Effect of the G·T Mismatch on Backbone and Sugar Conformations of Z-DNA and B-DNA: Analysis by Raman Spectroscopy of Crystal and Solution Structures of d(CGCGTG) and d(CGCGCG)[†]

J. M. Benevides,[‡] A. H.-J. Wang,[§] G. A. van der Marel,[∥] J. H. van Boom,[∥] and G. J. Thomas, Jr.*[,‡]

Division of Cell Biology and Biophysics, School of Basic Life Sciences, University of Missouri—Kansas City, Kansas City, Missouri 64110, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and Gorlaeus Laboratories, University of Leiden, 2300 RA Leiden, The Netherlands

Received May 2, 1988; Revised Manuscript Received July 14, 1988

ABSTRACT: The Z-DNA crystal structures of d(CGCGTG) and d(CGCGCG) are compared by laser Raman spectroscopy. Raman bands originating from vibrations of the phosphodiester groups and sensitive to the DNA backbone conformation are similar for the two structures, indicating no significant perturbation to the Z-DNA backbone as a result of the incorporation of G·T mismatches. Both Z structures also exhibit Raman markers at 625 and 670 cm^{-1}, assigned respectively to C3′-*endo*/*syn*-dG (internal) and C2′-*endo*/*syn*-dG conformers (3′ terminus). Additional Raman intensity near 620 and 670 cm^{-1} in the spectrum of the d(CGCGTG) crystal is assigned to C4′-*exo*/*syn*-dG conformers at the mismatch sites (penultimate from the 5′ terminus). A Raman band at 1680 cm^{-1}, detected *only* in the d(CGCGTG) crytal, is assigned to the hydrogen-bonded dT residues and is proposed as a definitive marker of the Z-DNA wobble G·T pair. For aqueous solutions, the Raman spectra of d(CGCGTG) and d(CGCGCG) are those of B-DNA, but with significant differences between them. For example, the usual B-form marker band at 832 cm^{-1} in the spectrum of d(CGCGTG) is about 40% less intense than the corresponding band in the spectrum of d(CGCGCG), and the former structure exhibits a companion band at 864 cm^{-1} not observed for d(CGCGCG). The simplest interpretation of these results is that the conventional B-form OPO geometry occurs for only 6 of the 10 OPO groups of d(CGCGTG). The remaining four OPO groups, believed to be those at or near the mismatch site, are in an "unusual B" conformation which generates the 864 cm^{-1} band. The B form of d(CGCGTG) also exhibits evidence of dG conformers differing from the usual C2′-endo/anti conformation. The data are consistent with C1′-*exo*/*anti*-dG conformers at the mismatch sites. The present findings indicate that the B helix (solution structure) is more significantly perturbed than the Z helix (crystal structure) by the incorporation of G·T mismatches. The high resolution of the experimental data and its enhancement by the method of Fourier deconvolution confirm that a Raman band of very low intensity occurs near 806 ± 2 cm^{-1} in the B-DNA solution structures of both d(CGCGCG) and d(CGCGTG). This weak band contrasts sharply with the 12-fold more intense band near 807–810 cm^{-1} that is associated with the A-DNA backbone geometry (Benevides et al., 1986). Nonetheless, in quantitative analytical applications, the weak 806 cm^{-1} band should be taken into account in order to estimate accurately the percentages of B- and A-type backbone geometries, in accordance with the recent proposal of Nishimura et al. (1986).

The detection of noncomplementary base pairs by proofreading enzymes is an obligatory first step in the repair process which assures the fidelity of DNA replication. Accordingly, there is considerable interest in the structural details of base pairs which deviate from the canonical Watson–Crick pattern. Information on the nature of hydrogen bonding between bases of the noncomplementary G·T, A·C, and A·G pairs has been obtained recently by methods of X-ray crystallography (Ho et al., 1985; Hunter et al., 1986; Privê et al., 1987). The structural consequences of incorporating such mismatched pairs within DNA double helices have also been discussed (Kennard, 1985). However, the structural attributes of DNA which are responsible for enzymatic recognition of mismatches are not known, and little information is available on aqueous DNA structures containing mismatched sequences.

The G·T mismatch has been the most extensively studied by X-ray crystallography. The results show that hydrogen bonding between G and T is of the wobble type in A-, B-, and Z-DNA crystal structures (Kennard, 1985; Ho et al., 1985). Wobble G·T pairing is also consistent with NMR data obtained in studies of aqueous B-DNA (Hare et al., 1986; Quignard et al., 1987; Kalnik et al., 1988a,b). Distance geometry methods applied to the NMR data provide information about glycosyl torsion and sugar pucker, but no information has yet been obtained about the phosphodiester geometry associated with the mismatched sequence in the B-DNA solution structures. Information of this type is not easily inferred from NMR measurements, and therefore, other methods are required to investigate the phosphodiester geometry associated with mismatched sequences of aqueous DNA.

The ease of application of laser Raman spectroscopy to solutions of nucleic acids has encouraged its use as a complement to X-ray and NMR methods. Raman spectroscopy is particularly advantageous for the study of 5′C–O–P–O–C3′ geometry in the backbone of aqueous DNA, as well as for identifying conformational properties of the nucleoside sugars,

---

[†] Supported by NIH Grant AI18758 (G.J.T.) and NSF Grant DMB-8612268 (A.H.-J.W.). This is paper 34 in the series Raman Spectral Studies of Nucleic Acids. Paper 33 in this series is Benevides and Thomas (1988).

* Author to whom correspondence should be addressed.

[‡] University of Missouri—Kansas City.

[§] Massachusetts Institute of Technology. Present address: Department of Physiology and Biophysics, University of Illinois, Urbana, IL 61801.

[∥] University of Leiden.

0006-2960/89/0428-0304$01.50/0

including pucker and glycosyl orientation. In previous work, we have obtained and interpreted Raman spectra of DNA and RNA oligonucleotide crystals for which high-resolution X-ray crystal structures were available (Benevides et al., 1984b, 1986, 1988). These empirical studies have facilitated the assignment of individual Raman bands in A-, B-, and Z-DNA to specific backbone geometries and nucleoside conformers, and the results have been applied to the diagnosis of conformation in aqueous DNA and viruses (Thomas et al., 1986, 1988). The Raman bands which are diagnostic of specific conformational features of DNA and RNA have been catalogued (Thomas, 1986; Thomas & Wang, 1988).

