Four-stranded DNA structure stabilized by a novel G : C : A : T tetrad

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · MAY 2003

Impact Factor: 12.11 · DOI: 10.1021/ja0344157 · Source: PubMed

6 AUTHORS, INCLUDING:

Núria Escaja
University of Barcelona
23 PUBLICATIONS **237** CITATIONS
SEE PROFILE

Josep L Gelpí
University of Barcelona
102 PUBLICATIONS **3,021** CITATIONS
SEE PROFILE

Enrique Pedroso
University of Barcelona
140 PUBLICATIONS **1,957** CITATIONS
SEE PROFILE

Carlos González
Spanish National Research Council
131 PUBLICATIONS **2,521** CITATIONS
SEE PROFILE

J|A|C|S
ARTICLES
Published on Web 04/16/2003

Four-Stranded DNA Structure Stabilized by a Novel G:C:A:T Tetrad

Núria Escaja,[†] Josep Lluis Gelpí,[‡] Modesto Orozco,[‡] Manuel Rico,[§] Enrique Pedroso,*[,†] and Carlos González*[,§]

Contribution from the Instituto de Química Física "Rocasolano", CSIC, C/, Serrano 119, 28006 Madrid, Spain, Departament de Química Orgànica, Universitat de Barcelona, C/, Martí i Franquès 1-11, 08028 Barcelona, Spain, and Departament de Bioquímica, Universitat de Barcelona, C/, Martí i Franquès 1-11, 08028 Barcelona, Spain

Received January 30, 2003; E-mail: cgonzalez@iqfr.csic.es; pedroso@qo.ub.es

Abstract: The solution structure of a cyclic oligonucleotide d⟨pCGCTCATT⟩ has been determined by two-dimensional NMR spectroscopy and restrained molecular dynamics. Under the appropriate experimental conditions, this molecule self-associates, forming a symmetric dimer stabilized by four intermolecular Watson−Crick base pairs. The resulting four-stranded structure consists of two G:C:A:T tetrads, formed by facing the minor groove side of the Watson−Crick base-pairs. Most probably, the association of the base-pairs is stabilized by coordinating a Na^+ cation. This is the first time that this novel G:C:A:T tetrad has been found in an oligonucleotide structure. This observation increases considerably the number of sequences that may adopt a four-stranded architecture. Overall, the three-dimensional structure is similar to those observed previously in other quadruplexes formed by minor groove alignment of Watson−Crick base pairs. This resemblance strongly suggests that we may be observing a general motif for DNA−DNA recognition.

Introduction

Four-stranded DNA structures have attracted considerable attention during the last years. It has been suggested that these structures may play a role in a number of biological processes such as telomere function, genetic recombination, transcription, or replication, and they have been also found in vitro in DNA sequences associated with human disease (for reviews see refs 1−6). Very recently, evidence of quadruplex formation in vivo has been found by antibody staining of protozoan nuclei.[7]

The most studied four-stranded motif in DNA is the G-quadruplex, where four guanines are paired through their Watson−Crick and Hoogsteen sides. However, other quadruplex structures with tetrads not exclusively composed by guanine bases have been found recently.[8−16] Among these new tetrads, those formed by association of Watson−Crick base-pairs are of particular interest, since they have been invoked in processes where homologous DNA recognition is required, such as genetic recombination.[17] These tetrads may be responsible for the four-stranded structures observed in poly(CA).poly(TG) fragments.[18]

Mixed tetrads consisting of two G−C base-pairs have been found in the solution structures of the fragile X syndrome d(C−G−G)n triplet repeat,[9] and in the G−G−G−C repeats of the adeno-associated viral DNA.[11,19] In both cases, the two G−C base-pairs interact through their major groove sides. In addition, other structural studies have identified G:C:G:C tetrads where the base pairs interact through their minor groove sides.[8,15,20] Mixed tetrads formed by two A−T base-pairs have been also observed in solution[15,16] and in crystallographic studies.[10,21] As in the case of G:C:G:C tetrads, the two possible alignments,

[†] Departament de Química Orgànica, Universitat de Barcelona.
[‡] Departament de Bioquímica, Universitat de Barcelona.
[§] Instituto de Química Física "Rocasolano", CSIC.

(1) Arthanari, H.; Bolton, P. H. *Chem. Biol.* **2001**, *8*, 221−230.
(2) Suhnel, J. *Biopolymers* **2001**, *61*, 32−51.
(3) Keniry, M. A. *Biopolymers* **2000**, *56*, 123−146.
(4) Feigon, J.; Koshlap, K. M.; Smith, F. W. *Methods Enzymol.* **1995**, *261*, 225−255.
(5) Patel, D. J.; Bouaziz, S.; Kettani, A.; Wang, Y. In *Oxford Handbook of Nucleic Acid Sructures*; Neidle, S., Ed.; Oxford University Press: New York, 1999; pp 389−453.
(6) Williamson, J. R. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 703−730.
(7) Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Pluckthun, A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8572−8577.
(8) Leonard, G. A.; Zhang, S.; Peterson, M. R.; Harrop, S. J.; Helliwell, J. R.; Cruse, W. B.; d'Estaintot, B. L.; Kennard, O.; Brown, T.; Hunter, W. N. *Structure* **1995**, *3*, 335−340.
(9) Kettani, A.; Kumar, R. A.; Patel, D. J. *J. Mol. Biol.* **1995**, *254*, 638−656.
(10) Salisbury, S. A.; Wilson, S. E.; Powell, 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.
(11) Kettani, A.; Bouaziz, S.; Gorin, A.; Zhao, H.; Jones, R. A.; Patel, D. J. *J. Mol. Biol.* **1998**, *282*, 619−636.
(12) Patel, P. K.; Hosur, R. V. *Nucleic Acids Res.* **1999**, *27*, 2457−2464.
(13) Patel, P. K.; Koti, A. S.; Hosur, R. V. *Nucleic Acids Res.* **1999**, *27*, 3836−3843.
(14) Patel, P. K.; Bhavesh, N. S.; Hosur, R. V. *Biochem. Biophys. Res. Commun.* **2000**, *270*, 967−971.
(15) Escaja, N.; Pedroso, E.; Rico, M.; González, C. *J. Am. Chem. Soc.* **2000**, *122*, 12 732−12 742.
(16) Zhang, N.; Gorin, A.; Majumdar, A.; Kettani, A.; Chernichenko, N.; Skripkin, E.; Patel, D. J. *J. Mol. Biol.* **2001**, *312*, 1073−1088.
(17) McGavin, S. *J. Mol. Biol.* **1971**, *55*, 293−298.
(18) Gaillard, C.; Strauss, F. *Science* **1994**, *264*, 433−436.
(19) Bouaziz, S.; Kettani, A.; Patel, D. J. *J. Mol. Biol.* **1998**, *282*, 637−652.
(20) Thorpe, J. H.; Teixeira, S. C.; Gale, B. C.; Cardin, C. J. *Nucleic Acids Res.* **2003**, *31*, 844−849.
(21) Parkinson, G. N.; Lee, M. P.; Neidle, S. *Nature* **2002**, *417*, 876−880.

