

The NMR structure of [Xd(C2)]₄ investigated by molecular dynamics simulations

Thérèse E. Malliavin,^{1*} Karim Snoussi² and Jean-Louis Leroy³

¹ Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, 13 rue P. et M. Curie, F-75005 Paris, France

² Department of Biophysical Chemistry, Lund University, Box 124, SE-22100 Lund, Sweden

³ Laboratoire de RMN à Haut Champ ICSN–CNRS, 91190 Gif sur Yvette, France

Received 18 April 2002; Revised 30 June 2002; Accepted 5 August 2002

The i-motif tetrameric structure is built up from two parallel duplexes intercalated in a head-to-tail orientation, and held together by hemiprotonated cytosine pairs. Two topologies exist for the i-motif structure, one with outermost 3' extremities and the other with outermost 5' extremities, called the 3'E and 5'E topology, respectively. Since the comparison of sugar and phosphate group interactions between the two topologies is independent of the length of the intercalation motif, the relative stability of the 3'E and 5'E topologies therefore should not depend on this length. Nevertheless, it has been shown that the 3'E topology of the [d(C2)]₄ is much more stable than the 5'E topology, and that the former is the only species observed in solution. In order to understand the reason for this atypical behavior, the NMR structure of the [Xd(C2)]₄ was determined and analyzed by molecular dynamics simulations. In the NMR structure, the width of the narrow groove is slightly smaller than in previously determined i-motif structures, which supports the importance of phosphodiester backbone interactions in the structure stability. The simulations show that the stacking of cytosines, essential for the i-motif stability, is produced by a similar and non-negative twisting of the phosphodiester backbones. The twisting is induced by an interaction between the backbones; the [Xd(C2)]₄ in 5'E topology, exhibiting very limited interaction between the phosphodiester backbones, is thus unstable. Copyright © 2002 John Wiley & Sons, Ltd.

KEYWORDS: NMR; DNA structure; [Xd(C2)]₄; i-motif; proton exchange; protonated cytidine; molecular dynamics

INTRODUCTION

Tetrameric DNA structures^{1,2} are thought to play an important role in the chromatin organization in the cell nucleus³. Two main types of tetrameric structures are known: the G-quartet,¹ observed in G-rich sequences, and the i-motif,⁴ formed by the anti-parallel intercalation of two parallel duplexes of cytosines. Centromeric⁵ and telomeric^{6,7} DNA sequences were shown *in vitro* to be folded into an i-motif structure.

Two intercalation topologies exist for the i-motif structure (Fig. 1), one with outermost 3' extremities and the other with outermost 5' extremities, called 3'E and 5'E topology, respectively^{8,9}. Both topologies are observed in the crystal¹⁰ and solution⁹ structures. The major difference between them (Fig. 1) lies in the fact that the sugar and phosphate groups belonging to neighboring phosphodiester backbones have different relative positions. Indeed, because of the asymmetric position of the phosphate group between two bases, the phosphate pairs (and consequently the sugar pairs)

can be grouped as close and remote pairs. Figure 1 shows that the 3'E topology exhibits an additional close phosphate pair and an additional close sugar pair, compared with the 5'E topology. Close phosphate pairs display larger electrostatic repulsion, whereas attractive interactions between sugars are observed only for close sugar pairs. As the comparison of the sugar and phosphate group interactions between the two topologies is independent of the length of the intercalation motif, the relative stability of the 3'E and 5'E topologies of [d(Cn)]₄ should therefore not depend on *n*. A systematic study of the end effects in the [d(Cn)]₄ tetramers¹¹ (*n* = 2–6) showed that both topologies exhibit similar stability for 3 ≤ *n* < 6. In the case of [d(C2)]₄, the 3'E topology is the only species observed in solution.

In the pH range required for the association of four d(C2) strands into an i-motif, i.e. around pH 4.3, the solubility of d(C2), c2 mM, is much lower than that of longer deoxycytidine oligomers¹¹. The charge of d(Cn) sequences is (1 – *n*/2), in the cases both of a hemiprotonated structure and of a single strand when the pH is equal to the cytidine p*K*, i.e. 4.3. Hence d(C2) is neutral under such conditions, which probably accounts for its poor solubility.

The i-motif concentration of oligoC-rich polynucleotide solutions increases as the fourth power of the monomer concentration⁹. The low solubility of d(C2) (about 2 mM at 0 °C, pH 4.3) therefore limits the formation of the i-motif. At

*Correspondence to: Thérèse E. Malliavin, Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, 13 rue P. et M. Curie, F-75005 Paris, France.
E-mail: therese.malliavin@ibpc.fr
Contract/grant sponsor: CNRS.
Contract/grant sponsor: IDRIS supercomputing Center;
Contract/grant number: 11410.

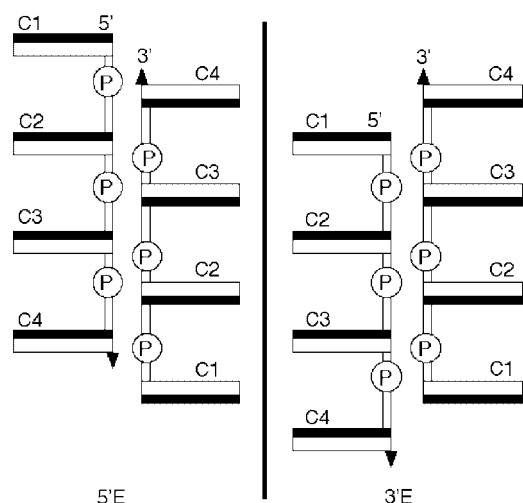


Figure 1. Schematic representation of the two maximally intercalated topologies 3'E and 5'E of [d(C4)]₄. The scheme shows two strands on either side of a narrow groove. The bases are represented by rectangles with their black face oriented in the 5' direction and their white face in the 3' direction. The crystal structure of [d(C4)]₄ adopts the 5'E topology. In solution, the two topologies co-exist in comparable proportions.

