Molecular Structure of a DNA Decamer Containing an Anticancer Nucleoside Arabinosylcytosine: Conformational Perturbation by Arabinosylcytosine in B-DNA†

Yi-Gui Gao, Gijs A. van der Marel,‡ Jacques H. van Boom,‡ and Andrew H.-J. Wang*
Department of Physiology and Biophysics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, and Gorlaeus Laboratories, Leiden State University, 2300RA Leiden, The Netherlands

Received May 14, 1991; Revised Manuscript Received July 29, 1991

ABSTRACT: Arabinosylcytosine (araC) is an important anticancer drug that has been shown to be misincorporated into DNA double helix. The incorporation of araC into DNA may have significant conformational consequences that could affect the function of DNA. In this paper, we present the high-resolution 3D structure of an araC-containing decamer d[CCAGGC(araC)TGG], as determined by X-ray diffraction analysis, and assess the possible DNA structural perturbation induced by araC. The modified decamer was crystallized in the monoclinic *C*2 (a = 31.97 Å, b = 25.56 Å, c = 34.62 Å and β = 114.50°) space group, the same as that from d(CCAGGCCTGG) [Heinemann, U., & Alings, C. (1989) *J. Mol. Biol. 210*, 369]. The structure of the araC-containing decamer was solved by the molecular replacement method and refined by the constrained least-squares refinement procedure to obtain a final R factor of 0.187 using 2349 [>2.0 $\sigma(F)$] observed reflections to a resolution of 1.6 Å. The overall conformation resembles that of the canonical decamer DNA structure, but with significant differences in regions close to the araC site. The O2′ hydroxyl groups of the araC residues lie in the major groove of the helix, and they are in close contact with the C5 methyl and C6 H6 atoms of the thymine on the 3′-side. This creates a higher buckle in the araC7–G14 base pair (14°), as compared to that found in the canonical decamer (9°). This may slightly destabilize B-DNA. No direct intramolecular hydrogen bond is formed, in contrast to the situation when araC is incorporated into Z-DNA. In the latter structure, an intramolecular hydrogen bond is formed between O2′ of araC and N2 of the 5′-guanine residues in the deep groove of Z-DNA. On the basis of a model building study, the incorporation of araC into A-DNA or RNA–DNA hybrid structures is found to be disfavored due to the severe steric hindrance resulting from the presence of the O2′ hydroxyl groups. The differential conformational perturbation due to the incorporation of araC between B-DNA and A-DNA (and structurally similar RNA–DNA hybrid) helices may explain the biological roles of araC.

Arabinosylcytosine (araC; Figure 1) is an effective chemotherapeutic agent in the treatment of various forms of human leukemia (Keating et al., 1982; Kufe & Spriggs, 1985). Using the therapy in combination with other agents, it was proven to be particularly useful in the treatment of acute lymphoblastic leukemia of childhood (Lauer et al., 1987). The mode of actions of araC appears to be associated, in part, with its ability to misincorporate into DNA, but not RNA, thereby interfering with the function of DNA (Kufe & Spriggs, 1985; Mikita & Beardsley, 1988). The consequence of its incorporation into DNA regarding the biological function thereof is not completely clear. Mikita and Beardsley (1988) have shown that both the replication and transcription activities are significantly affected by the misincorporation of araC in vitro. When a single araC residue is misincorporated into an internucleotide position of the template DNA, the replication is significantly retarded at or near the araC site. This observation may be associated with a number of factors, including the structural distortion caused by araC, its interference with polymerase action, the induced change in DNA properties (e.g., conformational rigidity), or a combination of these factors.

Recently, we have shown that the presence of an additional hydroxyl group at the C2′ position of the deoxyribose has a strong effect on the ability of nucleic acids (with an alternating dG–dC sequence) to form the left-handed Z-DNA conformation (Teng et al., 1989). The structures of two modified DNA hexamers, d(CG)r(CG)d(CG) and d(CG)(araC)d(GCG), have been determined by X-ray crystallographic analysis at 1.5-Å resolution (Teng et al., 1989). The structural analyses of these modified DNA hexamers have shown that the O2′ hydroxyl group of the ribosyl- or arabinosylcytosine nucleoside residues forms a hydrogen bond to the NH_2 amino group of the guanine in the 5′-(dG)p(rC) or 5′-(dG)p(araC) steps. This intramolecular hydrogen bond involving the araC sugar seems to stabilize the guanine in the syn glycosyl conformation and consequently the Z-DNA conformation.