10.1021/ja0344157 CCC: $25.00

through minor[10,15] or major groove[16,21] pairing of the two Watson−Crick A−T base-pairs, have been found.

We are interested in exploring other possible base-pair arrangements susceptible of forming tetrads as well as the sequence requirements for their formation. Previous studies have shown that cyclic oligonucleotides are good models to study these noncanonical DNA structures. The two cyclic octamers d⟨pTGCTCGCT⟩ and d⟨pCATTCATT⟩ form homodimers, which coexist in solution with monomeric dumbbell-like structures.[22] In that case, it was demostrated that self-association of these molecules occurs by forming intermolecular Watson−Crick base-pairs, in principle, however intramolecular base-pairs could also be formed. For DNA dimers formed by linear oligonucleotides, this uncertainty can be resolved by using pure and mixed labeled samples in ^{15}N and ^{13}C.[16,23−25] However, the synthesis of uniformly labeled cyclic oligonucleotides, in quantities needed for NMR studies, is not affordable with the current technology. For this reason, we decided to explore different sequences, like that of the cyclic oligonucleotide d⟨pCGCTCATT⟩ reported here, that can form dimers, but not stable intramolecular base-pairs.

In this paper, we report on the three-dimensional solution structure of d⟨pCGCTCATT⟩ as obtained by restrained molecular dynamics calculations based on NMR derived experimental constraints. Whereas in diluted conditions all experimental evidence indicates that this oligonucleotide does not adopt a defined structure, at high oligonucleotide concentration this molecule dimerizes, by forming intermolecular Watson−Crick base-pairs (two G−C and two A−T). The two resulting G:C:A:T tetrads are formed by association of G−C and A−T base-pairs through their minor groove sides.

Methods

Experimental Details. The cyclic octamers were synthesized as reported by Alazzouzi et al.[26] Samples were suspended (in Na^+ salt form) in 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 XWIN NMR software. In the experiments in D_2O, presaturation was used to suppress the residual H_2O signal. A jump-and-return pulse sequence[27] was employed to observe the rapidly exchanging protons in 1D H_2O experiments. NOESY[28] spectra in D_2O were acquired with mixing times of 100, 200, and 300 ms. TOCSY[29] spectra were recorded with the standard MLEV-17 spin-lock sequence and a mixing time of 80ms. In 2D experiments in H_2O, water suppression was achieved by including a WATERGATE[30] module in the pulse sequence prior to acquisition. ^{31}P resonances were assigned from proton-detected heteronuclear correlation spectra.[31] The spectral analysis program XEASY[32] was used for semiautomatic assignment of the NOESY cross-peaks, and quantitative evaluation of the NOE intensities.

NMR Constraints. Despite the intrinsic ambiguity between inter- and intramolecular distances in dimeric structures, most of the cross-peaks could be directly assigned on the basis of a rough model of the molecule. Trial assignments were made for the few remaining cross-peaks, and used in preliminary structure 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. Any remaining cross-peaks, that were still ambiguous, were not included in the subsequent calculations. These initial calculations were carried out with qualitative distance constraints (classified as 3, 4, or 5 Å).

Refined structures were calculated employing more accurate distance constraints. These distances were obtained from NOE cross-peak intensities by using a complete relaxation matrix analysis with the program MARDIGRAS.[33] 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 interprotonic 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 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 non-exchangeable protons, and at a single mixing time (200 ms) for labile protons. Final constraints were obtained by averaging the upper and lower distance bounds in all the MARDIGRAS 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. Because 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[34] and further refined with the SANDER module of the molecular dynamics package AMBER 5.0.[35] 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. All counterions were place around the exterior of the molecule and near the phosphate groups. The electrostatic term was calculated using a distance-dependent dielectric constant, and the cutoff value for nonbounded interactions was 10 Å. The temperature and the relative weights of the experimental constraints were varied during the simulations according to standard annealing protocols used in our group.[15,36]

(22) González, C.; Escaja, N.; Rico, M.; Pedroso, E. *J. Am. Chem. Soc.* **1998**, *120*, 2176−2177.
(23) Kettani, A.; Bouaziz, S.; Skripkin, E.; Majumdar, A.; Wang, W.; Jones, R. A.; Patel, D. J. *Structure Fold Des.* **1999**, *7*, 803−815.
(24) Kuryavyi, V.; Kettani, A.; Wang, W.; Jones, R.; Patel, D. J. *J. Mol. Biol.* **2000**, *295*, 455−469.
(25) Zhang, N.; Gorin, A.; Majumdar, A.; Kettani, A.; Chernichenko, N.; Skripkin, E.; Patel, D. J. *J. Mol. Biol.* **2001**, *311*, 1063−1079.
(26) Alazzouzi, E.; Escaja, N.; Grandas, A.; Pedroso, E. *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 1506−1508.
(27) Plateau, P.; Güeron, M. *J. Am. Chem. Soc.* **1982**, *104*, 7310−7311.
(28) Kumar, A.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1−6.
(29) Bax, A.; Davies, D. J. *J. Magn. Reson.* **1985**, *65*, 355−360.
(30) Piotto, M.; Saudek, V.; Sklenar, V. *J. Biomol. NMR* **1992**, *2*, 661−665.
(31) Sklenar, V.; Miyashiro, H.; Zon, G.; Miles, H. T.; Bax, A. *FEBS Lett.* **1986**, *208*, 94−98.
(32) Bartels, C.; Xia, T.; Billeter, M.; Güntert, P.; Wüthrich, K. *J. Biomol. NMR* **1995**, *6*, 1−10.
(33) Borgias, B. A.; James, T. L. *J. Magn. Reson.* **1990**, *87*, 475−487.
(34) Guntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol.* **1997**, *273*, 283−298.
(35) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; III, T. E. C.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A.; 5 ed.; University of California: San Francisco, 1997.

Analysis of the average structures as well as the MD trajectories was carried out with the programs Curves V5.1,[37] and MOLMOL.[38] Reliability factors (R-factors) were calculated with the program CORMA.[33]

CMIP Calculations. Classical Molecular Interaction (cMIP) calculations[39] were carried out to evaluate the shape of the cavity formed in the center of the dimeric structure of d⟨pCGCTCATT⟩, and its ability to accommodate a Na^+ counterion. cMIP calculations were performed on a grid centered inside of the oligonucleotide. Interaction energies at each grid point were computed by using the full interaction energy (electrostatic + van der Waals) between the oligonucleotide and the Na^+ ion (both defined using standard AMBER force-field).[40,41] The electrostatic energy was determined using the linear form of the Poisson−Boltzmann equation (for a salt concentration of 100 mM NaCl), using internal and external dielectric constants of 2 and 80, respectively. The focusing procedure (with a final grid spacing of 0.6 Å) was used to improve the definition quality of the electrostatic potential inside the oligonucleotide.[42]

Calculations for the G:C:A:T tetrad were carried out considering the global averaged structure (single grid calculations; SGC), and also using all the MD-averaged structures to define a Boltzman-averaged consensus grid (Boltzman-averaged grid calculations; BGC), which allows us to include flexibility considerations in the determination of the reactive characteristics of the central cavity of a molecule. Reference calculations for the G:C:G:C and A:T:A:T tetrads were computed following the same SGC procedure for the experimental structures (pdb codes: 1eu2 and 1eu6).[15]

The SGC procedure computes the total interaction energy (E_{TOT}) directly from eq 1

$$E^i_{TOT} = E^i_{vw} + E^i_{ele} \quad (1)$$

where E_{vw} and E_{ele} are the van der Waals and electrostatic energies, and i stands for a grid point.

