Determination of G*G·C Triple-Helix Structure by Molecular Modeling and Vibrational Spectroscopy

Mohammed Ouali,[†] Richard Letellier,[†] Jian-Sheng Sun,[‡] Alain Akhebat,[†] Frédéric Adnet,[†] Jean Liquier,[†] and Eliane Taillandier*,[†]

Contribution from the Laboratoire CSSB-URA CNRS 1430, UFR Santé-Médecine-Biologie Humaine, Université Paris XIII, 74 Rue Marcel Cachin, 93012 Bobigny, France, and Laboratoire de Biophysique, Muséum National d'Histoire Naturelle, INSERM U 201-URA CNRS 481, 43 rue Cuvier, 75231 Paris Cédex 05, France

Received August 17, 1992

Abstract: A structural study has been performed on dG_n*dG_n·dC_n and rG_n*dG_n·dC_n triple helices by molecular modeling, Fourier transform infrared spectroscopy, and Raman spectroscopy. Triple-helix geometries have been obtained by molecular mechanics calculations. Stereochemically reasonable structures in agreement with vibrational spectroscopy data and models of third-strand base pairing and third-strand orientation can be proposed from these simulations. Our results show the existence of both S- and N-type sugars and of an anti glycosidic torsion angle conformation. The dC_n strand has S-type sugars, the dG_n strand of the Watson–Crick duplex N-type sugars, and the third strand either N-type sugars (rG_n strand) or S-type sugars (dG_n strand). The orientation of the third strand is parallel with respect to the homopurine strand involved in Hoogsteen hydrogen bonding.

Introduction

Triple-helix formation can occur through the binding of a third nucleic acid strand in the major groove of a double-helice DNA. The formation of a triple helix between two strands of dG_n and one strand of dC_n was described initially by Lipsett[1] and previously studied by CD spectroscopy.[2] More recently possible hydrogen-bonding schemes have been proposed for the G*G·C triplets formed with homopolymers[3] or in supercoiled plasmids[4] (the * notation concerns the third strand and means that the third G strand is bound to the purine strand of the G·C Watson–Crick duplex). Two different positions of the third strand have been suggested corresponding respectively to the so-called Hoogsteen[3] or reverse-Hoogsteen[4] hydrogen-bonding pattern, by analogy with the model proposed by Hoogsteen for hydrogen bonding between T and A. In addition, the orientation of the third strand, either parallel or antiparallel with respect to the purine Watson–Crick strand, has been correlated with the possible hydrogen-bonding patterns in the triplex formation at G·C-rich sites of natural DNA.

The formation of a triple helix, in which the third strand consisting of 27 purine-rich bases was designed parallel to the purine strand of a sequence in the promoter region of the c-myc gene, was shown to inhibit c-myc expression in vitro.[5] However, triplex formation has also been described with an antiparallel third strand orientation for natural DNA targets in the promoter regions of human c-myc, the epidermal growth factor receptor genes, and the mouse insulin receptor genes.[6] In addition it has been reported by several authors that the third strand should be antiparallel to the purine-rich strand of the Watson–Crick duplex in different intramolecular (Pur)*(Pur)·(Pyr) triplexes formed upon folding of a single-stranded oligomer[7,8] as well as in intermolecular triple helices formed between three distinct oligonucleotides.[9-12] Molecular modeling studies and footprinting experiments have indicated that the third-strand orientation might depend on the sequence according to the number of ApG and GpA steps in the homopurine sequence.[13] Finally in the case of a target duplex which contains mixed purine–pyrimidine sequences, in order to maintain hydrogen-bonding interactions between the third-strand bases and the purine residues of the duplex, the oligonucleotide may have to alternate its binding between the duplex strands. In such "strand-switching" models, both parallel and antiparallel G*G·C triplets could be involved in the triple-helical structure.[12,14]

In the present work, we have studied by molecular mechanics calculations and vibrational spectroscopy the most simple structures of (Pur)*(Pur)·(Pyr) triple helices: dG_n*dG_n·dC_n and rG_n*dG_n·dC_n. The sugar conformation of the energy-minimized structures which have the lowest total energy has been confirmed by FTIR and Raman spectroscopy experiments. Infrared marker bands and Raman marker peaks are detected for both N-(C3′-endo) and S-(C2′-endo) type sugars. The spectra show that the dC_n strand contains S-type sugars and the dG_n strand of the Watson–Crick duplex N-type sugars, while the third strand has either N-type sugars (rG_n strand) or S-type sugars (dG_n strand). All the guanine residues of both triple helices have been detected in an anti conformation. The obtention of energy-minimized structures which are stereochemically reasonable and in agreement with the vibrational data allows us to propose models for the rG_n*dG_n·dC_n and dG_n*dG_n·dC_n triple helices, in which the third strand is Hoogsteen hydrogen bonded to the Watson–Crick duplex and its orientation is parallel with respect to the homopurine strand of the dG_n·dC_n duplex.

* Author to whom all correspondence should be addressed.
† Université Paris XIII.
‡ Muséum National d'Histoire Naturelle.

(1) Lipsett, M. N. *J. Biol. Chem.* **1964**, *239*, 1256–1260.
(2) Marck, C.; Thiele, D. *Nucl. Acids Res.* **1978**, *5*, 1017–1028.
(3) Letai, A. G.; Palladino, M. A.; Fromm, E.; Rizzo, V.; Fresco, J. R. *Biochemistry* **1988**, *27*, 9108–9112.
(4) Kohwi, Y.; Kohwi-Shigematsu, T. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 3781–3785.
(5) Cooney, M.; Czernuszerwicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, *241*, 456–459.
(6) Durland, R. H.; Kessler, D. J.; Gunnell, S.; Duvic, M.; Pettitt, B. M.; Hogan, M. E. *Biochemistry* **1991**, *30*, 9246–9255.
(7) Chen, F. M. *Biochemistry* **1991**, *30*, 4472–4479.
(8) Jayasena, S. D.; Johnston, B. H. *Biochemistry* **1992**, *31*, 320–327.
(9) Pilch, D. S.; Levenson, C.; Shafer, R. H. *Biochemistry* **1991**, *30*, 6081–6087.
(10) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360–1363.
(11) Blume, S. W.; Gee, J. E.; Shresta, K.; Miller, D. M. *Nucl. Acids Res.* **1992**, *20*, 1777–1784.
(12) Mc Shan, W. M.; Rosen, R. D.; Laughter, A. H.; Trial, J. A.; Kessler, D. J.; Zendegui, J. G.; Hogan, M. E.; Orson, F. M. *J. Biol. Chem.* **1992**, *267*, 5712–5721.
(13) Sun, J. S.; De Bizemont, T.; Duval-Valentin, G.; Montenay-Garestier, T.; Hélène, C. *C. R. Acad. Sci. Paris* **1991**, *313*, Série III, 585–590.
(14) Horne, D. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 2435–2437.

