Conformations in crystals and solutions of d(CACGTG), d(CCGCGG) and d(GGCGCC) studied by vibrational spectroscopy

L.Urpi, J.P.Ridoux1, J.Liquier1, N.Verdaguer, I.Fita, J.A.Subirana, F.Iglesias2, T.Huynh-Dinh2, J.Igolen2 and E.Taillandier1*

Unitat de Quimica Macromolecular, ETS Enginyers Industrials, Diagonal 647, 08028 Barcelona, Spain, Laboratoire de Spectroscopie Biomoléculaire UFR Santé Médecine et Biologie Humaine, 74 rue Marcel Cachin, 93012 Bobigny Cedex and U. de Chimie Organique, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris Cedex 15, France

Received April 26, 1989; Revised and Accepted July 17, 1989

ABSTRACT
Crystals of self complementary DNA hexamers d(CACGTG), d(CCGCGG) and d(GGCGCC) were grown by vapour diffusion technique and studied by microRaman and microIR spectroscopies. The oligonucleotides were studied in parallel in solution by vibrational spectroscopy. A B->Z transition was detected by Raman spectroscopy during the crystallization process for d(CACGTG). Vibrational spectroscopy shows that the d(GGCGCC) crystals adopt a B geometry. On the contrary the d(CCGCGG) sequence which is shown to be able to undergo in solution or in films quite easily the B->Z transition, remains trapped in crystals in a geometry which may correspond to an intermediate conformation often proposed in models of the B->Z transition. The crystals used in this study were characterized by X-ray diffraction. The unit cell and space group have been determined.

INTRODUCTION
Fiber diffraction studies gave the first evidence of DNA polymorphism due to sequence and environmental conditions. In the last few years single crystal X-ray structures have shown that DNA helices are variable even in the same conformational family and many molecular details are now known, but difficulties persist to obtain information about some sequences due to the problem of crystallizing them (for review see 1-3). The results of X-ray crystal diffraction studies can be correlated with those obtained by other techniques and in particular by vibrational spectroscopy. A major advantage both of Raman and IR spectroscopies is the possibility to study oligo or polynucleotides under a wide variety of physical states: solutions, hydrated films or fibers, crystals. Thus for example the identity of the structures of the d(C-G)6 hexamer in the crystal and of high salt poly d(G-C) in solution was obtained by Raman spectroscopy (4). Several Raman studies of crystal conformations have been already presented (5-9). On the contrary microIR studies of crystallized oligonucleotides have not been published up to now. We report here results obtained by micro IR and micro Raman spectroscopies on three crystals d(CACGTG), d(GGCGCC) and d(CCGCGG). In parallel these oligonucleotides have been studied in solution by vibrational spectroscopy. The crystallized oligonucleotide d(CACGTG) has been shown very recently by X-ray diffraction to adopt a Z geometry (10). We present here the Raman study of this crystal, allowing us to confirm the marker lines of left-handed A-T
base pairs. The results are in excellent agreement with previously proposed marker lines in the case of poly d(A-C).poly d(G-T) or poly d(A-T), but obtained in particular ionic strength conditions, in presence of nickel ions. The comparative study by Raman spectroscopy of the d(CAGGTG) crystal and solution shows that the hexamer is in B form in solution and in Z form in the crystal. The d(GGCGCC) sequence was considered as a good candidate for a potential B geometry in crystalline state. The reverse 5'->3' sequence d(CCGCGG), should theoretically be able to adopt a Z geometry. Base pairs out of purine-pyrimidine alternance were used so as to avoid the formation of concateners. Infrared spectra of d(CCGCGG) solutions and hydrated films show that this oligonucleotide is able to undergo quite easily the B->Z transition either by increase of the ionic strength (solutions) or by decrease of the relative humidity (films). However microRaman and microIR results obtained on the d(CCGCGG) crystal may reflect an intermediate structure between B and Z geometries.