0 °C, the NMR spectrum of a saturated d(C2) solution shows that only 20% of the strands are involved in i-motif tetramers. In order to improve the oligonucleotide solubility and therefore to shift the [‡]d(C2)–[d(C2)]₄ equilibrium towards the tetrameric species, we synthesized the oligonucleotide Xd(C2), where X stands for the HO(CH₂)₂O(CH₂)₂O(CH₂)₂O group linked by a phosphodiester bond to the O5' oxygen of C1. At pH 4.3, Xd(C2) is soluble up to 15 mM and forms a fairly stable i-motif. In this work, we solved by NMR the i-motif structure of [Xd(C2)]₄ and, in order to understand what is unusual about its atypical behavior, we used this structure as a model in a molecular dynamics (MD) study, together with that of the crystal structure of [d(C4)]₄.¹⁰

EXPERIMENTAL

Synthesis of Xd(C2) and structure determination

The Xd(C2) and d(C2) oligomers were synthesized, as reported previously,⁸ on a 15 μM scale using dC- and 9-O-dimethoxytrityltriethylene glycol 1-[(2-cyanoethyl)-(N,N-diisopropyl)]phosphoramidites. They were purified by chromatography on a DEAE Hiload Q-Sepharose column by elution with an ammonium carbonate gradient. The pool collected after purification was lyophilized. Most of the ammonium carbonate was removed during the lyophilization. Nevertheless, the sample was dissolved in 0.1 M NaCl, pH 4.3, and dialyzed against water for 15 min, a time short enough to avoid an important loss of the oligomer but allowing the exchange of a substantial amount of the NH₄⁺ counter-ion. The NMR experiments were performed using a 360 MHz laboratory-built spectrometer at 0 and –3 °C. The d(C2) strand concentration was 12 mM. Experiments in H₂O solution used the 'jump and return' method for solvent suppression.¹² The methods used for the proton exchange

study and for data acquisition and processing have been described.⁹

During the structure computation, the intra-residue distances were restrained to the measured value with upper and lower bounds 10 to 20% above and below the value, depending on the spectral resolution. The inter-residue distance restraints were sorted into three categories with upper and lower limits of 1.8–2.9, 1.8–3.7 and 2.9–4.7 Å. Spin diffusion may be fast between geminal protons (i.e. H2'–H2'' and internal–external amino proton pairs), even at short mixing times. For this reason, we used only the strongest cross peak as a restraint for the nuclear overhauses effects (NOEs) detected between H2'–H2'' or internal–external amino proton and a third partner. The distance derived from the weakest cross peak was considered as a lower bound. H-bond restraints were applied between the N3 nitrogens and the O2 and the H-bonded amino protons of pairs C1.C1⁺. The structure was computed as reported previously,⁷ using the simulated annealing protocol¹³ implemented in the X-PLOR program.¹⁴

Molecular dynamics simulations

The starting conformations for the simulations were generated in the following way. Hydrogens were added to the 5'E topology of the [d(C4)]₄ crystallographic structure,¹⁰ which was the initial conformer for the MD simulations of the 5'E topology of [d(C4)]₄. A model for the [d(C2)]₄ complex in 3'E topology was one conformer of the NMR structure of the [Xd(C2)]₄ complex. The 3'E topology of [d(C4)]₄ and the 5'E topology of [d(C2)]₄ were built from the other topology by translating one duplex with respect to the other along the main molecular axis. The cytosine protonation was modeled by adding an H3 hydrogen to the N3 atom of one cytosine in each base pair. Two strands separated by a wide groove were protonated on the atom N3.

The molecular dynamics and molecular mechanics calculation was performed using the program AMBER 6.0,¹⁵ with the parm98 parameter set¹⁶ and the TIP3P model for water.¹⁷ The additional H3 hydrogen in protonated cytosine was of AMBER type H (i.e. H bonded to the nitrogen atoms).¹⁸ The partial charges of the protonated cytosines were the same as in Ref. 19.

The tetrameric structures were hydrated by a box of water molecules in a pseudo-equilibrated configuration. A cutoff was used along the three axes to discard any water molecule further than 11 Å from any solute molecule. The box dimensions are given in Table 1, for each simulation run. Since the [d(C2)]₄ complex is neutral, no counter-ion was added in the simulations performed on [d(C2)]₄. For [d(C4)]₄, four sodium ions were used to neutralize the charge of the tetramer, corresponding to an ionic concentration of 86 mM. Long-range electrostatic interactions were calculated using the Particle Mesh Ewald protocol,²⁰ with a 9 Å cutoff for direct space non-bonded calculations and with a 0.00001 Ewald convergence tolerance for the long-range electrostatic interactions.

All the MD simulations were initiated with several rounds of semi-restrained and eventually unrestrained minimizations of the entire system followed by heating,

Table 1. The parameters of the different MD simulations^a

| Simulation ID | Solute | Topology | Length (ns) | Water box size (Å) | Pucker restraint | Phosphate restraint (Å) |
|---------------|----------|----------|-------------|--------------------|---------------------|-------------------------|
| c23-free | [d(C2)]4 | 3'E | 4 | 40, 31, 28 | No | No |
| c25-free | [d(C2)]4 | 5'E | 4 | 40, 33, 28 | No | No |
| c23-3endo | [d(C2)]4 | 3'E | 2 | 40, 31, 28 | N type | 3, 4, 7, 8 |
| c25-3endo | [d(C2)]4 | 5'E | 2 | 40, 33, 28 | N type | 3, 4, 8, 9 |
| c23-delta | [d(C2)]4 | 3'E | 2 | 40, 31, 28 | N type (δ) | 3, 4, 7, 8 |
| c23-pp | [d(C2)]4 | 3'E | 2 | 40, 31, 28 | No | 3, 4, 7, 8 |
| c23-nopp | [d(C2)]4 | 3'E | 0.2 | 40, 31, 28 | N type | No |
| c23-parm99 | [d(C2)]4 | 3'E | 1 | 40, 31, 28 | No | No |
| c43-free | [d(C4)]4 | 3'E | 2 | 48, 46, 35 | No | No |
| c45-free | [d(C4)]4 | 5'E | 2 | 48, 46, 35 | No | No |
| c43-3endo | [d(C4)]4 | 3'E | 2 | 48, 46, 35 | N type | 3, 4, 7, 8 |
| c45-3endo | [d(C4)]4 | 5'E | 2 | 48, 46, 35 | N type | 3, 4, 8, 9 |

^a The N type restraint is a series of harmonic restraints applied on the ribose dihedral angles ν_i in order to restrain the pucker phase in the 0–40° range. The N type (δ) restraint is a harmonic restraint applied to keep the ribose dihedral angle δ in the 70–100° range. The shape of the restraint potential is a well with a square bottom, parabolic sides out to a defined distance or angle value, and then linear sides beyond that. The parameters defining the inter-phosphate restraints are given in the following order: the upper and lower bounds of the square bottom interval are in the middle, and the interval limits for the parabolic sides are on each side.

equilibration and unrestrained MD in the (T,P,N) ensemble.²¹ Harmonic restraints of 100 kcal mol⁻¹ Å⁻² (1 kcal = 4.184 kJ) were placed on the positions of the DNA atoms and of the sodium counter-ions during the first round of 500 steps of minimization. They were relaxed on the ions more quickly than on the DNA atoms over the course of a subsequent 500-step minimization, followed by three subsequent 250-step minimizations. The final round of 500 steps of minimization involved all atoms of the system.