In this paper, we report the results of a Raman spectroscopic analysis of the self-complementary hexamer d(CGCGTG) in both crystalline and aqueous solution forms. The crystal structure is a Z-DNA double helix known to contain two G·T mismatches of the wobble type (Ho et al., 1985), while the aqueous solution structure incorporates the mismatches into a B-DNA double helix. Comparison of the Raman spectrum of the d(CGCGTG) crystal with spectra of two previously studied Z-DNA crystals, d(CGCGCG) and d(CGCATGCG), reveals the extent to which the G·T mismatch affects the conformation-sensitive Raman bands of the left-handed Z helix, and permits identification of Raman bands which are diagnostic of the G·T wobble pair in Z-DNA. Comparison of the spectra of aqueous solutions of these oligomers provides the corresponding information for the G·T wobble pair in right-handed B-DNA.

EXPERIMENTAL PROCEDURES

Materials. The oligonucleotides d(CGCGTG), d(CGCGCG), and d(CGCATGCG) were synthesized by an improved phosphotriester technique in which 1-hydroxybenzotriazole was used as an activating reagent (van der Marel et al., 1981). In each case, after the fully protected molecule was deblocked, the fragment was purified by Sephadex G50 column chromatography and converted into the ammonium salt. The purity of each oligomer was greater than 95%, as judged by HPLC analysis. The crystallizations followed procedures described in detail elsewhere (Wang et al., 1982).

Raman Spectroscopy. Raman spectra of crystals and solutions were obtained by methods previously described in detail (Prescott et al., 1984; Benevides et al., 1984b) and summarized below.

Crystals of d(CGCGTG) or d(CGCGCG) were transferred with approximately 10 μL of mother liquor [2-methyl-2,4-pentanediol (40%) + 25 mM sodium cacodylate (pH 7) + 2 mM spermine + 10 mM $MgCl_2$] to a Raman sample cell (Kimax 34507 glass capillary), which was thermostated at 32 °C. The 514.5-nm line from a Coherent Model CR-2 argon laser was focused on the crystal, and the Raman scattering at 90° was collected and analyzed by a Spex Ramalog spectrometer under the control of a North Star Horizon II microcomputer. Spectral data were collected at 1.0 cm^{-1} intervals with an integration time of 1.5 s. A slit width of 8 cm^{-1} was employed for each scan. Similar procedures were employed for d(CGCATGCG) (Benevides et al., 1984b).

For obtaining solution spectra, d(CGCGTG) was dissolved to a concentration of 24 mM in 0.2 M NaCl at pH 7.0 ± 0.2; d(CGCGCG) was similarly dissolved in 0.1 M NaCl. Raman sample cells were thermostated at 12 °C, and spectra were recorded as described above.

Table I: Selected Raman Line Frequencies Diagnostic of A-, B-, and Z-DNA Backbones[a]

A-DNA	B-DNA	Z-DNA	assignment
706 ± 5[b]			C–O
807 ± 3[c]	790 ± 5	745 ± 3	O–P–O
	832 ± 7[d]		O–P–O
1099 ± 1	1092 ± 1	1095 ± 2	PO_2^-
1418 ± 2	1422 ± 2	1425 ± 2	CH_2

[a] Frequencies in cm^{-1} units are determined from Raman spectra of DNA and RNA crystals and fibers of known structure [reviewed in Thomas and Wang (1988)]. [b] This band occurs at 712 ± 3 cm^{-1} in A-RNA structures. [c] This band occurs at 812 ± 2 cm^{-1} in A-RNA structures. A very weak band also occurs at ca. 810 cm^{-1} in Z-DNA structures. [d] The band is centered close to 830 cm^{-1} for DNA containing G and C and close to 840 cm^{-1} for DNA containing A and T.

Each Raman spectrum displayed in the figures is the average of several scans, of 1.5 cm^{-1} or better repeatability. In each case, the fluorescent background and scattering by the solvent or mother liquor were removed by using computer subtraction techniques described previously (Benevides et al., 1984b). The spectrum of solvent or mother liquor was always recorded with the same instrument conditions employed for solution or crystal. When indicated in the figure legends, the noise was smoothed by a least-squares fit of third-order polynomials to overlapping 15-point regions. This procedure did not measurably alter the Raman line frequencies or intensities. Otherwise, the data have not been smoothed or altered by refinement techniques. Raman frequencies are believed accurate to within ±2 cm^{-1}. Fourier deconvolution (Thomas & Agard, 1984) was employed to separate or sharpen partially overlapped bands in the 600–900 cm^{-1} region that could not be further resolved by narrowing the spectral band-pass. Generally, a Gaussian–Lorentzian product function of 20 cm^{-1} half-width was used to deconvolve the spectral envelope.

In the computation of difference spectra of DNA crystals and solutions, the minuend and subtrahend were normalized to give the same integrated intensity of the strong Raman band of dG near 1575 cm^{-1}. Previous studies have indicated that the integrated intensity of this band is not sensitive to changes in DNA conformation, even though the position of the band maximum may shift slightly from one secondary structure to another (Prescott et al., 1984; Benevides & Thomas, 1983). For the crystals and solutions compared here, the difference spectra revealed that the band near 1091–1095 cm^{-1} due to the PO_2^- group is invariant in intensity and may also be used as a reliable intensity standard.