0002-7863/93/1515-4264$04.00/0

Calculations

Energy minimization calculations of $dG_n*dG_n{\cdot}dC_n$ and $rG_n*dG_n{\cdot}dC_n$ triple helices were performed using the JUMNA IV (Junction Minimization of Nucleic Acid) computer program developed by R. Lavery.[15] This approach is fully adapted for nucleic acids and allows optimization of nucleic acid segments using helicoidal coordinates. Solvent water and counterions were not explicitly included in the calculation, and in order to mimic their effects, the net charge on each phosphate group was set to −0.5 e and a sigmoidal distance-dependent dielectric function was used.[16] Calculations were carried out on a Fujitsu VP 200 computer. The JUMNA IV computer program has been vectorized and adapted to VP 200 by D. Giroud from CIRCE (Centre Inter Regional de Calcul Electronique, CNRS, Orsay, France). It now runs 20 times faster when the array processor is activated.

Energy-minimized structures were analyzed with the CURVES computer program,[17] which allows calculation of helicoidal parameters for nucleic acids with respect to an optimized global helical axis obtained by a least square fit. The molecular structures were visualized with the interactive computer graphics program MAD (Lahana, R., Copyright Laboratoire Pierre Fabre, Castres, France). Initial geometries were model-built from a library of nucleotides included in the JUMNA IV package by specifying the helicoidal parameters according to the Cambridge convention.[18] Triple helices $dG_n*dG_n{\cdot}dC_n$ and $rG_n*dG_n{\cdot}dC_n$ were constructed with twelve base triplets ($n = 12$). This choice was made to simulate experimental samples where the third strand was twelve nucleotides in length.

Starting geometries for triple helices $dG_n*dG_n{\cdot}dC_n$ were generated using the following procedure: Xdisp, Ydisp, Rise, Inc, Tip, and Twist,[16-18] which characterize the helix parameters and the base positions with respect to the helix axis, were derived from crystallographic data obtained by Arnott et al.[19]

As suggested by previous authors,[3,4,6,10] models corresponding to Hoogsteen and reverse-Hoogsteen third-strand hydrogen bonding were considered (Figure 1). To this end the parameters (Xdisp, Ydisp, Inc, Tip, Twist) were varied from the values given by Arnott et al.[19] and fitted empirically. Third-strand orientation has been taken either parallel or antiparallel to the purine strand of the Watson–Crick duplex $dG_n{\cdot}dC_n$. Assumptions on base-pairing models and third-strand orientations led us to consider four kinds of initial triple-helix structures, and in each case, models were constructed with the sugar-backbone conformation parameters derived from the B-form double helix[20] or the A-form triple helix[19] (Table I). Structures where the three strands have the same initial conformation and structures in which the Watson–Crick duplex and the third strand are in two different conformations have been examined.

In early stages of minimization, the helicoidal variables were locked in order to maintain triple helicity while the sugar–phosphate-backbone parameters were free to change. Next, minimizations were carried out by successively relaxing the constraints. In order to avoid end effects, a mononucleotide symmetry with respect to all helicoidal and internal parameters of each strand was used.

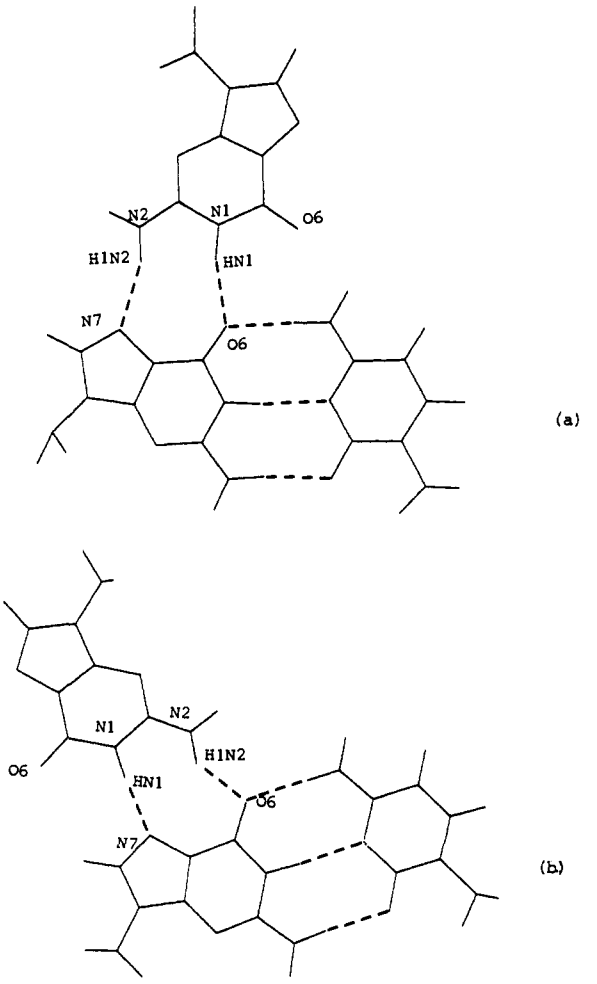

Figure 1. Models of G*G·C triplets: (a) Hoogsteen hydrogen-bond pattern; (b) Reverse-Hoogsteen hydrogen-bond pattern.

The secondary structures of the $dG_n*dG_n{\cdot}dC_n$ triple-helix energy-minimized geometries were compared to the sugar conformation information obtained by FTIR and Raman spectroscopies. For the calculated $dG_n*dG_n{\cdot}dC_n$ triple helix geometry which best fits with vibrational spectroscopy data, the deoxyribose sugar third strand was substituted by a ribose sugar strand in our theoretical model. The perturbations induced by a ribose-strand replacement in the triple-helix energy-minimized structure were compared to the spectral changes observed in IR and Raman spectra upon substitution of the third dG_n strand by a rG_n strand.

