to 117 kcal/mol
av total energy	−2235 kcal/mol	−2462 kcal/mol
range of total energies	−2193 to −2268 kcal/mol	−2391 to −2490 kcal/mol

[a] Sixth-root *R*-factor for all mixing times. [b] All except residues 1, 5, 9, and 13.

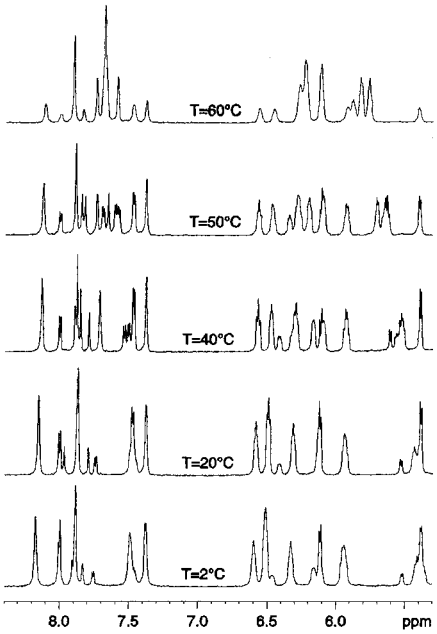


Figure 3. One-dimensional NMR spectra of d<pTGCTCGCT> in D_2O at different temperatures (2.4 mM oligonucleotide concentration, pH = 7). The intensity of the dimeric signals (the major ones at low temperature) decreases with the temperature, but their chemical shifts do not change, indicating that the dimer−monomer equilibrium is slow in the NMR time scale in the whole range of temperatures.

which clearly point to a minor groove base-pair alignment. In conclusion, the NMR spectra of the dimeric forms of d<pTGCTCGCT> and d<pCATTCATT> are only consistent with a minor groove association of G−C or A−T base pairs. As mentioned above, this alignment can only occur if the Watson−Crick base pairs are intermolecular.

Experimental Constraints and Structure Calculations. Direct assignment of cross-peaks to specific distance constraints was impeded by the intrinsic ambiguity between inter- and intramolecular cross-peaks. The situation is even more complicated because the sequence is repetitive (or almost repetitive in the case of d<pTGCTCGCT>). Therefore, all the resonances of residues 1−4 and 5−8 in d<pCATTCATT>, and between 2 and 4 and 6 to 8 in d<pTGCTCGCT>, are completely degenerate. As a consequence, four different distance constraints between pairs of protons are possible for each cross-peak. Inter- and intramolecular NOEs could be distinguished by analyzing the change in NOE intensities between pure labeled and mixed labeled samples. However, this requires samples uniformly enriched in ^{15}N and ^{13}C. At present, the synthesis of uniformly labeled cyclic oligonucleotides is not straightforward, and therefore, we have to employ a different strategy. Fortunately, the small size of these molecules allows for a rapid calculation of the structure with different sets of constraints. The assignment of each NOE to specific distance constraints could be done by carrying out several structure calculations with trial assignments (see Materials and Methods for details). After several cycles of assignment-calculation runs, a large number of unambiguous distance constraints could be obtained. Similar strategies have been used in the structural determination of other symmetric molecules, like dimeric proteins,[27,28] peptides,[29] and in nucleic acids.[30] Moreover, a complete relaxation matrix analysis of the 2D NOE cross-peaks was carried out, providing more accurate distance values for the experimental constraints. A summary of the final constraints is shown in Table 2. Inter-residual distances are indicated schematically in Figure 4.

In addition to the NOE-derived information, an analysis of the *J*-coupling constants obtained from DQF-COSY spectra was carried out. In all cases the *J*-couplings are consistent with sugar puckers predominantly in the S-domain. The population of the major S conformer, estimated from the sum of $J_{1'2'}$ and $J_{1'2''}$,[31] was in all cases greater that 90%, except in deoxyriboses of residues 1 and 5 where this population is around 60%. In d<pTGCTCGCT>, this analysis could not be carried out for residue C3(7) since H2′ and H2" are degenerated. Vicinal *J*-coupling constants between H4′ and H5′ and H5" were particularly small in the case of residues 2 and 4. These *J*-values together with the pattern of intra-residual NOEs between H3′/H4′ and H5′/H5" at short mixing time, allowed for stereospecific

(27) Clore, G. M.; Appella, E.; Yamada, M.; Matsushima, H.; Gronenborn, A. M. *Biochemistry* **1990**, *29*, 1689−1701.
(28) Rico, M.; Jiménez, M. A.; González, C.; De Filippis, V.; Fontana, A. *Biochemistry* **1995**, *33*, 14834−14847.
(29) Kay, L. E.; Forman-Kay, J. D.; McCubbin, W. D.; Kay, C. M. *Biochemistry* **1991**, *30*, 4323−4333.
(30) Nonin, S.; Phan, A. T.; Leroy, J. L. *Structure* **1997**, *5*, 1231−1246.