When the BGC procedure is used the total interaction energy (E_{TOT}) is computed as shown in eq 2, where the van der Waals energy is determined as shown in eq 3, and the Boltzman factor (ζ) used to compute the average electrostatic energy is determined as shown in eq 4

$$E^i_{TOT} = E^i_{vw} + \sum_s \zeta^s_i E^{i,s}_{ele} \quad (2)$$

where i and s stand for a grid point defined from structure S

$$E^i_{vw} = -RT \ln N^{-1} \sum_{s=1}^{N} \exp(-E^{i,s}_{vw}/RT) \quad (3)$$

where N is the number of structures considered

$$\zeta^s_i = \frac{\exp(-E^{i,s}_{vw}/kT)}{\sum_s \exp(-E^{i,s}_{vw}/kT)} \quad (4)$$

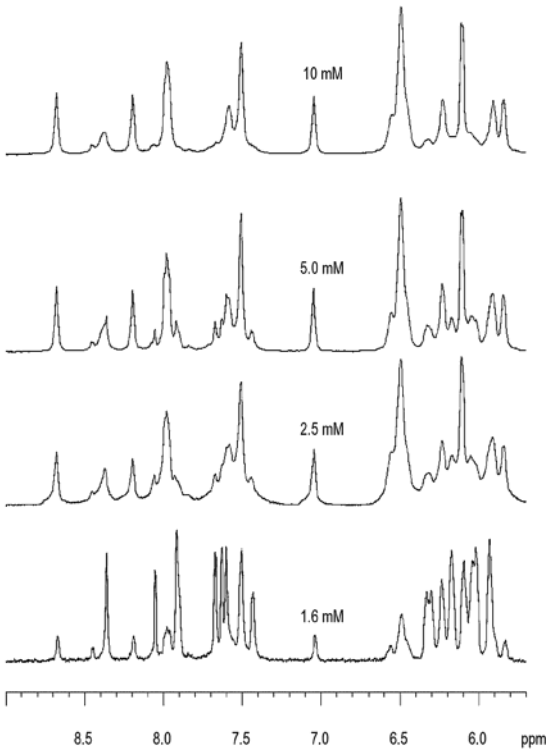

Figure 1. One-dimensional NMR spectra of d⟨pCGCTCATT⟩ in D_2O at 5 °C and different concentrations.

Results

Monomer−Dimer Equilibrium. The NMR spectra of d⟨pCGCTCATT⟩ depends strongly on oligonucleotide and salt concentration. At low temperature and salt concentration, an interconverting equilibrium between two species is observed. This equilibrium is slow on the NMR time scale, so that resonances from both species can be observed simultaneously. In these circumstances, the equilibrium constants at different temperatures can be determined from the ratio of the peak areas. Because the populations of both species depend on the oligonucleotide concentration (see Figure 1), and all CD and NMR data show that the major conformer at low concentration does not have a well-defined structure,[43] it can be concluded that this is an association equilibrium between a dimeric structure and a random coil monomeric form. Thermodynamic parameters for this dimer-random coil equilibrium were estimated from a van't Hoff analysis of the equilibrium constants at different temperatures giving a ΔG^0_{298} value of −9 kJ/mol in 25 mM salt concentration.

The population of the dimeric form is higher when salt concentration is increased, and no variation in the NMR signals of this form is observed upon addition of NaCl. For this reason, most of the NMR experiments were recorded at 100 mM NaCl and 5 mM oligonucleotide concentration, where the population of the dimeric form at 5 °C is essentially 100%.

NMR Assignment. Sequential assignments of exchangeable and nonexchangeable protons and ^{31}P resonances were conducted following standard methods. Fragments of the two-dimensional NOESY spectrum of d⟨pCGCTCATT⟩ in D_2O are shown in Figure 2. All intra-nucleotide H1′-base NOEs are

(36) Soliva, R.; Monaco, V.; Gomez-Pinto, I.; Meeuwenoord, N. J.; Marel, G. A.; Boom, J. H.; Gonzalez, C.; Orozco, M. *Nucleic Acids Res.* **2001**, *29*, 2973−2985.
(37) Lavery, R.; Sklenar, H.; 3.0 ed.; Laboratory of Theoretical Biochemistry CNRS: Paris, 1990.
(38) Koradi, R.; Billeter, M.; Wuthrich, K. *J. Mol. Graphics* **1996**, *14*, 29−32.
(39) Gelpí, J. L.; Kalko, S.; de la Cruz, X.; Barril, X.; Cirera, J.; Luque, F. J.; Orozco, M. *Proteins* **2001**, *45*, 428−437.
(40) Cheatham, T. E., 3rd; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−862.
(41) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(42) Orozco, M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187−4225.
(43) Escaja, N.; Gómez-Pinto, I.; Rico, M.; Pedroso, E.; González, C. *ChemBioChem* **2003**, in press.

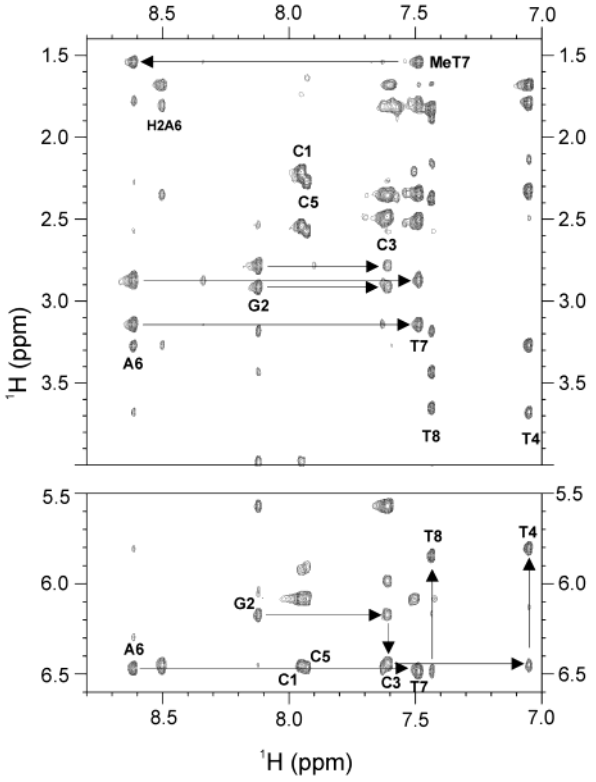

Figure 2. Regions of the NOESY spectrum (300 ms mixing time) of d⟨pCGCTCATT⟩ in D_2O (5 mM oligonucleotide concentration, 100 mM NaCl, $T = 5$ °C, pH = 7). Sequential assignment pathways are drawn in the H1′-aromatic region.