RESULTS AND DISCUSSION

The A, B, and Z conformations of DNA are easily distinguished on the basis of their characteristic Raman spectra (Benevides et al., 1988; Thomas & Wang, 1988). Two types of Raman spectral lines or "bands" are especially well established as diagnostic of the DNA conformation. One type, referred to as a *backbone* conformation marker, originates in the 5′C–O–P–O–C3′ network and reflects the unique phosphodiester torsions (α, ζ) associated with a particular backbone geometry. These are listed in Table I. A second type, referred to as a *nucleoside* conformation marker, originates in the base–sugar network and reflects sensitivity of the base vibration to the furanose pucker (torsion angle δ) and glycosyl orientation (χ) of the attached sugar. Tables II and III contain representative nucleoside conformation markers of dG and dT, respectively. More comprehensive listings are given elsewhere (Thomas et al., 1986; Nishimura et al., 1986).

Comparison of Wobble and Watson–Crick Structures of Z-DNA

Raman spectra of the d(CGCGTG) and d(CGCGCG) crystal structures are shown in Figure 1. Reference to the data of Tables I and II shows that each of these crystal

Table II: Selected Conformation-Sensitive Raman Lines of Deoxyguanosine[a]

C1′-exo/anti	dG	670 ± 2
		1343 ± 2
C2-endo/anti	dG	682 ± 2
		1316 ± 2[b]
		1333 ± 2
C3′-endo/anti	dG	664 ± 2
		1318 ± 2[b]
C2′-endo/syn	dG	671 ± 2
		1316 ± 2
C3′-endo/syn	dG	625 ± 2
		1316 ± 2
C4′-exo/syn	dG	620 ± 5
		670 ± 5
		1316 ± 3

[a] Basis for assignments is discussed in Thomas et al. (1986) and Thomas and Benevides (1985). [b] This band exhibits very low intensity compared to the band of similar frequency in syn conformers. [c] $S\alpha\sqrt{P}$

Table III: Selected Conformation-Sensitive Raman Lines of Thymidine[a]

C2′-endo/anti	dT	668 ± 2
		748 ± 2
		1142 ± 2
		1208 ± 2
		1360 ± 2[b]
C3′-endo/anti	dT	665 ± 2[c]
		745 ± 2[d]
		777 ± 2
		1239 ± 2

[a] Basis for assignments is discussed in Thomas et al. (1986) and Thomas and Benevides (1985). [b] This band is diagnostic of C2′-*endo*/*anti*-thymidine in Z-DNA, specifically. In B- and A-DNA, the band shifts to 1373 ± 2 cm^{-1}. [c] Very low intensity compared to the 668 cm^{-1} band of the C2′-endo isomer. [d] Very low intensity compared to the 748 cm^{-1} band of the C2′-endo isomer.

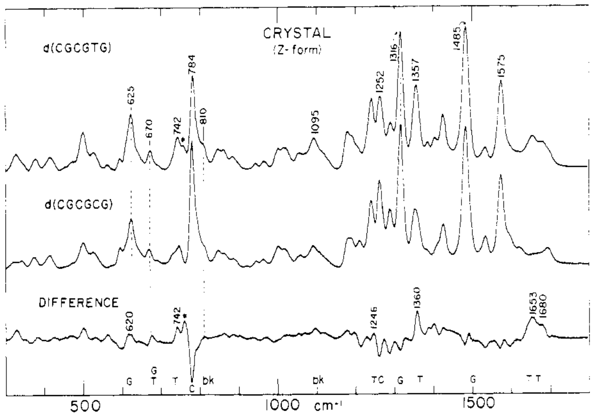

FIGURE 1: Raman spectra in the region 300–1800 cm^{-1} of Z-DNA crystals. (Top) d(CGCGTG); (middle) d(CGCGCG); (bottom) computed difference spectrum, d(CGCGTG) minus d(CGCGCG). The labels along the abscissa indicate band assignments to the bases (C, G, or T) or Z backbone (bk). Frequencies (cm^{-1} units) are labeled for several of the bands discussed in the text. The asterisk indicates the position of a strongly interfering band from the mother liquor, most of which has been eliminated by spectral subtraction (Benevides et al., 1984b).

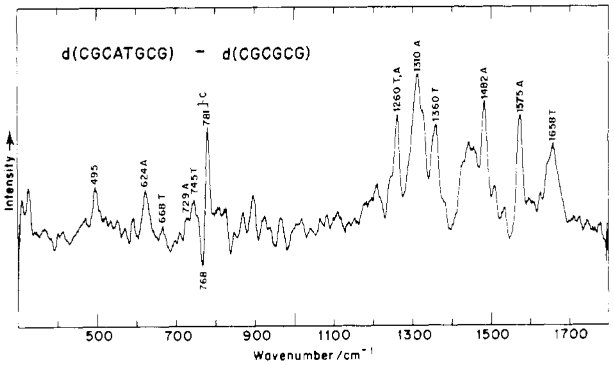

FIGURE 2: Raman spectrum in the region 300–1800 cm^{-1}, computed by difference for a hypothetical Z-DNA crystal containing only Watson–Crick AT pairs. The spectrum was generated by subtraction of the spectrum of the Z-DNA crystal d(CGCGCG) from that of the Z-DNA crystal d(CGCATGCG). [See also Benevides et al., (1984b).]

structures contains the key Raman bands characteristic of the phosphodiester geometry and *syn*-dG conformations of Z-DNA. In particular, both crystals contain the principal marker for the Z backbone (745 cm^{-1}) and lack markers for the A and B backbones (Table I). The C2′-*endo*/*anti*-dT marker of d(CGCGTG), which is expected at ca. 748 cm^{-1} (Table III), is revealed in the difference spectrum of Figure 1 as a positive band at 742 cm^{-1} and thus overlaps the principal Z-backbone marker. This accounts for the considerably higher spectral intensity of d(CGCGTG) near 742–750 cm^{-1}. For both d(CGCGTG) and d(CGCGCG), the bands at 625 and 1316 cm^{-1} (C3′-*endo*/*syn*-dG), and the band at 670 cm^{-1} (C2′-*endo*/*syn*), are indicators of the *syn*-dG conformers of Z-DNA (Thamann et al., 1981; Benevides et al., 1984b). In other respects, the spectra are significantly different from one another, as is evident in the Figure 1 difference spectrum.