The heating and initial system equilibration was performed as follows: 10 ps of heating up to 300 K was applied at constant volume while restraining the DNA tetramer and the ion atom locations with a restraint of 25 kcal mol⁻¹ Å⁻²; 5 ps of MD at constant volume and four MD rounds of 2.5 ps at constant pressure, during which the restraints on the counter-ions were reduced more quickly than on the DNA atoms; and a last MD round of 10 ps with a restraint of 1.5 kcal mol⁻¹ Å⁻² on the DNA atoms and no restraints on the ions. The isothermal–isobaric unrestrained simulation then started using the Berendsen algorithm²² for temperature bath coupling and a 2 fs time step. A SHAKE tolerance of 0.00001 Å was used on all the covalent bonds involving hydrogen atoms. All the MD simulations were run at a temperature of 273 K and a pressure of 1 atm. The simulation lengths are given in Table 1. The atom coordinates were saved every 1 ps.

For each tetramer, the MD simulations were performed (Table 1) in the 3'E and 5'E intercalation topologies. The names given to each MD simulation for the sake of reference are made explicit in Table 1. Two sets of simulations were run, either unrestrained or with restraints on the sugar pucker phases and/or on the inter-phosphate distances. The restraints were harmonic with force constants equal to 30 kcal mol⁻¹ Å⁻², similar to those used in MD simulations of

A-DNA;²³ the restraint on sugar pucker phase was applied through all the ν_i ring dihedral angles. The pucker phase was restrained in the 0–40° range. The shape of the restraint potential was a well with a square bottom, extended by parabolic sides up to a defined distance or angle value, and then linear sides beyond that. The interval values for the inter-phosphate restraints are given in Table 1; for ν_i dihedral angles, the square bottom has a width of 5°, and parabolic sides were used for two subsequent intervals of 5°.

The MD simulation analysis was mainly performed with the program PTRAJ.¹⁵ For the visualization of the trajectories, the program VMD²⁴ was used. The atomic fluctuations and the root-mean-square deviations (r.m.s.d.s) with respect to the starting conformer were calculated after superimposing each MD snapshot with the starting conformation. If not stated otherwise, the DNA helicoidal parameters were calculated from MD snapshots, NMR and x-ray structures, using the program CURVES 6.0²⁵ on each of the two duplexes forming the i-motif structure.

RESULTS

Proton identification and determination of the intercalation topology of [Xd(C2)]4

The 1D spectrum of Xd(C2) displayed in Fig. 2 exhibits a single imino proton (15.8 ppm) and two amino proton pairs in the frequency ranges characteristic of the *cis*- and *trans*-amino protons of hemiprotonated pairs. The observation of a single set of proton peaks for each residue indicates a fully symmetrical structure on the NMR time-scale.

The cytidine at the 3' end, C2, was assigned by the high-field shift of its H3' proton (cross peak N, Plate 1). The imino proton of the outer C2. C2⁺ pair is broadened out owing to fast exchange with water. The imino proton at 15.8 ppm was assigned to C1(H3) by its NOE with C1(NH2). As expected

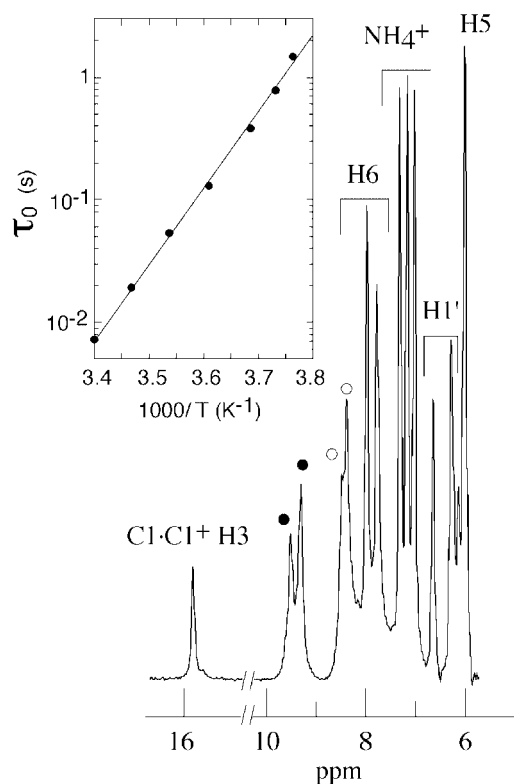


Figure 2. The proton spectrum of $[\text{Xd}(\text{C}2)]_4$ in H_2O solution indicates a fully symmetrical i-motif structure. The imino proton of $\text{C}2.\text{C}2^+$ is exchange broadened beyond detection. The *cis*- and *trans*-amino protons peaks are labeled by filled and open circles, respectively. Experimental conditions: strand concentration 12 mM, $T = -3^\circ\text{C}$, $\text{pH} = 4.3$. Inset: the base-pair lifetime of pair $\text{C}1.\text{C}1^+$ vs temperature as determined from the imino proton exchange measured by magnetization transfer from water.

for the imino proton of base-paired protonated cytidine,²⁶ the exchange rate of $\text{C}1(\text{H}3)$ is controlled by the opening of the pair. The lifetime of a pair $\text{C}1.\text{C}1^+$ determined from imino proton exchange experiments, 0.3 s at 0°C , varies with respect to temperature according to an activation energy of 116 kJ mol^{-1} (inset in Fig. 2). The i-motif structure of $\text{Xd}(\text{C}2)$ is fairly stable: at a strand concentration of 12 mM, its melting temperature is 12°C at $\text{pH} 4.3$ and the mid-titration pH at 0°C is 5.3. The stability of the tetrameric i-motif is nevertheless much smaller than that of the intramolecular i-motif structure of $\text{d}(5\text{mCCTTTCCTTTCCTTTC})$, which has a concentration-independent melting temperature of about 45°C and a mid-titration pH at 0°C equal to 7.4.²⁷

The intensities of the $\text{C}1(\text{NH}2) - \text{C}1(\text{H}2'/\text{H}2'')$ cross peaks in the 250 ms NOESY spectrum (cross peaks α , β and γ , inset in Plate 1) correspond to inter-proton distances in the range 3.2–4.4 Å, which are shorter than the minimum intra-residue distance (5 Å). In the i-motif structure, the intercalation in a head-to-tail orientation of two hemiprotonated duplexes generates black–black and white–white contacts in alternation (Fig. 1) between the bases, where the 'black' ('white') face of a nucleoside was the face oriented towards the 5' (3') direction. The amino- $\text{H}2'/\text{H}2''$ NOEs are thus characteristic of the i-motif structure.

They establish the connectivities of two $\text{C}1.\text{C}1^+$ pairs stacked by their white faces and thus reveal the $\text{C}2 \text{ C}1 \text{ C}1 \text{ C}2$ intercalation topology of the tetramer.