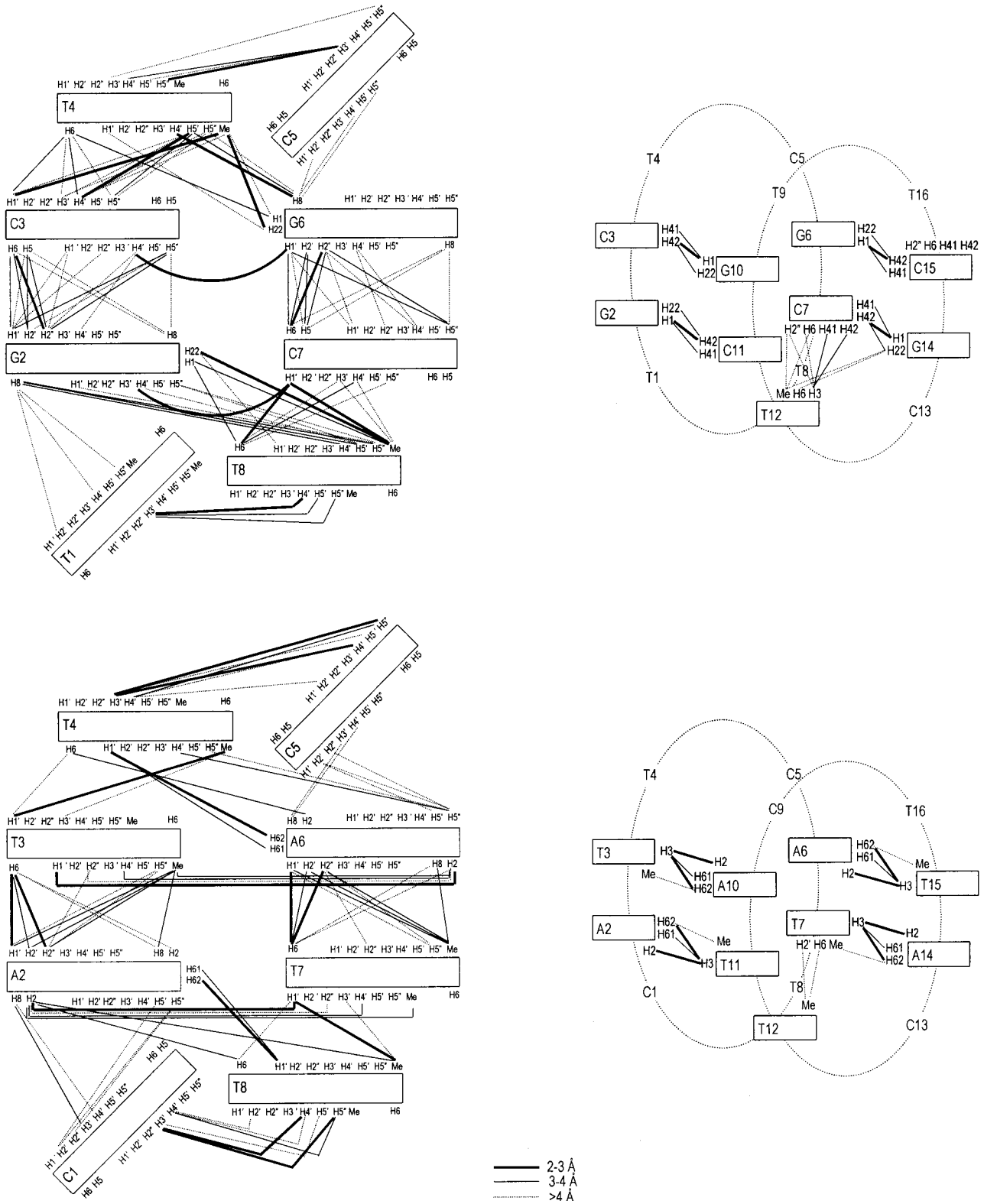


Figure 4. Schematic representation of intra- and intermolecular distance constraints of d<pTGCTCGCT> (top) and d<pCATTCATT> (bottom). Constraints are classified in three categories according to their upper distance limit.

assignment of H5′/H5″,[32] and indicates that the backbone angle γ in these two nucleotides is in a single γ^+ conformation.