Since both d(CGCGTG) and d(CGCGCG) crystallize in the same space group ($P2_12_12_1$) with similar hydration, all of the bands in the Figure 1 difference spectrum are a consequence of the replacement of two of the Watson–Crick G·C pairs by wobble G·T pairs. Further, most of the positive bands in the difference spectrum can be assigned to dT, and most of the negative bands can be assigned to dC, as labeled in Figure 1. This generalization occurs without exception for the bands in the 900–1475 cm^{-1} interval, indicating that the different primary structures account for most of the features of the observed difference spectrum. Thus, the dC → dT substitution in Z-DNA does not significantly affect the structure beyond the substitution site itself.

It is interesting to compare the difference spectrum of Figure 1 with the difference spectrum (Figure 2) computed by subtraction of the Raman spectrum of d(CGCGCG) from that of the Z-DNA octamer d(CGCATGCG) (Benevides et al., 1984b). Such comparison reveals a close correspondence between thymidine bands in d(CGCGTG) and d(CGCATGCG), which suggests a similarity in thymidine conformation and thymidine interactions in the two structures despite the different purine hydrogen-bonding partners. A similar conclusion has been reached from the results of X-ray diffraction analysis (Ho et al., 1985). A striking similarity between d(CGCGTG) and d(CGCATGCG), apparent in the difference spectra of Figure 1 and 2, is the prominent band of dT at 1360 cm^{-1}. This band is assigned to a thymine ring vibration, and it characterizes thymine in Z-DNA. In B-DNA and A-DNA structures, the corresponding thymine band is greatly shifted from this position (Thomas & Benevides, 1985; Benevides & Thomas, 1988).

In the 600–900 and 1480–1700 cm^{-1} regions, the data of Figure 1 indicate small but significant perturbations to Raman frequencies of dT, dC, and dG. These perturbations are the result of different secondary and tertiary structural features in the two Z-DNA crystals, as will be discussed below. Again, the Raman frequencies and intensities associated with dG residues appear to be the least affected by the switch from Watson–Crick to wobble pairing, as is evident by direct comparison of the two Z-DNA crystal spectra of Figure 1. No-table exceptions are considered below.

Region from 600 to 900 cm^{-1}. In this section, we propose an interpretation of features in the difference spectrum of Figure 1 which appear to be distinct from those associated strictly with the change in primary structure discussed above. We consider first the bands of dG in the 600–900 cm^{-1} interval.

Conspicuous as *positive* bands in the Figure 1 difference spectrum are bands near 620 and 670 cm^{-1}, indicating that both arise from greater relative intensity in d(CGCGTG) than in d(CGCGCG). In principle, several factors could contribute to the observed intensity enhancement in the mismatched structure, including (i) hidden bands of dT, (ii) changes in dG sugar pucker and/or glycosyl torsion, (iii) other changes in dG residue conformation apart from those associated with sugar pucker and glycosyl torsion, (iv) altered base stacking, and (v) the different hydrogen bonding of G·T in comparison to that of the G·C pair. Among these, (iii) can be eliminated on the basis of X-ray crystallographic results which provide direct evidence to the contrary (Ho et al., 1985), and (iv) can be eliminated on the basis of previous Raman studies which show that dG marker bands are not very sensitive to changes in base stacking geometry (Benevides et al., 1984a). Also, we consider (v) unlikely in view of the fact that the distinguishing donor (N–H) and acceptor (C═O) groups for hydrogen bonding do not generate bands in this region of the spectrum. We note also that the difference in interbase hydrogen bonding between G·T and G·C pairs is not likely to affect the guanine ring vibrations in a significant way, since only the exocyclic amino hydrogen at 2C is involved. Either of the remaining two possibilities, (i) or (ii), or a combination of them, could account for the observed spectral differences. In order to distinguish between them, we note the following.

The crystal structure of d(CGCGTG) shows that the dG residue of each G·T pair has the C4′-exo/syn conformation, which is significantly different from the C3′-*endo*/*syn*-dG conformation of each corresponding G·C pair in d(CGCGCG) (Ho et al., 1985). Therefore, it is likely that the two difference bands observed at 620 and 670 cm^{-1} (Figure 1) are due to C4′-*exo*/*syn*-dG. In support of this assignment, we note that bands near 620 and 670 cm^{-1} have been reported in two independent studies of cyclic rGMP, a ribosyl phosphodiester analogue which is known to contain the C4′-*exo*/*syn*-rG conformation (Benevides et al., 1984a; Nishimura et al., 1986). Therefore, we assign the 620 and 670 cm^{-1} bands observed here to C4′-*exo*/*syn*-dG at the mismatch sites. Of course, we cannot categorically rule out the possibility that underlying bands of dT may also contribute to the observed 620 and 670 cm^{-1} bands in the Figure 1 difference spectrum. In fact, it is likely that dT contributes at least partially to the difference band at 670 cm^{-1} for the following reason.

Previous studies indicate that there are several conformation-sensitive bands of dT in the 600–900 cm^{-1} region (Thomas & Benevides, 1985). In B-DNA and in the B form of poly(dA-dT)·poly(dA-dT) (Table III), the C2′-*endo*/*anti*-dT conformer generates a moderately intense Raman band at 668 cm^{-1}. Therefore, we expect the single dT residue (C2′-*endo*/*anti* conformer) in each strand of d(CGCGTG) to yield weak Raman scattering at this frequency, which could be masked by the proposed 670 cm^{-1} marker of C4′-*exo*/*syn*-dG. As indicated in Table III, other conformation-sensitive bands of dT might also contribute to the observed spectral profile in the 600–900 cm^{-1} interval. A recent normal coordinate analysis supports the conformation dependence of the thymidine ring modes in this region of the vibrational spectrum (Letellier et al., 1987). The calculations propose in-plane deformation of the thymidine 4C═O group as a contributory motion to the normal mode.