Other characteristic inter-residue cross peaks expected for this intercalation topology are the rectangular patterns of cross peaks formed across one diagonal by the intra-residue cross peaks [e.g. $\text{C}1(\text{H}6) - \text{C}1(\text{H}1')$ and $\text{C}2(\text{H}6) - \text{C}2(\text{H}1')$], and across the other diagonal by the corresponding inter-residue cross-peaks [e.g. $\text{C}1(\text{H}6) - \text{C}2(\text{H}1')$ and $\text{C}2(\text{H}6) - \text{C}1(\text{H}1')$], the cross peaks a and b in Plate 1].

The strong cross peak between the two $\text{H}1'$ protons (cross-peak L, Plate 1) is also an i-motif signature. The $\text{H}4' - \text{H}4'$ cross peak (cross peak M) reflects the very short $\text{C}1(\text{H}4') - \text{C}2(\text{H}4')$ distance across the narrow grooves between the residues stacked by their black faces (right side of Plate 2). The intensities of the $\text{H}3' - \text{H}6$ cross peaks (N and O, Plate 1), corresponding to distances close to the shortest possible, indicate N sugar puckers.

The $\text{CH}2$ protons of the residue linked at the 5' end of $\text{Xd}(\text{C}2)$ were identified on the TOCSY spectrum. The absence of NOE connectivities between the 'X' and the cytidine protons and the identity of the proton chemical shifts of $[\text{d}(\text{C}2)]_4$ and $[\text{Xd}(\text{C}2)]_4$ argue in favor of similar structures with and without X. Moreover, the observation that the $[\text{Xd}(\text{C}2)]_4$ and $[\text{d}(\text{C}2)]_4$ concentrations are the same in a solution containing a total strand concentration of 2 mM shows that the 'X' extension does not stabilize the tetramer.

$[\text{Xd}(\text{C}2)]_4$ structure analysis

The structure was determined on the basis of the inter-proton distances obtained from the build-up of NOE cross-peaks collected with mixing times of 50, 70, 90 and 250 ms in D_2O at 0°C and of 50, 120 and 200 ms in H_2O at -3°C . The statistics on the distance restraints are given in Table 2 together with the NOE violations, the deviations from ideal geometry along with the pair-wise r.m.s.d.s determined from 10 conformers selected for their low NOE-related energy, among 45 computed structures.

The small standard deviation values observed for the helicoidal parameters and phosphodiester backbone angles in the $[\text{Xd}(\text{C}2)]_4$ NMR structure (Table 3) are not surprising, according to the large number of restraints (124) relative to the structure size. The structure is formed of four stacked $\text{C}.\text{C}^+$ pairs intercalated in the 3'E topology (Plate 2). The pairs are relatively flat, with deviations from the planarity smaller than 11° (Table 3). Almost all the structural parameters (Table 3) are close to the values reported for the i-motif core of other tetrameric,⁹ dimeric²⁸ or monomeric²⁷ structures containing four stacked $\text{C}.\text{C}^+$ pairs. In particular, the dihedral angles do not depart from values typical for the i-motif.²⁹ It is noteworthy that the stacking interval between the $\text{C}1.\text{C}1^+$ pairs may be overestimated because of the lack of restraints, due to symmetry, at the step between the two $\text{C}1.\text{C}1^+$ pairs.

The widths of the narrow groove at the black step between $\text{C}1$ and $\text{C}2$ (Table 3) display a slightly smaller value (4.2 Å) than those observed for $[\text{d}(\text{TC}2)]_4$ (4.8 Å) and for the monomeric i-motif of the human telomeric sequence (5.4 Å).⁷ The small width of the $[\text{Xd}(\text{C}2)]_4$ narrow grooves results from $\text{CH}1' - \text{O}4'$ distances shorter than in

Table 2. Distance restraints, violations, deviations from ideal geometry and root mean square deviations (r.m.s.d.s) of the computed structures

| Distance restraints ^a | | | |
|--|------|----------|-----------|
| Number of inter-residue restraints | 13 | | |
| Number of intra-residue restraints | 10 | | |
| Number of base-pairing restraints ^b | 6 | | |
| Number of repulsive restraints ^c | 2 | | |
| Violations | | r.m.s.d. | |
| Largest NOE violation (Å) | 0.15 | 0.015 | |
| Largest angle violation (°) | 4 | 1.1 | |
| Largest improper violation (°) | 2 | 0.4 | |
| Number of bond violations largest than 0.5 Å | 0 | 0.001 | |
| Pair-wise r.m.s.d. (Å) | Base | Sugar | Phosphate |
| | 0.25 | 0.5 | 1.1 |

^a The NMR spectrum implies an identical environment for the four Xd(C2) strands. Hence, the intra-residue restraints were imposed to each strand and the inter-residue restraints were imposed between the four couples of strands.

^b Three base-pairing restraints for each C1.C1⁺ pair.

^c The repulsive restraints were applied with lower bounds of 4.2 and upper bounds of 30 Å.

other *i*-motif structures: 2 ± 0.15 Å, as compared with 2.3 (2.6) Å in [d(TC2)]₄ ([d(C)4]₄). It has been proposed that the CH1'...O4' hydrogen bonds can span the *i*-motif narrow groove at the black steps;³⁰ this may favor the 3'E topology, which has two more black steps than the 5'E topology, and therefore displays two more possible CH1'...O4' hydrogen bonds. The particularly short CH1'...O4' distances observed in [Xd(C2)]₄ may indicate a stronger hydrogen bonding favorable to the 3'E topology of [Xd(C2)]₄.

Fit of the *i*-motif simulated conformation to the NMR structures

During all the MD simulations, the temperature and the system total energy were stable, and the conformers exhibited r.m.s.d. values from the starting conformer smaller than 2 Å. After a few hundred picoseconds, the r.m.s.d. values reach a plateau, which means that the simulation stabilizes after the first nanosecond.

In the c23-free, c25-free, c43-free and c45-free simulations, the phases of sugar pucker were shifted out of the range of values derived from the NMR and x-ray structures to the 150–200° range.⁴ The tendency to push the sugar pucker values towards the type S values is a well-known feature³¹ of the AMBER force fields. The unrestrained MD simulation c23-parm99, run with the parm99 force field,³² produces similar shifts towards the S-type pucker. In order to keep the *i*-motif conformations closer to the NMR structure, a second series of simulations (c23-3endo, c25-3endo, c43-3endo and c45-3endo) were run with restrained sugar pucker phases as described in the Experimental section. The inter-phosphate distances were also restrained during these simulations as explained below. The sugar pucker amplitudes were observed in the range 30 to 40 and –40 to –30° for unrestrained and restrained simulations, respectively; this is in agreement with observations in the literature.³³