In contrast, the disposition of the C2′ atom in the B-DNA conformation places the O2′ hydroxyl group (from either a misincorporated ribosyl or arabinosyl nucleoside) in a position that precludes this type of intramolecular hydrogen bond. In order to better understand the effect of the misincorporation of araC on B-DNA conformation, we synthesized several DNA decamers with sequences related to the d(CCAGGCCTGG) (abbreviated HA-DECA) but with araC replacing dC at different locations. One reason to choose a decamer with this sequence is that it has been crystallized in a high-resolution crystal lattice, and its structure adopts a more "standard" B-DNA conformation (Heinemann & Alings, 1989). We believe that a study of the incorporation of araC into this sequence would provide more representative structural information pertaining polymeric DNA incorporating araC. One of the decamers, d[CCAGGC(araC)TGG] (abbreviated

† This work was supported by Grants GM-41612 and CA-52506 from the NIH to A.H.J.W.
* To whom correspondence should be addressed.
‡ Leiden State University.

0006-2960/91/0430-9922$02.50/0

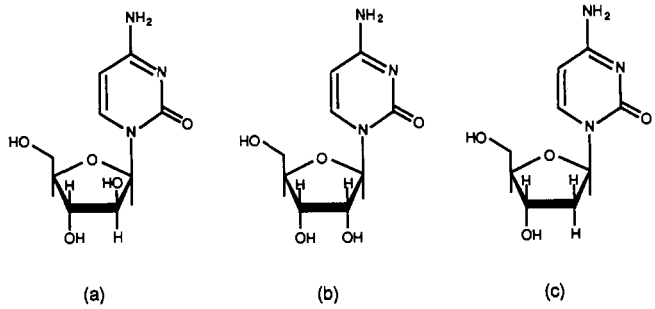


FIGURE 1: Molecular formula of (a) arabinosylcytosine (araC), (b) cytidine, and (c) deoxycytidine.

araC-DECA), was crystallized isomorphously with HA-DECA, and its crystal structure has been determined at high resolution. In this paper, we describe its molecular structure in detail and compare it with the native HA-DECA B-DNA structure. In addition, we compare its structure with the left-handed Z structures of DNA molecules containing arabinose (Teng et al., 1989) sugar moieties.

Materials and Methods

Several decamer nucleotides containing arabinsoylcytosines were synthesized by the hydroxybenzotriazole phosphotriester method described earlier (de Vroom et al., 1988). They were subjected to crystallization experiments using the procedure described previously (Wang & Gao, 1990). Thus far only the decamer d[CCAGGC(araC)TGG] has been crystallized in a useful crystal form. The crystallization solution contained 1.0 mM decamer (single strand concentration), 5 mM Tris-HCl buffer at pH 7.5, 40 mM $MgCl_2$, and 8% 2-methyl-2,4-pentanediol (2-MPD), and it was equilibrated against 40% 2-MPD by the vapor diffusion technique in a cold room (4 °C). Large crystals of somewhat irregular shape appeared after 2 weeks. It is interesting to note that both HA-DECA and araC-DECA were crystallized from solutions in the absence of spermine or spermidine. Addition of polyamines in the crystallization solution appeared to inhibit the formation of crystals. Surprisingly, no crystal could be obtained from crystallization setups at room temperature.

These crystals, grown under cold room conditions, lasted for about 3 days under the exposure of X-rays during data collection at room temperature. Consequently, two crystals were required for collecting a complete data set to the highest possible resolution. Each crystal was mounted in a thin-walled capillary and sealed with a droplet of the crystallization mother liquor. They are in the monoclinic space group $C2$ and have unit cell dimensions of a = 31.97 (2) Å, b = 25.56 (2) Å, c = 34.62 (2) Å, and β = 114.50° (2). The diffraction data were collected at room temperature on a Rigaku AFC-5R rotating-anode diffractometer, using an ω-scan mode at 20 °C with Cu K_α radiation (1.5406 Å with graphite monochromater) at a power of 50 KV and 180 mA. One crystal of size 0.3 × 0.5 × 0.7 mm was used to collect data from 20- to 2.0-Å resolution, and another crystal of comparable size was used to collect data from 2.0- to 1.54-Å resolution. Some reflections in the overlapping resolution range were collected for scaling purposes. After merging the two data sets together, a total of 2349 unique reflections to 1.6-Å resolution were considered to be observable at a 2.0 $\sigma(F)$ level above background after Lorentz polarization, absorption, and decay corrections. These were used in the refinement.