Distance and torsion angle constraints were used to calculate the dimeric structures of d<pTGCTCGCT> and d<pCATTCATT> by using restrained molecular dynamics methods. Initial structures were calculated with the program DYANA and then refined with the AMBER package. The final structures exhibit good *R*-factors and convergence statistics (see Table 2). Only the residues in positions 1 and 5, and the corresponding ones in the symmetry-related subunit are relatively disordered (see Figures 5 and 6). All others are very well-defined, with RMSDs below 0.5 and 0.9 Å for d<pTGCTCGCT> and d<pCATT-

(31) Rinkel, L. J.; Altona, C. *J. Biomol. Struct. Dyn.* **1987**, *4*, 621−649.
(32) Wijmenga, S. S.; Mooren, W.; Hilbers, C. W. In *NMR in Macromolecules*; *Roberts, G. C., Ed.*; IRL Press: Oxford; 1993; pp 217−288.

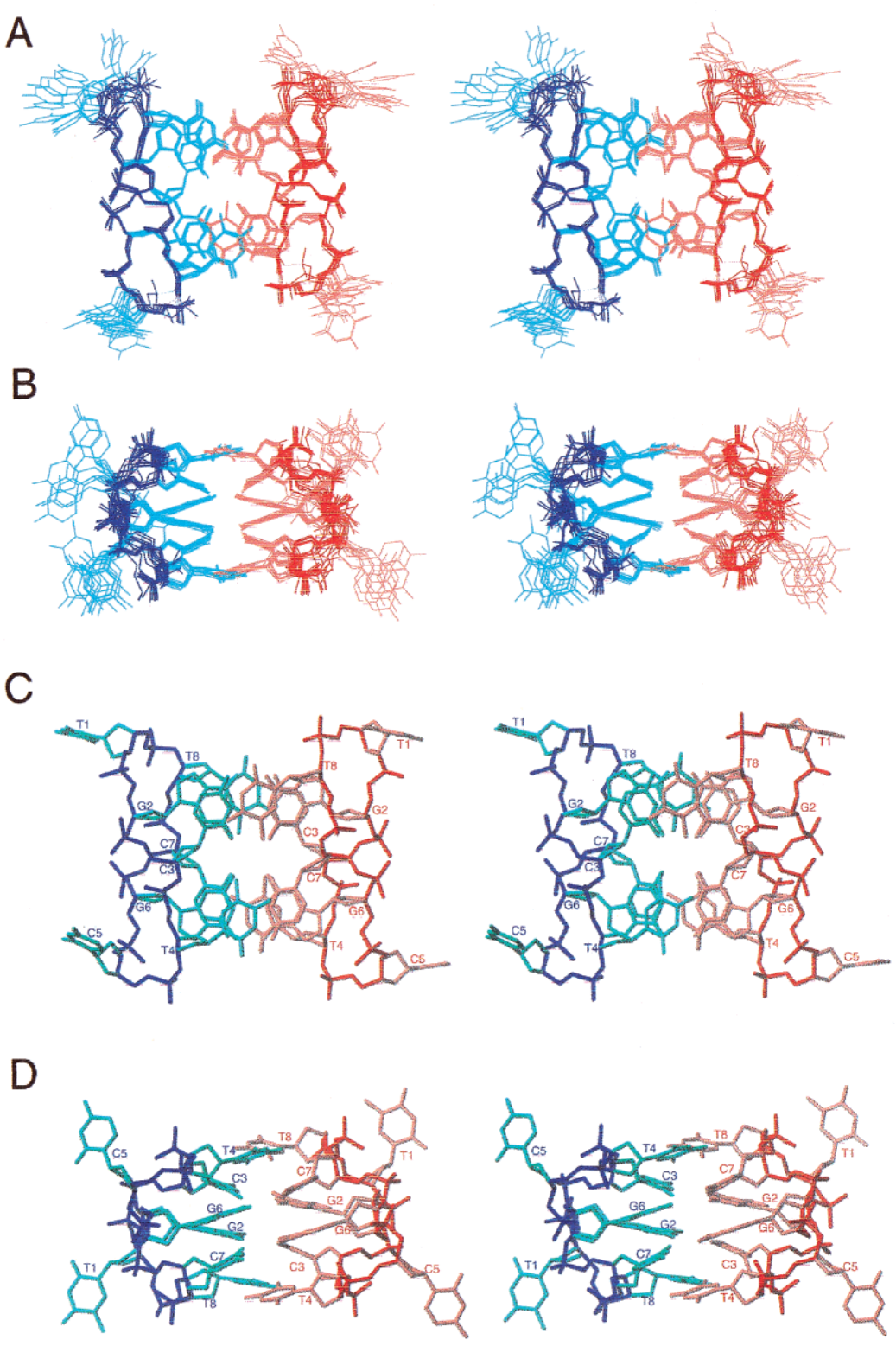


Figure 5. (A and B) Two strereoscopic views of the superposition of the 10 refined structures of d<pTGCTCGCT>. (C and D) The same views of the average structure of d<pTGCTCGCT>. Red and blue indicate different molecules. The phosphate backbone is indicated in darker colors.