Table 3. Geometric parameters of the [Xd(C2)]₄ NMR structure, and of the c23-3endo trajectory

| Helical twist (°) | | 15.5 ± 2 | | | | | | |
|------------------------|----------------------------------|--------------------|--------------------|------------------------------------|--------------------------------------|--------------------|--------------------|--------------------|
| Residue | Propeller twist (°) ^a | Buckle (°) | Step | Stacking interval (°) ^b | Narrow groove width (Å) ^c | CH1'–O4' (Å) | Ω (°) ^d | |
| C1 | –7 ± 4 | –6 ± 2 | C1/C2 | 3.1 ± 0.2 | 4.2 ± 0.2 | 2.0 ± 0.15 | 92.5 | |
| C2 | 4 ± 4 | 11 ± 2 | C1/C1 | 3.5 ± 0.2 | 6.2 ± 0.5 | 5.3 ± 0.5 | 77 | |
| | α (°) ^e | β (°) ^e | γ (°) ^e | δ (°) ^e | ε (°) ^e | ζ (°) ^e | ρ ^e | χ (°) ^e |
| C1 | — | — | — | 108 ± 29 | –147 ± 14 | –80 ± 45 | 23 ± 19 | –126 ± 1.5 |
| C2 | –179 ± 52 | 187 ± 10 | –151 ± 20 | 95 ± 10 | — | — | 6 ± 9 | –125 ± 1 |
| c23-3endo ^f | –166 ± 25 | 162 ± 34 | –163 ± 28 | 84 ± 7 | –158 ± 26 | –81 ± 36 | 36 ± 8 | –122 ± 17 |

^a The propeller twist and the buckle are the complements of the angle between the axis perpendicular to the average base plan and vectors C6N3 and C4N2 respectively. Positive values indicate that the vectors point in the 5' direction.

^b The stacking interval is defined by the vertical distance between the average base planes.

^c The narrow groove widths are defined as the average distance of the inter-strand distances across each groove between the C1', C2', C3', C4' and O4' atoms. The small (large) width and CH1'–O4' values correspond to close (remote) sugar pairs (Fig. 1).

^d The Ω angle specifies the relative orientation of the stacked base pairs. It is determined by the projection, in the average plane parallel to the CC' pairs, of the C6N3 vectors of two stacked base pairs on either side of the narrow grooves.

^e Dihedral angles in the NMR structure of [X(dC2)]₄ *i*-motif. The angles were computed using nine conformers selected for their small NOE violations, and are displayed when at least eight out of the 10 values fall in a sector range smaller than 60°.

^f Mean values and standard deviations of the dihedral angles and of the pucker phase in the c23-3endo trajectory.

The agreement with the [Xd(C2)]₄ NMR structure of the [d(C2)]₄ i-motif conformers obtained for simulations run in the 3'E topology (c23-free, c23-3endo, c23-pp, c23-nopp, c23-delta, c23-opened, c23-parm99) was checked by comparing the distances observed during the molecular dynamics simulation, with the 100 NOE restraints used to calculate the NMR structure. The mean value R of a given distance r along the trajectory was calculated as $1/R^6 = \langle 1/r^6 \rangle$ where $\langle \rangle$ denotes the mean value calculated along the trajectory. A distance restraint was considered as violated for an R value larger than $R_u + 1.0 \text{ \AA}$, where R_u is the upper bound of the restraint interval.

Three kinds of NMR restraints (Table 4) are violated: intra-strand NMR restraints H6–H3', inter-strand NMR restraints between hydrogen sugars (H1', H4', H2', H2'') through the narrow groove and inter-strand NMR restraints observed through the wide groove between H42 and H2' (only for c23-3endo). The simulation run only with restrained inter-phosphate distances (c23-pp) shows the largest number of violations (12), followed by the c23-free unrestrained simulation (eight). On the other hand, the simulation run with restraint on the pucker phases and no restraint on the inter-phosphate distances (c23-nopp), displays nine violated NMR restraints between the ribose hydrogens. Hence both restraints on the sugar pucker phases and on the inter-phosphate distances are essential to minimize the number of violations.

The sugar pucker phase was also restrained through the dihedral angles δ of the phosphodiester backbone (c23-delta); the trajectory displays the same number of inter-ribose violated NMR restraints as c23-free (four violations), but with a smaller number of violated intra-ribose-base restraints (one violation). The unrestrained trajectory c23-parm99 run with the parm99 force field³² displays six violations of NMR restraints (Table 4); this number is close to that observed for c23-delta.

Table 4. Number of violated NMR restraints during the simulations run on the 3'E topology of the tetramer [d(C2)]₄^a

| ID | Intra-ribose | | Inter-ribose | | Total |
|------------|--------------|--------------|--------------|--|-------|
| | -base | Inter-ribose | -base | | |
| c23-free | 4 | 4 | 0 | | 8 |
| c23-3endo | 0 | 2 | 2 | | 4 |
| c23-delta | 1 | 4 | 0 | | 5 |
| c23-parm99 | 4 | 2 | 0 | | 6 |
| c23-pp | 5 | 7 | 0 | | 12 |
| c23-nopp | 0 | 9 | 0 | | 9 |

^a An NOE was considered as violated if the mean distance R ($1/R^6 = \langle 1/r^6 \rangle$, where $\langle \rangle$ is the mean value calculated on the trajectory) is larger than $R_u + 1.0 \text{ \AA}$, where R_u is the upper bound of the restraint interval. 'Intra-ribose-base' refers to the intra-strand NMR restraints between ribose H3' and base H6 hydrogens; 'inter-ribose' refers to the inter-strand NMR restraints between ribose hydrogens H1', H4', H2', H2'' across the narrow groove; 'inter-ribose-base' refers to the inter-strand NMR restraints between ribose H2' and base H42 hydrogens across the major groove; 'ID' is the simulation ID, defined in Table 1.

The c23-3endo simulation displays the smallest number of violated NMR restraints (four violations); this number is small with respect to the total number of 100 NOEs observed experimentally. The number of violated NMR restraints and the violation amplitudes compare well with other MD trajectories performed on NMR structures.³⁴ The mean values of the restraint energy during the simulations corresponded to about 0.2% of the total energy of the solute, and may be considered to be negligible. In addition, comparison of the phosphodiester backbone dihedral angles explored during the c23-free simulation (Table 3^f) with the values measured on the NMR structure (Table 3^e) shows that the [X(dC2)]₄ conformations along the trajectory do not depart significantly from the NMR structure.

Helical parameters of the mean conformers and structures

The mean values and standard deviations of the DNA helical parameters were calculated on the mean conformer of the c23-free, c25-free, c43-free, c45-free, c23-3endo, c25-3endo, c43-3endo and c45-3endo trajectories and compared with the mean values measured in the x-ray crystallographic structure of [d(C4)]₄¹⁰ and in the NMR structure of [X(dC2)]₄. Two parameters, the twist and the rise, display significant differences (Fig. 3) between the analyzed conformations.