As is evident from the unit cell dimensions, this crystal form is closely related to other B-DNA decamer crystal lattices (Heinemann & Alings, 1989). In order to obtain an unbiased structure, an idealized B-DNA decamer was constructed as the starting model for refinement. The model was placed in the same position as that in the HA-DECA crystal, and it was refined using the Konnert–Hendrickson constrained refinement procedure (Hendrickson & Konnert, 1979; Westhof et al., 1985). The decamer duplex lies on the crystallographic 2-fold axis in the $C2$ space group, i.e., there is only one strand of decamer in the crystallographic asymmetric unit. Therefore, the two strands of the decamer duplex are identical in their conformation. However, the entire duplex was used during the refinement to provide the necessary base-pairing hydrogen-bonding constraints, but with the 2-fold symmetry maintained throughout the refinement. After many cycles of refinement, the R factor was ~33% at 2.0-Å resolution. Solvent molecules located from the $(2F_o - F_c)$ Fourier maps were gradually included in the subsequent refinements with increasing resolution. The final R factor was 18.7% at 1.6-Å resolution with one decamer strand and 74 water molecules per asymmetric unit. Most of the water molecules are associated with well-defined electron densities.

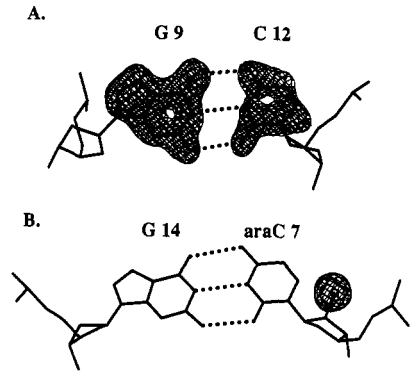


FIGURE 2: Portion of the $(F_o - F_c)$ difference Fourier electron density maps of the araC-containing DNA decamer showing (A) the G9–C12 base pair and (B) the O2′ hydroxyl group of the araC7 residue. The high-resolution (1.6 Å) nature of the structure is evident. The maps were calculated by removing portions of the structure from the model, which is then refined for a few cycles to adjust the scale factor.

Figure 2 shows the difference Fourier electron density maps of the structure obtained by removing different portions of the DNA from the phase contribution. In this figure, it can be seen that the molecular fragment of the G9–C12 base pair fits nicely in the respective electron density envelopes, which are well resolved due to the high resolution of the crystal structure. In Figure 2B, the difference electron density due to the O2′ hydroxyl group of araC7 residue stands out clearly in the map. This unambiguously supports the conclusion that araC was indeed successfully incorporated (without any side reaction or modification) into the DNA oligomer during the chemical synthesis. However, no metal ion, like the hydrated magnesium ions seen in the d(CCAACGTTGG) decamer structure (Prive et al., 1991) or the triethylamine ions found in the structure of d(CCAGGCCTGG) (Heinemann & Alings, 1989), could be unambiguously located. In the present refined structure, the RMSD from the ideal bond length is 0.014 Å. The final atomic coordinates of the structure have been deposited in the Brookhaven Protein Data Bank.

Results

An araC-Containing B-DNA Structure

The overall structure of this araC-containing DNA decamer resembles the B-DNA structure of the canonical d(CCAGGCCTGG) decamer (Heinemann & Alings, 1989). This is expected as the two decamers crystallized in nearly isomorphous crystal lattices. Figure 3 shows the stereoscopic skeletal view of the araC-DECA duplex from an orientation

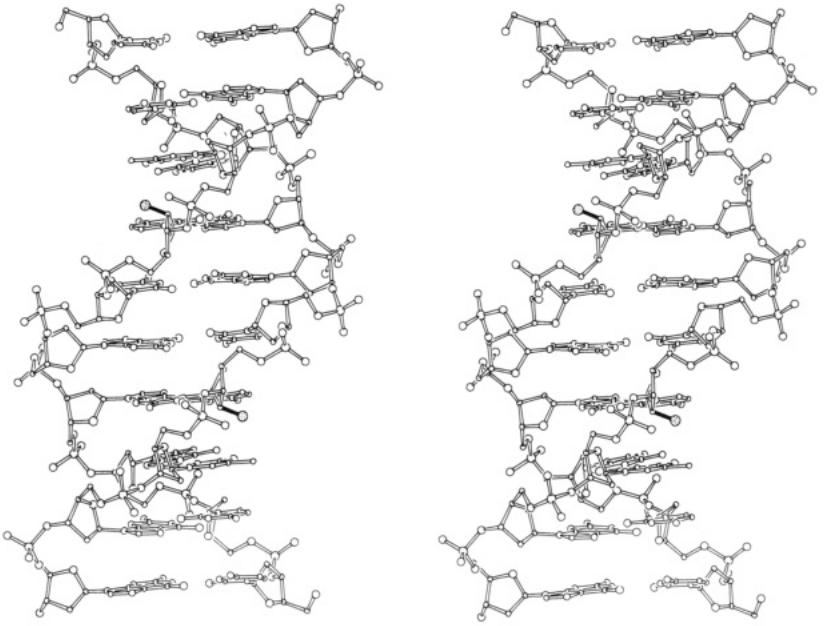


FIGURE 3: Stereoscopic skeletal drawings of d[CCAGGC(araC)TGG] B-DNA structure looking into the minor groove. In the crystal, decamers are stacked end-over-end along the helix axis (which is coincided with the *c*-axis), forming an infinite polymer. The O2′ hydroxyl oxygen atoms are stippled with the C2′–O2′ bonds in black.