CATT>, respectively. These values are even lower when only atoms in the bases are considered. The final AMBER energies and NOE terms are reasonably low in all the structures, with no distance constraints violation greater than 0.5 Å.

Description of the Structures. All of the experimental constraints are consistent with an antiparallel arrangement of the two octameric subunits. With the exception of the disordered residues, the dimers are symmetric. This symmetry is reflected in the geometrical parameters of both molecules, as can be seen in Figure 7 and Table 3 (additional geometrical parameters are shown in Tables S1 and S2). The two subunits are related by two perpendicular dyad axes. A third dyad axis relates the two well-defined sequence equivalent parts in each octamer (residues 2, 3, 4 with 6, 7, 8). The core of the dimers consists of two tetrads formed by purines and pyrimidines, R:Y:R:Y, (where R: purine, Y: pyrimidine). All glycosidic angles in these tetrads

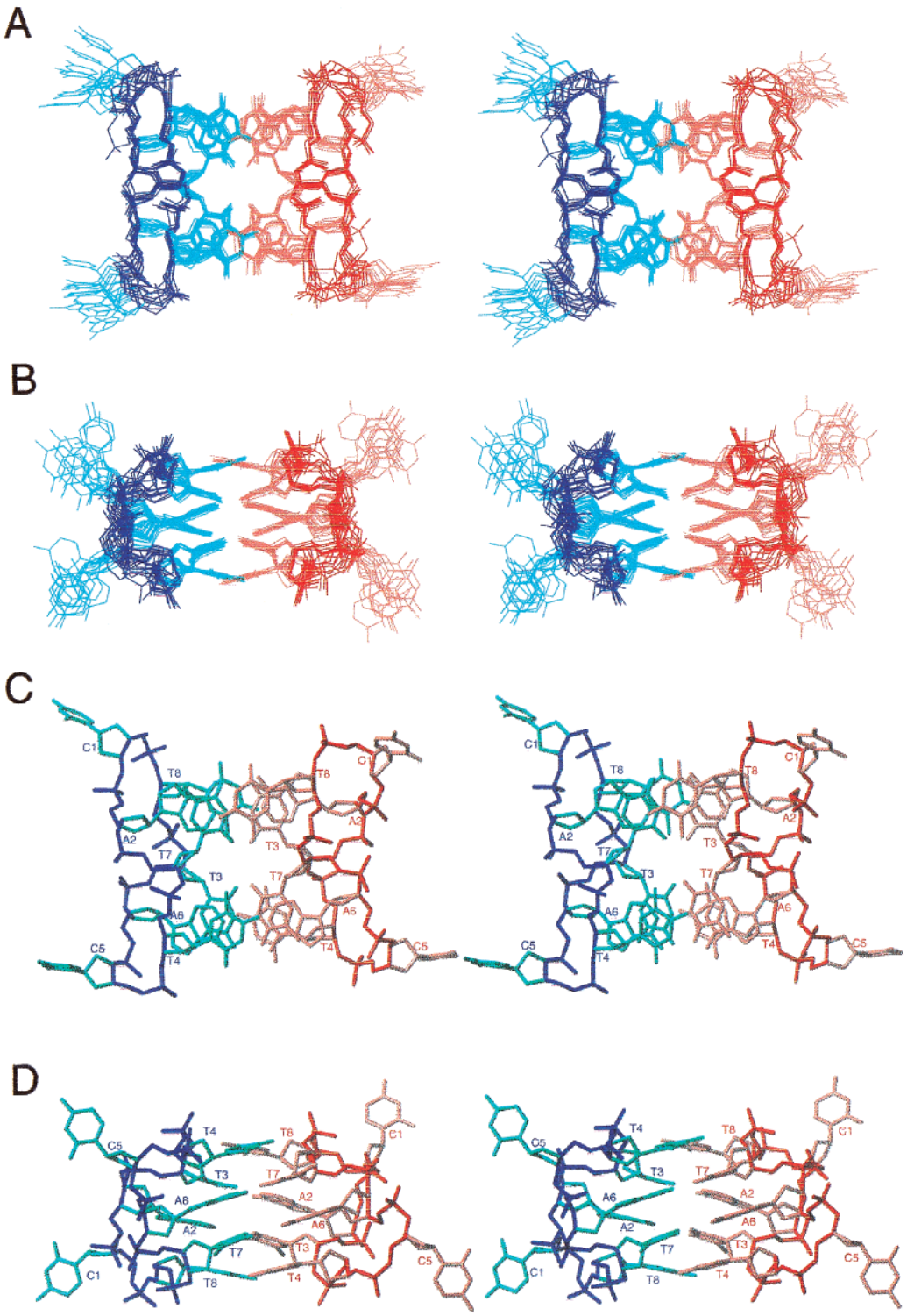


Figure 6. (A and B) Two stereoscopic views of the superposition of the 10 refined structures of d<pCATTCATT>. (C and D) The same views of the average structure of d<pCATTCATT >. Red and blue indicate different molecules. The phosphate backbone is shown in darker colors.