The stability of $dG_n*dG_n{\cdot}dC_n$ and $rG_n*dG_n{\cdot}dC_n$ triple-helix refined structures was verified by imposing stretching–compression and twisting–untwisting deformations.[21] The stretching–compression and twisting–untwisting constraints were applied by progressive variations of rise or twist parameters. This allows us to explore the energy potential surface around a local minimum and to obtain a smooth quadratic energy curve in response to these constraints, ensuring the stability of the local minimum-energy conformation.

Materials and Methods

Polynucleotides, $dG_n{\cdot}dC_n$ (lot AA7890P12) and rG_n (lot AA4257P07), and the oligomer dG_{12-18} (lot AG7785202) were purchased from Pharmacia and used without purification. Concentrations were estimated by UV absorbance measurements using a Beckman 3600 spectrophotometer and assuming ϵ_{max} around 260 nm: $dG_n{\cdot}dC_n$, 7400; dG_{12-18}, 9700; and rG_n, 10400.

Triple-stranded structures $dG_n*dG_n{\cdot}dC_n$ and $rG_n*dG_n{\cdot}dC_n$ were obtained by equimolar mixing of double-stranded $dG_n{\cdot}dC_n$ with the

(15) Lavery, R. *Structure and Expression Volume 3 DNA bending and curvature*; Olson, W. K., Sarma, M. H., Sarma, R. H., Sundaralingam, M., Eds.; Adenine Press: Guilderland, NY, 1988, 191–211.
(16) Lavery, R.; Sklenar, H.; Zakrzewska, K.; Pullman, B. *J. Biomol. Struct. Dyn.* **1986**, *3*, 929–1014.
(17) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 655–667.
(18) Dickerson, R. E.; Bansal, M.; Calladine, C. R.; Dieckman, S.; Hunter, N. N.; Kennard, O.; Lavery, R.; Nelson, H. C. M.; Olson, W. K.; Saenger, W.; Shakked, Z. H. S.; Soumpasis, D. M.; Thung, O. S.; Von Kitzing, E.; Wang, A. H. J.; Zhurkin, V. B. *J. Biomol. Struct. Dyn.* **1989**, *6*, 627–634.
(19) Arnott, S.; Bond, P. J.; Selsing, E. J.; Smith, P. J. C. *Nucl. Acids Res.* **1976**, *3*, 2459–2470.
(20) Arnott, S.; Chandrasekharan, R.; Birdsall, D. L.; Leslie, A. G. W.; Ratliff, R. L. *Nature* **1980**, *283*, 743–746.
(21) Sun, J. S.; Mergny, J. L.; Lavery, R.; Montenay-Garestier, T.; Helene, C. *J. Biomol. Struct. Dyn.* **1991**, *9*, 411–424.

Table I. Molecular Mechanics Total Energy Values (kcal/mol) of the $dG_n*dG_n{\cdot}dC_n$ Triple Helix

conf no.	starting geometry	third-strand base pairing	third-strand orientation	tot energy	refined-struct sugar conf[a]
I	N/anti	Hoogsteen	parallel	−1185.06	S/anti
	N/anti				N/anti
	N/anti				S/anti
II	S/anti	Hoogsteen	parallel	−1154.14	S/anti
	S/anti				S/anti
	S/anti				S/anti
III	N/anti	Hoogsteen	antiparallel	−1038.37	S/anti
	N/anti				S/anti
	N/anti				N/syn
IV	S/anti	Hoogsteen	antiparallel	−1047.41	S/anti
	S/anti				S/anti
	S/anti				N/syn
V	N/anti	reverse Hoogsteen	parallel	−913.40	S/anti
	N/anti				S/anti
	N/anti				E/syn
VI	S/anti	reverse Hoogsteen	parallel	−1001.79	S/anti
	S/anti				S/anti
	S/anti				E/syn
VII	N/anti	reverse Hoogsteen	antiparallel	−1149.13	S/anti
	N/anti				S/anti
	N/anti				S/anti
VIII	S/anti	reverse Hoogsteen	antiparallel	−1132.33	S/anti
	S/anti				S/anti
	S/anti				S/anti

[a] The last column on the right species the sugar pucker and glycosidic torsion angle of (top) the dC strand; (middle) the dG strand; and (bottom) the dG strand where the second G is the third non-Watson–Crick strand. The S (south), N (north), and E (east) conformations correspond to C2′-endo, C3′-endo, and O1′-endo, respectively.

required single strand (dG_{12-18} or rG_n) heated at 70 °C for 10 min prior to addition so as to avoid aggregation of the third strand. Magnesium chloride was then added so as to obtain a final Mg^{2+} per base ratio of 1 in the studied samples. The formation of the triplexes has been followed by infrared spectroscopy (study of the stretching vibrations of the base double bonds in their plane, discussed in detail in the Results section, FTIR Spectroscopy).

For FTIR measurements samples were dissolved either in H_2O or D_2O. Solution spectra were obtained in ZnSe cells, at a concentration around 100 mM base. FTIR spectra were run using a Perkin Elmer 1760 Fourier transform spectrophotometer coupled to a PE7700 computer. Usually 10 scans were performed. Data transferred to an IBM PS2 computer were treated with the Galaxy Spectra Calc. Program. This treatment included multiple base-line correction, smoothing using the Savitsky and Golay procedure, and spectral normalization using the phosphate symmetric stretching vibration found around 1090 cm^{-1} as internal standard.

Raman spectra were recorded using a DILOR OMARS 90 multichannel spectrophotometer coupled to an IBM AT3 computer. The 514-nm line of a Spectra-Physics 2025 ionized-argon laser was used for excitation. The power at the source was 400 mW. Integration time was usually 10–15 s, and an average of 250 spectra were accumulated. The data treatment involved only base-line correction.

Results

Stereochemically Valid Models for $dG_n*dG_n{\cdot}dC_n$ Triple-Helix Geometry. The energy minimization calculations on $dG_n*dG_n{\cdot}dC_n$ triple helices were performed using A-form[19] and B-form[20] DNA backbone parameters as starting geometries. In order to take into account remarks of previous authors on triple-helix structures, we have constructed our initial models with Hoogsteen and reverse-Hoogsteen base pairing. The third-strand orientation has been taken either parallel or antiparallel to the Watson–Crick purine strand (see Calculations section).