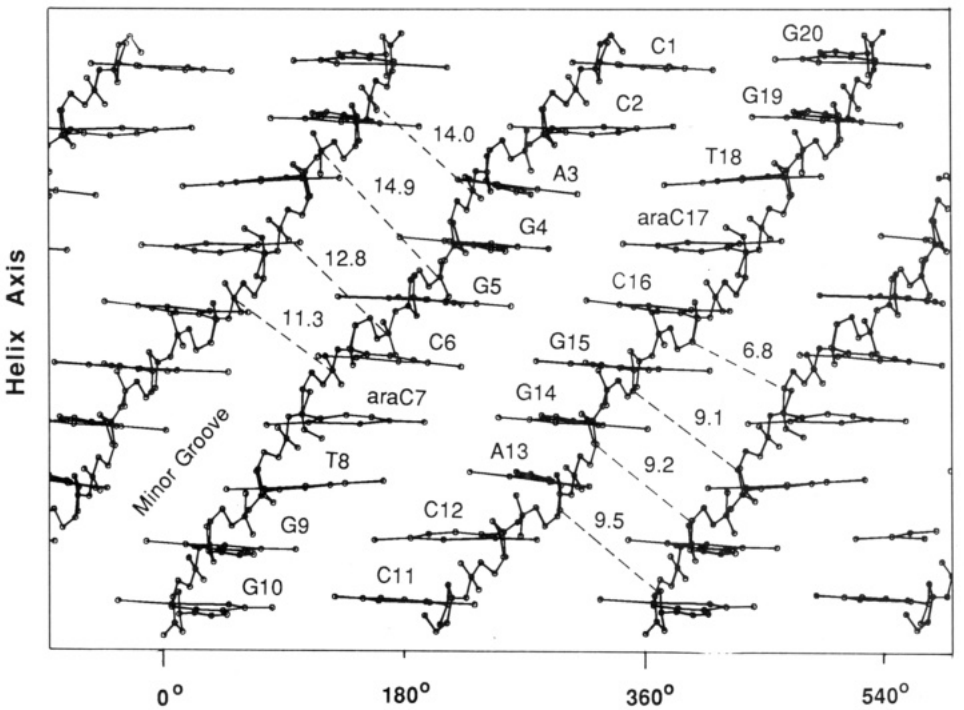


FIGURE 4: Cylindrical plot of the DNA decamer duplex. The variations in the minor groove width, as measured by the P–P distances (shown in the left side of the figure in angstroms) or O4′–O4′ distances (shown in the right side of the figure in angstroms) are marked. The vertical axis represents the helix axis, and the horizontal axis is the viewing angle seen from within the center of the helix.

viewing into the minor groove. The characteristic features of wide major groove and narrow minor groove are evident. Upon closer inspection, the minor groove width is found to vary somewhat in going from one end of the molecule to the other. This is more readily seen in the cylindrical projection of the araC-DECA double helix, shown in Figure 4. As indicated by the numbers in the figure, the groove width of the closest phosphorous–phosphorous distances across the minor groove ranges from 11.3 Å (P7–P17) to 14.9 Å (P5–P19). In fact, the closest P–P distance occurs between P8 and P17 (10.0 Å). To calculate the effective groove width in terms of van der Waals radii, it is customary to subtract 5.8 Å (the sum of the van der Waals radius of the two phosphates) from the P–P distances. Therefore, the effective minor groove of this decamer is 4.2 Å at its narrowest point. Another way to gauge the minor groove width is to measure the O4′–O4′ distance across the groove, as indicated in Figure 4. Consistent with result from the measurement of the P–P distance, the groove has a narrow width of 6.8 Å between C6 and araC17 nucleotides as measured by the O4′–O4′ distance. These data strongly suggest that, despite the appearance of a normal B-DNA conformation of this decamer, it has a subtle sequence-specific variation of DNA conformation. In fact, it is surprising to note that the area having the narrowest minor groove occurs in the central GGCC sequence. This is in contrast to the observations in DNA dodecamers related to

