Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 303 (2003) 81–90
BBRC
www.elsevier.com/locate/ybbrc

NMR study of hexanucleotide d(CCGCGG)$_2$ containing two triplet repeats of fragile X syndrome☆

Daniel Monleón, Vicent Esteve, and Bernardo Celda*

Departamento de Química Física, Universitat de València, Edifici d'Investigació, Lab 3-054, Cl Dr. Moliner, 50, Burjassot, 46100 Valencia, Spain

Received 10 February 2003

Abstract

Long repeated stretches of d(CCG) and tri-nucleotide are crucial mutations that cause hereditary forms of mental retardation (fragile X-syndrome). Moreover, the alternating (CG) di-nucleotide is one of the candidates for Z-DNA conformation. Solution NMR structure of d(CCGCGG)$_2$ has been solved and is discussed. The determined NMR solution structure is a distorted highly bent B-DNA conformation with increased flexibility in both terminal residues. This conformation differs significantly from the Z-DNA tetramer structure reported for the same hexamer in the crystal state at similar ionic strength by Malinina and co-workers. Crystal structure of d(CCGCGG)$_2$ at high salt concentration includes a central alternating tetramer in Z-DNA conformation, while the initial cytosine swings out and forms a Watson–Crick base-pair with the terminal guanine of a symmetry-related molecule. In solution, NMR data for sugar ring puckering combined with restrained molecular dynamics simulations starting from a Z-DNA form show that terminal furanose residues could adopt the conformation required for aromatic bases swinging out. Therefore, tetramer formation could be considered possible once the hexanucleotide had previously adopted the Z-DNA form. This work gives some insight into correlations between anomalous crystal structures and their accessibility in the solution state.

Keywords: NMR DNA structure; Oligonucleotide conformation; Recombination; Fragile X-syndrome; CCG repeat

DNA has been proved to form a wide variety of tridimensional structures in many different conditions. B-DNA, characterized by a right-handed antiparallel double-stranded conformation, is the preferred conformation in solution and has been shown to play a fundamental role in many biological processes [1]. A and Z regular conformations have also been observed and correlated to particular biochemical situations [2]. DNA structure in its several forms has been extensively studied by crystal X-ray diffraction or NMR techniques to elucidate many biological questions related to gene molecular biology. Recently, many unusual DNA conformations such as single-stranded hairpins [3], triplexes [4,5], and quadruplexes [6,7] have been observed and well characterized. Different hydrogen bonding schemes are responsible in many cases for the stabilization of otherwise unstable global conformations. The large amount of possible hydrogen bond donors and acceptors in the chemical structure of DNA makes feasible a wide variety of stable conformations. Several internal and external factors govern DNA polymorphism such as base sequence, concentration, temperature, pH, ionic strength, and other solvent conditions [8]. Modifications of several external conditions can lead to transformations from one conformation to another.

Regular DNA conformations can be detected unambiguously by observation of characteristic structural patterns such as inter-protonic distances, glycosidic conformation, sugar ring pseudo-rotation angles, etc. Many of these patterns are suitable for determination by NMR spectra detailed analysis [9]. Particularly, sugar

☆ *Abbreviations:* NMR, nuclear magnetic resonance; NOE, nuclear overhauser effect; NOESY, NOE spectroscopy; DQF-COSY, double quantum filtered correlation spectroscopy; TOCSY, total correlation spectroscopy; AMBER, assisted model building with energy refinement; RMSD, root-mean-square deviation.

* Corresponding author. Fax: +34963543156.
E-mail address: bernardo.celda@uv.es (B. Celda).

0006-291X/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0006-291X(03)00304-8

ring pseudo-rotation angle has been shown to be a critical conformational feature to distinguish among different DNA conformational states. Proper analysis of DQF-COSY spectra by peak shape simulation makes it possible to obtain conformational exchange between two states and their population ratios [10].

Moreover, NMR can allow detecting even local mobility at the atomic level by signal broadening and other techniques. On the other hand, crystal-packing forces can be responsible for some supramolecular aggregate formations that probably are not stable in solution. Therefore, in order to check the stability of unusual DNA conformations in physiological conditions, solution NMR structural analysis may be preferred over crystal state determination. Likewise, oligonucleotide structural studies by X-ray single crystal diffraction can be complemented and expanded in some cases by low complexity NMR data analysis of DNA spectra. Therefore, NMR structural information applied to theoretical methods, like molecular dynamics simulations, can give deep insights into the structure and conformational flexibility of DNA fragments in solution.

The solution conformation of the hexanucleotide $d(CCGCGG)_2$ has been investigated. This palindromic oligonucleotide includes the CCG and CG subunits related to some serious pathological conditions and Z-DNA conformation, respectively. Particularly, long repeated stretches of d(CCG) and/or d(GGC) are critical mutations that cause some hereditary forms of mental retardation (fragile X-syndrome) [11]. Crystal structure of $d(CCGCGG)_2$ has already been solved at high sodium chloride concentration level. Reported structure is a DNA double dimer with a central alternating tetramer in Z-DNA conformation, while initial cytosine bases swing out to form base-pair hydrogen bonds with a symmetry-related molecule [12]. Despite the fact that solution structure would be expected to be a B-DNA, previous studies pointed out the possibility of a Z-DNA form based on circular dichroism spectroscopy data in solution at high NaCl concentration [13]. Moreover, vibrational spectroscopy studies on $d(CCGCGG)_2$ hexa-nucleotide show that transitions from crystal form to solution B-DNA form are not complete and that the oligonucleotide remains trapped in an intermediate conformation that is often proposed in models of B-greater than Z transitions [14]. On the other hand, cation dependence for $d(TGGCGGC)_2$ containing two triplet repeats d(CGG) also responsible for fragile X-syndrome has been studied by NMR [15]. Particularly, $d(TGGCGGC)_2$ at high Na^+ concentrations shows a strong dependence on pH. Low pH values seem to stabilize the formation of DNA quadruplex and neutral values of pH stabilize the usual B-DNA duplex. This dependence has not been observed for $d(CCGCGG)_2$. Therefore, conditions and mechanisms for quadruplex formation required for fragile X-syndrome effect may be different in the case of d(CCG) triple repeats and in the case of d(GGC) triple repeats. As conclusion, $d(CCGCGG)_2$ conformation cannot be established exclusively from previous studies on the molecule.

All these sequential and structural features make $d(CCGCGG)_2$ an excellent subject for a NMR solution structure study. The short length of the hexanucleotide makes the NMR analysis relatively straightforward. Moreover, it contains interesting di- and tri-nucleotide sequences and it shows a very peculiar tridimensional structure in the crystalline form. Despite the fact that crystal structure has not been observed in the NMR solution comparative study carried out here, other conformational and dynamic particularities have been detected by restrained molecular dynamics simulations.