The energy-refined $dG_n*dG_n{\cdot}dC_n$ triple-helix structure which gives the lowest total conformational energy corresponds to a dG_n third-strand Hoogsteen base-paired structure, and its orientation is parallel to the dG_n strand of the Watson–Crick duplex $dG_n{\cdot}dC_n$ (conformation I in Table I). This structure has been obtained starting from an N-type sugar conformation for the three strands (Arnott A-form triple helix). The optimized geometry shows an S-type/anti conformation for the dC_n strand, an N-type/anti conformation for the Watson–Crick dG_n strand, and an S-type/anti conformation for the Hoogsteen dG_n strand (Tables I and II). The helicoidal parameters (noted respectively Inc and Tip in Table II) used to characterize the base inclination with respect to the helical axis[15-17] keep small values for the guanine third strand. The three bases involved in a G*G·C triplet are nearly situated in the same plane.

Table II. Helicoidal Parameters and Conformational Angles of the $dG_n*dG_n{\cdot}dC_n$ Optimized Geometry Triplex[a]

strand	Xdisp	Ydisp	Rise	Inc	Tip	Twist
I	−2.94	0.35	3.26	4.45	2.28	33.75
II	−2.97	−0.25	3.26	1.19	−11.16	33.75
III	−0.65	1.61	3.26	−7.92	−8.09	33.75

strand	χ	α	β	γ	δ	ε
I	246	−71	−178	56	142	−172
II	203	−77	176	65	83	−160
III	251	−61	172	57	143	−170

[a] The helicoidal parameters Xdisp, Ydisp, Inc, and Tip give the position of the base in each strand with respect to a reference point defined as the intersection where the helical axis would cut the base plane in the standard B conformation. The parameters Xdisp and Ydisp are respectively the translations on the local dyad axis pointing toward the minor groove and on the long axis oriented along the base pair axis. The Inc and Tip parameters are the rotation parameters around the same two respective axes.[36] Twist, Rise, and conformation torsion angles refer to the usual definitions.[37]

Another energetically favored geometry with all strands containing S-type sugars has been obtained for the $dG_n*dG_n{\cdot}dC_n$ triple helix (conformation II in Table I). However, this structure has a total molecular mechanics energy higher than that of the structure discussed above. This refined geometry has a third-strand Hoogsteen base-paired structure and a third-strand orientation parallel to the dG_n Watson–Crick strand. These two criteria are in agreement with the model presented by van Vlijmen et al.[22] in their molecular dynamics studies on triple-helix structures involving G*G·C base triplets. As suggested by these authors, we have extended our study to the reverse-Hoogsteen base-pairing model and also considered an antiparallel third-strand direction.

The refined-structure geometries obtained in the present work show a correlation between the third-strand orientation, the type of third-strand hydrogen bonding (Hoogsteen or reverse-Hoogsteen), and the glycosidic torsion angle conformation of the third-strand nucleosides (syn or anti). In agreement with previous authors' remarks,[6,10,23] a parallel third-strand orientation requires that the glycosidic torsion angle of the third-strand nucleosides has an anti conformation in Hoogsteen-bonded triplets (see conformations I and II in Table I) and a syn conformation in reverse-Hoogsteen-bonded triplets (see conformations V and VI in Table I). The inverse situation is observed if the third-strand orientation is antiparallel to the guanine strand of the Watson–Crick duplex. In such cases, the glycosidic torsion angle of the third-strand nucleosides in the Hoogsteen-bonded triplet is in the syn conformation (conformations III and IV in Table I) and that of reverse-Hoogsteen bonded triplets is in the anti conformation (conformations VII and VIII in Table I).

The molecular mechanics total energy is lower for the two refined structures involving Hoogsteen-bonded triplets and a parallel third-strand orientation (conformations I and II in Table

(22) van Vlijmen, H. W. T.; Ramé, G. L.; Pettitt, B. M. *Biopolymers* **1990**, *30*, 517–532.

(23) Cheng, Y. K.; Pettitt, B. M. *J. Am. Chem. Soc.* **1992**, *114*, 4465–4474.

I). The conformations VII and VIII which contain reverse-Hoogsteen-bonded triplets and an antiparallel-oriented third strand have a higher total energy than the previous cases. However, the difference of total energy between the reverse-Hoogsteen–antiparallel conformations VII and VIII and the Hoogsteen–parallel conformation II is small, and the three conformations (II, VII, and VIII in Table I) may be considered equally probable.

The two other families of triple-helix structures which involve, on the one hand, Hoogsteen-bonded triplets with antiparallel third-strand orientations and, on the other hand, reverse-Hoogsteen-bonded triplets with parallel third-strand orientations have higher total energies (conformations III–VI in Table I). Moreover, the syn conformation is always maintained for the glycosidic torsion angle of the third-strand nucleosides. The interstrand backbone distance is shorter in the syn glycosidic configuration than in the anti configuration. Consequently the electrostatic repulsion between the charged phosphate groups results in conformational energy higher than that for the anti conformers.

In light of the present calculations, even if the parallel–Hoogsteen model is energetically the most favored (conformations I and II), the antiparallel–reverse-Hoogsteen model cannot be ruled out. The two other families (antiparallel–Hoogsteen and parallel–reverse-Hoogsteen) which are characterized by the emergence of a glycosidic torsion angle in the syn conformation are less stable and thus appear to be unfavored.

These theoretical results have been compared to data obtained by vibrational spectroscopy experiments performed on dG_n*dG_n·dC_n and rG_n*dG_n·dC_n triple helices. Simulations have given stereochemically reasonable structures for dG_n*dG_n·dC_n triple helices. The Raman and infrared data allow us to confirm experimentally the sugar conformation of the most stable energy-refined geometry (conformation I).

Other structures were calculated starting with geometries formed with an A-form duplex and an S-type sugar third strand or with a B-form duplex and an N-type sugar third strand. The refined structures were either stereochemically unreasonable, or the obtained sugar conformations were not compatible with the experimental results.

Raman Spectroscopy. Two spectral domains are of particular interest concerning the sugar conformations (N- or S-type) and their orientation with respect to the bases (anti or syn). The first region is found between 600 and 900 cm^{-1}, the second one between 1200 and 1400 cm^{-1}.