Table I: Torsion Angles (deg) of d[CCAGGC(araC)TGG][a]

base	α	β	γ	δ	ε	ζ	χ	ε−ζ	helix type
C1			57	120	−157	−95	−123	−63	B_I
			68	112	−157	−99	−134	−58	B_I
C2	−69	180	46	140	−107	175	−85	78	B_{II}
	−56	169	42	137	−94	157	−88	109	B_{II}
A3	−66	146	47	143	−174	−92	−81	−82	B_I
	−79	152	51	132	178	−70	−88	−112	B_I
G4	−47	178	20	151	−131	176	−75	53	B_{II}
	−79	180	43	141	−118	161	−84	81	B_{II}
G5	−57	142	44	146	−163	−128	−112	−35	B_I
	−57	149	37	140	165	−102	−104	−93	B_I
C6	−36	−179	22	137	−170	−90	−117	−81	B_I
	−42	−176	36	132	−162	−103	−102	−59	B_I
araC7	−48	167	42	105	−172	−88	−125	−84	B_I
	−49	168	38	107	178	−93	−114	−89	B_I
T8	−66	173	48	136	−99	163	−89	99	B_{II}
	−61	169	51	133	−96	161	−89	103	B_{II}
G9	−59	152	33	144	178	−88	−81	−90	B_I
	−57	146	50	159	−156	−109	−81	−47	B_I
G10	−76	168	56	120			−103		
	−72	171	46	112			−99		

[a] Torsion angles along the backbone of the oligonucleotide are defined as $P\underline{^{\alpha}}O5'\underline{^{\beta}}C5'\underline{^{\gamma}}C4'\underline{^{\delta}}C3'\underline{^{\epsilon}}O3'\underline{^{\zeta}}P$ and χ is the glycosyl angle. The nucleotides are numbered from C1 to G10 in one strand and C11 to G20 in the other strand. Torsion angles are calculated using the program NEWHELIX by Dickerson (1990). The numbers in the top row are from the araC-DECA structure, and those in the second row are from the HA-DECA structure.

the d(CGCGAATTCGCG) sequences in which the narrowest minor groove region is found in the central A–T sequences (Coll et al., 1987; DiGabriele et al., 1989; Fratini et al., 1981; Nelson et al., 1987; Wang & Teng, 1990). Therefore narrow minor grooves may occur either in the A–T sequence or in the G–C sequence. A common structural feature associated with the narrow minor groove appears to be the high propeller twist of the base pairs in these nucleotide segments. In the araC-DECA structure, many base pairs have significant propeller twisting, which is clearly visible in Figure 3. The details will be discussed more extensively later.

The objective of the present study is to determine the structural consequence due to the insertion of araC in B-DNA. Therefore, we are particularly interested in the disposition of the extra O2′ in B-DNA and its influence on the structure of B-DNA. In Figure 3, the O2′ hydroxyl groups of the araC7 and araC17 residues are seen to lie at the edges of the major groove. The C2′–O2′ vector is nearly on the horizontal plane that is perpendicular to the helix axis. The O2′ hydroxyl group is nestled in a small cavity, surrounded by two bases (araC7 and T8) and two phosphate groups (P7 and P8). This disposition of the O2′ hydroxyl group is better seen in the van der Waals drawing of the araC-DECA duplex (Figure 5). The O2′ oxygen of araC is in close contact (3.2 Å) with the methyl group of the thymine base of the T8 residue on the 3′-side of the DNA strand. This close contact, which is a little shorter than the sum of the van der Waals radii of methyl group (2 Å) and oxygen atom (1.4 Å), appears to push the T8 base slightly away from araC7 base, causing a small destacking between the two bases.

The overall similarity of the conformation of the two decamers (DECA and araC-DECA) is revealed by the small root-mean-square difference (RMSD = 0.45 Å) between the two structures. The helix type (B_I or B_{II} conformation) associated with each phosphate in the backbone remains unchanged. Most of the torsion angles do not vary by more than 20° (Tables I and II). However, there are some important deviations due to the presence of the extra O2′ hydroxyl group in the backbone. The major conformational changes are associated with the torsion angles of the internucleotide linkages between A3 and G4, and between G5 and C6. Specifically, ζ of A3, α of G4, and ε and ζ of G5 have variations greater than 20°. Interestingly, the conformation of the araC7 residue itself did not vary to a significant degree. The only notable change may be in its glycosyl χ angle, from −114° in HA-DECA to −125° in araC-DECA. It is interesting to note that the glycosyl torsion angle χ in all decamer structures in the *C*2 space group falls in two distinct narrow regions: the high-anti region near −84° (±5°) and the more normal anti