MATERIALS AND METHODS

Synthesis and crystallization

The oligonucleotides d(CAGGTG), d(CCGCGG) and d(GGCGCC) were synthesized by the phosphotriester method in solution from dimers with tris(isopropylsulfonyl)nitrotriazole as the coupling reagent (11). They were purified by Sephadex G-25 gel column chromatography and subsequently by preparative high-performance liquid chromatography on reverse-phase Zorbax OMS 9.3 mm column. Crystals were grown using the vapor diffusion method. The oligonucleotide d(CAGGTG) was crystallized in standard conditions, both in presence and absence of spermine (10). On the other hand the sequences d(CCGCGG) and d(GGCGCC) were crystallized with counterions not previously used in oligonucleotide monocrystals. The counterions employed with d(CCGCGG) were alaninamide, isopropylamide and serinol, while d(GGCGCC) has been crystallized both in standard conditions and with the basic counterions Ac(Arg)NHEt.

Characterization of crystals by X-ray diffraction

Crystals of the sequence d(CAGGTG) had two different cell dimensions depending on the presence or absence of spermine in the crystallization solution. Both types of crystals examined by still and precession X-ray diffraction photographs were found to have the same orthorhombic space group P2_12_1. Cell dimensions were: a = 16.1, b = 29.7, c = 41.6 Å with spermine and a = 17.6, b = 31.1, c = 44.4 Å without spermine (10). All the d(CCGCGG) crystals obtained had a similar morphology, independently of the counterion present. They were tetragonal plates, always smaller than 0.2 x 0.2 x 0.05 mm. The precession photographs clearly indicated a tetragonal system with a P422 space group. However the c axis cell dimension showed two distinct values for different crystals grown under the same conditions. Complete data sets to 2.5 Å resolution were collected for the two kinds of crystals using a FAST area detector with synchrotron radiation at Daresbury. The unit cell parameters determined were a = b = 40.9, c = 35.6 Å and a = b =
Figure 1: Raman spectra of d(CACGTG) a: Crystal; b: Solution; marker lines characteristic of B form and of Z form.

40.9, c = 31.2 Å respectively, therefore confirming the presence of two subtly different crystal forms. Crystals of the sequence d(GGCGCC) are thin plates. The precession photographs showed two 2-fold axis perpendicular to each other. The cell dimensions were a = 46, b = 36, c = 110 Å, one of the longest axis found for oligonucleotides. Precession photographs showed systematic absences for hk0 when h + k = 2n + 1 and also for 001 when l = 2n + 1, thus the lattice must be centred (C) and the 2-fold axis parallel to the c axis must be a 2-fold screw axis, clearly indicating the space group C222. No sign of stacking was seen in the precession photographs. A complete data set to 2.3 Å resolution was collected in a CAD4-Enraf Nonius with Cu Kα radiation. The accurate cell dimensions were a = 46.15 (1), b = 36.90 (2), c = 110.03 (3) Å and the space group C222, was confirmed. From packing considerations three hexamers or eighteen base pairs should be placed in the asymmetric unit, approximately aligned along the c crystal axis though no clear relationship between neighbouring hexamers can be inferred at the present stage.

Raman spectroscopy

Samples were exposed to the 514.5 nm line from a Spectra-Physics model 2025 argon laser. The output power used was 400 mW at the source. Raman spectra were recorded with a Dilor Omars 89 multichannel spectrophotometer coupled to an IBM AT3 computer. The microscope used was an Olympus BH-2 model.
with a 100 X objective. About 20mW was focused down to an area of 2-4 microns in diameter. Integration time was usually 10 s. Each spectrum is an average of about 250 integrations. Solvent background correction was performed by subtracting the solvent spectrum recorded in the same conditions.