A. Spectral Region 600–900 cm^{-1}. Figure 2b,c shows the Raman spectra of the rG_n*dG_n·dC_n and dG_n*dG_n·dC_n triple helices recorded in the spectral region 600–900 cm^{-1}. This spectral region contains a characteristic conformational marker mode of the sugar conformation and of the glycosidic torsion angle of the dG residue. This mode arises from the guanine ring breathing mode coupled to a deoxyribose vibration. The behavior of this mode versus the glycosidic torsion angle and the sugar conformation has been studied in double-stranded DNA.[24-28] This mode is observed around 689 cm^{-1} in B-form DNA (dG: S-type/anti), 665 cm^{-1} in A-form DNA (dG: N-type/anti), and 625 cm^{-1} in Z-form DNA (dG: N-type/syn). In the present work, the dG_n*dG_n·dC_n triple-helix spectrum shows two strong Raman lines respectively located at 689 and 662 cm^{-1} (Figure 2b). The former peak at 689 cm^{-1} is characteristic of dG residues in an S-type/anti conformation while the latter one at 662 cm^{-1} indicates the presence of dG residues in an N-type/anti conformation. No study has been published concerning the intensity conservation between the 689- and 662-cm^{-1} lines on the B to A conformational transition. However such an intensity conservation has been shown to exist between the corresponding 689- and 625-cm^{-1} lines on the B to Z transition.[27] It thus seems reasonable to consider that the similarity of the relative intensities of the two 689- and 662-cm^{-1} peaks on the dG_n*dG_n·dC_n spectrum suggests an equal amount of dG residues in N-type/anti and S-type/anti conformations in the triple helix.

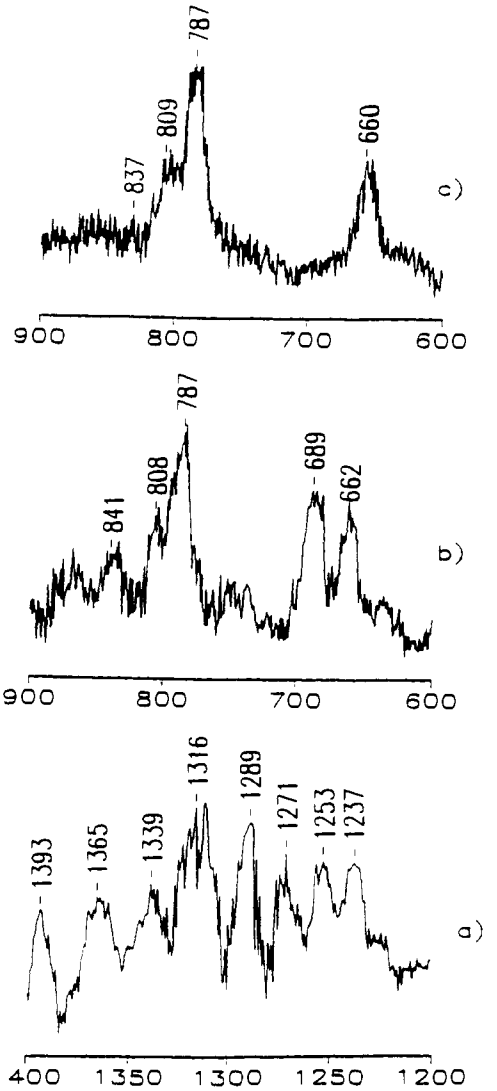

Figure 2. Raman spectra of triple helices: (a) dG_n*dG_n·dC_n spectral region, 1400–1200 cm^{-1}; (b) dG_n*dG_n·dC_n spectral region, 900–600 cm^{-1}; (c) rG_n*dG_n·dC_n spectral region, 900–600 cm^{-1}.

The Raman spectrum of the rG_n*dG_n·dC_n triple helix (Figure 2c) presents in this region only one peak at 660 cm^{-1}, showing that the guanosines of both strands adopt an N-type/anti geometry. It is important to notice that an N-type/syn geometry would have been detected on the Raman spectrum by the presence of a line around 625 cm^{-1}. The absence of such a peak allows us to reject the eventuality of such a conformation in the rG_n or dG_n strands of the rG_n*dG_n·dC_n triple helix. Similarly the absence of a Raman line around 689 cm^{-1} (characteristic of an S-type/anti geometry) shows that there are no more sugars bound to guanines in an S-type/anti conformation in the rG_n*dG_n·dC_n triple helix. The comparison with the dG_n*dG_n·dC_n spectrum (Figure 2b) shows, as will be discussed below in the FTIR section, that in this latter case the third dG_n strand was in an S-type/anti geometry.

(24) Thamann, T. J.; Lord, R. C.; Wang, A. H. J.; Rich, A. *Nucl. Acids Res.* **1981**, *9*, 5443–5457.
(25) Letellier, R.; Ghomi, M.; Taillandier, E. *J. Biomol. Struct. Dyn.* **1986**, *3*, 671–687.
(26) Nishimura, Y.; Torigoe, C.; Tsuboi, M. *Nucl. Acids Res.* **1986**, *14*, 2737–2748.
(27) Wang, Y.; Thomas, G. A.; Peticolas, W. L. *J. Biomol. Struct. Dyn.* **1987**, *5*, 249–274.
(28) Peticolas, W. L.; Kubasek, W. L.; Thomas, G. A.; Tsuboi, M. *Biological applications of Raman Spectroscopy*; Spiro, T. G., Ed.; J. Wiley: 1987; Vol. 1, 81–133.

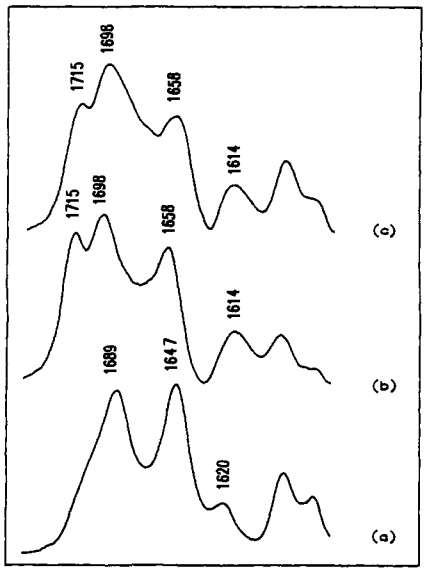

Figure 3. FTIR spectra in D_2O solutions in the spectral region 1800–1550 cm^{-1} of (a) $dG_n{\cdot}dC_n$ duplex, (b) $dG_n{*}dG_n{\cdot}dC_n$ triplex, and (c) $rG_n{*}dG_n{\cdot}dC_n$ triplex.