Table II: Some Helical Parameters of d[CCAGGC(araC)TGG][a]

base	τ_m	*P*	sugar pucker	ω	κ	prop. twist
C1	37	130	C1′-exo	29	7	−10
	46	118	C1′-exo	28	9	−6
C2	43	159	C2′-endo	50	−3	−7
	47	146	C2′-endo	51	0	−9
A3	36	186	C3′-exo	24	0	−8
	28	171	C2′-endo	24	2	−8
G4	49	165	C2′-endo	34	−14	−5
	54	145	C2′-endo	37	−9	−5
G5	49	157	C2′-endo	44	−8	−15
	33	167	C2′-endo	41	−3	−12
C6	31	171	C2′-endo	34	8	−15
	32	169	C2′-endo	37	3	−12
araC7	40	111	C1′-exo	24	14	−5
	37	109	C1′-exo	24	9	−5
T8	48	152	C2′-endo	50	0	−8
	42	151	C2′-endo	51	−2	−8
G9	34	197	C3′-exo	29	3	−7
	36	188	C3′-exo	28	0	−9
G10	40	127	C1′-exo		−7	−10
	26	106	C1′-exo		−9	−6

[a] Nomenclature of the helical parameters follows that of Dickerson et al. (1990): ω is the helical twist angle, and κ is the base pair buckle; prop. twist is the propeller twist angle; and τ_m and *P* are the pseudorotation amplitude and angle, respectively.

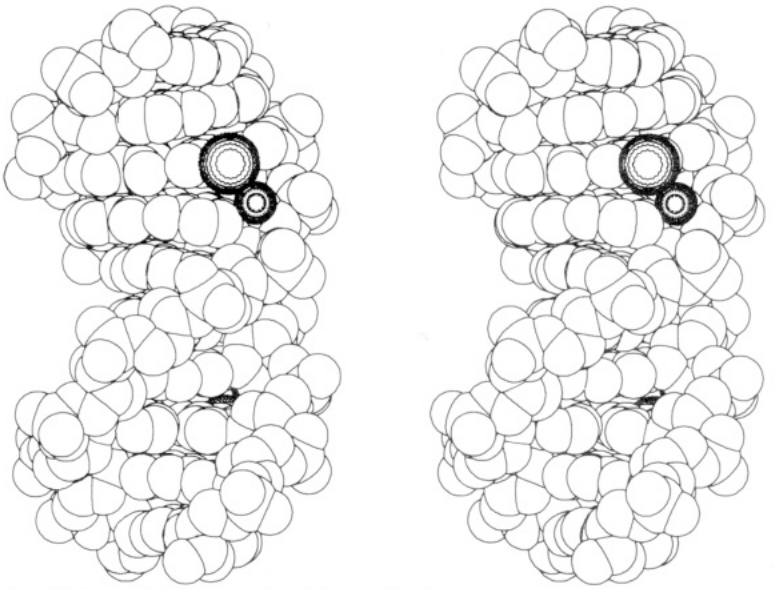


FIGURE 5: van der Waals drawing of the araC-DECA duplex. The O2′ hydroxyl groups (small heavily shaded circles) from the araC residues are located in the major groove of the B-DNA double helix. They are in close contact with the methyl groups (large circles with jagged lines) of the thymine bases.

FIGURE 6: Superposition of the two decamer structures (araC-DECA in thick bonds and HA-DECA in thin bonds) showing the structural perturbation due to the presence of the araC nucleotides. The two structures are superposed by a least-squares fitting using all atoms except the extra O2′ atoms in the araC-DECA structure. The O2′ hydroxyl groups are emphasized as large circles.

region of −113° (±8°). There is no apparent correlation (or anticorrelation) between the χ angles with other backbone torsion angles as has been suggested previously (Drew & Dickerson, 1981). The only pattern we could detect is that the high anti (ca. −84°) glycosyl angle seems to be associated with the nucleotide that is on the 5′-side of a purine nucleotide. However, this pattern is violated in the decamer d-(CCAACGTTGG) with the C5pG6 dinucleotide step. In that structure, the C5 residue is on the 5′-side of a purine nucleotide G6, and it has a χ value of −124°. Therefore, the exact cause of the clustering of the χ values in two regions remains to be determined.

araC Alters DNA Conformation

The effect of the conformational perturbation in the decamer DNA due to the incorporation of araC is more clearly seen by superimposing the HA-DECA and araC-DECA decamers onto each other (Figure 6). It can be seen that the largest movements in araC-DECA relative to HA-DECA occur in three areas. The araC7 residue is shifted away from the T8 residue, presumably due to the close contact between the O2′ of araC7 and the methyl of T8, as already shown in Figure 5. This has a domino effect of pushing its base-paired partner G14 away from the original position in the HA-DECA structure. The deoxyribose of the G14 residue along with its 3′-phosphate group move significantly by an average of 0.15 Å. This has a further rippling effect onto the next nucleotide, G15. The RMSD of the position of the G15 nucleotide (base, sugar, and phosphate) between the two structures is 0.24 Å.