**FTIR spectroscopy**

Infrared spectra were recorded using a Perkin Elmer 1750 spectrophotometer coupled to a PE 7700 minicomputer. Solution spectra were obtained by depositing a droplet of the oligomer solution between two ZnSe windows. DNA concentrations were around 2 OD /µl. Solution spectra were run at two different ionic strengths, 1M and 0.4M NaCl. Film spectra and relative humidity monitoring (H2O or D2O) were obtained as previously explained (12). Crystal transmission spectra were recorded using a Spectra-Tech. microscope attachment with mirror focusing. In this latter case an MCT detector was used. Data treatment included subtraction of the solvent spectrum, base line correction, smoothing following the Savitsky and Golay procedure (usually on 13 points).

**RESULTS AND DISCUSSION**

*d(CACGTG)*

In figure 1 are shown the Raman spectra of d(CACGTG) in solution (1b) and in crystal (1a). The spectra are quite obviously very different reflecting two geometries for the oligonucleotide depending on its physical state. In the crystal the structure has been shown by X-ray diffraction studies to belong to the Z family (10). The Raman spectrum (1a) contains the marker lines both of AT and GC base pairs in a left-handed geometry, which we have previously characterized in the case of polynucleotides (13,14). We present in figure 2 with an expanded scale the Raman spectra of the left-handed geometries of poly d(G-C) (2a) and poly d(A-T) (2d) in two regions containing marker lines, i.e. in the base fingerprint region between 1400 and 1150 cm⁻¹ (left) and in the low wavenumber region between 850 and 600 cm⁻¹ (right). The spectrum of the d(CACGTG) crystal (2c) can be satisfactorily compared either with the experimental spectrum of Z form poly d(A-C).poly d(G-T) (13) or with a simulated spectrum obtained by a linear combination of Z form spectra of poly d(G-C) (2a) and poly d(A-T) (2d) taking into account the base composition of the oligonucleotide i.e. 2 GC base pairs per 1 AT base pair. The computed spectrum is presented in the same spectral regions in Figure 2b. The experimental spectrum of the d(CACGTG) crystal has the features of the Z geometries of both AT and GC base pairs. The line observed at 1360 cm⁻¹ can be assigned to contributions of guanosines and adenosines in a syn geometry, which have Raman lines respectively at 1354 and 1362 cm⁻¹ in the Z polymers poly d(G-C) and poly d(A-T). The strong line observed at 1312 cm⁻¹ is due to the overlap of the contributions of guanosines (strong line at 1314 cm⁻¹ in poly d(G-C)) and adenosines (lines located at 1332 and 1314 cm⁻¹ in poly d(A-T)). Similarly the strong line observed at 1264 cm⁻¹ in the crystal spectrum is due to a cytosine vibration detected at the same position in the case of Z poly d(G-C)(6).
Figure 2: Raman spectra in two conformation sensitive spectral regions; left: 1400-1200 cm\(^{-1}\); right: 850-600 cm\(^{-1}\) of a) poly d(G-C) Z form (solution); b) computer spectrum 2a + Id; c) d(CAGCTG) crystal; d) poly d(A-T) Z form (solution); bases backbone vibrations