The NMR structure of [X(dC2)]₄ and the X-ray structure of [d(C4)]₄ exhibit similar rise and twist values. Analogous standard deviations are observed for the twist in the NMR and x-ray structures, but the standard deviation for the rise is larger in the NMR than in the x-ray structure, which probably reflects the lack of distance restraints at the symmetrical C1–C1 step.

The twist values from unrestrained (restrained) trajectories are located in the upper (lower) part of the twist range, between 10 and 25° (–10 and 8°) values. The pairs of twist and rise values calculated for each duplex of c23-free and c25-free are dispersed, indicating a misfit between the two duplexes; this may induce the increase in rise observed for these two trajectories. In all other simulations, they can be grouped by pairs, which is a sign of a better fit between the two duplexes.

Large rise values imply a decrease in stability of the i-motif. Indeed, during the loss of stacking, the favorable Lennard–Jones interactions between cytosine bases decrease along with the favorable dipole–dipole interactions between the N4–H4 and the C2–O2 bonds, as the mobility of the bases increases and the angles between the bonds become less favorable. Because of the distance dependence of the non-bonded interactions ($1/r^6$ for Lennard–Jones, $1/r^3$ for the dipolar and $1/r$ for the Coulombic interaction), the decrease in attractive interactions is much larger than the parallel decrease in the repulsive Coulombic interaction between the protonated cytosines. Consequently, the large rise values of the c25-3endo, c23-free and c25-free mean conformers result in a lower stability. The difference in stability between c23-3endo and c25-3endo is in complete agreement with experimental observations.¹¹ As the standard deviations of the rise (twist) values are mainly in the 0.08–0.33 Å (0.01–5.43°) range (Fig. 3), the twist and rise values observed for c23-3endo, c25-free and c23-free are significantly different

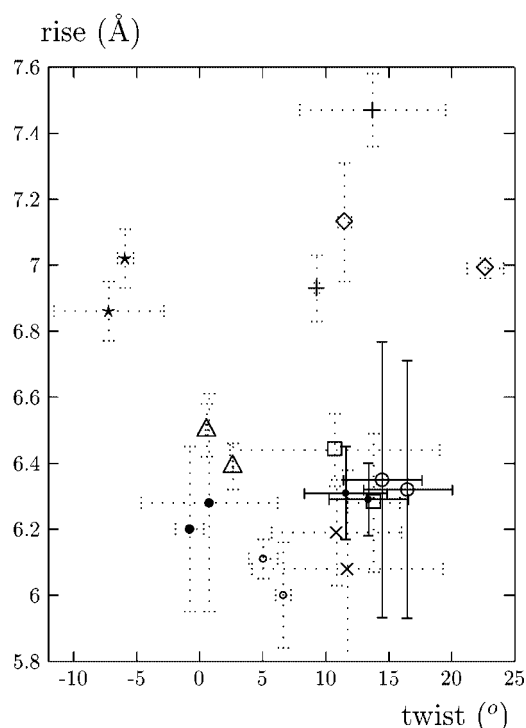


Figure 3. Mean rise values (Å) displayed as a function to the mean twist values (degrees) for the mean conformers of the NMR structure of $[d(C2)]_4^9$ (○), the crystallographic structure of $[d(C4)]_4^{10}$ (● and solid line), the mean conformer of c23-free (◇), c25-free (+), c43-free (□), c45-free (×), c23-3endo (△), c25-3endo (*), c43-3endo (small ○), c45-3endo (● and dotted line). As the calculations were performed separately on each of the two duplexes making up the i-motif, they provide, for each structure or trajectory analyzed, two pairs of twist and rise values. The standard deviations on the twist and rise values are in solid lines for the NMR and x-ray structures and in dotted lines for the mean conformers of MD simulations.

from the other twist and rise pairs. The stability of the c23-3endo, c25-free and c23-free conformers is therefore significantly different from the other conformer stabilities.

The conformers whose rise values are larger than those observed in NMR and x-ray structures display either (i) twist values more negative than -10° (c25-3endo) or (ii) different twist values for the two duplexes (c23-free, c25-free). On the other hand, conformers with rise values similar to those of the NMR and x-ray structures also display twist values in the 0 – 15° range. It is difficult to determine precisely the twist values from MD simulations, but one can see in Fig. 3 that the longer the cytosine stretches, the smaller are the rise values; the smaller rise values are observed along with twist values and closer to 10° . The 0 – 15° interval of twist values can therefore be considered as the range of values corresponding to stable i-motif tetramers.

The width of the narrow groove gives values 5.1 ± 0.6 , 4.9 ± 0.1 , 5.3 ± 0.3 and 5.5 ± 0.3 Å for c23-free, c23-3endo, c25-free and c25-3endo, respectively. It is correlated with the mean values of rise (Fig. 3); indeed, all mean conformers (c23-free, c25-free and c25-3endo) with a rise value larger than those observed in the NMR and x-ray structures also exhibit the widest narrow grooves and the largest

standard deviations. On the other hand, c43-free, c43-3endo, c45-free and c45-3endo simulations, whose rise values are similar to those of the NMR and x-ray structures, display smaller widths: 3.6 ± 0.4 , 3.1 ± 0.6 , 3.2 ± 0.9 and 3.6 ± 1.0 Å, respectively. Small rise values are thus related to a smaller width of the narrow groove and closer interactions of phosphodiester backbones.

DISCUSSION

The determination of the NMR structure of the $[Xd(C2)]_4$ tetramer has been presented, along with MD simulations run on both 3'E and 5'E topologies of the i-motif structure. The mean conformers obtained for each simulation were analyzed and related to the NMR structure.

A correct twist of the phosphodiester backbone is essential to keep the rise to a value allowing good stacking of the bases. Indeed, the MD simulations performed showed that the conformers displaying large rise values (Fig. 3) also exhibit heterogeneous or very negative twist values. Nevertheless, the exact value of the twist is difficult to estimate from MD simulations: a large range of twist values is accessible and seems compatible with good stacking of bases (Fig. 3).

As seen for c23-free and c25-free (Fig. 3), in order to keep a good base stacking, the twist of the phosphodiester backbone should be the same for the four DNA strands, and an attractive interaction between the strands is certainly required to enforce it. The best candidate for this attractive interaction is the interaction between sugars, whatever its origin (van der Waals or $CH \cdots O$ hydrogen bonds). This model can also explain the greater stability of the 3'E topology with respect to the 5'E topology, observed for $[Xd(C2)]_4$. Indeed, because of the elasticity of the phosphodiester backbone, two close sugar–sugar pairs are necessary to twist the two backbones in the same way. For all the $d[(Cn)]_4$ ($3 \leq n < 6$) tetramers, in either the 3'E or 5'E topology, the number of close sugar pairs across each narrow groove is larger than two. This results in correct twist and good base stacking of both topologies. In contrast, for the 5'E topology of $[Xd(C2)]_4$, the single close sugar pair observed for each backbone pair is not sufficient to produce the backbone twist and good base stacking. This topology is unstable and not observed experimentally.