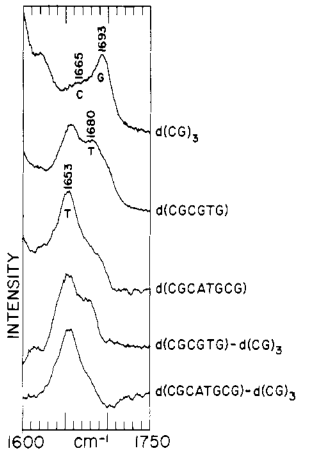

FIGURE 3: Raman spectra in the region 1600–1750 cm^{-1} of the following Z-DNA crystals (from top to bottom): d(CGCGCG); d(CGCGTG); d(CGCATGCG); difference spectrum [d(CGCGTG) minus d(CGCGCG)] showing Raman bands diagnostic of the G·T wobble pair of Z-DNA; difference spectrum [d(CGCATGCG) minus d(CGCGCG)] showing Raman bands diagnostic of the A·T Watson–Crick pair of Z-DNA.

In summary, the band near 620 cm^{-1} in the Figure 1 difference spectrum is assigned exclusively to the C4′-*exo*/*syn*-dG conformer, and the band near 670 cm^{-1} is assigned primarily to C4′-*exo*/*syn*-dG with a possible additional contribution from C2′-*endo*/*anti*-dT. The observed frequencies and intensities are fully consistent with known model compound studies and normal mode analyses.

Region from 1600 to 1800 cm^{-1}. Figure 3 shows the region 1600–1800 cm^{-1} from Raman spectra of three Z-DNA crystals: d(CGCGCG), d(CGCGTG), and d(CGCATGCG). This region of the spectrum, often called the "double-bond region", contains bands due to stretching vibrations of double bonds including carbonyl groups of C, G, and T. However, the bands in this region also include contributions from in-plane deformations of N–H groups of G (amino and imino), C (amino), and T (imino). No contributions are expected from O–H deformations of either water of crystallization or mother liquor, since these are compensated by subtraction (Benevides et al., 1984b). Although the band origins are complex, reasonable assignments can be made by reference to appropriate model compound data (Lord & Thomas, 1967; Benevides et al., 1984b; Thomas & Benevides, 1985; Thomas et al., 1986). These show that the 1693 cm^{-1} band in all three crystals is due to base-paired G and that the 1653 and 1680 cm^{-1} bands in d(CGCGTG) and d(CGCATGCG) are due predominantly to base-paired T. The band of base-paired C is very weak in all crystals and is seen only as a broad and weak shoulder near 1660–1670 cm^{-1} in d(CGCGCG).

The prominence of the Raman bands of T in this spectral region may seem surprising in view of the low dT content in d(CGCGTG) and d(CGCATGCG). However, their extraordinary Raman intensity in comparison to bands of G and C is well-known (Lord & Thomas, 1967; Thomas & Benevides, 1985). This is confirmed by the difference spectrum of Figure 1, which shows that nearly all of the Raman intensity between 1650 and 1690 cm^{-1} is due to dT residues of d(CGCGTG). The present results (Figure 3) demonstrate further that in d(CGCGTG) the 1653 and 1680 cm^{-1} intensities are comparable, when measured over a base-line tangent to the wings of the doublet. On the other hand, in d(CGCATGCG) the 1680 cm^{-1} band is discernible only as a weak shoulder to the much more intense 1653 cm^{-1} companion

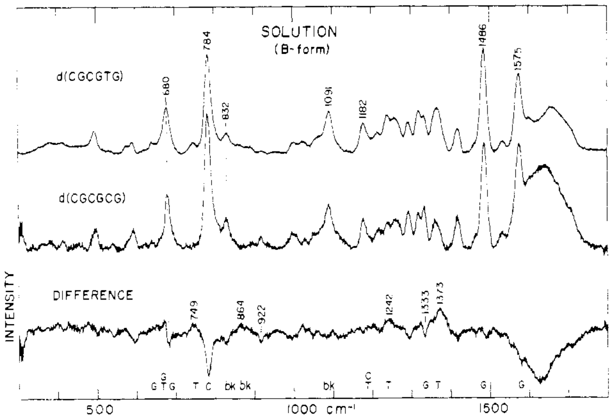

FIGURE 4: Raman spectra in the region 300–1800 cm^{-1} of B-DNA solutions. (Top) d(CGCGTG); (middle) d(CGCGCG); (bottom) Computed difference spectrum, d(CGCGTG) minus d(CGCGCG). Labels are as defined in Figure 1, except that bk here indicates assignment to OPO groups of the B-DNA backbone.

band. Evidently, the intensity ratio of the 1653 and 1680 cm^{-1} bands distinguishes the wobble G·T pair from the Watson–Crick A·T. Thus, the prominent 1680 cm^{-1} band serves as a convenient indicator of the G·T wobble base pair of Z-DNA.

Comparison of Wobble and Watson–Crick Structures of B-DNA

Figure 4 shows Raman spectra in the 300–1800 cm^{-1} region obtained from aqueous solutions of d(CGCGTG) and d(CGCGCG). Both hexamers have prominent Raman bands near 680 cm^{-1} indicating C2′-*endo*/*anti*-dG conformers, as well as bands at 832 cm^{-1} associated with the backbone conformation of B-DNA (Tables I and II). However, as seen in the difference spectrum of Figure 4, the intensity of the 832 cm^{-1} band of d(CGCGTG) is about 40% (±5%) weaker than that of d(CGCGCG). The inequivalent intensities at 832 cm^{-1} are clearly revealed when measured either in relation to the 1091 cm^{-1} band (PO_2^- group vibration, Figure 4) or in relation to the 784 cm^{-1} band (C ring vibration, not shown) as the intensity standard. The results show that about 40% of OPO groups of d(CGCGTG) are conformationally different from those of d(CGCGCG). Therefore, we conclude that the wobble base pairing at the mismatch sites imposes a configurational change in the B backbone. The present data do not permit us to identify which OPO groups undergo the conformation change, but those at or near the mismatch positions are the obvious candidates. Since each strand contains five phosphodiesters, as many as two in each strand may be altered from the usual B conformation. It is reasonable to assume that some or all of the affected OPO groups are directly linked to the mismatched pairs.