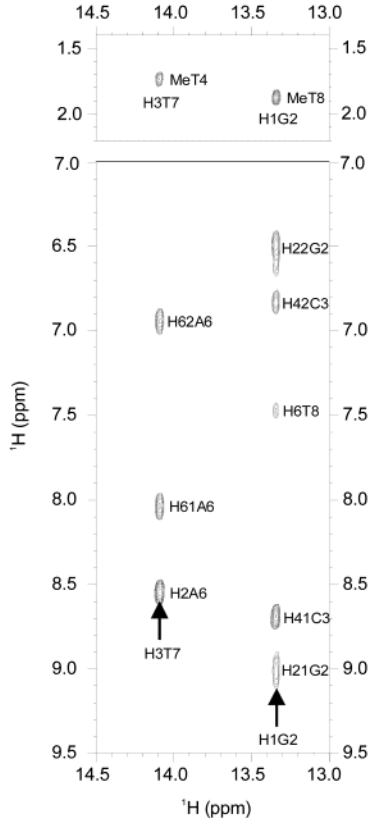

Figure 3. Regions of the NOESY spectrum (150 ms mixing time) of d⟨pCGCTCATT⟩ in H_2O (5 mM oligonucleotide concentration, 100 mM NaCl, $T = 5$ °C, pH = 7). Watson−Crick base pairing can be established from H1G2−H42C3 and H1G2−H41C3 cross-peaks for the GC pair, and H3T7−H2A6, H3T7−H62A6, and H3T7−H61A6 for the AT pair.

medium or weak, indicating that the glycosidic angle in all the nucleotides is in an anti conformation. Strong sugar-base sequential connections were observed between residues 2 → 3, and 6 → 7. Sequential NOEs between residues 3 → 4 and 7 → 8 are clearly observed in the H1′-base region, but are very weak in the H2′/H2″-base region (see Figure 2). Only intra-residual NOEs were observed for the bases of C5 and C1. Almost all resonances were identified, including the stereoespecific assignment of some H5′/H5″ protons (see Table S1). Unusual values are observed for the chemical shifts of the H4′ and H5′/H5″ protons of T4 and T8 (see Figure 2). These shifts are due to the proximity of the purine bases of A6 and G2, respectively, in the three-dimensional structure.

The exchangeable proton spectra are particularly informative (see Figure 3). Amino protons of all cytosines were assigned from their NOE cross-peaks with the H5 protons. In the case of C1 and C5, these resonances are broad and completely degenerated, but in case of C3, the two amino protons present strong cross-peaks with the imino signal observed at 13.2 ppm. This signal was assigned to G2, and its amino resonances could be also identified. This NOE pattern is characteristic of a GC Watson−Crick base-pairs. The other narrow imino signal (14.1 ppm) presents a number of cross-peaks that are typical of an A−T Watson−Crick base-pair (cross-peaks with adenine H2 and amino protons). This signal was assigned to T7. Two more imino protons are observed in the 10−11 ppm region, indicating that the other two thymines are not base-paired.

Only one of the amino protons of the base-paired cytosines resonates at low field (8.67 ppm), indicating that the other amino (6.83 ppm) is not hydrogen bonded. However, the two guanine amino protons present rather narrow signals, suggesting that neither of them are exposed to the solvent. These amino protons show the same behavior as that observed in the dimeric form of d⟨pTGCTCGCT⟩.[15] A number of NOEs involving guanine amino protons, like the NOE with the H1′ of T7, clearly indicates that these protons are in the core of the dimer, and therefore, the two base-pairs are facing each other through their minor grooves. This arrangement is also supported by other contacts, like the strong NOE between H2A6 and H1′C3.

Table 1. Experimental Constraints and Calculation Statistics

experimental distance constraints		
total no.		334
intraresidue		210
sequential		50
range > 1.		74
intramolecular		300
intermolecular		34
RMSD (Å)		
all well-defined bases[+]		0.4 ± 0.1 Å
all well-defined heavy atoms[+]		0.6 ± 0.1 Å
backbone.		1.1 ± 0.2 Å
all heavy atoms.		2.4 ± 0.6 Å
residual violations	**average**	**range**
sum of violation (Å)	9.1	8.64.. 9.48
max. violation (Å)	0.5	0.39.. 0.49
R-factor (×100)[&]	6.02	5.85.. 6.22
NOE energy (kcal/mol)	66	59.. 71
total energy (kcal/mol)	−2333	−2384.. −2269

[+] All except residues 1, 5, 9 and 13. [&] Sixth-root R-factor.

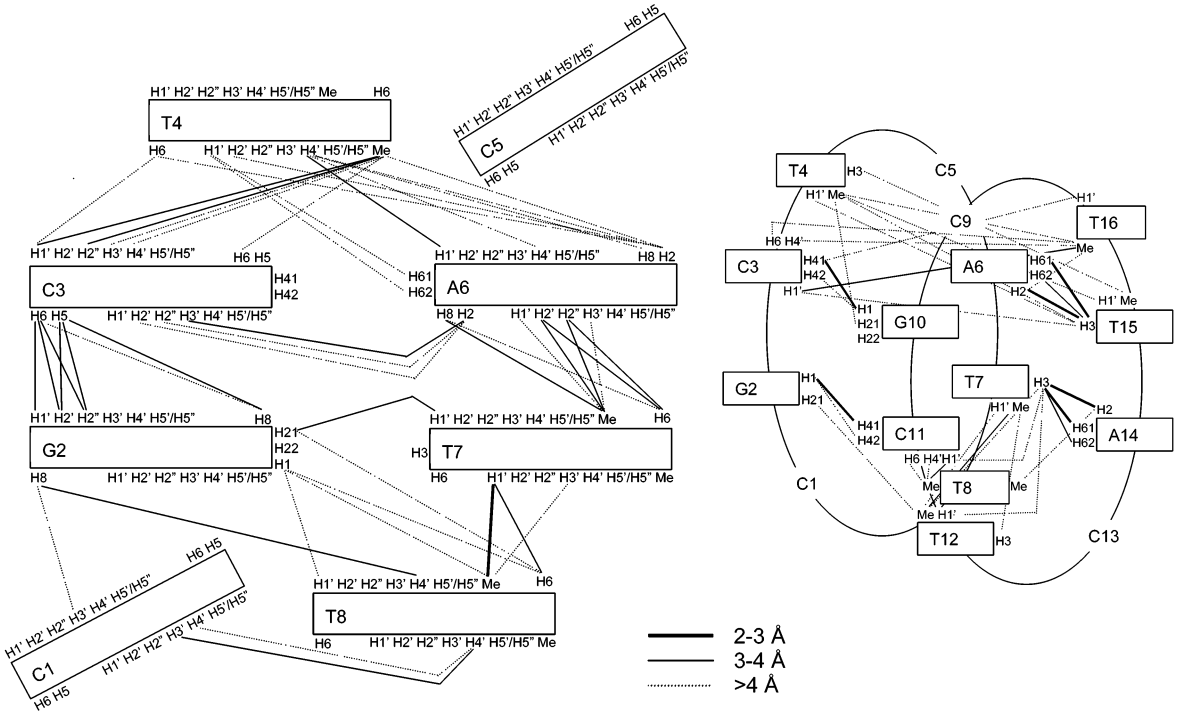

Figure 4. Schematic representation of (left) intra- and (right) intermolecular distance constraints. Constraints are classified in three categories according to their upper distance limit.