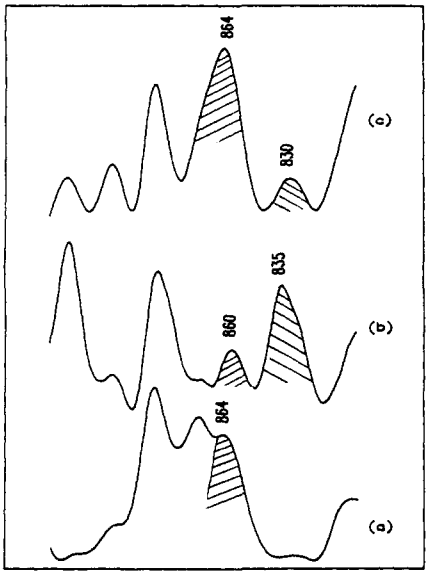

Figure 4. FTIR spectra between 950 and 750 cm^{-1} in D_2O solution of (a) $dG_n{\cdot}dC_n$ duplex, (b) $dG_n{*}dG_n{\cdot}dC_n$ triplex, and (c) $rG_n{*}dG_n{\cdot}dC_n$ triplex: \\\\, S-type sugar ////, N-type sugar.

B. Spectral Region 1200–1400 cm^{-1}. This region (Figure 2a) is well-known to contain several conformational Raman marker peaks of dG and dC residues. The positions of the dG Raman marker peaks reflect conformations for these residues in agreement with the previous assignment in the 600–900-cm^{-1} spectral region. The Raman peak at 1393 cm^{-1} is well-known to arise from a vibration mode of a dG residue in an N-type/anti conformation.[26] The broad peak centered around 1316 cm^{-1} results from the superimposition of two peaks characteristic of N-type/anti dG respectively located at 1319 and 1309 cm^{-1}. Moreover, this 1316-cm^{-1} broad line involves the 1317-cm^{-1} peak due to S-type/anti dG residue vibrations.[26] The existence of dG with an S-type sugar is confirmed by the presence of two peaks at 1365 and 1339 cm^{-1}. While the 600–900-cm^{-1} spectral region gives information on the dG residue conformations, the 1200–1400-cm^{-1} region allows moreover a determination of the dC residue conformation in the studied triple helices. The triplet present at 1271–1253–1237 cm^{-1} (Figure 2a) is characteristic of S-type/anti dC.[26] This shows that the dC_n strand contains only S-type sugars.

FTIR Spectroscopy. The formation of the (Pur)*(Pur)·(Pyr) triplexes has been followed in the spectral region of the in-plane double-bond stretching vibrations of the bases between 1800 and 1500 cm^{-1}. This spectral region had been shown to be sensitive to the formation of T*A·T, U*A·U, and C^+*G·C base triplets.[29-33] Figure 3 presents the FTIR spectra recorded in D_2O solution of the $dG_n{*}dG_n{\cdot}dC_n$ and $rG_n{*}dG_n{\cdot}dC_n$ triplexes and of the $dG_n{\cdot}dC_n$ duplex. In the double-helix $dG_n{\cdot}dC_n$ spectrum (Figure 3a) two intense IR bands located at 1689 and 1647 cm^{-1} are classically assigned to the C6=O6 of guanine and C2=O2 of cytosine stretching vibrations. However, a strong vibrational coupling exists between the guanine and cytosine residues forming the Watson–Crick base pairs.[29]

When the triplex is formed, both carbonyl stretching vibrations of the Watson–Crick duplex are shifted toward higher wavenumbers (at 1698 and 1658 cm^{-1}, respectively) and a new band is observed at 1715 cm^{-1}. A C=O group usually gives a strong IR band near 1720 cm^{-1}; however its wavenumber can be lowered when the carbonyl bond has much of a single bond nature or when it is involved in hydrogen bonding. The new 1715-cm^{-1} infrared absorption band might be assigned to the unbound C6=O6 stretching vibration of the third-strand guanosine residues invoved in the $dG_n{*}dG_n{\cdot}dC_n$ and $rG_n{*}dG_n{\cdot}dC_n$ triplexes.

A. Spectral Region 800–900 cm^{-1}. IR marker bands of sugar conformations are found in the 800–900-cm^{-1} region presented in Figure 4. In B family form nucleic acids, an absorption characteristic of S-type sugars is detected around 835 cm^{-1} while in A family geometry a band is observed around 860 cm^{-1}. We observe on Figure 4a in the duplex spectrum an absorption band at 864 cm^{-1} which shows that all the 2′-deoxyribose sugars have an N-type conformation[32,34] in highly concentrated $dG_n{\cdot}dC_n$ solution. In contrast, in the $dG_n{*}dG_n{\cdot}dC_n$ triplex spectrum (Figure 4b) the relative intensity of the absorption band located at 835 cm^{-1} is about twice as large as that of the 860-cm^{-1} absorption band. This fact shows a predominance of S-type sugars for the $dG_n{*}dG_n{\cdot}dC_n$ triplex with a relative proportion of two S-type sugars for one N-type. The ratio of the relative intensities of the 864- and 830-cm^{-1} IR absorption bands in $rG_n{*}dG_n{\cdot}dC_n$ and $dG_n{*}dG_n{\cdot}dC_n$ triplex spectra is inverted (Figure 4c,b).

B. Spectral Region 1250–1450 cm^{-1}. The spectral region between 1250 and 1450 cm^{-1} contains several marker bands which reflect the sugar conformations and the base–sugar orientations. Thus in B-form DNA an absorption is observed around 1420 cm^{-1} whereas the same vibrational mode is detected around 1405 cm^{-1} in the A- and Z-conformations. This mode is assigned to a coupling between vibrations of the purine imidazole ring and sugar deformations. Here it is observed at 1403 cm^{-1} in the duplex spectrum (Figure 5a) (Table III) and is characteristic of a guanosine residue with an N-type sugar pucker ring.[34] The anti conformation of the guanosine residues in double-stranded DNA is evidenced by the presence of an absorption band at 1376 cm^{-1}. This band arises from a guanosine residue vibration mode involving the glycosidic bond deformation and is an efficient marker of the glycosidic torsion angle conformation (Table III). It is shifted to 1355 cm^{-1} in the case of a syn conformation of the purine nucleoside.[25,34,35]

The conformational IR marker bands (1403 and 1376 cm^{-1}) are still present in the $dG_n{*}dG_n{\cdot}dC_n$ triplex IR spectrum presented

(29) Howard, F. B.; Frazier, J.; Miles, H. T. *J. Biol. Chem.* **1971**, *246*, 7073–7086.
(30) Thomas, G. A.; Peticolas, W. L. *J. Am. Chem. Soc.* **1983**, *105*, 993–996.
(31) Ohms, J.; Ackerman, T. *Biochemistry* **1990**, *29*, 5237–5244.
(32) Liquier, J.; Coffinier, P.; Firon, M.; Taillandier, E. *J. Biomol. Struct. Dyn.* **1991**, *9*, 437–445.
(33) Akhebat, A.; Dagneaux, C.; Liquier, J.; Taillandier, E. *J. Biomol. Struct. Dyn.* **1992**, *10*, 577–588.
(34) Ghomi, M.; Letellier, R.; Liquier, J.; Taillandier, E. *Int. J. Biochem.* **1990**, *22*, 691–699.
(35) Taboury, J. A.; Liquier, J.; Taillandier, E. *Can. J. Chem.* **1985**, *63*, 1904–1909.
(36) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63–91.