are *anti*, with values of −100° for the purines, and −120° for the pyrimidines (−130° for the Ts in d<pCATTCATT>). The two base pairs forming the tetrads are not in the same plane (see Figures 5 and 6). Although the bases in each pair present very little buckle and propeller twist, the two Watson−Crick base pairs forming the tetrad have a mutual inclination of 37°−40°. The displacement of the bases along the axis defined by their own stack is very small, with a little twist of around 25°, giving rise to very favorable stacking interactions. The average rise between the base pairs in each stack is 2.8 Å in d<pTGCTCGCT> and 3.2 Å in d<pCATTCATT>.

In both molecules, the quadruplex is formed through the minor groove side of the Watson−Crick base pairs. In the case of d<pCATTCATT>, each tetrad is stabilized by only four hydrogen bonds. In the case of d<pTGCTCGCT>, in addition to the six Watson−Crick hydrogen bonds per tetrad, two more intramolecular H-bonds are formed (see Figure 9A and 9B). The guanine amino proton resonating at 9.25 ppm is forming

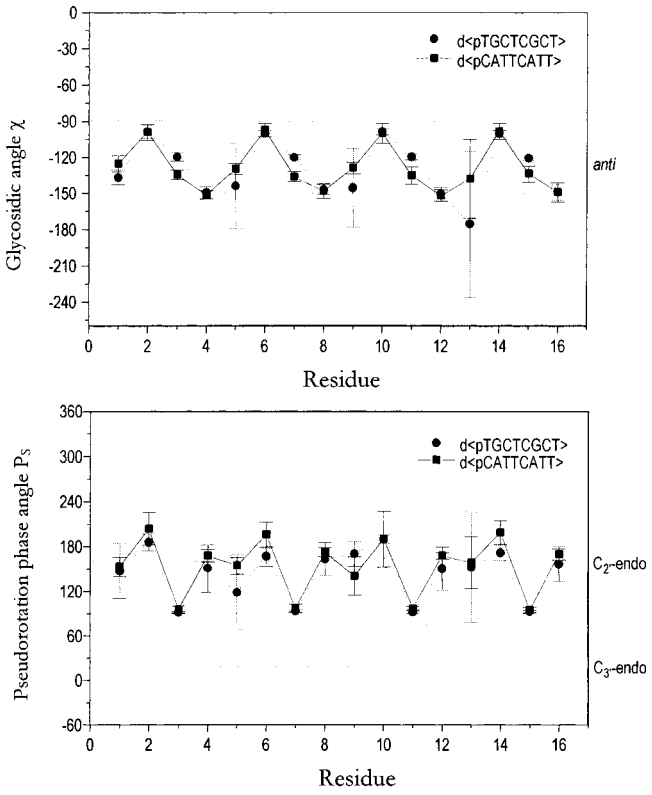


Figure 7. Glycosidic torsion angles (χ) and sugar puckers (P_S) for d<pTGCTCGCT>, dashed lines, and d<pCATTCATT>, solid lines. Residues 9−16 correspond to the symmetry-related molecule.

Table 3. Local Helical Parameters for the Dimeric Forms of d<pCATTCATT> and d<pTGCTCGCT>

step	shift D_x (Å)	slide D_y (Å)	rise D_z (Å)	tilt τ (deg)	roll ρ (deg)	twist Ω (deg)
			d<pCATTCATT>			
A2/*T*3	0.1	−1.5	3.2	0.6	−0.5	25
T3/*T*4	−10.8	−1.4	4.7	−29.9	−2.3	−41
A6/*T*7	−0.1	−1.4	3.1	0.8	−0.3	24
T7/*T*8	−10.8	−1.4	4.7	−30.2	−2.0	−40
T12/*T*11	10.8	−1.4	4.7	30.3	−3.7	−41
T11/A10	0.1	−1.5	3.2	−0.6	−1.8	24
T16/*T*15	10.8	−1.3	4.7	30.4	−1.0	−41
T15/A14	−0.1	−1.4	3.1	−1.3	0.1	26
			d<pTGCTCGCT>			
G2/C3	0.2	−0.4	2.8	9.8	−4.5	26
C3/*T*4	−10.2	−1.9	4.7	−38.4	7.8	−35
G6/C7	0.3	−0.3	2.7	10.9	−3.0	26
C7/*T*8	−9.7	−2.0	5.1	−41.1	12.5	−31
T12/C11	10.1	−1.9	4.8	38.2	7.1	−35
C11/G10	−0.3	−0.4	2.8	−10.3	−4.2	26
T16/C15	9.8	−1.9	5.0	39.7	10.6	−32
C15/G14	−0.3	−0.4	2.7	−11.2	−2.3	26