Materials and methods

Oligonucleotide obtention and sample preparation. Oligonucleotide $d(CCGCGG)_2$ was synthesized by standard solid phase methods and purified by reverse phase HPLC to a purity higher than 95%. Composition of the oligonucleotides was confirmed by HPLC analysis of the deoxynucleotides generated by enzymatic hydrolysis. Sample concentration was determined by UV absorption spectroscopy at 25° using $\varepsilon_{260} = 6400\,M^{-1}\,cm^{-1}$. Double helipticity of DNA was confirmed by observation of typical hyperchromicity and thermic denaturalization transition. Two different NMR samples were prepared both with a DNA duplex concentration of 1.1 mM. DNA was dissolved in 50 mM phosphate buffer, 0.1% NaN_3, and 50 and 2000 mM NaCl for preparation of samples 1 and 2, respectively. The pH of both samples was adjusted to ~6 and to match conditions reported for crystal structure. Additional sample of $d(CCGCGG)_2$ at high salt concentration and low pH was prepared to evaluate the possible quadruplex formation at low pH values as reported in previous work [15]. Unfortunately, buffer used in the crystal structure determination was not suitable for DNA NMR studies because of the high amount of overlapping resonances in the aliphatic region of the spectra. Ninety five percent of D_2O and 5% H_2O were used as solvent for both samples for better elimination of water line interference.

NMR spectroscopy. Most of NMR spectra were recorded on Varian Unity spectrometers at 300 and 400 MHz at 25 °C. In order to solve overlapping problems for 3′-terminal guanidines some additional experiments were recorded on a Bruker DRX 600 MHz at 20 °C on the sample at high sodium chloride concentration. One-dimensional imino proton spectra were also recorded on a Bruker DRX 600 MHz at 1 °C on samples at high and low sodium chloride concentrations. DSS was used as external reference by measuring 1D spectra of standard sample between different experiments. One-dimensional ^{31}P spectra were recorded at 300 MHz to confirm the structural symmetry of the DNA duplex. DQF-COSY, 2D-H,H-TOCSY, and 2D-NOESY spectra were acquired, processed, and analyzed using standard methodology for nucleic acid NMR structure determination [9]. Assigned NOE cross-peaks were classified by intensities with respect to intra geminal $H_{2'}$–$H_{2''}$ and intra aromatic H_5–H_6 cytosine NOE cross-peak average intensities in strong, medium, and weak, and then upper limit NOE distance constraints were assigned to 2.5, 3.5, and 5.0 Å, respectively. Mixing times for NOESY spectra were 50, 100, and 200 ms and mixing times for TOCSY spectra were 40, 80, and 120 ms for both samples. Presaturation during recycling delay was used in all spectra for water line suppression. Time proportional phase increment (TPPI) was used

as quadrature detection method in all 2D experiments. Five hundred and twelve t_1 increments were collected in all 2D experiments and linear predicted to 1024 final points to get high digital resolution 2D spectra.

DQF-COSY cross-peak simulation was performed for $H_{1'}$–$H_{2''}$, $H_{1'}$–$H_{2'}$, and $H_{2'}$–$H_{3'}$ peaks of C1, G5, and G6 using LINSHA software for theoretical FID calculation and SPHINX software for processing and plotting [10,16]. Spectral windows of 60 by 40 Hz equivalent to 0.15 by 0.1 ppm at 400 MHz were used for simulation. Identical processing and contour plotting parameters were used for experimental and simulated FIDs. Two stages of optimization were performed over shape peak simulation. In a first stage, a grid of J-coupling values for $^3J_{1'2'}$ and $^3J_{1'2''}$ was used with increments of 0.5 Hz to generate initial sets of simulated peaks. $^3J_{1'2'}$ and $^3J_{1'2''}$ are the J-coupling constants with the largest influence in peak shape and size. Automatic comparison over several hundreds of cross-peak simulations was performed by minimal difference detection on a set of 20 digital points in the peak graph. Fine tuning of $^3J_{1'2'}$ and $^3J_{1'2''}$ was performed by empirical comparison of simulated and experimental cross-peaks. A second and final stage of fine tuning for $^3J_{1'2''}$, $^3J_{2'2''}$, $^3J_{3'4'}$, and $^3J_{3'P}$ was performed by manual modification of standard values for pseudo-rotation angle *P* estimated from $^3J_{1'2'}$ and $^3J_{1'2''}$ and direct fitting of the corresponding graphs to the experimental peak shapes. DQF-COSY cross-peak features of residues C2, G3, and C4 were close to typical features of a standard B-DNA form. Therefore, no optimization was required to extract corresponding J-coupling values. Finally, accurate values for *P* and sugar puckers were obtained and two dihedral angle constraints (–$O_{2'}$–$C_{3'}$–$C_{4'}$–$O_{4'}$– and –$C_{1'}$–$C_{2'}$–$C_{3'}$–$C_{4'}$–) for each nucleotide were derived for molecular dynamics simulations.

Helicoid parameter analysis. Statistical and geometrical analysis of final DNA structures was performed with NDBSTAT software [17] and NEWHELIX software [18]. Agreement in a margin of 5% was always observed between values obtained by NDBSTAT and NEWHELIX. However, NDBSTAT was preferred over NEWHELIX and other usual software packages because of the improved precision in helix axis calculation. Global axis calculation in NDBSTAT is independent from local helix axis. Fitting of helix projection over a regular ellipse is performed to obtain global and local helix axis vector and position. This method gives very reliable results, even for very distorted geometries. Moreover, helix axes are calculated at global and local level and therefore geometrical parameters can be evaluated with respect to the helix coordinate frame. NDBSTAT software can be obtained at no cost upon request to the author (vicent@mol.biol.ethz.ch).

Molecular dynamics simulation. InsightII (Molecular Simulations, San Diego, CA) was used to generate initial models of $d(CCGCGG)_2$ for A, B, and Z-DNA conformations. DISCOVER (Molecular Simulations, San Diego, CA) software was used for all molecular dynamics calculations. AMBER [19] force field was used to calculate the energy of the system. This forcefield has been extensively tested on many oligonucleotide structure calculations [20,21]. All calculations were repeated without NMR conformational restraints to check whether structural distortions in final models were the result of intrinsic forcefield effects. No significant distortion was observed for final theoretical models calculated in the absence of structural restraints. Electrostatic interactions were calculated using a coulombic potential with a distance-dependent dielectric constant to simulate solvent effects. Standard restrained simulated annealing protocols for DNA calculations distributed with InsightII software package were used to calculate $d(CCGCGG)_2$ NMR structures. Distance constraints and dihedral angle constraints were obtained from a calibration of NOE cross-peaks and DQF-COSY peak shape analysis as explained above. Thirty degrees were added and subtracted to estimate values for dihedral angle in order to get upper and lower limits for the equivalent restraint. Additional weak reinforcements of hydrogen bonds equivalent to one NOE constraint were included in the initial heating stages of simulated annealing calculations for all residues showing interstrand NOE connectivities. Twenty and 5 kcal mol^{-1} $Å^{-2}$ force constants were used for intra and inter-nucleotide restraints, respectively. Fifty kcal mol^{-1} deg^{-2} constant force was used for dihedral angle constraints. Fifty structures were calculated from simulated annealing calculations and the best 10 were selected for further energy optimization by a 10 ps nonrestrained molecular dynamics including Na^+ counter-ions empirically positioned.