In view of the diminished intensity of the 832 cm^{-1} band in d(CGCGTG), corresponding positive intensity due to the altered OPO conformers should appear in the same spectrum. Examination of Figure 4 shows in fact that d(CGCGTG) contains a weak band near 864 cm^{-1} which is absent from d(CGCGCG). The band in question is of the correct intensity and frequency to support its assignment to the altered OPO conformers of d(CGCGTG). In a previous study of A-DNA crystals (Benevides et al., 1986), a band at 852 cm^{-1} was assigned to the unusual OPO conformation at the CpG step (5′→3′) in d(CCCCGGGG). Therefore, there exists a precedent for assigning the higher frequency phosphodiester band to an unusual conformation of the OPO group. X-ray crystallography of d(CCCCGGGG) has revealed α, ζ torsions of −204, −59 at this CpG step (Haran et al, 1987), and we propose a similar conformation for the altered OPO groups of d(CGCGTG).

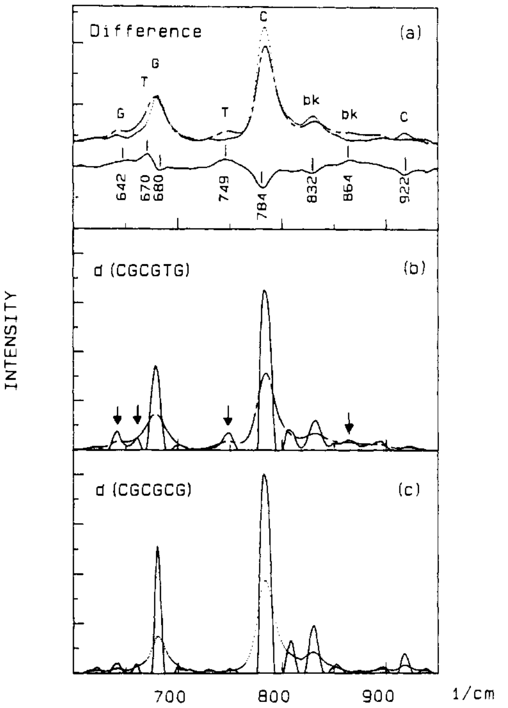

FIGURE 5: (a) Raman spectra in the region 600–950 cm^{-1} of aqueous solutions of B-form d(CGCGTG) (– – –) and B-form d(CGCGCG) (···) and their computed difference (—). The experimental data are those shown in Figure 4, after smoothing by a 9-point moving-average method. Principal bands in the difference spectrum are labeled, and assignments are abbreviated as in Figures 1 and 4. (b) Observed (– – –) and deconvolved (—) spectra of d(CGCGTG). Deconvolution was achieved by using the iterative Fourier deconvolution method of Thomas and Agard (1984) and a Gauss–Lorentz desmearing function of 20 cm^{-1} half-width. Arrows indicate bands which are significantly more intense in this spectrum than in the corresponding d(CGCGCG) spectrum shown in (c). (c) Observed (···) and deconvolved (—) spectra of d(CGCGCG).

The most conspicuous feature in the Figure 4 difference spectrum is the strong negative band at 784 cm^{-1}. As discussed elsewhere (Thomas et al., 1986), this band contains two distinct components: a C ring vibration (ca. 780 cm^{-1}) and a phosphodiester group vibration characteristic of B-DNA (ca. 790 cm^{-1}). Thus, the negative 784 cm^{-1} band in the Figure 4 difference spectrum is considered to reflect both the difference in base composition and the different backbone conformation in d(CGCGTG). As in the case of the corresponding crystals (cf. Figure 1), the difference spectrum of Figure 4 displays several positive bands assignable to dT and negative bands assignable to dC, in accordance with the dC → dT substitution.

Bands of dG in the 600–700 and 1300–1500 cm^{-1} intervals which are recognized as diagnostic of specific dG conformations (Nishimura et al., 1986; Thomas et al., 1986; Benevides et al., 1988) are listed in Table II. Several of these can be recognized in the spectra of Figure 4. Intensity differences between the two solution spectra for the region 600–900 cm^{-1}, which are of potential conformational significance, are best revealed by direct overlay of the spectra (Figure 5a) and are discussed in detail in the following section.

Deconvolution of the Region from 600 to 900 cm^{-1}. Previously, we employed Fourier deconvolution of the observed Raman profile in the 600–900 cm^{-1} region in order to compare quantitatively the Raman band shapes and to search for hidden or partially resolved bands in this interval (Thomas et al., 1986; Benevides et al., 1986, 1988). Curve-fitting methods, introduced by Wartell and co-workers (Wartell & Harrell, 1986), may also be employed for this purpose (Kubasek et al., 1986; Thomas & Wang, 1988). Both the deconvolution and curve-fitting results provide evidence that the Raman contour in the 800–870 cm^{-1} region comprises several overlapping bands. A complete understanding of these bands has not yet been elucidated, but the contour itself, like its deconvolution or curve fit, provides an empirical basis for distinguishing different structures from one another. As an example, we show in Figure 5 the computed difference spectra (Figure 5a) and deconvolutions (Figure 5b,c) within the 600–950 cm^{-1} intervals of spectra of d(CGCGTG) and d(CGCGCG). Both the difference spectra and the deconvolution results show the following features unique to d(CGCGTG): (i) Enhanced intensity at 642 and 670 cm^{-1} attributable to C1′-*exo/anti*-dG and C2′-*endo/anti*-dT; (ii) a Raman band at 749 cm^{-1} attributable to C2′-*endo/anti*-dT; (iii) a Raman band at 864 cm^{-1}, in lieu of the expected intensity at 832 cm^{-1}, attributable to unusual OPO geometry at mismatch sites; and (iv) reduced intensity at 784 and 922 cm^{-1}, attributable to the replacement of dC by dT.