[a] Only well-defined residues are included (all except 1, 5, 9, and 13).

the Watson−Crick hydrogen bond, while the other proton (6.3 ppm) is located within hydrogen-bonding distance to the cytosine O2 in all the calculated structures (the H−O distance is around 2.0 Å, and the N−H−O angle is around 32°). This different number of hydrogen bonds is probably the main reason for difference in ΔG values between the two dimers. An additional reason for their different stability, is the proximity between the O2 atoms of the base-paired thymines across the tetrads in d<pCATTCATT>. In the crystallographic structure of this molecule, the electrostatic repulsion is compensated by the presence of a metal ion in the center of the tetrad. Although no ion was placed in that position in our calculations, a sodium ion could fit in the center of the dimer, according to our structures. Interestingly, although the solution structures of d<pTGCTCGCT> and d<pCATTCATT> are very similar (see Figure 8A), the later present a larger rise between the central base pairs (see Table 3), and a larger separation between the two stacks (around 1 Å more). These differences may have implications in the cation binding properties of the two molecules.

In addition to the hydrogen bonds, the structures are stabilized by a number of hydrophobic contacts between the deoxyribose moieties and by the favorable stacking interactions of the thymines in positions 4 and 8. These thymines are stacked against the neighboring layer and show a tilt value of around 30°. A shift of 10 Å places the thymine base in the middle of the underlying base pair, forming a cap at both ends of the stacks. The glycosidic angles of these residues are also *anti*, with an average value of −150°. The two thymines located at the same side of the dimer are not base-paired, but they interact with each other via a hydrophobic contact between their methyl groups (see Figure 9D). This orientation is well supported, among others, by the NOEs between the methyl group of T4 and the exchangeable protons of the neighboring base pair (see Figures 1 and 2).

In both dimers, several short phosphorus−phosphorus distances are observed in each subunit. Values of 5.5−6.0 Å are obtained for the distances between phosphorus in positions 3−4 and 7−8, located in the turns of the backbone. Moreover, short interstrand distances are observed between phosphates 3 and 7 (∼5.5 Å). The repulsion between these nearby phosphates is partially compensated by favorable hydrophobic contacts between deoxyriboses.

Comparison with Other Four-Stranded Structures. The most studied four-stranded motif of DNA is the G-quadruplex, where four guanines are paired through their Watson−Crick and Hoogsteen sides. However, other pyrimidine containing tetrads have been observed recently in oligonucleotides with the G-quadruplex[7,33,34] or i-motif folding.[35] The most related to our case are the G:C:G:C tetrads found in the structure of the fragile X syndrome d(C−G-G)n triplet repeat,[5] and in the G−G-G−C repeats of the adeno-associated viral DNA.[6,36] In both cases, the quadruplex arises from dimerization of two DNA hairpins, stabilized by several G-tetrads. The layers with G:C:G:C tetrads are formed by intramolecular Watson−Crick base pairs which associate through the major groove side of the base pairs. Intermolecular bifurcated hydrogen bonds are formed between the cytosine NH4 and the guanine O6 and N7 atoms (see Figure 9C).

In the structure of d<pTGCTCGCT> the topology of the G:C:G:C tetrad is different. Dimerization occurs through formation of intermolecular Watson−Crick base pairs, which form additional hydrogen bonds through their minor groove edge. Contrary to the previous cases, the quadruplex is formed entirely by G:C:G:C tetrads without any G-tetrad layer. This different arrangement has important consequences in the width of the grooves. Two of the four grooves are extremely narrow, with O4′−O4′ distance as small as 4.5−5.5 Å, while the other two are very wide (14.5 Å).

(33) Patel, P. K.; Hosur, R. V. *Nucl. Acids Res.* **1999**, *27*, 2457−2464.
(34) Patel, P. K.; Koti, A. S. R.; Hosur, R. V. *Nucl. Acids Res.* **1999**, *27*, 3836−3843.
(35) Gallego, J.; Chou, S.; Reid, B. R. *J. Mol. Biol.* **1997**, *273*, 840−856.
(36) Bouaziz, S.; Kettani, A.; Patel, D. J. *J. Mol. Biol.* **1998**, *282*, 637−652.