Results and discussion

NMR data analysis

Only six unique sugar proton spin systems and three unique H_5–H_6 cross-peaks were observed (Fig. 1A). Two signals were observed in the imino NH_2 proton region of 1D spectrum at low salt concentration and at 298 K, indicating hydrogen bond formation for at least two of the three cytosines. However, at high salt concentration and low temperatures, three signals can be observed. Nevertheless, these results confirm the proposed symmetry of the palindromic nucleotide sequence. Sequence-specific resonance assignment of non-exchangeable protons in $d(CCGCGG)_2$ was performed following well-established strategies proposed by Wüthrich and co-workers [9,22]. DQF-COSY and 2D-H,H-TOCSY at medium and long mixing times were used to separate sugar proton's spin systems. 2D-NOESY spectra at different mixing times were used to locate sequential NOE interactions and therefore to identify and map spin systems to unique nucleotides in the sequence. Fig. 1B shows the H_8/H_6–$H_{2'}/H_{2''}$ region of a NOESY spectrum recorded at 25 °C with a mixing time of 100 ms with connectivities used for sequential assignment. Observed connectivities correspond to those of a right-handed B-DNA symmetric duplex and can be followed from residue C1 to residue G6 without interruption. Absence of G $H_{2'}$–C H_5 NOE cross-peaks is indicative of a non-Z-DNA conformation in solution, different from the structure reported in crystal state. Distinction between $H_{2'}$ and $H_{2''}$ protons was achieved by analysis of the $H_{1'}$–$H_{2'}/H_{2''}$ region of the DQF-COSY spectrum shown in Fig. 2 and comparison to intensity ratios of equivalent cross-peaks in 2D-NOESY spectra. Most of the $H_{2'}$ proton resonances are upfield with respect to $H_{2''}$ of the same residue. However, inversion in $H_{2'}$ and $H_{2''}$ chemical shift values is observed for residue G6. $H_{2''}$ resonance is upfield with respect to $H_{2'}$ in this case. This anomaly is confirmed by the inversion observed in the hyperfine coupling structure of the $H_{1'}$–$H_{2'}/H_{2''}$ intra-residue DQF-COSY cross-peaks as shown in Fig. 2, and by the observation of both $H_{3'}$–$H_{2'}$ and $H_{3'}$–$H_{2''}$ cross-peaks in the NOESY spectra. This can be related to conformational differences in the sugar ring conformation between G6 and the rest of nucleotides. G6 in reported

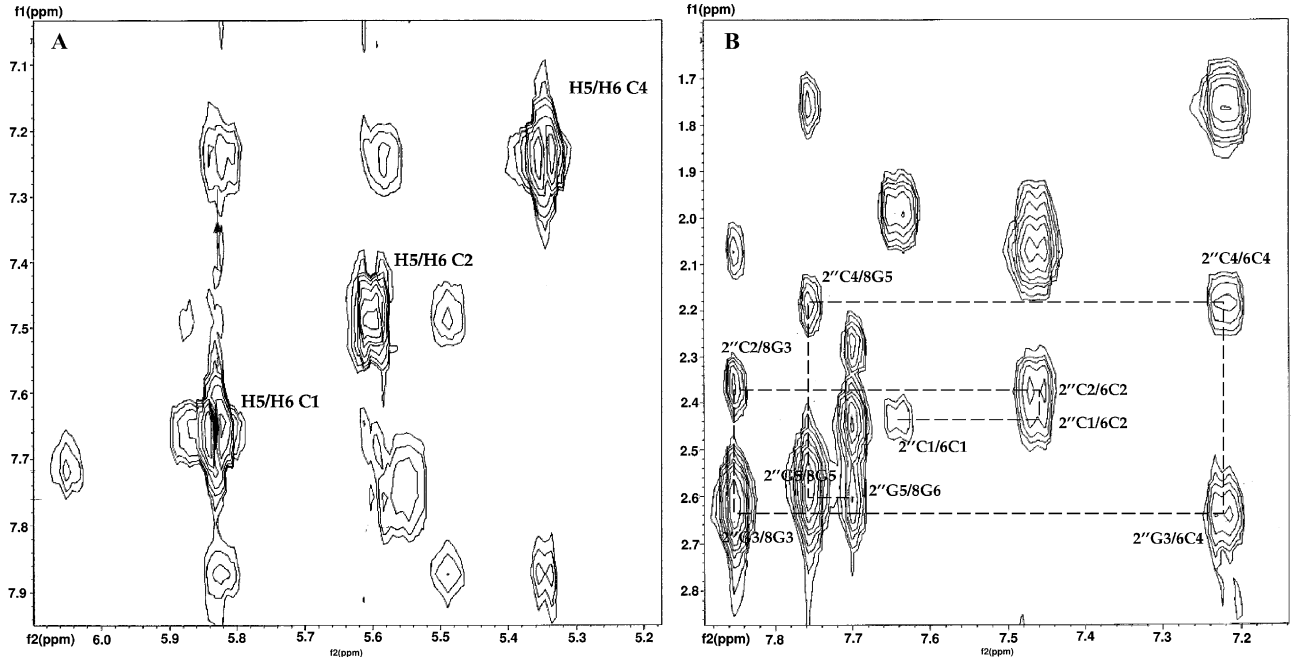

Fig. 1. (A) Expanded contour plot of the aromatic H_5–H_6 region of the 2D-NOESY spectrum (100-ms mixing time) in D_2O at 400 MHz and 25 °C. (B) Expanded contour plot of the $H_{8/6}$–$H_{2'/2''}$ region of 2D-NOESY spectrum (100-ms mixing time) in D_2O at 400 MHz and 25 °C, showing connectivities used for the sequential assignment of $d(CCGCGG)_2$.