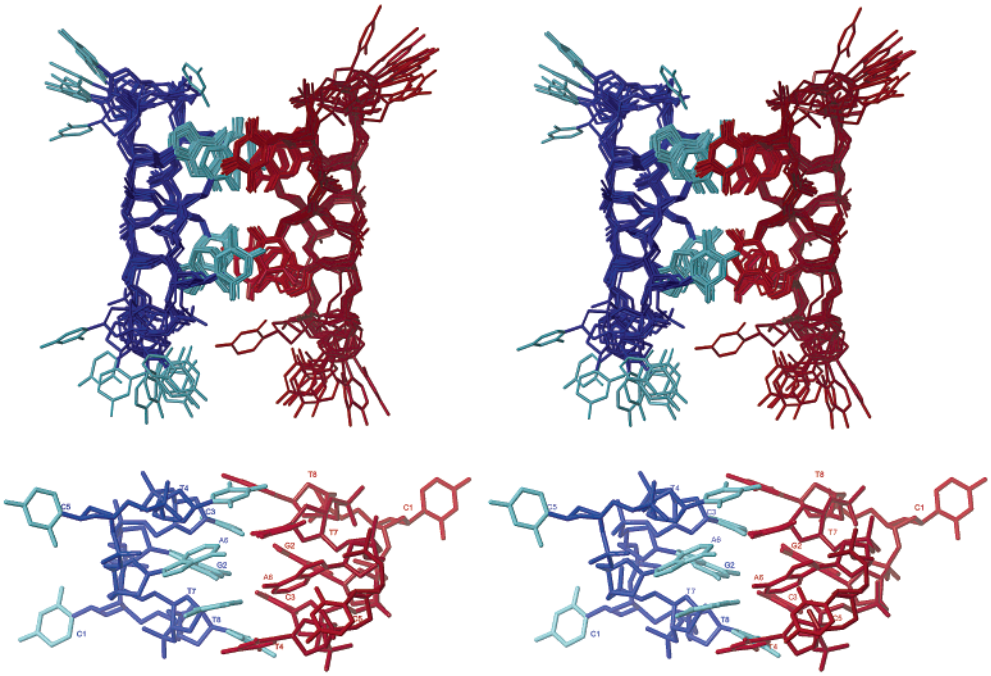

Figure 5. Top: Stereoscopic views of the superposition of the 10 refined structures of d⟨pCGCTCATT⟩. The axis of the molecule runs perpendicular to the view plane. Bottom: Stereoscopic view of the average structure (lateral view). Red and blue indicate different molecules. The sugar−phosphate backbone is indicated in darker colors.

Experimental Constraints and Structure Calculations. In all dimeric molecules, there is an intrinsic ambiguity between inter- and intramolecular cross-peaks. In the present case, the oligonucleotide was designed so that only intermolecular base-pairs can be formed. This permitted us to build an initial model of the dimer that served to eliminate most of the ambiguities. In the few cases where several distances might correspond with a NOE cross-peak, the final assignment could be done by carrying out several structure calculations with trial assignments (see Materials and Methods for details). After the complete relaxation matrix calculation, a total of 334 distance constraints where obtained. A summary of these constraints is shown in Table 1 and 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. The population of the major S conformer, estimated from the sum of $J_{1'2'}$ and $J_{1'2''}$,[44] was in all cases greater that 90%. Vicinal J-coupling constants between H4′ and H5′

Table 2. Local Helical Parameters for the Dimeric Structure of d⟨pCGCTCATT⟩

step	shift Dx (Å)		slide Dy (Å)		rise Dz (Å)		tilt τ (°)		roll ρ (°)		twist Ω (°)	
G2/C3	0.1	±0.1	−1.5	±0.3	3.1	±0.1	2	±2	3	±3	−27	±1
C3/*T*4	−9.9	±0.3	−1.3	±0.2	5.6	±0.4	−34	±5	−19	±3	36	±2
A6/*T*7	0.2	±0.2	0.6	±0.4	4.4	±0.4	−8	±4	−21	±2	28	±2
T7/*T*8	−10.3	±0.3	2.2	±0.1	3.6	±0.7	−32	±4.	7	±5	−38	±2
T12/C11	−9.6	±0.4	1.3	±0.2	6.0	±0.7	−36	±7	18	±5	34	±3
C11/G10	0.2	±0.1	1.5	±0.3	3.1	±0.1	2	±1	−3	±2	−28	±2
T16/*T*15	10.5	±0.3	2.2	±0.2	3.4	±0.8	30	±4	8	±4	−40	±4
T15/A14	−0.2	±0.6	0.8	±0.3	4.3	±0.3	7	±2	−22	±3	29	±6

and H5″ were particularly small in the case of residues 2, 4, 6, and 8. These J-values together with the pattern of intra-residual NOEs between H3′/H4′ and H5′/H5″ obtained using a short mixing time, permitted the stereospecific assignment of some of the H5′/H5″,[45] and indicates that the backbone angle γ in these nucleotides is in a single γ^+ conformation.

All of these experimental constraints were used to calculate the structure of d⟨pCGCTCATT⟩ 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 1). Except for the C1 and C5, and the corresponding ones in the symmetry related subunit, all residues are very well defined, with an RMSD of 0.75 Å (see Figure 5). These values are even lower when only atoms in the bases are considered (see Table 1). The final AMBER energies and NOE terms are reasonably low in all the structures, with no distance constraint violation greater than 0.5 Å.

Description of the Structures. The resulting structure is a dimer consisting of two molecules of d⟨pCGCTCATT⟩ arranged in an antiparallel way. With the exception of the disordered residues (C1 and C5), the dimer is symmetric, as reflected in the geometrical parameters (Tables 2 and S2). The two octamers associate with each other by forming four intermolecular Watson−Crick base-pairs (see Figure 5). These base-pairs form two G:C:A:T tetrads by aligning their minor groove sides. All glycosidic angles in these tetrads are anti, with values ranging from −88° to −118°. In addition to the five Watson−Crick hydrogen bonds, the tetrad may be stabilized by an additional intramolecular H-bond between one of the amino protons of G2 and the O4′ of the T7 deoxyribose. The guanine amino proton resonating at 8.98 ppm is forming the Watson−Crick hydrogen bond, whereas the other proton (6.47 ppm) is very close to the thymine O4′ in all the calculated structures (the H−O distance is around 2.2 Å, and the N−H−O angle is around 145°, almost within the hydrogen bonding distance). Like in other minor grove tetrads, the two base-pairs are not in the same plane (see Figure 5), but have a mutual inclination of around 40°. The base-pairs are arranged in two stacks, one formed by G−C pairs and the other by A−T pairs. In both cases the stacking is very favorable, with almost no displacement of the base-pairs along the axis defined by their own stack (see Figure 6). However, in the A−T stack the base-pairs have an slight propeller twist of around 20°. The average rise between the base pairs is 3.1 Å in the GC stack, 4.3 Å in the AT one.