Figure 5 shows that both the conventional Watson–Crick B-DNA structure and the mismatched B-DNA structure yield by deconvolution a complicated band pattern in the 800–870 cm^{-1} interval. In addition to the principle B marker (ca. 832 cm^{-1}) and the previously discussed anomalous marker (864 cm^{-1}), both hexamers contain a deconvolved component near 806 cm^{-1}. A similar spectral component generated by curve fitting (Wartell & Harrell, 1986; Nishimura et al., 1986) or deconvolution (Thomas et al., 1986) of Raman spectra has been discussed previously in studies of high molecular weight DNA. Two interpretations may be advanced to explain such a band: It arises from a small percentage of A-type OPO conformation; or it is intrinsic to the B conformation but is much weaker than the A marker of similar frequency. Although it is not possible to categorically exclude the presence of small amounts of A-type OPO conformers in large B-DNA molecules, it seems unlikely that such conformers would be present in the hexamers studied here. Clearly, there are no indications in the spectra of Figure 4 of any C3′-endo/anti nucleosides as are usually associated with A-form backbone geometry. The present results therefore support the conclusion first advanced by Tsuboi and co-workers (Nishimura et al., 1986) that the OPO conformation of B-DNA generates not only the strong Raman band near 790 cm^{-1} and the moderately intense Raman band near 832 cm^{-1} but also a weak Raman band near 806 cm^{-1}. It should be noted that the weak band shifts to higher frequency (ca. 815 cm^{-1}) in DNA composed only of AT pairs (Thomas et al., 1986). If we ascribe to the 806 (or 815) cm^{-1} band of B-DNA the intensity suggested here by deconvolution, i.e., roughly half the intensity of the 832 cm^{-1} band, then this band is about 12-fold weaker than the band near 807 cm^{-1} common to spectra of A-DNA.

SUMMARY AND CONCLUSIONS

The mismatch Z-DNA crystal d(CGCGTG), in common with the Watson–Crick Z-DNA crystal d(CGCGCG), exhibits nucleoside conformation markers identified previously as those of C3′-*endo/syn*-dG and C2′-*endo/syn*-dG. In addition, the d(CGCGTG) crystal contains Raman intensity near 620 and 670 cm^{-1} which is not present in the spectrum of the d(CGCGCG) crystal. In accordance with the results of X-ray crystallography, we assign the additional Raman bands of d(CGCGTG) to the C4′-*exo/syn*-dG conformers at the mismatch sites. These bands overlap those of the C3′-*endo/syn*- and C2′-*endo/syn*-dG markers, respectively, but may be useful for identifying the C4′-exo/syn conformer of dG in the absence of one or both of the former. Raman bands diagnostic of backbone OPO geometry show that the Z-DNA backbones of d(CGCGTG) and d(CGCGCG) crystals are barely distinguishable from one another, indicating no significant effect of the wobble base pairs upon the OPO geometry of the Z helix. A major difference in the spectra of the two Z-DNA crystals is a prominent band at 1680 cm^{-1} for d(CGCGTG), which is largely a carbonyl group stretching vibration and is assigned by analogy with model structures to thymidine of the G·T wobble base pair. This band serves as a definitive marker of the wobble G·T pair of Z-DNA.

Comparison of the Raman spectra of aqueous solutions of d(CGCGTG) and d(CGCGCG) shows that both structures are in the B-DNA family. However, the backbones of the two solution structures differ from one another more significantly than do the backbones of the corresponding crystal (Z-DNA) structures. Thus, G·T wobble pairs appear to perturb the B-helix structure more than the Z-helix structure. In the mismatch B helix, we find approximately 40% of the OPO Raman intensity shifted from the position normally associated with B-DNA of the same base composition. This indicates that as many as 4 of the 10 phosphodiester groups in the duplex do not contain the α, ζ torsions associated with conventional B-DNA. The four anomalous conformers generate instead a Raman band near 864 cm^{-1}, well above the frequency range ordinarily associated with OPO groups of B-DNA. Among previously studied models, only d(CCCCGGGG) has exhibited a similar OPO conformation marker, attributed to the strained CpG step at the center of its sequence (Benevides et al., 1986). The conformation of the deoxyribose ring of dG is also affected by wobble base pairing. A fraction of the dG residues of d(CGCGTG) appears to exhibit the C1′-exo/anti conformation, not evident in d(CGCGCG). The present data do not allow a firm quantitative estimate of the number of nucleosides so affected, but the data are consistent with one-third of the dG residues—very likely those which are wobble base paired to dT.

The present results confirm the greater variability of molecular geometry in B-DNA, as compared with Z-DNA. This appears to be true both for the classical Watson–Crick double helix of d(CGCGCG) and for the wobble double helix of d(CGCGTG). The data obtained here also show that the B-DNA backbone generates a Raman band near 806 cm^{-1}, which although 12-fold weaker than the corresponding Raman band of A-DNA should be taken into account in quantitative analyses of secondary structure.

Registry No. d(CGCGTG), 101479-87-2; d(CGCGCG), 58927-26-7; d(CGCATGCG), 89991-81-1; guanine, 73-40-5; thymine, 65-71-4.

REFERENCES

Benevides, J. M., & Thomas, G. J., Jr. (1983) *Nucleic Acids Res. 11*, 5747–5761.

Benevides, J. M., & Thomas, G. J., Jr. (1988) *Biochemistry 27*, 3868–3873.

Benevides, J. M., LeMeur, D., & Thomas, G. J., Jr. (1984a) *Biopolymers 23*, 1011–1024.

Benevides, J. M., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., Rich, A., & Thomas, G. J., Jr. (1984b) *Nucleic Acids Res. 12*, 5913–5925.

Benevides, J. M., Wang, A. H.-J., Rich, A., Kyogoku, Y., van der Marel, G. A., van Boom, J. H., & Thomas, G. J., Jr. (1986) *Biochemistry 25*, 41–50.

Benevides, J. M., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., & Thomas, G. J., Jr. (1988) *Biochemistry 27*, 931–938.

Hare, D. R., Shapiro, L., & Patel, D. J. (1986) *Biochemistry 25*, 7445-7456.

Haran, T. E., Shakked, Z., Wang, A. H.-J., & Rich, A. (1987) *J. Biomol. Struct. Dyn. 5*, 199-217.

Ho, P. S., Frederic, C. A., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J., & Rich, A. (1985) *EMBO J. 4*, 3617-3623.