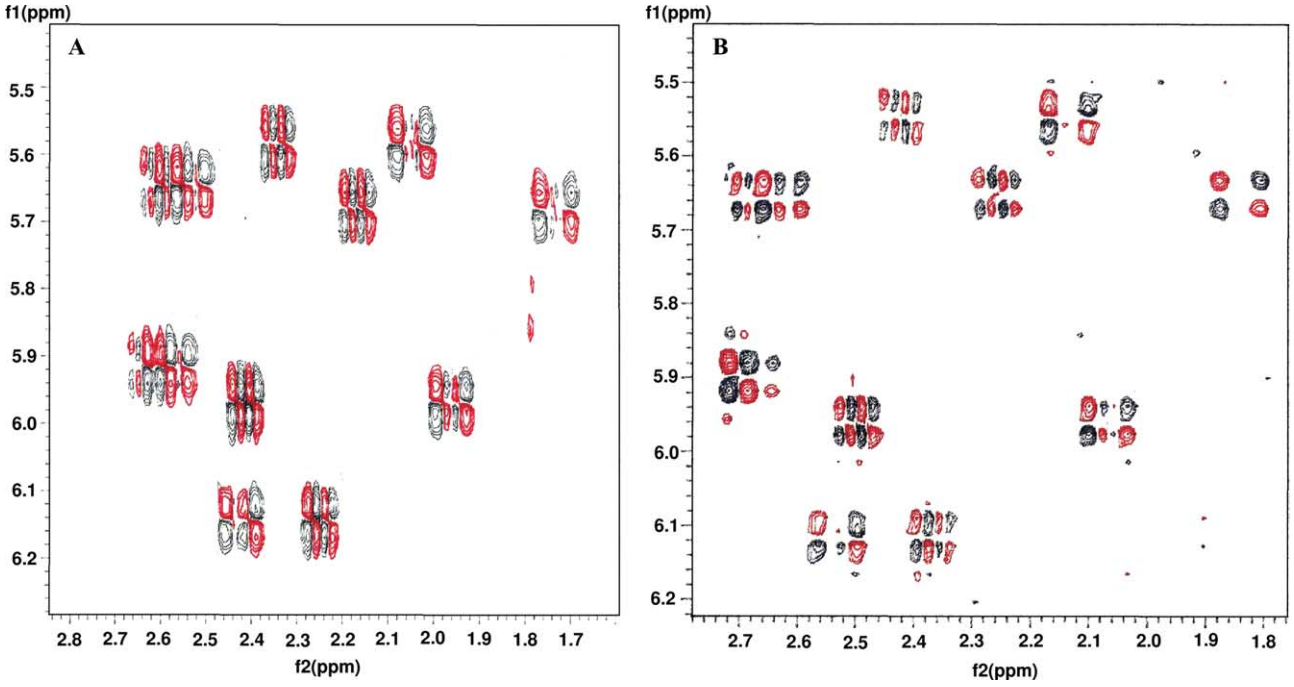

Fig. 2. Expanded contour plots of the $H_{1'}$–$H_{2'/2''}$ region of DQF-COSY spectrum of $d(CCGCGG)_2$ at 400 MHz in D_2O at 25 °C and 50 mM NaCl (A) and at 600 MHz in D_2O at 25 °C and 2 M NaCl (B).

X-ray crystallographic structure of $d(CCGCGG)_2$ is one of the most stretched residues because of intermolecular hydrogen pairing [12]. $H_{3'}$ and $H_{4'}$ resonances were assigned by cross-peak analysis in the corresponding regions of DQF-COSY and 2D-H,H-TOCSY spectra. $H_{5'}$ and $H_{5''}$ proton resonances for most residues were not unambiguously assigned at 400 MHz because of severe overlapping in the corresponding regions and absence of necessary cross-peaks. $H_{5'}$ and $H_{5''}$ proton resonances were unambiguously

assigned by DQF-COSY and 2D-NOESY spectra at 600 MHz with the exception of residue C4. Therefore, all non-exchangeable proton resonances but C4 $H_{5'}$ and $H_{5''}$ have been completely assigned. Chemical shift resonances for all assigned protons have been deposited and can be accessed at the BioMagnetic Resonance Databank (http://www.bmrb.wisc.edu) (BMRB entry 5562).

The general intensity patterns observed in many different regions of NOESY spectra are indicative of a particular conformational feature. For example, intensities of $H_{6/8}$–$H_{2'}$ intra-residue cross-peaks are larger than the intensities of $H_{6/8}$–$H_{2''}$ ones. Whereas H_6/H_8–$H_{1'}$ intra-residue cross-peak intensities are very weak. These NOE intensities are directly related to the glicosydic dihedral angle χ defined by the relative orientation of the aromatic base with respect to the sugar ring. The particular pattern observed for most residues in $d(CCGCGG)_2$ indicates an *anti* configuration for χ. On the other hand, alternating *syn*/*anti* patterns typical of Z-DNA form structures are not observed, indicating an A or B-DNA conformation for $d(CCGCGG)_2$ in solution. $H_{2'}$–$H_{3'}$ intra-residue cross-peaks observed for residues C1 and G6 are related to some degree of interconversion between *N* and *S* conformations for the deoxyribose ring. In principle, this corresponds well with larger flexibility expected for terminal residues.

Nucleotide sugar puckering is directly related to tridimensional DNA conformation and can be estimated by *P*. *P* range values can be obtained by complete simulation of $H_{1'}$–$H_{2'}$ and $H_{1'}$–$H_{2''}$ DQF-COSY cross-peaks [10]. DQF-COSY cross-peaks were simulated for half of the hexanucleotide residues. Hyperfine structures of $H_{1'}$–$H_{2'}$ and $H_{1'}$–$H_{2''}$ DQF-COSY cross-peaks of C2, G3, and C4 residues for $d(CCGCGG)_2$ were directly compared to well-known cross-peak patterns previously reported [10] and an extensive matching was observed. For the other three nucleotides, C1, G5, and G6, DQF-COSY cross-peaks were successfully simulated. However, extensive overlapping of DQF-COSY $H_{1'}$–$H_{2'}$, and $H_{1'}$–$H_{2''}$ cross-peaks for residue G5 made it very difficult to analyze them separately, so additionally grouped J-coupling values were used for the simulations of these cross-peaks. Comparison between final simulation and experimental data for DQF-COSY $H_{1'}$–$H_{2'}$ cross-peaks of residues C1, G5, and G6 is shown in Fig. 3. ^3J coupling constant values were obtained for sugar protons in all residues and are reported in Table 2. Relative populations of *S* type sugar conformations were estimated from the summatory $\Sigma_{1'} = J(N)_{1'2'} + J(S)_{1'2'}$ [23]. All sugar puckering of non-terminal residues in this oligonucleotide was found to be at least 90% in the *S* conformation. In residues C1 and G6, only 70% of deoxyriboses were found to be in *S* conformation. Therefore, ratios between *N* and *S* conformation calculated from J-coupling values in the sugar ring indicate a larger than usual amount of chemical exchange in these terminal residues, as observed in other oligonucleotide NMR studies [10]. *P* average values for most of the residues indicate a sugar puckering conformation close to $C_{2'}$-endo, typical of B-DNA conformation, with the exception of G6 whose furanose conformation is close to $O_{4'}$-endo. *P* value for terminal residues C1 and G6 can be related to strong conformational variations. Moreover, *P* values for C1 and G6 combined with changes observed in hyperfine coupling structure for $H_{1'}$–$H_{2'/2''}$ DQF-COSY cross-peaks may be related to some type of large conformational exchange for these residues. Despite flexibility which is always expected for terminal residues in short oligonucleotides like $d(CCGCGG)_2$, NMR data suggest unusually increased mobility for C1 and G6.