The adjustment of the conformation in the araC-DECA structure seems to involve the movement of the entire motif of the nucleotide components. This is accomplished by a

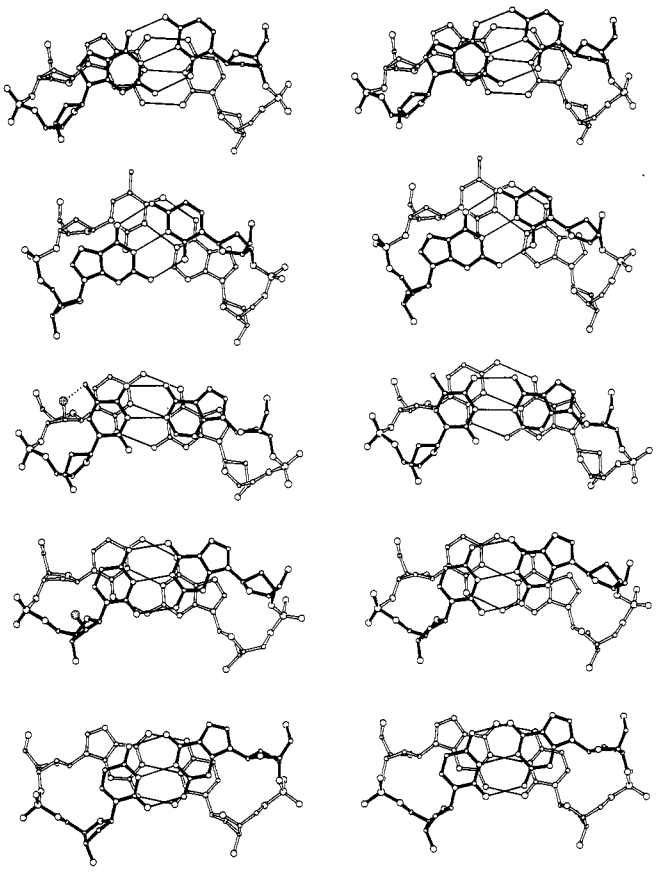


FIGURE 7: Stacking interactions in araC-DECA (left) and HA-DECA (right) B-DNA structures. There are only five unique dinucleotide steps in the decamer duplex because of the 2-fold symmetry in the molecule. The steps are, from top to bottom, C1pC2, C2pA3, A3pG4, G4pG5, and G5pC6.

combination of twisting the internucleotide phosphate linkage and changing the base-pair conformation. Therefore, in the former case (of phosphate conformation) the largest changes are located in the A3pG4 and G5pC6 steps (Table I). Another way to see it is that relatively small changes are found in the individual nucleoside conformations, such as sugar puckers (consequently, δ angle) or glycosyl angle χ (Tables I and II). All sugar moieties remain in the C2′-endo family conformation. The only significant changes are in the pseudorotation angles of A3 and G4 and the pseudorotation amplitude (τ_m) of G5. Additionally, the nucleosides reposition themselves as a result of the crank-shaft motion of the phosphate linkage. This tends to cause the base-pair conformation (e.g., propeller twist, buckle, etc.) to change. For example, the base pair associated with araC, araC7–G14, has a larger buckle (κ = 14°) relative to that of the corresponding base pair in the HA-DECA decamer (κ = 9°). Similarly, the C6–G15 base pair in the araC-DECA structure has both higher buckle (κ = −8°) and propeller twist (ω = −15°), as compared to the corresponding values of −3° and −12° in the canonical HA-DECA structure.

Base Stack. The changes in conformation can be comprehended more easily by comparing the stacking interactions of the two structures side by side, as shown in Figure 7. In both structures, there are sequence-dependent stacking interactions. For example, in this decamer there are three independent purine–purine steps, namely A3pG4, G4pG5, and G9pG10. It can be seen that the A3pG4 and G9pG10 steps have large ring–ring stacking areas between the two neighboring purines, but less ring–ring interactions between the two adjacent pyrimidines in the complementary strand. In contrast, the G4pG5–C16pC17 step has a diminished ring–ring stacking interaction in the purine–purine strand, while it has a similar stacking interaction in the pyrimidine–pyrimidine strand with the C4–N4 bond of C17 lying over the next cytosine ring of the C16 residue (Figure 7).