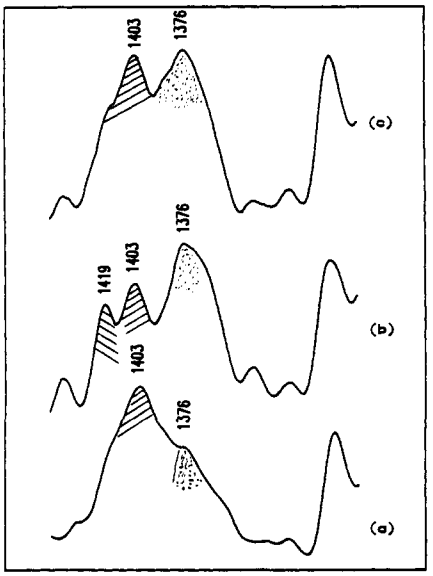

Figure 5. FTIR spectra between 1500 and 1250 cm^{-1} in H_2O solution of (a) dG_n·dC_n duplex, (b) dG_n*dG_n·dC_n triplex, and (c) rG_n*dG_n·dC_n triplex: \\\\, S-type sugar; ////, N-type sugar; dotted area, anti.

Table III. IR Marker Bands of Glycosidic Torsion and Sugar Conformations

struct	anti orientation of guanine	N-type sugar linked to a guanine	S-type sugar linked to a guanine	N-type sugar	S-type sugar
dG_n·dC_n duplex	1376	1403		864	
dG_n*dG_n·dC_n triplex	1376	1403	1419	860	835
rG_n*dG_n·dC_n triplex	1376	1403		864	830

in Figure 5b, but we observe also a band at 1419 cm^{-1} (Table III). The 1419-cm^{-1} absorption band indicates the presence of dG residues with an S-type sugar conformation[25,34,35] even if some dG residues of the triplex keep an N-type conformation (presence of the 1403-cm^{-1} IR band). Moreover, the existence of a band at 1376 cm^{-1} in this spectral region and the absence of any significative absorption around 1355 cm^{-1} show that all the guanine residues of the triplex are in anti conformations.[34,35]

The substitution of the dG_n third strand in the dG_n*dG_n·dC_n triple helix by an rG_n third strand induces the disappearance of the S-type sugar IR marker at 1419 cm^{-1} in the rG_n*dG_n·dC_n triplex spectrum (Figure 5c) (Table III). This shows that the dG_n third strand had an S-type conformation in the dG_n*dG_n·dC_n triple helix. According to this evidence, the presence of the N-type dG IR marker at 1403 cm^{-1} on the triplex spectrum leads us to propose an N-type conformation for the sugars of the dG_n strand involved in the Watson–Crick duplex. The presence of an S-type IR marker at 830 cm^{-1} on the rG_n*dG_n·dC_n triplex spectrum reflects an S-type conformation for the Watson-Crick dC_n strand. The IR data, in agreement with the previously discussed Raman results, confirm an rG_n*dG_n·dC_n triple-helix model containing two N-type/anti guanine strands and one S-type/anti cytosine strand. In the dG_n*dG_n·dC_n triple helix, the dG_n Watson–Crick strand also adopts an N-type/anti conformation and the dC_n strand an S-type/anti conformation but the third dG_n strand has an S-type/anti conformation.

Discussion

The Raman and IR spectra show that in the dG_n*dG_n·dC_n triple helix the dC_n strand contains S-type sugars while the dG residues exist with both S- and N-type sugar conformations in the same proportion. Moreover the N-type sugars belong to the dG_n strand of the Watson–Crick duplex while the third dG_n strand contains S-type sugars. The vibrational data on the dG_n*dG_n·dC_n triple helix allow us to assign the triple-helix sugar conformations in good agreement with the computed structure which has the lowest total energy (see conformation I in Tables I and II).

Table IV. Helicoidal Parameters and Conformational Angles of the rG_n*dG_n·dC_n Optimized Geometry Triplex[a]

strand	Xdisp	Ydisp	Rise	Inc	Tip	Twist
I	−3.31	0.22	3.35	−0.06	4.12	32.60
II	−3.49	−0.12	3.35	−0.62	−6.09	32.60
III	−0.62	1.13	3.36	−16.97	−23.07	32.60

strand	χ	α	β	γ	δ	ε	sugar conf
I	240	−70	−177	55	140	−173	S
II	198	−76	178	64	82	−158	N
III	207	−80	169	78	84	−164	N

[a] Abbreviations are defined in the footnote of Table II.

In order to check the ability of our dG_n*dG_n·dC_n theoretical model to interpret the spectral changes observed in vibrational spectra upon substitution of the third dG_n strand by a rG_n strand, we have studied the conformation of the rG_n*dG_n·dC_n triple helix. The starting geometry is derived from the optimized dG_n*dG_n·dC_n triple-helix geometry which has the lowest total energy. The S-type dG_n third strand of the dG_n*dG_n·dC_n triple helix was substituted by an N-type rG_n strand in the theoretical model. This rG_n*dG_n·dC_n triple-helix geometry was then energy minimized (total conformational energy: −987 kcal/mol).

Obviously, as for the dG_n*dG_n·dC_n triple helix (see conformation I in Table I), we retain Hoogsteen hydrogen bonding between the rG_n strand and the duplex, and the third rG_n strand orientation is also parallel to the dG_n strand of the dG_n·dC_n Watson–Crick duplex. The main differences between dG_n*dG_n·dC_n and rG_n*dG_n·dC_n concern the sugar conformations. In the optimal dG_n*dG_n·dC_n triple helix, N-type sugars are assigned to the dG_n strand of the Watson–Crick duplex and S-type sugars to the other strands. In the rG_n*dG_n·dC_n triple helix, calculations assign an N-type sugar conformation to the third rG_n strand and to the Watson–Crick dG_n strand, while the Watson–Crick dC_n strand has an S-type sugar (Table IV), in agreement with Raman and IR data.