In the low wavenumber spectral region (between 850 and 600 cm\(^{-1}\)) the agreement between the experimental d(CAGCTG) crystal spectrum (2c right) and the computed one (2b right) is also excellent. The experimental spectrum contains the main Z form marker line located at 621 cm\(^{-1}\) involving a purine (guanine as well as adenine) breathing motion coupled through the glycosidic bond to a deoxyribose vibration. This mode is observed at 621 cm\(^{-1}\) when the purine nucleotides are in a C3'-endo/\(\gamma\)-syn geometry instead of 682 cm\(^{-1}\) for guanosines and 669 cm\(^{-1}\) for adenosines in C2'-endo/anti geometry (B form). The line observed at 674 cm\(^{-1}\) in the crystal may be assigned to a thymidine vibration (669 cm\(^{-1}\) in poly d(A-T) Z form). The Z backbone modes in the crystal are observed at 746 and 807 cm\(^{-1}\) at positions comparable to those of the same modes in the Z form polymers (748 and 810 cm\(^{-1}\) for poly d(G-C), 746 and 815 cm\(^{-1}\) for poly d(A-T)). We must notice that this crystal is a relatively rare example of an oligonucleotide containing A-T base pairs without any chemical modifications such as methylation or bromination, which is found in a Z geometry by X-ray diffraction (10). The Raman markers of Z
<table>
<thead>
<tr>
<th>Poly d(G-C)</th>
<th>Poly d(A-C)</th>
<th>Poly d(A-T)</th>
<th>dCACGTG solution</th>
<th>Assignments</th>
<th>Poly d(G-C)</th>
<th>Poly d(A-C)</th>
<th>Poly d(A-T)</th>
<th>dCACGTG crystal</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1363</td>
<td>1374</td>
<td>1374</td>
<td>1367</td>
<td>dA anti</td>
<td>1354</td>
<td>1354</td>
<td>1362</td>
<td>1360</td>
<td>dA, dG syn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>dG anti</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>dG syn</td>
</tr>
<tr>
<td>830</td>
<td>832</td>
<td>832</td>
<td>830</td>
<td></td>
<td>810</td>
<td>804</td>
<td>815</td>
<td>807</td>
<td></td>
</tr>
<tr>
<td>787</td>
<td>790</td>
<td>786</td>
<td></td>
<td>ϕ Backbone</td>
<td>748</td>
<td>742</td>
<td>746</td>
<td></td>
<td>ϕ Backbone</td>
</tr>
<tr>
<td>682</td>
<td>682</td>
<td>727</td>
<td>737</td>
<td>dA anti</td>
<td>625</td>
<td>622</td>
<td>622</td>
<td></td>
<td>dG, dA syn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>dG anti</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Poly d(G-C) from ref. (21), poly d(A-C), poly d(G-T) ref. (13), Poly d(A-T) ref. (14)
Figure 3: MicroRaman spectra of crystals a) d(GGC9CC) + Ac(Arg)NHEt B form \///; b) d(GGC9CC). B form \///; c) d(CCGCGG) + alaminamide intermediate form ///; d) d(CCGCGG) + serinol intermediate form ///

form for A-T base pairs which we have detected here on the d(CACGTG) crystal spectrum confirm our previous assignments concerning the left-handed markers of A-T base pairs obtained in the case of polymers (poly d(A-C), poly d(G-T) and poly d(A-T)) but in particular experimental conditions i.e. in the presence of divalent metal ions (nickel) (13,14). The addition of these ions, which have been shown to stabilize the syn geometry of purines by interacting on the H7 site of the bases, and therefore the Z form of the sequences (15), does not modify the position of the Z form marker lines of A-T base pairs which we detect here for d(CACGTG) at the same positions without the presence of the metal ions.

The spectrum of d(CACGTG) obtained in solution (Fig. 1b) is quite different from the previously discussed one and clearly reflects another geometry. The Raman marker lines of a right-handed B form can be easily detected, as previously pointed out in the case of polynucleotides, oligonucleotides and native DNAs (16-19). In particular we observe a strong line at 682 cm⁻¹ reflecting the existence of guanosines in C2'-endo/anti.
Figure 4: FTIR spectra of d(CCGGG) solutions: top: 0.4 M NaCl B form; bottom: 1 M NaCl Z form; marker bands of B form — marker bands of Z form

goometry, with a shoulder to the lower wavenumbers (669 cm\(^{-1}\)) due to the thymidine contribution. The backbone vibrations are reflected by two lines located at 830 and 786 cm\(^{-1}\), the latter being superimposed on peaks due to cytosine as well as thymidine vibrations. In the base fingerprint region we detect the C2'-endo/anti guanosine vibration line at 1367 cm\(^{-1}\). All these results are summarized in table 1.