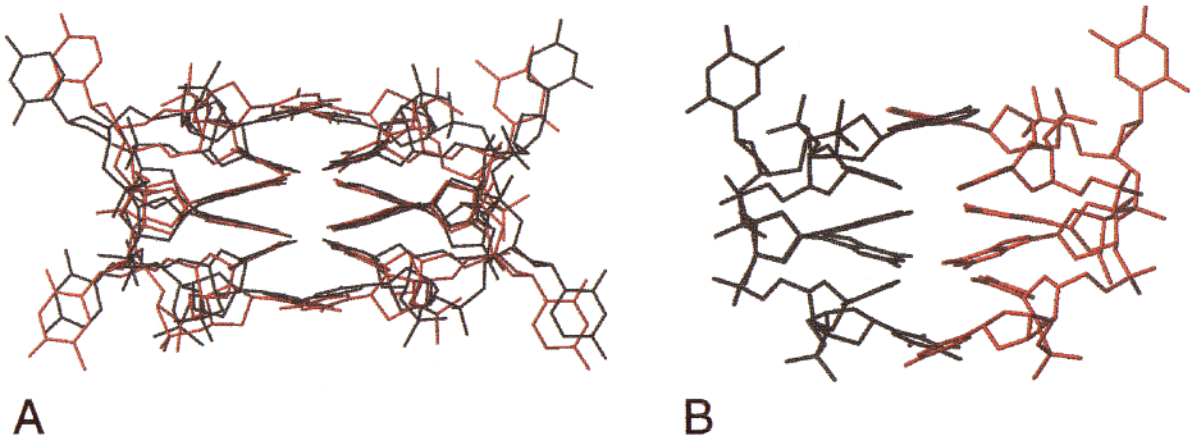


Figure 8. (A) Superposition of the average solution structures of d<pTGCTCGCT> and d<pCATTCATT>. (B) Crystal structure of the linear heptamer d(GCATGCT).

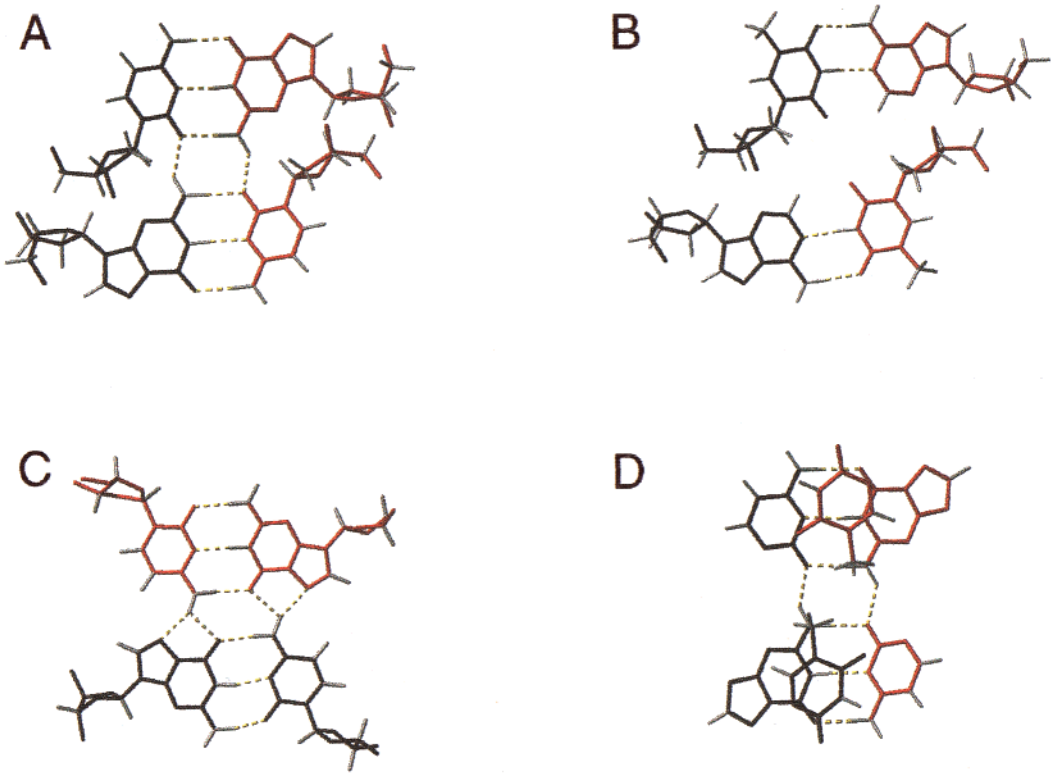


Figure 9. Detail of interactions. (A) G:C:G:C tetrad of d<pTGCTCGCT>; (B) A:T:A:T tetrad in d<pCATTCATT>; (C) G:C:G:C tetrad of d(GCGGT$_3$GCGG) by Kettani et al.;[5] (D) hydrophobic interaction between the capping thymines in the structure of d<pTGCTCGCT>. Each subunit is shown in a different color. Hydrogen bonds are indicated with dashed lines.