In addition to the hydrogen bonds, the structure is stabilized by a number of hydrophobic contacts between the deoxyribose moieties, and by the favorable stacking interactions of the T4 and T8 (and the symmetry related ones), which form two caps at both ends of the stacks (see Figure 6). The glycosidic angles of these residues are also anti, with an average value of −135° and −145° for T4 and T8, respectively. The two thymines located on the same side of the dimer are not base-paired, but they interact each other via a hydrophobic contact between their methyl groups (see Figure 6). Several NOEs between the methyl groups of the thymines with the exchangeable protons of the neighboring base-pair provide strong evidence for this orientation (see Figure 3).

All deoxyribose rings are in the S-domain, adopting in most cases a C2′-endo conformation. However, the deoxyriboses of nucleotides in position 3 and 7 adopt an O4′-endo conformation, with pseudorotation phase angles of around 90°. This conformation is confirmed by the larger $J_{3'4'}$ observed in these residues. Backbone angles are provided in the Supporting Information (Table S2). Most of the well-defined angles (order parameter

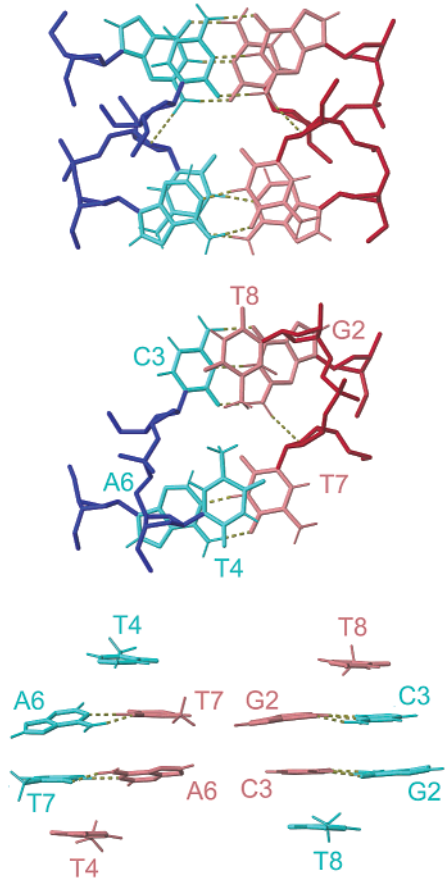

Figure 6. Top: Detail of the stacking interaction between the two G:C:A:T tetrads. Middle: Hydrophobic contacts between the two thymines located at the same side of the stacks. Bottom: Detail of the AT (left) and GC stacks (right).

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

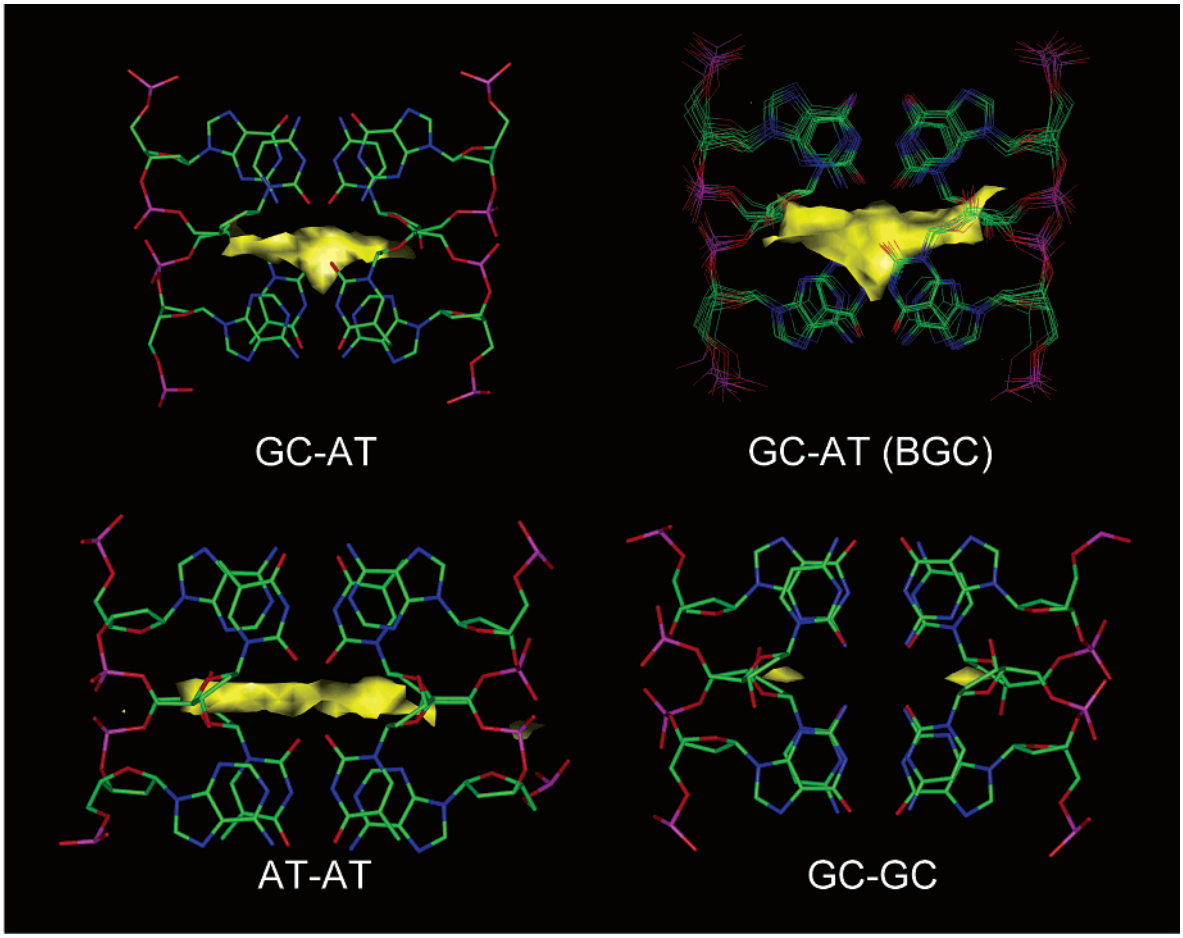

Figure 7. CMIP contour for the interaction of Na^+ with the dimeric structures of d⟨pCGCTCATT⟩, d⟨pCATTCATT⟩, and d⟨pTGCTCGCT⟩. In the case of d⟨pCGCTCATT⟩, the BGC is also reported. In all cases displayed contour correspond to −10 kcal/mol.

larger than 0.9) are within the usual values found in right-handed double stranded DNA. Exceptions are the ζ angles in residues T4 and T8, and the ϵ angles of C3 and T7, which are in a *gauche*-conformation. Other torsion angles with unusual mean values are not well-defined in the ensemble of structures (Table S2).