Thus the d(CACGTG) sequence crystallizes in a Z form while the B geometry is the conformation in solution. This type of effect had been first observed in the case of the d(C-G)\(_6\) crystal which was obtained from a solution in which the hexamer was in B form (20). The structural change under the crystallization process has been relatively frequently observed, either reflected by a B->Z transition or by a B->A transition. However this is not systematically the case as will be shown by the study of the following oligonucleotide.

d(GCGGCC)

The sequence d(GCGGCC) has been studied by microRaman spectroscopy. It has been reported that sequences with initial purines and terminating pyrimidines will not go into the Z form (7). Thus this sequence was expected to be a good candidate for a still rare geometry of crystallized
oligonucleotides: B form. Figure 3 shows the Raman spectra of d(GGCCGCC) crystallized in standard conditions (3b) and in presence of ActArg_{2}NHHet (3a). Both spectra are very similar and reflect a B form of the oligonucleotide: line at 835 cm\(^{-1}\) due to the antisymmetrical phosphate stretching vibration in B geometry, strong marker line at 684 cm\(^{-1}\) involving the guanosine breathing motion in a C2'-endo/anti geometry, characteristic profile in the 1200-1400 cm\(^{-1}\) with positions and relative intensities of the lines identical to those previously published for B form d(C-G)n sequences (21).

d(CCGCGG)

Figure 4 shows the IR spectra of d(CCGCGG) solutions in 0.4 and 1M NaCl. They can be assigned using the marker IR bands of the different helices previously determined for poly d(G-C) (22) or d(CCGCGG) (23). The characteristic spectroscopic features of a B geometry are detected at low ionic strength and of a Z geometry at high ionic strength. The main spectral modifications observed under the B->Z transition of d(CCGCGG) are: (I) shift of a band located at 1420 cm\(^{-1}\) assigned to a C2'-endo deoxyribose vibration to 1410 cm\(^{-1}\) reflecting a C3'-endo sugar pucker; (II) shift of the band located at 1372 cm\(^{-1}\) assigned to a ring vibration of the guanine coupled to a deoxyribose motion via the glycosidic linkage to 1354 cm\(^{-1}\), indicative of the anti->syn reorientation of the purine nucleotides; (III) presence of a new strong band at 1320 cm\(^{-1}\) assigned to a syn geometry of the guanosine residue; (IV) presence of a band at 1264 cm\(^{-1}\) (shoulder on the strong phosphate band) also due to a guanosine vibration; (V) shift of the phosphate antisymmetric stretching vibration to 1213 cm\(^{-1}\); (VI) drastic decrease of the relative intensity of the absorption located at 1086 cm\(^{-1}\) assigned to the symmetric stretching vibration of the phosphate groups; (VII) enhancement of the relative intensities of the 1061 and 1020 cm\(^{-1}\) bands; (VIII) presence of a band at 924 cm\(^{-1}\) characteristic of left-handed helices and (IX) decrease of the relative intensity of the band at 895cm\(^{-1}\) characteristic of right-handed helices.

The B->Z transition of this oligonucleotide can also be induced in films by varying the water content of the sample. The figure 5 shows the IR absorption spectra of d(CCGCGG) films exposed to solutions of saturated salts in D_{2}O corresponding to high (98%) and low (58%) relative humidities. The results obtained on films exposed to H_{2}O saturated salts are identical to those obtained for solutions and the corresponding spectra will not be presented here. The decrease of RD induces in the 1750-1550 cm\(^{-1}\) region shifts of the two bands located at 1681 and 1649 cm\(^{-1}\) to lower wavenumbers (respectively 1666 and 1635 cm\(^{-1}\)) similarly to what had been previously observed for poly d(G-C) (12), characteristic of a B->Z transition for the sequence. In the low wavenumber region between 1000 and 750 cm\(^{-1}\) a decrease of the relative intensity of the 892 cm\(^{-1}\) band and the emergence of a strong band at 927 cm\(^{-1}\) reflect the B->Z transition induced by the decrease of the RD. All these modifications show that the d(CCGCGG) sequence is able...
Figure 5: FTIR spectra of d(CCGCGG) films exposed to D$_2$O saturated solutions. Left between 1750 and 1550 cm$^{-1}$. Right between 1000 and 750 cm$^{-1}$. Top: high RH, B form. Bottom: low RH, Z form. Bases backbone vibrations.

to undergo a right->left handed helical transition in solution as well as in films.