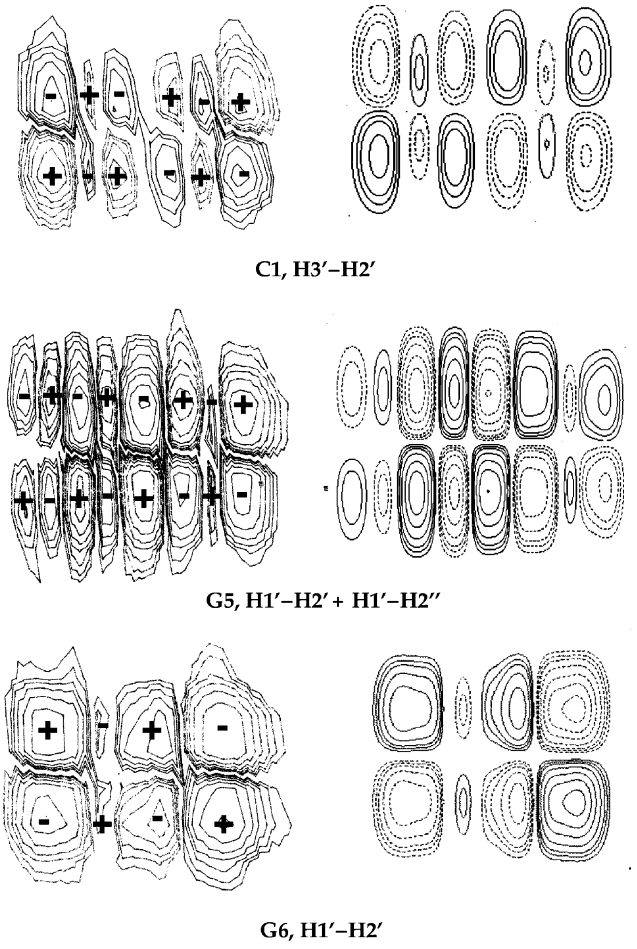

Fig. 3. Comparison between experimental (right) and LINSHA simulated (left) DQF-COSY $H_{1'}$–$H_{2''}$ cross-peaks for residues C1, G5, and G6 with a particularly complex hyperfine structure. Discontinuous lines indicate negative sign of correspondent signal.

Opposite to what was expected, differences between low and high salt concentration spectra were very small. This could be expected as previous work indicates that Na^+ effect on DNA quadruplex can only be seen at low

pH values. However, 1D-WATERGATE spectrum of imino protons for $d(CCGCGG)_2$ in H_2O at high salt concentration and pH ~3 does not present significant differences with respect to the sample at the same salt concentration and pH ~6 spectrum (data not shown). Only a slight resonance shifting for some protons was observed as opposite to what was reported previously for similar oligonucleotides responsible for fragile X syndrome [15]. Similar or identical DQF-COSY hyperfine structure patterns are observed for samples at both salt concentrations as shown in Fig. 2. Minimal differences are observed in some regions of the 2D-NOESY spectra as a result of some additional overlapping produced by resonances shifting associated to ionic strength effects. Although ^{31}P resonances were not completely assigned, good dispersion was observed and ^{31}P chemical shift values were 0.1 ppm or more upfield with respect to that observed for standard B-DNA. This could suggest additional exposure to the solvent of phosphate groups in $d(CCGCGG)_2$ [24].

At high salt concentration and low temperature, the complete set of three imino NH cross-peaks can be observed. Signal broadening produced by the low temperature required for imino proton detection makes it difficult to evaluate whether the additional peak resulted from resonance shifting or stabilization of a new hydrogen bond. On the other hand, increased flexibility expected for terminal C1 residue could make that hydrogen bond less stable than the others. Moreover, there are no evidences in all the rest of NMR data, indicating structural changes in $d(CCGCGG)_2$ at high salt concentration. This made resonance shifting by counter-ion effect the most probable cause for the additional peak observed at high salt concentration. Possible stabilization of a particular hydrogen bond produced by a combined effect of temperature and salt concentration could not be confirmed. Therefore, only low salt concentration experimental data were used to perform one set of molecular dynamics calculations. Overall, the high similarity among spectra collected in high and low salt concentrations and high and low pH conditions seems to indicate that no conformational switch exists for $d(CCGCGG)_2$ in these conditions. However, crystal structure suggests that this switch could be possible in different conditions and therefore justifies the reported role of this oligonucleotide in the fragile X syndrome [12].

Molecular modeling

A total of 44 meaningful inter-proton distance constraints were obtained from the 2D-NOESY spectra. This amount of distance restraints is typical of short symmetric oligonucleotides. As a result of the high symmetry of the molecule all 40 intra-strand NOEs were applied twice, once in each strand. Only four inter-strand NOEs were observed for residues C2, G3, C4, and G5 in the 2D-NOESY spectra at 200-ms mixing time. Therefore, only on those residues the hydrogen bond was reinforced as detailed in Materials and methods. Twelve dihedral angle constraints derived from the hyperfine structure DQF-COSY cross-peak analysis were applied to each strand. Final set of structural restraints was considered satisfactory because of the large contribution of dihedral angle constraints to calculations convergence and the high symmetry of the molecule. Ten final structures were obtained with these structural restraints molecular dynamics starting from A, B, and Z-DNA conformations. A summary of the structural restraints and violations of final models is shown in Table 1.

Only molecular dynamics calculations starting from B-DNA form converged to a well-fitted set of structures as expected from all NMR qualitative information discussed above. Superposition of heavy atoms for best 10 models is shown in Fig. 4. RMSD value for all heavy atoms in the 10 models was 0.44 Å. Final conformation for $d(CCGCGG)_2$ shows some important distortions versus typical B-DNA standard conformation. Molecular dynamics calculations starting from A-DNA form quickly change their structural characteristics to a conformation close to B-DNA. However, protocols used for this work do not provide enough conformational accessibility to achieve a complete transition from A to B-DNA. Final models calculated from A-DNA form are semi-stable distorted B-DNA forms. Other protocols were not tested because study of transitions from A to B-DNA form was beyond the scope of this work. Only 1 and 4 NOE constraint violations larger than 0.2 Å are observed in the final structures of molecular dynamics calculations starting

Table 1
Summary of meaningful structural restraints used in the restrained molecular dynamics calculation of $d(CCGCGG)_2$ and statistics of restraint violations for final best 10 models selected from calculations starting at A, B, and Z-DNA forms

NOE distance restraints	
Total	84
Intra-nucleotide	56
Sequential	24
Inter-strand	4
Dihedral angle restraints	24
Hydrogen bond restraints	12
Total number of distance restraints violations >0.1 Å	
A-DNA form starting point	4 ± 2
B-DNA form starting point	1 ± 1
Z-DNA form starting point	44 ± 12
Total number of dihedral angle restraints violations >5°	
A-DNA form starting point	0 ± 0
B-DNA form starting point	0 ± 0
Z-DNA form starting point	1 ± 1

NOEs used for sequential assignment of resonances but lacking structural information were not applied.