An interesting feature of the three-dimensional structure of d⟨pCGCTCATT⟩ is the central cavity located in the middle of the molecule. This cavity is very narrow and resembles other oligonucleotidic cavities binding cations such as those found in G-DNA.[46] Despite its small size, it has an excellent ability to interact with Na^+ as noted in the −10 kcal/mol contour plot shown in Figure 7. The optimum position for the Na^+ is the center of the cavity. Interaction values around −36 kcal/mol are reached, indicating that there is a strong electrostatic interaction between the tetraplex and the cation. When flexibility is taken into account using the Boltzman-averaged grids, there is only a slight increase in the size of the cavity, and a very modest change in the best interaction energy (from −36 to −39 kcal/mol). This demonstrates that the cavity shape is quite rigid, and well defined by the experimental restraints.

Na^+-cMIP contours can be compared with those exhibited by similar tetrads where the presence/absence of a cation has been confirmed by direct experimental evidence. For this purpose we repeat single grid calculations for the dimeric structures of d⟨pCATTCATT⟩ and d⟨pTGCTCGCT⟩, stabilized by A:T:A:T and G:C:G:C tetrads, respectively. Clearly, the Na^+−cMIP contour for G:C:A:T is very similar to that of the A:T:A:T tetrad (see Figure 7), where the presence of Na^+ has been experimentally proved,[10] and quite different than that of the G:C:G:C tetrad. Comparison of the best Na^+−oligonucleotide interaction energies confirms the similar affinity of G:C:A:T and A:T:A:T tetrads for Na^+. The interaction energies are −36 and −38 kcal/mol, respectively. In both cases, the best position for Na^+ is at the center of the cavity. When the same optimization is performed for the G:C:G:C tetrad, the best interaction energy is only −23 kcal/mol and the optimal position of the Na^+ is shifted to the sides of the molecule (see Figure 7). Interestingly, in the crystallographic structure of d(GCATGCT),[8] which was crystallized in the presence of Mg^{2+}, all the cations were found outside of the molecule close to the phosphate backbone. In summary, cMIP calculations strongly support that in solution the center of the G:C:A:T is occupied by a Na^+, which interacts with the nucleobases and helps to stabilize the system.

Discussion

The most prominent feature of the dimeric structure of d⟨pCGCTCATT⟩ is the novel G:C:A:T tetrad, resulting from the interaction of a G−C and an A−T base-pair through their minor groove sides. This is not the first example of quadruplex structures stabilized by minor groove aligned tetrads. Minor

(46) Phillips, K.; Dauter, Z.; Murchie, A. I.; Lilley, D. M.; Luisi, B. *J. Mol. Biol.* **1997**, *273*, 171−182.

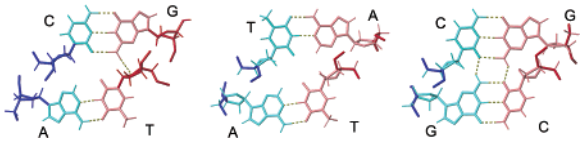

Figure 8. Detail of the three minor groove tetrads observed in the dimeric solution structures of d⟨pCGCTCATT⟩ (left), d⟨pCATTCATT⟩ (middle), and d⟨pTGCTCGCT⟩ (right).

groove G:C:G:C tetrads were first observed in the crystallographic structure of the linear heptamer d(GCATGCT).[8] A similar structure was observed in solution for the cyclic octamer d⟨pTGCTCGCT⟩.[15] Minor groove A:T:A:T tetrads have been found in the crystal[10] and solution[15] structure of d⟨pCATTCATT⟩. These tetrads are shown in Figure 8 together with the novel G:C:A:T tetrads found in the structure of d⟨pCGCTCATT⟩. Although in all cases the two Watson−Crick base-pairs interact through the minor groove, their alignment geometries are slightly different. In the G:C:G:C tetrad, the two G−C pairs align directly opposite each other, whereas in A:T:A:T and in G:C:A:T tetrads the two base pairs are shifted about 1 Å along the axis defined by the Watson−Crick base-pairs. Also, the displacement between the two base-pairs along the perpendicular axis is slightly larger in A:T:A:T and G:C:A:T than in G:C:G:C. These small changes in the geometry have important implications in the interaction that stabilizes each tetrad. In G:C:G:C, the tetrad is stabilized by two additional hydrogen bonds between one of the guanine amino protons and the O2 of the cytosine. In A:T:A:T, the tetrad is stabilized by coordinating a sodium cation between the O2 atoms of the four thymines.[10] Presumably this is also the case in the G:C:A:T tetrad. Although we do not have a direct experimental evidence of cation position, cMIP calculations indicate that the ability of the dimeric structures of d⟨pCGCTCATT⟩ and d⟨pCATTCATT⟩ to interact with a Na^+ cations is very similar. The small cavity in the center of the dimer allows for a sodium ion to fit between the O2 atoms of thymines and cytosines, contributing to stabilize the structure. In the case of G:C:A:T, the hydrogen bond between the guanine amino and the O4′ of the deoxyribose of the adjacent thymine may also play a role in the stability of the tetrad.

Experimental ΔG values have been estimated for the dimeric structures of d⟨pTGCTCGCT⟩, d⟨pCATTCATT⟩,[15] and d⟨pCGCTCATT⟩. These values are −32, −13 and −9 kJ/mol, respectively. Because the three structures are very similar, most probably their relative stability derives from different interactions within each tetrad. G:C:G:C, with direct hydrogen bonds between the base pairs, is more stable than A:T:A:T and G:C:A:T, which are stabilized by coordinating a Na^+ ion. CMIP calculations also show that the interaction energy between the Na^+ and the tetrad is higher in A:T:A:T than in C:G:A:T. This is in agreement with the different stabilities observed experimentally.