We have crystallized this oligonucleotide in presence of various counterions and the crystals have been studied by Raman and FTIR microspectroscopies. The crystals are not in a Z geometry. The figure 3 shows

Figure 6: Micro FTIR spectrum of d(CCGCGG) crystal + isoproplamine intermediate geometry between B and Z forms. Z genus B genus
the Raman spectra of the crystals obtained in presence of alaninamide (3c) and serinol (3d). The absence of a line around 625 cm⁻¹ characteristic of guanosines in a C3'-endo/syn geometry and in contrast the existence of an intense line at 684 cm⁻¹ reflecting a C2'-endo/anti conformation of the purine nucleosides clearly shows that the structure does not belong to the left-handed Z form family. These spectra have mainly B form family characteristics; however several points show that the structure which is studied does not belong to a classical B geometry. Thus we observe a decrease of the relative intensity of the phosphate line located at 835 cm⁻¹ in B geometry, an enhancement of a line at 644 cm⁻¹ and a very intense doublet at 1245-1271 cm⁻¹. These three lines have been assigned to cytosine vibrations (7,16) and their intensities increase when destacking of the bases occurs (24). Such strong intensities of the 1245-1271 cm⁻¹ lines have been observed for a crystal of d(CpG) where base stacking should not be as important as in the case of longer oligonucleotides (7).

The infrared spectrum of a d(CCGCGG) crystal with isopropylamine (Figure 6) is different from the previously discussed B form spectrum recorded in solution or in films. Several absorptions involving the deoxyribose are affected. In particular we detect a strong band at 1065 cm⁻¹, an important enhancement of the deoxyribose contribution at 1010 cm⁻¹, a new band at 946 cm⁻¹ and a shift of the band located at 1420 cm⁻¹ to 1413 cm⁻¹ which would indicate a C3'-endo pucker. A decrease of the 894 cm⁻¹ band relative intensity is also detected. The comparison of the crystal spectrum (figure 6) and of the B and Z form solution spectra (figure 4) shows that the guanosines remain in an anti geometry (band at 1378 cm⁻¹, no bands at 1320 and 1264 cm⁻¹).

It may thus be proposed that the microRaman and microIR spectra of d(CCGCGG) crystals obtained in presence of these different cations reflect a modified B geometry, may be an intermediate step in the B→Z transition. Such an intermediate step in the case of d(C-G)ₙ sequences has been proposed after CD experiments. In ethanolic solutions, induction of the B→Z transition by increase of temperature allows to isolate an intermediate form characterized by its particular CD spectrum (25). Similarly a B' form induced by increase of the ionic strength has been interpreted in terms of unfolding or premelting of the structure (26). Intermediates in the B→Z transition have also been detected by Raman spectroscopy. Spectra recorded at regular time intervals after addition of salt show the emergence of a Raman line around 640 cm⁻¹ prior to that of the characteristic Z form marker at 625 cm⁻¹ (21). The spectrum recorded in 1.5 M NaCl does not present the 625 cm⁻¹ line but clearly the existence of a line at 640 cm⁻¹ (27).

The d(CCGCGG) sequence which is in B form in solution at low ionic strength was expected to adopt a Z form in crystal, as the B→Z transition of this oligonucleotide is easily induced either by increase of the ionic strength in solution or by decrease of the hydration in films. Rather surprising is the result that d(CCGCGG) crystals are detected by micro IR and micro Raman neither in a classical B form nor in Z geometry. It seems thus
that the expected B→Z transition which should have occurred by crystallization has possibly been stopped by the presence of the counterions used in the crystallization process: isoproplamine, alaninamide or serinol.

*To whom correspondence should be addressed

REFERENCES