To our knowledge, this is the first time that a four-stranded DNA oligonucleotide, containing only Y:R:Y:R tetrads, has been observed in solution. However, similar structures have been found in solid state. In addition to the previously mentioned X-ray structure of d<pCATTCATT>, the linear oligonucleotide d(GCATGCT) adopts a very similar fold in the crystal (see Figure 8). In both cases, the structures are stabilized by a number of intermolecular interactions between symmetry-related units. For example, in the case of d<pCATTCATT>, the crystal is stabilized by continuous stacking interactions between the cytosines (which are disordered in solution) along different cell units. In the case of d(GCATGCT), the thymines in position 4 are forming wobble TT base pairs with crystal symmetry-related molecules. The presence of these contacts may cast some doubt on the occurrence of such structures in vivo. The NMR structure of d<pCATTCATT>, reported in this study, confirms the main features of the crystal structure, indicating that the bi-loop motif is also stable in solution conditions.

More interesting is the comparison between the solution structure of d<pTGCTCGCT> and the crystal structure of d(GCATGCT). The core of both dimers consists of two G:C:G:C tetrads. However, the bases capping the stacks are different. The linear heptamer is less symmetric, with one end of the stacks capped by thymines (T7), and the other by adenines (A3). The T7 in both subunits interact with each other through a hydrophobic methyl contact, whereas an A3 forms an AA intermolecular base pair. The similarity between these two structures indicates that both belong to the same DNA structural motif. The different identity of the capping bases shows that neither the AA hydrogen bond nor the hydrophobic contact

between the thymines is essential for the stability of the motif. Most probably, the stacking interaction between the capping residues and the rest of the stack plays an important role in the stability of the structure, since, in the three cases, the geometry of these bases is very similar.

Biological Implications. The structures of d<pTGCTCGCT> and d<pCATTCATT> are good examples of self-recogniziton between DNA molecules, and present features that may be relevant to chromosomal DNA organization or tight packing of DNA in single-stranded DNA virus. Our finding of four-stranded structures consisting entirely of G:C:G:C and A:T:A:T tetrads may have implications in our understanding of those biological processes where four-stranded structures are being invoked. Thus, additional sequences should be considered in processes such as genetic recombination, telomerase inhibition, meiosis, and so forth, where four-stranded DNA structures may be involved.

For many years, it has been speculated that homologous recombination could be initiated by formation of quadruplex structures.[3] The arrangement of G:C:G:C and A:T:A:T tetrads could serve to bring into register homologous DNA fragments. Thus, formation of such structures could be a first step in the recombination process prior to strand exchange through Holliday junctions. Feasible models of such structures were proposed as early as 1979.[37] More recently, the observation of four-stranded structures formed by poly(CA)·poly(TG) fragments[38] and their recognition by the non-histone high mobility group proteins (HMG) has stimulated the investigation of more sophisticated models by using modern theoretical methods.[39] In these models, tetrad formation occurs through alignment of the major groove side of the GC or AT base pairs. Such orientation is in agreement with the two G:C:G:C tetrads reported by Patel's group. However, these two tetrads have been observed in the framework of G-quadruplexes, and their orientation could be affected by the adjacent G-tetrads. In this paper, we show that tetrad formation through the minor groove side of Watson−Crick base pairs is also possible. This different orientation should be taken into account in the strand- exchange models.

Coordinates. Atomic coordinates have been deposited in the Protein Data Bank (accession numbers 1EU2 and 1EU6).

Acknowledgment. We gratefully acknowledge Dr. Doug Laurent and Dr. José Gallego for careful reading of the manuscript. This work was supported by the DGICYT Grants PB97-0941-C02-01/02 and Generalitat de Catalunya (Centre de Referència en Biotecnologia).

Supporting Information Available: Two figures with regions of the NOESY spectrum of d<pCATTCATT> in H_2O, and plots of K_{eq} vs T for d<pTGCTCGCT> and d<pCATTCATT>; two tables with average values and order parameters of the dihedral torsion angles of d<pTGCTCGCT> and d<pCATTCATT> (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA002778Q

(37) Wilson, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 3641−3645.
(38) Gaillard, C.; Strauss, F. *Science* **1994**, *264*, 433−6.
(39) Lebrun, A.; Lavery, R. *J. Biomol. Struct. Dyn.* **1995**, *13*, 459−464.