It is worth mentioning the striking similarities in the overall three-dimensional structures of the four molecules where minor groove tetrads have been found so far (see Figure 9). In all cases, the structure consist of two stacks of two base-pairs. As a consequence of the lack of planarity in the tetrad, the two stacks have a mutual inclination of around 30°−40°. This inclination makes it impossible for these stacks to extend more than two contiguous tetrads. A second important similarity worth noting is that the capping residues at the end of the stacks, which are thymines in most of the cases, but also adenines in one of the

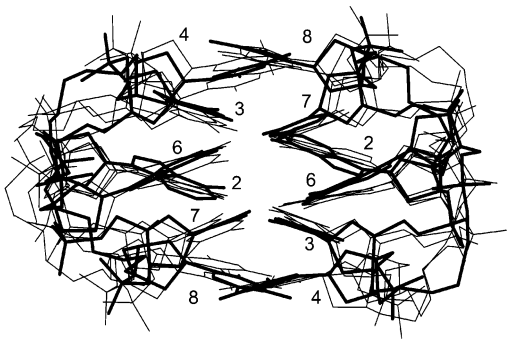

Figure 9. Superposition of the average solution structures of d⟨pTGCTCGCT⟩, d⟨pCATTCATT⟩, and d⟨pCGCTCATT⟩. Bases at positions 1 and 5 (and the symmetry related ones) are not shown, since they are mainly disordered in the three cases.

ends of d(GCATGCT),[8,20] adopt almost identical conformations in all the cases. This indicates that the exact identity of the capping residues is not determinant for the formation of the motif, although it may affect the stability of the structure. Another common feature of all these structures is the great proximity of the backbones of the two strands in the same subunit. This gives rise to a large number of contacts between deoxyribose atoms and very short distances between phosphates.

In all cases where minor groove alignment of Watson−Crick base-pairs occurs, the resulting tetrads have a very similar geometry. In contrast, major groove alignment appears to place less restrictions on the structure of the tetrad, judging from the examples of quadruplex structures resolved to date. As mentioned before, major groove G:C:G:C tetrads have been observed in the solution structures of the fragile X syndrome d(C−G−G)n triplet repeat,[9] and in the G−G−G−C repeats of the adeno-associated viral DNA.[11] In the latter case, two conformations of the mixed tetrad are observed, depending on the cation present in the sample.[19] In the sodium form the two G−C pairs align directly opposite to each other, whereas in the presence of potassium the alignment of the two based-pairs is shifted, with the two major groove edge of the guanines coordinating the K^+ cation. These two alignments have been also observed in major groove A:T:A:T tetrads. The slipped alignment was found in the dimeric solution structure of the octamer d(GACGAGGT),[16] and, more recently, the direct A:T:A:T alignment has been observed in the crystallographic structure of an oligonucleotide with four consecutive human telomeric repeats.[21]

It must be emphasized that major groove tetrads have only been observed in quadruplexes containing other pure guanine tetrads. The same caveat applies to other unusual tetrads found in some guanine quadruplexes.[12−14] In contrast, minor groove alignment of Watson−Crick base-pairs in quadruplexes have always been found in structures where no other kind of tetrad is present. It may be that the major groove alignment is preferred when Watson−Crick base-pairs associate within the scaffold of a guanine quadruplex. On the other hand, minor groove tetrads appear to induce a distinct DNA folding motif, that may not be compatible with pure guanine tetrads. The term "bi-loop" has been suggested for this motif.[10] More examples of quadruplex structures containing mixed tetrads are necessary to verify this hypothesis.

The particular features observed in the dimeric structure of d⟨pCGCTCATT⟩, and the other quadruplexes forming analogous

structures, may be relevant to some biological processes. For example, the close proximity between phosphate groups suggest that similar structures could be formed during the tight packing of DNA in virus. Also, the association of DNA duplexes through minor groove contacts has been proposed as an initial step in the recombination process prior to strand exchange.[47] A number of extra-helical interactions involving minor groove association between DNA duplexes have been reported in crystallographic structures.[48,49] For example, in the structure of a DNA−RNA chimeric duplex, a G:C:G:C tetrad has been observed to form between two base-pairs in symmetry related duplexes.[50] Although quite distorted, this tetrad is formed by two minor groove aligned G−C Watson−Crick base-pairs. All these findings and those presented here suggest that formation of minor groove tetrads may be a mechanism of general recognition between DNA helices. Although the mutual inclination between the two stacks makes it impossible to extend the tetrads more than two consecutive base pairs, selective recognition may be achieved by the multiple occurrence of this motif in discontinuous segments of DNA.

Interestingly, there are some similarities between the G:C:A:T tetrad reported here and the A-minor interaction observed in RNA structures.[51,52] The A-minor motif seems to be a general mode of recognition between RNA helices, that has been found in many large RNA structures with tertiary contacts. In this motif, an extra-helical adenine base interacts with the minor groove of a Watson−Crick G−C base pair. The adenine base approaches the G−C base pair through its minor groove and can adopt different conformations, some of which resembles the G:C:A:T tetrad described here. However, the interactions in the two systems are different since in RNA the adenine always forms direct hydrogen bonds, either with the G−C pair or with the ribose 2′-OH groups.

Quadruplex formation in DNA sequences containing triplet repeats has been proposed to lead to errors in replication and cause the expansion of the triplet sequence.[53,54] However, this hypothesis has been criticize since some repeats cannot form classical guanine quadruplexes. Our observation that stable four-stranded structures can form without guanine tetrads suggest that these structures may be relevant for the mechanism of expansion of triplet repeats. The occurrence of such structures within the cell may disturb the DNA metabolism and be responsible of the genetic instabilities associated with DNA repeats, and the subsequent disease.

Conclusions

The dimeric structure of d⟨pCGCTCATT⟩ is a new case of self-recognition between DNA molecules. This recognition is mediated by intermolecular Watson−Crick base-pairs, and results in the formation of two minor groove aligned G:C:A:T tetrads. This is the first time that a G:C:A:T tetrad has been found. The observation of this novel tetrad increases considerably the number of sequences that can adopt a quadruplex structure. Although stabilized by an entirely novel tetrad, the quadruplex structure of d⟨pCGCTCATT⟩ is very similar to other dimeric structures stabilized by minor groove tetrads. This indicates that minor grove association of Watson−Crick base-pairs is a robust motif for DNA−DNA recognition, and strongly suggests that we may be facing a motif that is common in nature.

Coordinates. Atomic coordinates have been deposited in the Protein Data Bank (accession number PDB: 1N96). The complete assignment list has been deposited at the BMRB (accession number 5681).

Acknowledgment. This work was supported by the DGICYT grants BQU2001-3693-C02-01/02 and Generalitat de Catalunya (2001SGR49 and Centre de Referència en Biotecnologia). We gratefully acknowledge Dr. Doug Laurents for careful reading of the manuscripts and his useful comments.

Supporting Information Available: Two tables with the assignment list and average values and order parameters of the dihedral torsion angles. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0344157

(47) Wilson, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 3641−3645.
(48) Ramakrishnan, B.; Sundaralingam, M. *Biochemistry* **1993**, *32*, 11 458−11 468.
(49) Subirana, J. A.; Abrescia, N. G. *Biophys. Chem.* **2000**, *86*, 179−189.
(50) Ban, C.; Ramakrishnan, B.; Sundaralingam, M. *J. Mol. Biol.* **1994**, *236*, 275−285.
(51) Doherty, E. A.; Batey, R. T.; Masquida, B.; Doudna, J. A. *Nat. Struct. Biol.* **2001**, *8*, 339−343.
(52) Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, P. B.; Steitz, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4899−4903.
(53) Fry, M.; Loeb, L. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 4950−4954.
(54) Usdin, K. *Nucleic Acids Res.* **1998**, *26*, 4078−4085.