Table 2
J-coupling values of sugar protons as measured by DQF-COSY cross-peak simulations with LINSHA and SPHINX and comparison with experimental cross-peaks

	$J_{1'2'}$	$J_{1'2''}$	$J_{2'2''}$	$J_{2'3'}$	$J_{3'4'}$	$J_{3'P}$
C1	7.6 ± 0.2	6.1 ± 0.3	-12.5 ± 0.5	5.5 ± 0.5	4.0 ± 0.5	<3.5
C2	10.0 ± 0.5	5.2 ± 0.5	-12.5 ± 0.5	5.0 ± 0.5	<3.0	<2.0
G3	10.8 ± 0.8	5.2 ± 0.5	-13.5 ± 0.5	5.0 ± 0.5	<3.0	<2.0
C4	10.7 ± 0.3	5.4 ± 0.3	-13.5 ± 0.5	8.3 ± 0.3	<3.0	<3.0
C5	9.8 ± 0.7	5.8 ± 0.6	-13.5 ± 0.5	4.5 ± 0.5	4.0 ± 0.5	<3.5
G6	7.7 ± 0.5	6.0 ± 0.5	-13.5 ± 0.5	6.0 ± 0.8	<4.5	<3.5

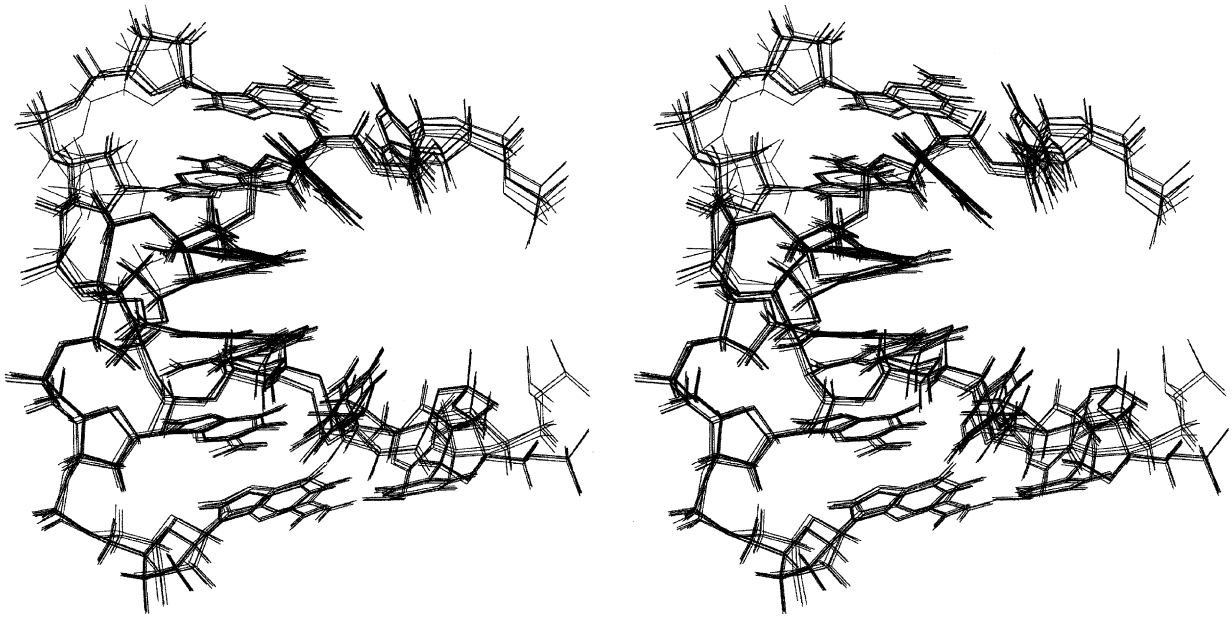

Fig. 4. Stereoscopic view of the 10 lowest energy structures simulated by NMR restrained molecular dynamics calculations for $d(CCGCGG)_2$ hexanucleotide in solution. RMSD value for backbone heavy atoms of all residues with respect to the average structure is 0.44 Å as calculated by MOLMOL [26].

from B and A-DNA, respectively. Table 1 shows a summary of NMR structural restraints used in the calculations and restraint violations observed in final models. AMBER energy values for final structures are in the range expected for a stable hexanucleotide in solution. Total energies of -118 and $-105\,\mathrm{kcal\,mol^{-1}}$, and van der Waals energies of -95 and $-82\,\mathrm{kcal\,mol^{-1}}$ are observed for structures coming from B and A-DNA conformations, respectively, as reported by DISCOVER. No convergent set of structures was obtained from molecular dynamics calculations starting from Z-DNA standard form. Very high AMBER energy values and van der Waals repulsion terms were observed in all models obtained from molecular dynamics simulations starting from Z-DNA form. Moreover, more than 50% of total constraints were violated, indicating a general disagreement between this model and experimental data. However, some swinging out of the terminal residue C1 was observed with almost null residual violation energy for dihedral angle constraints in that residue (Fig. 5). This suggests some slight possibility for C1 to reach the conformation observed in the crystal structure without an energy barrier larger than the transition of the hexanucleotide to Z-DNA form. Therefore, in conditions where a Z-DNA form could be stable for $d(CCGCGG)_2$ in solution, previously reported crystal tetramer structure could be adopted. However, transition between B and Z-DNA forms is not easily achieved. This fact and the high amount of NOE constraint violations for Z-DNA models make tetramer formation in solution at the conditions studied still very improbable.