The coordinates of the lowest NOESY-related energy conformer $[Xd(C2)]_4$ have been submitted to the protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY, USA (acquisition number not yet available).

Acknowledgments

T.E.M. thanks Dr Jose Gallego for the partial charges for protonated cytosines, and helpful discussions with Drs Richard Lavery, Brigitte Hartmann, David Kombo and Felicia Petici are acknowledged. CNRS is acknowledged for funding. The calculations presented here were supported by grant 11410 from the IDRIS Supercomputing center.

REFERENCES

1. Keniry M. *Biopolymers* 2000; **56**: 123.
2. Guéron M, Leroy JL. *Curr. Opin. Struct. Biol.* 2000; **10**: 326.
3. Arthanari H, Bolton P. *Chem. Biol.* 2001; **8**: 221.
4. Gehring K, Leroy JL, Guéron M. *Nature (London)* 1993; **363**: 499.

| | | |
|----|---|----|
| 1 | 5. Nonin-Lecomte S, Leroy JL. <i>J. Mol. Biol.</i> 2001; 309 : 491. | 41 |
| 2 | 6. Ahmed S, Kintanar A, Henderson E. <i>Nature Struct. Biol.</i> 1994; 1 : | 42 |
| 3 | 83. | 43 |
| 4 | 7. Phan A, Guéron M, Leroy JL. <i>J. Mol. Biol.</i> 2000; 299 : 123. | 44 |
| 5 | 8. Phan A, Leroy JL. <i>J. Biomol. Struct. Dyn.</i> 2000; 299 : 123. | 45 |
| 6 | 9. Leroy JL, Guéron M. <i>Structure</i> 1995; 3 : 101. | 46 |
| 7 | 10. Chen L, Cai L, Zhang X, Rich A. <i>Biochemistry</i> 1994; 33 : 13 540. | 47 |
| 8 | 11. Leroy JL, Snoussi K, Guéron M. <i>Magn. Reson. Chem.</i> 2001; 39 : | 48 |
| 9 | S171. | 49 |
| 10 | 12. Plateau P, Guéron M. <i>J. Am. Chem. Soc.</i> 1982; 104 : 7310. | 50 |
| 11 | 13. Clore GM, Nilges M, Brünger AT, Karplus M, Gronenborn AM. | 51 |
| 12 | Q1 14. Brünger AT. Yale University Press: New Haven, CT, 1990.● | 52 |
| 13 | 15. Case D, Pearlman DA, Caldwell JW, Cheatham TE, Ross W, | 53 |
| 14 | Simmerling C, Darden T, Merz K, Stanton R, Cheng A, | 54 |
| 15 | Q2 Vincent J, Crowley M, Tsui V, Radmer R, Duan Y, Pietera J, | 55 |
| 16 | Massova I, Seibel G, Singh U, Weiner P, Kollman PA. <i>Amber 6</i> , | 56 |
| 17 | 1999.● | 57 |
| 18 | 16. Cheatham TE, Cieplak P, Kollman PAJ. <i>Biomol. Struct. Dyn.</i> 1999; | 58 |
| 19 | 16 : 845. | 59 |
| 20 | 17. Jorgensen W. <i>J. Am. Chem. Soc.</i> 1981; 103 : 335. | 60 |
| 21 | 18. Cornell W, Cieplak P, Bayly C, Gould I, Merz K, Ferguson D, | 61 |
| 22 | Spellmeyer D, Fox T, Cadwell J, Kollman PA. <i>J. Am. Chem. Soc.</i> | 62 |
| 23 | 1995; 117 : 5179. | 63 |
| 24 | | 64 |
| 25 | | 65 |
| 26 | | 66 |
| 27 | | 67 |
| 28 | | 68 |
| 29 | | 69 |
| 30 | | 70 |
| 31 | | 71 |
| 32 | | 72 |
| 33 | | 73 |
| 34 | | 74 |
| 35 | | 75 |
| 36 | | 76 |
| 37 | | 77 |
| 38 | | 78 |
| 39 | | 79 |
| 40 | | 80 |
| | 19. Gallego J, Luque F, Orozco M, Gago F. <i>J. Biomol. Struct. Dyn.</i> | |
| | 1994; 12 : 111. | |
| | 20. Darden T, York D, Pedersen L. <i>J. Chem. Phys.</i> 1993; 98 : 10 089. | |
| | 21. Young MA, Ravishanker G, Beveridge DL. <i>Biophys. J.</i> 1997; 73 : | |
| | 2313. | |
| | 22. Berendsen H, Postma J, van Gunsteren W, DiNola A, Haak J. | |
| | <i>J. Chem. Phys.</i> 1984; 81 : 3684. | |
| | 23. Cheatham TE, Kollman PA. <i>J. Am. Chem. Soc.</i> 1997; 119 : 4805. | |
| | 24. Humphrey W, Dalke A, Schulten K. <i>J. Mol. Graphics</i> 1996; 14 : 33. | |
| | 25. Lavery R, Sklenar H. <i>J. Biomol. Struct. Dyn.</i> 1989; 6 : 63. | |
| | 26. Leroy JL, Gehring K, Kettani A, Guéron M. <i>Biochemistry</i> 1993; 32 : | |
| | 6019. | |
| | 27. Han X, Leroy JL, Guéron M. <i>J. Mol. Biol.</i> 1998; 278 : 949. | |
| | 28. Nonin S, Phan A, Leroy JL. <i>Structure</i> 1997; 5 : 1231. | |
| | 29. Snoussi K, Nonin-Lecomte S, Leroy JL. <i>J. Mol. Biol.</i> 2001; 309 : | |
| | 139. | |
| | 30. Berger I, Egli M, Rich A. <i>Proc. Natl. Acad. Sci. USA</i> 1996; 93 : | |
| | 12 116. | |
| | 31. Cheatham TE, Kollman PA. <i>J. Mol. Biol.</i> 1996; 259 : 434. | |
| | 32. Wang J, Cieplak P, Kollman PA. <i>J. Comput. Chem.</i> 2000; 21 : 1049. | |
| | 33. Altona C, Sundaralingam M. <i>J. Am. Chem. Soc.</i> 1972; 94 : 8205. | |
| | 34. Reyes CM, Nifosi R, Frenkel AD, Kollman PA. <i>Biophys. J.</i> 2001; | |
| | 80 : 2833. | |