Hunter, W. N., Brown, T., Anand, N. N., & Kennard, O. (1986) *Nature* (*London*) *320*, 552-555.

Kalnik, M. W., Kouchakdjian, M., Li, B. F. L., Swann, P. F., & Patel, D. J. (1988a) *Biochemistry 27*, 100-108.

Kalnik, M. W., Kouchakdjian, M., Li, B. F. L., Swann, P. F., & Patel, D. J. (1988b) *Biochemistry 27*, 108-115.

Kennard, O. (1985) *J. Biomol. Struct. Dyn. 3*, 205-226.

Kubasek, W. L., Wang, Y., Thomas, G. A., Patapoff, T. W., Schoenwaelder, K.-H., van der Sande, J. H., & Peticolas, W. L. (1986) *Biochemistry 25*, 7440-7445.

Letellier, R., Ghomi, M., & Taillandier, E. (1987) *Eur. Biophys. J. 14*, 243-252.

Lord, R. C., & Thomas, G. J., Jr. (1967) *Spectrochim. Acta, Part A 23A*, 2551-2591.

Nishimura, Y., Tsuboi, M., Sato, T., & Aoki, K. (1986) *J. Mol. Struct. 146*, 123-153.

Prescott, B., Steinmetz, W., & Thomas, G. J., Jr. (1984) *Biopolymers 23*, 235-256.

Privê, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L., & Dickerson, R. E. (1987) *Science* (*Washington, D.C.*) *238*, 498-504.

Quignard, E., Fazakerely, G. V., van der Marel, G., van Boom, J. H., & Guschlbauer, W. (1987) *Nucleic Acids Res. 15*, 3397-3409.

Thamann, T., Lord, R. C., Wang, A. H. J., & Rich, A. (1981) *Nucleic Acids Res. 9*, 5443-5457.

Thomas, G. J., Jr. (1986) *Adv. Spectrosc. 13*, 233-309.

Thomas, G. J., Jr., & Agard, D. A. (1984) *Biophys. J. 46*, 763-768.

Thomas, G. J., Jr., & Benevides, J. M. (1985) *Biopolymers 24*, 1101-1105.

Thomas, G. J., Jr., & Wang, A. H.-J. (1988) in *Nucleic Acids and Molecular Biology* (Eckstein, F., & Lilley, D. M. J., Eds.) Vol. 2, Springer-Verlag, Berlin.

Thomas, G. J., Jr., Prescott, B., & Benevides, J. M. (1986) *Biomol. Stereodyn. 4*, 227-254.

Thomas, G. J., Jr., Prescott, B., Opella, S. J., & Day, L. A. (1988) *Biochemistry 27*, 4350-4357.

van der Marel, G. A., van Boeckel, C. A. A., Wille, G., & van Boom, J. H. (1981) *Tetrahedron Lett.* 3887.

Wang, A. H.-J., Fujii, S., van Boom, J. H., van der Marel, G. A., van Boeckel, S. A., & Rich, A. (1982) *Nature* (*London*) *299*, 601-604.

Wartell, R. M., & Harrell, J. T. (1986) *Biochemistry 25*, 2664-2671.

Molecular Structure of the Netropsin-d(CGCGATATCGCG) Complex: DNA Conformation in an Alternating AT Segment[†]

Miquel Coll,[‡] Juan Aymami,[‡] Gijs A. van der Marel,[§] Jacques H. van Boom,[§] Alexander Rich,[‡] and Andrew H.-J. Wang*[,‡]

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and Gorlaeus Laboratory, Leiden State University, 2300 RA Leiden, The Netherlands

Received May 23, 1988; Revised Manuscript Received August 18, 1988

ABSTRACT: The molecular structure of the complex between a minor groove binding drug (netropsin) and the DNA dodecamer d(CGCGATATCGCG) has been solved and refined by single-crystal X-ray diffraction analysis to a final R factor of 20.0% to 2.4-Å resolution. The crystal is similar to that of the other related dodecamers with unit cell dimensions of a = 25.48 Å, b = 41.26 Å, and c = 66.88 Å in the space group $P2_12_12_1$. In the complex, netropsin binds to the central ATAT tetranucleotide segment in the narrow minor groove of the dodecamer B-DNA double helix as expected. However, in the structural refinement the drug is found to fit the electron density in two orientations equally well, suggesting the disordered model. This agrees with the results from solution studies (chemical footprinting and NMR) of the interactions between minor groove binding drugs (e.g., netropsin and distamycin A) and DNA. The stabilizing forces between drug and DNA are provided by a combination of ionic, van der Waals, and hydrogen-bonding interactions. No bifurcated hydrogen bond is found between netropsin and DNA in this complex due to the unique dispositions of the hydrogen-bond acceptors (N3 of adenine and O2 of thymine) on the floor of the DNA minor groove. Two of the four AT base pairs in the ATAT stretch have low propeller twist angles, even though the DNA has a narrow minor groove. Alternating helical twist angles are observed in the ATAT stretch with lower twist in the ApT steps than in the TpA step.

The conformational diversity of DNA has been under intense study in recent years with a variety of techniques. One of the intriguing questions related to DNA conformation is its relation to nucleotide base sequence and different environmental factors that affect the equilibrium between alternative conformers. For example, the left-handed Z-DNA double helix is favored by alternating CG sequence (Wang et al., 1979). The interconversion between Z-DNA and the righ-handed B-DNA is influenced by metal ions, ionic strength, supercoiling, and Z-DNA binding proteins as reviewed elsewhere (Rich et al., 1984). More recently, another conformational

---

[†] This work was supported by grants from NIH, NSF, NASA, and ONR to A.R. and from NSF (DMB 8612286) and Burroughs-Wellcome to A.H.-J.W. M.C. is a Fulbright Fellow. J.A. is a NATO Fellow.

[*] Address correspondence to this author at the Department of Physiology and Biophysics, University of Illinois at Urbana-Champaign, Urbana, IL 61801.

[‡] Massachusetts Institute of Technology.

[§] Leiden State University.

0006-2960/89/0428-0310$01.50/0