NMR data show that $d(CCGCGG)_2$ conformation in solution is a symmetric distorted B-DNA structure with increased flexibility and distortion in terminal residues C1 and G6 (Fig. 4). Final best 10 refined models for $d(CCGCGG)_2$ have been deposited in the PDB database (http://www.pdb.org) (PDB entry 1N1K) with the corresponding structural constraints. Helicoid parameters measured with NDBSTAT on the mean structure obtained from the best 10 B-DNA

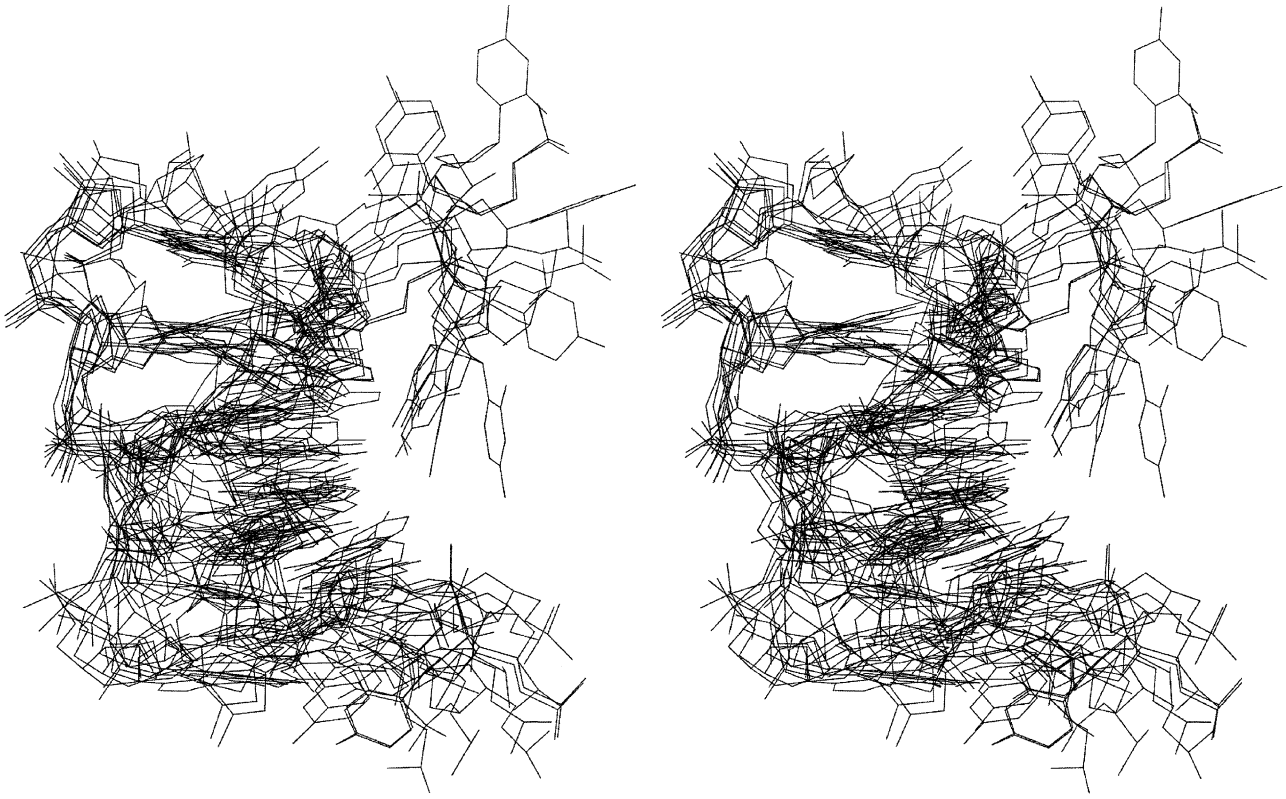

Fig. 5. Stereoscopic view of the 10 lowest energy structures simulated by NMR restrained molecular dynamics calculations for $d(CCGCGG)_2$ starting from Z-DNA conformation. High dispersion in atomic coordinates indicates poor agreement between experimental data with this type of conformation. Swinging out for terminal cytosines, similar to that reported in crystal structure, can be observed at both ends.

models for nonterminal residues are those expected for highly distorted B-DNA structures but are inside the statistical range observed in other reported DNA structural models. One of the most important structural features to be noted in $d(CCGCGG)_2$ solution structure is the high level of helix curvature (92° difference in the local axis orientation from C2 to G5). Increased solvent exposure suggested by ^{31}P NMR chemical shift values is in good agreement with this structural feature. This highly contrasts with the lower than usual curvature reported for this hexanucleotide in the crystal structure. However, the twin DNA molecule forming the tetramer in the crystal may require more parallel relative orientation, and therefore less curved helices for dimer stabilization. $d(CCGCGG)_2$ base-pair per turn value is 10.94 very close to 10.5 of standard B-DNA but typical of slightly stretched oligonucleotides. The high symmetry of the molecule makes a very well-centered set of global and local helical axes. High curvature produces similar sizes for major and minor grooves. Complexes in the minor groove probably are more stable than those in non-distorted DNA conformations. This could explain the minor groove interactions observed for this oligonucleotide in other studies [21].

The present results extend the previous work on DNA sequences containing triple repeats associated to fragile X syndrome. $d(CCGCGG)_2$ does not show any conformational switches under the environmental conditions reported in this work. However, the same conditions have been reported to produce quadruplex DNA formation for other oligonucleotides containing triple repeats associated with fragile X syndrome. Comparison between crystal structure and solution NMR structure seems to indicate that quadruplex formation is possible but that appropriate conditions are different to those observed in other fragile X syndrome oligonucleotide studies to date [15,25]. There is no pH dependence for quadruplex formation for $d(CCGCGG)_2$ in the conditions explored here. Final structure for $d(CCGCGG)_2$ and related observations presented here point to the relevance of environmental conditions for the fragile X syndrome.

Conclusions

We have obtained the tridimensional structure in a solution of $d(CCGCGG)_2$. The duplex is thermodynamically stable in solution at low and high sodium chloride concentrations and low and high pH values.

Significant differences have not been observed in the NMR data recorded at all sample conditions. Solution structure of this hexanucleotide is significantly different from that previously obtained by X-ray diffraction in the crystal state. Moreover, $d(CCGCGG)_2$ does not show the pH dependence observed in other oligonucleotide including triple repeats responsible for fragile X-syndrome. This seems to indicate a different mechanism or different conditions required for fragile X-syndrome effect in d(CCG) than in d(CGG) triple repeats. No experimental data indicating tetramer DNA conformation have been observed. Solution conformation appears to be a highly bent B-DNA form versus the tetramer Z-DNA form reported in the crystal structure. Increased flexibility and non-standard geometry are observed in terminal residues. Local geometry for non-terminal residues shows small differences versus standard B-DNA conformation but accumulation of small local distortions raises unusually large amount of bending for this short oligonucleotide. NMR has been used to elucidate large differences between the structure of an oligonucleotide in solution and crystal states and to evaluate the possible ways of interconversion between both conformations. However, this interconversion would occur in different environmental conditions than those explored in the present work.