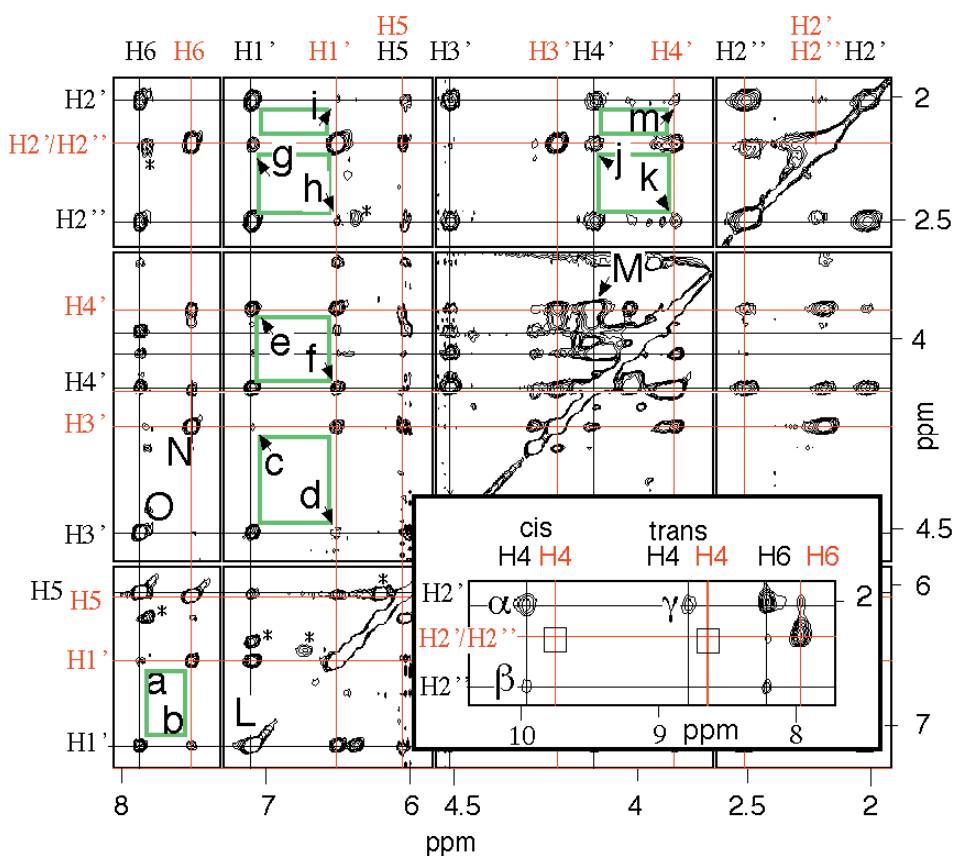


Plate 1. Expanded NOESY contour plot in D₂O solution at 0 °C (250 ms mixing time) of [Xd(C₂)]₄. The proton frequencies are color coded: black for C1 and red for C2. The rectangular sets of cross-peaks connecting the bases stacked by their black faces are indicated by green rectangles. The diagonally opposed inter-residue cross-peaks are labeled: (a) C1(H₆)–C2(H₁'), (b) C2(H₆)–C1(H₁'), (c) C1(H₁')–C2(H₃'), (d) C2(H₁')–C1(H₃'), (e) C1(H₁')–C2(H₄'), (f) C2(H₁')–C1(H₄'), (g) C1(H₁')–C2(H₂'/H₂''), (h) C2(H₁')–C1(H₂''), (i) C2(H₁')–C1(H₂'), (j) C1(H₄')–C2(H₂'/H₂''), (k) C2(H₄')–C1(H₂'), (m) C2(H₄')–C1(H₂'). L = C1(H₁')–C2(H₁'), M = C1(H₄')–C2(H₄'). The H₆–H₃' intra-residue cross peaks are labeled O (C1) and N (C2). NOE cross peaks from the non-structured Xd(C)₂ are indicated by stars (*). Inset: amino-H₂'/H₂'' region of the NOESY spectrum in H₂O solution at –3 °C (50 ms mixing time). The cross peaks (α , β and γ) reflect the close inter-residue contacts between the NH₂ and H₂'/H₂'' protons of two C1.C1⁺ pairs stacked by their faces oriented in the 3' direction. Note the absence of cross peaks between C2 (NH₂) and C2 (H₂'/H₂'') (empty rectangles) showing that the intra-residue NOE contribution is negligible.

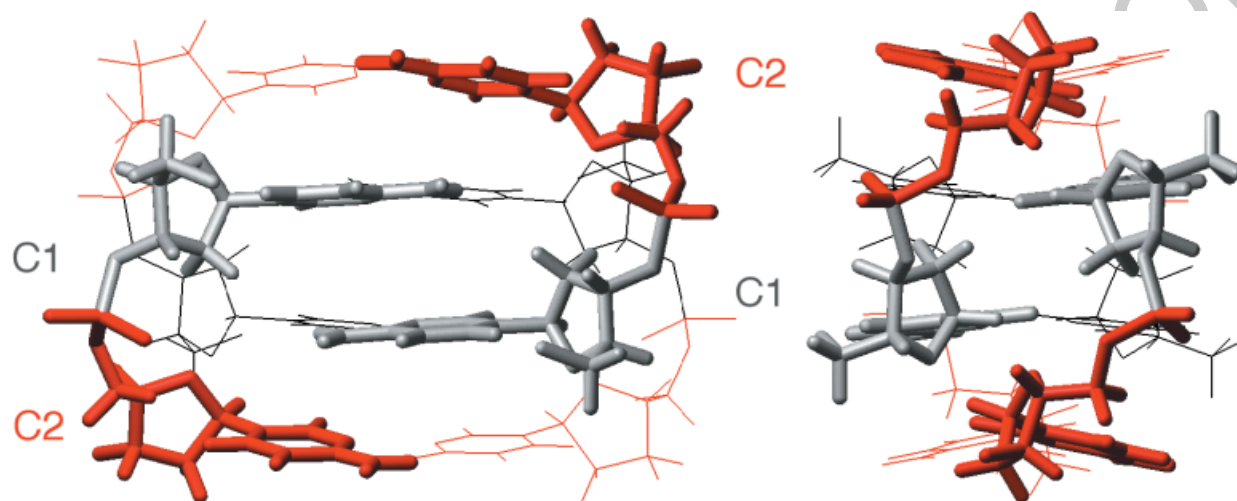


Plate 2. The computed structure of the [Xd(C2)]₄ tetramer. Left, view into the wide groove; right, narrow groove view. The background residues are drawn in thin lines.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

QUERIES TO BE ANSWERED BY AUTHOR (SEE MARGINAL MARKS)

IMPORTANT NOTE: Please mark your corrections and answers to these queries directly onto the proof at the relevant place. Do NOT mark your corrections on this query sheet.

| Query No. | Query |
|-----------|-------------------------------|
| Q1 | Please give book title. |
| Q2 | Give organisation + location. |

61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120

UNCORRECTED PROOFS