As stated in the Introduction, several recent works have reported that the orientation of a G-rich third strand is antiparallel to the purine-rich strand of the Watson–Crick duplex.[7-12] Our calculations show, in agreement with other studies,[6,10,23] that an antiparallel third-strand orientation associated with Hoogsteen-bonded triplets requires a syn conformation for the glycosidic torsion angle of the third-strand nucleosides (conformations III and IV). This model is in contradiction with FTIR spectra where only the IR marker band of the anti conformation is observed around 1374 cm^{-1} (Figure 5) and Raman spectra where marker peaks around 689 and 660 cm^{-1} characteristic of the anti conformation of the dG residue are observed (Figure 2b). We must however notice that the Raman marker of the S-type/syn conformation is not known, as such a geometry has up to now never been observed in DNAs. S-type sugars in guanosines have been detected by the presence of a peak at 689 cm^{-1} (classical position of the S-type/anti conformation marker) only in the spectrum of the dG_n*dG_n·dC_n triple helix. If the S-type/syn geometry existed in the dG_n*dG_n·dC_n triple helix, this would mean that this geometry has a Raman peak at the same position as that corresponding to the S-type/anti conformation. However the hypothesis of a syn guanosine conformation in the third strand seems to be excluded if the spectrum obtained with a third rG_n strand instead of a dG_n strand is considered. Riboses having an N-type geometry, the N-type/syn guanosines, should give rise to a Raman peak located around 625 cm^{-1},[24,25,27,28] In the

(37) IUPAC–IUB Joint Commission on Biochemical Nomenclature. Abbreviations and symbols for the description of conformations of polynucleotide chains. *Eur. J. Biochem.* **1983**, *131*, 9–15.

600–900-cm^{-1} spectral region, we observe on the Raman spectrum of the rG_n*dG_n·dC_n triple helix the presence of a unique peak at 660 cm^{-1}. No significative peak is detected around 625 cm^{-1}. This rules out a possible existence of syn conformers in the purine third strand.

Moreover we have turned to syn the glycosidic torsion angle in the best minimized conformation of the dG_n*dG_n·dC_n (conformation I). The so-obtained structure has again been energy minimized. In the resultant geometry the glycosidic torsion has gone back to anti, and the triple helix is destabilized (E = −1046.3 kcal/mol) and stereochemically unfavored. These experimental and theoretical results lead us to propose an anti conformation for the third-strand dG residues.

So taking into account these data alone, reverse-Hoogsteen-bonded triplets (anti conformation) could be consistent with an antiparallel third-strand orientation. However, our simulations show that even if such structures are stereochemically allowed (conformations VII and VIII), the S-type sugar conformation of the refined geometries fails to interpret the emergence of N-type markers in the vibrational spectra. In order to verify that this result was not due to falling into a poor local minimum, we have substituted the third strand in the refined structure which is in agreement with the vibrational spectroscopy data (conformation I) by an antiparallel-oriented strand containing guanines able to form reverse-Hoogsteen bonds with the Watson–Crick duplex. This starting conformation has been energy refined. We were particularly interested in the evolution of the sugar conformation during minimization. It was found to be impossible to keep the sugar conformation in agreement with Raman and IR data, and only S-type sugars are observed in the energy-minimized geometry (E = −1040 kcal/mol). We can notice that the same sugar conformation has been obtained using as initial geometries the Arnott A-form triple helix[19] and B-DNA[20] (conformations VII and VIII). The calculations show that the description of an intermolecular dG_n*dG_n·dC_n triple helix by an antiparallel–reverse-Hoogsteen model does not give rise to an energy-minimized structure in agreement with vibrational spectroscopy results.

The formation of a single triplet where the third dG residue is reverse-Hoogsteen base paired and has an orientation antiparallel to the dG nucleoside of the Watson–Crick pair dG·dC is however stereochemically allowed, and even if our calculations and vibrational spectroscopy show that in an intermolecular dG_n*dG_n·dC_n triple helix the presence of such a triplet is unfavored, we do not exclude the possibility of finding a sequence of such triplets within a longer triple helix depending on the flanking triplets. As an example, we have verified that an antiparallel and reverse-Hoogsteen base-paired triplet G*G·C (conformations VII and VIII) can be stabilized by parallel and Hoogsteen base-paired triplets G*G·C (conformation I) in the triple helix d(GGGGG-*G*-GGGGGG)*d(GGGGG-*C*-GGGGGG)·d-(CCCCC-*G*-CCCCCC) (conformational energy: −1163 kcal/mol). In this "strand-switching" model[12] the guanine of the third strand alternates its binding between the two duplex strands in order to form a G*G·C triplet. One of the base triplets is antiparallel and reverse-Hoogsteen bonded, and this does not prevent the formation of a stereochemically reasonable triple-helical structure. However, such modification of one base triplet induces an increase of energy of about 25 kcal/mol, and the total energy of the triple-helix structure will increase with the number of modified triplets.

Our results on the third-strand orientation are consistent with previous experimental and molecular mechanics studies[13] which showed that the third strand adopts an antiparallel orientation with respect to the purine strand of the underlying duplex in the triple helices containing T*A·T and G*G·C triplets if the number of ApG and GpA steps is more than three. Otherwise a parallel orientation is energetically favored for the G-rich third strand.

In conclusion, we have obtained stereochemically reasonable models for dG_n*dG_n·dC_n and rG_n*dG_n·dC_n triple helices based upon molecular mechanics calculations and Raman and FTIR spectroscopy experiments. Such an approach, combining these three complementary techniques, is an original and efficient way to determine the triple-helix structure. Vibrational spectroscopy has allowed us to determine the sugar conformation of each strand of the triple helix and the glycosidic torsion angle of the dG residues. The energy-minimized structures in which the sugar conformations are in agreement with Raman and IR data allow us to interpret the experimental data and to determine other geometrical parameters, i.e. third-strand base pairing and third-strand orientation, which are not directly detectable by vibrational spectroscopy.

Acknowledgment. We wish to thank the CIRCE (Centre Inter Regional de Calcul Electronique, CNRS, Orsay, France) for computational facilities. Thanks are also due to D. Giroud and J. M. Teuler for their suggestions on the use of the computer facilities. We thank Dr. R. Lavery for helpful discussions and for kind advice in preparing the manuscript.