Supplementary material

Chemical shift assignments can be retrieved at the BioMagnetic Resonance Databank (http://www.bmrb.wisc.edu, entry code 5562) and as supplementary material. PDB coordinates for $d(CCGCGG)_2$ in solution are available at the Protein Data Bank (http://www.pdb.org, entry code 1N1K). Inputs, shell scripts, and other utilities used for LINSHA and SPHINX simulations as well as J-coupling values tables are available upon request to daniel.monleon@uv.es.

Acknowledgments

We want to acknowledge Prof. E. Sletten for the help provided in 600 MHz measurements. D.M. acknowledges the Generalitat Valenciana for a pre-doctoral fellowship. Bruker España S.A. is acknowledged for the financial support. We finally acknowledge SCSIE of the University of Valencia for access to 300 and 400 MHz NMR spectrometers.

References

[1] J.D. Watson, F.H. Crick, Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid, Nature 171 (1953) 737–738.

[2] L.P.M. Orbons, G.A. van der Marcel, A.H. van Boom, C. Altona, Hairpin and duplex formation of the DNA octamer d(m5C-G-m5C-G-T-G-m5C-G) in solution. An NMR study, Nucleic Acids Res. 14 (1986) 4187–4196.

[3] J. Barciszewski, J. Jurczak, S. Porowski, T. Specht, V.A. Erdmann, The role of water structure in conformational changes of nucleic acids in ambient and high-pressure conditions, Eur. J. Biochem. 260 (1999) 293–307.

[4] J.S. Lee, D.A. Johnson, A.R. Morgan, Complexes formed by $(pyrimidine)_n \cdot (purine)n$ DNAs on lowering the pH are three-stranded, Nucleic Acids Res. 6 (1979) 6659.

[5] S.R. Bhaumik, K.V.R. Chary, G. Govil, K. Liu, H.T. Miles, NMR characterisation of a triple stranded complex formed by homo-purine and homo-pyrimidine DNA strands at 1:1 molar ratio and acidic pH, Nucleic Acids Res. 23 (1995) 4116–4121.

[6] F. Aboul-Ela, A.I.H. Murchie, D.G. Norman, D.M. Lilley, Solution structure of a parallel-stranded tetraplex formed by d(TG4T) in the presence of sodium ions by NMR spectroscopy, J. Mol. Biol. 243 (1994) 458–471.

[7] K. Phillips, Z. Dauter, A.I. Murchie, D.M. Lilley, B. Luisi, The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Å resolution, J. Mol. Biol. 273 (1997) 171–182.

[8] J. Volker, N. Makube, G.E. Plum, H.H. Klump, K.J. Breslauer, Conformational energetics of stable and metastable states formed by DNA triple repeat oligonucleotides: implications for triplet expansion diseases, Proc. Natl. Acad. Sci. USA. 99 (2002) 14700–14705.

[9] K. Wüthrich, NMR of Proteins and Nucleic Acids, John Wiley, New York, NY, 1986.

[10] B. Celda, H. Widmer, W. Leupin, W.J. Chazin, W.A. Denny, K. Wüthrich, Conformational studies of d(AAAAATTTTT)2 using constraints from nuclear Overhauser effects and from quantitative analysis of the cross-peak fine structures in two-dimensional 1H NMR spectra, Biochemistry 28 (1989) 1462–1471.

[11] V. Morell, The puzzle of the triple repeats, Science 260 (1993) 1422–1423.

[12] L. Malinina, L. Urpí, X. Salas, T. Huynh-Dinh, J.A. Subirana, Recombination-like structure of d(CCGCGG), J. Mol. Biol. 243 (1994) 484–493.

[13] M. Vorlickova, J.A. Subirana, J. Chladkova, I. Tejralova, T. Huynh-Dinh, L. Arnold, J. Kypr, Comparison of the solution and crystal conformations of (G + C)-rich fragments of DNA, Biophys. J. 71 (1996) 1530–1538.

[14] L. Urpi, J.P. Ridoux, J. Liquier, N. Verdaguer, I. Fita, J.A. Subirana, F. Iglesias, T. Huynh-Dinh, J. Igolen, E. Taillandier, Conformations in crystals and solutions of d(CA CGTG), d(CCGCGG) and d(GGCGCC) studied by vibrational spectroscopy, Nucleic Acids Res. 17 (1989) 6669–6680.

[15] P.K. Patel, N.S. Bhavesh, R.V. Hosur, Cation-dependent conformational switches in d(TGGCGGC) containing two triplet repeats of fragile X-syndrome: NMR observations, Biochem. Biophys. Res. Commun. 278 (2000) 833–838.

[16] H. Widmer, K. Wüthrich, Simulation of two-dimensional NMR experiments using numerical density matrix calculations, J. Magn. Reson. 70 (1986) 270–279.

[17] V. Esteve, Analysis of geometrical parameters in DNA structures (Análisis de parámetros geométricos en estructuras de ADN), Chemistry degree dissertation, Universitat de València, Valencia, Spain, 1996.

[18] M.A. Hassan, C.R. Calladine, The assesment of the geometry of dinucleotide steps in double-helical DNA: a new local calculation scheme, J. Mol. Biol. 251 (1995) 648–664.

[19] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, G.W. Caldwell, P.A.

Kollman, A second generation force field for the simulation of proteins, nucleic acids and organic molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197.
[20] K.J. McConnell, D.L. Beveridge, Molecular dynamics simulation of B'-DNA: sequence effects on A-tract-induced bending and flexibility, J. Mol. Biol. 314 (2001) 23–40.
[21] D. Monleón, Conformación tridimensional y reconocimiento molecular de biopolímeros, Aplicación de RMN multidimensional y desarrollo de metodología de cálculo y estructural, PhD Thesis dissertation, Universitat de València, Valencia, Spain, 1998.
[22] W.J. Chazin, K. Wüthrich, S. Hyberts, M. Rance, W.A. Denny, W. Leupin, 1H nuclear magnetic resonance assignments for d-(GCATTAATGC)2 using experimental refinements of established procedures, J. Mol. Biol. 190 (1986) 439–453.
[23] J. van Wijk, B.D. Huckriede, J.H. Ippel, C. Altona, in: D.M.J. Lilley, J.E. Dahlberg (Eds.), Methods in Enzymology, DNA Structures 2a: Spectroscopic Methods, Academic Press, Orlando, FL, 1992, pp. 286–306.
[24] J. Ott, F. Eckstein, 31P NMR spectral analysis of the dodecamer d(CGCGAATTCGCG), Biochemistry 24 (1985) 2530–2535.
[25] P.K. Patel, N.S. Bhavesh, R.V. Hosur, NMR observation of a novel C-tetrad in the structure of the SV40 repeat sequence GGGCGG, Biochem. Biophys. Res. Commun. 270 (2000) 967–971.
[26] R. Koradi, M. Billeter, K. Wüthrich, MOLMOL: a program for display and analysis of macromolecular structures, J. Mol. Graph. 14 (1996) 51–55.