This variation of the stacking interactions of similar dinucleotide steps (in this case, purine–purine step) definitively affects the respective helical twist angles. Specifically, the three purine–purine steps, A3pG4, G4pG5 and G9pG10, have helical twist angles of 24° (24°), 34° (37°), and 29° (28°), respectively. Notice that the steps with low helical twist angles (<30°) have good purine–purine ring–ring stacks, and the step with normal twist angle (~36°) has a modest purine–purine ring–ring stack. The low helical twist angles of the A3pG4 and C1pC2 (i.e., G19pG20) steps may have a profound effect on the helical twist angle of the dinucleotide step C2pA3 that is sandwiched between them.

It has been pointed out that the C2pA3 step in the AH-DECA decamer has an unusually high twist angle of 51°, which was suggested as an intrinsic property of the CpA step (Heinemann & Alings, 1990). However, this should be taken cautiously in the context of surrounding sequences. For example, in the B-DNA structure of d(CGCAAATTTGCG), the CpA step has a rather normal twist angle of 35° (Coll et al., 1987). In the dodecamer case, the CA step is embedded in the tetranucleotide sequence of GCAA, whereas in the decamer case it is embedded in the CCAG sequence. It is likely that in the latter case the flanking CC and AG steps prefer to adopt low twist angles in order to enhance the ring–ring stacking interactions between two adjacent purines. To maintain an average twist angle of ~36° in B-DNA, the CpA step overwinds to a surprisingly high angle (ω = 51°) to compensate the underwound helical twists of the two flanking steps. The sequence-context effect of the helical twist angle associated with a particular dinucleotide step in B-DNA has recently been discussed extensively (Yanagi et al., 1991).

This type of alternating high-twist/low-twist step was first reported in the A-DNA structure of the DNA octamer d(GGCCGGCC) in which the CpG step has an unusually low twist angle of 20° (Wang et al., 1982). The flanking CC (and symmetry-related GG) step has a twist angle of 40°, substantially higher than the average twist angle of 33° in A-DNA. This was subsequently observed and reaffirmed in several more A-DNA structures, including d(CCCCGGGG) (Haran et al., 1987), d(GGGCGCCC) (Rabinovich et al., 1988), and d(ACCGGCCGGT) (Frederick et al., 1989). From the comparison of these two families of DNA conformation, it is interesting to note that the GpG step in B-DNA and A-DNA adopts a contrasting stacking pattern, resulting a lower than average helical twist angle ω in B-DNA, but a higher than average ω in A-DNA.

Environment around araC. The detailed environment surrounding the araC residue in the decamer is depicted in Figure 8. As mentioned eariler, the O2′ hydroxyl group is situated in a cavity enclosed by the bases and the sugar–phosphate backbone. The distance between O2′ and the H6 of its own cytosine base is 2.47 Å. This close contact may be responsible for the increase of the glycosyl torsion angle in the araC7 residue, from −110° to −125°. Furthermore, the H2′ and H4′ hydrogens of its own sugar are also close to the O2′ hydroxyl group. The distances between the O2′ and the methyl (on the C5 position) and the H6 of the T8 residue on the 3′-side of the nucleotide chain are 3.2 and 3.3 Å, respectively. This causes a kinking of the T8 base relative to the araC7

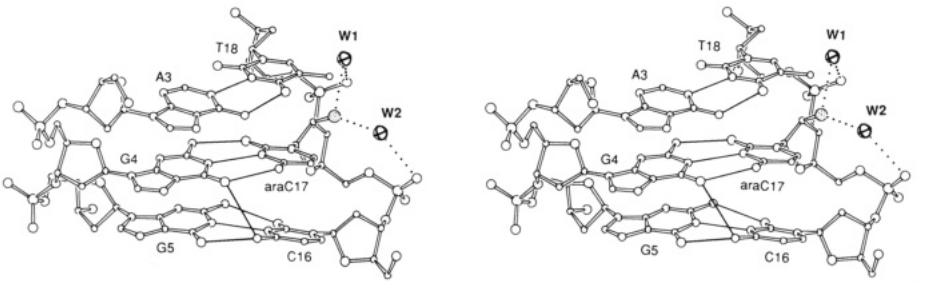


FIGURE 8: Detailed structure surrounding the araC residue in the d[CCAGGC(araC)TGG] B-DNA duplex. Two bridging first-shell water molecules (represented as ellipsoids) are shown. Notice that three-centered hydrogen bonds exist between the N4 of C16 and the O6 oxygen atoms of the two guanines (G5 and G4) in the complementary strand.

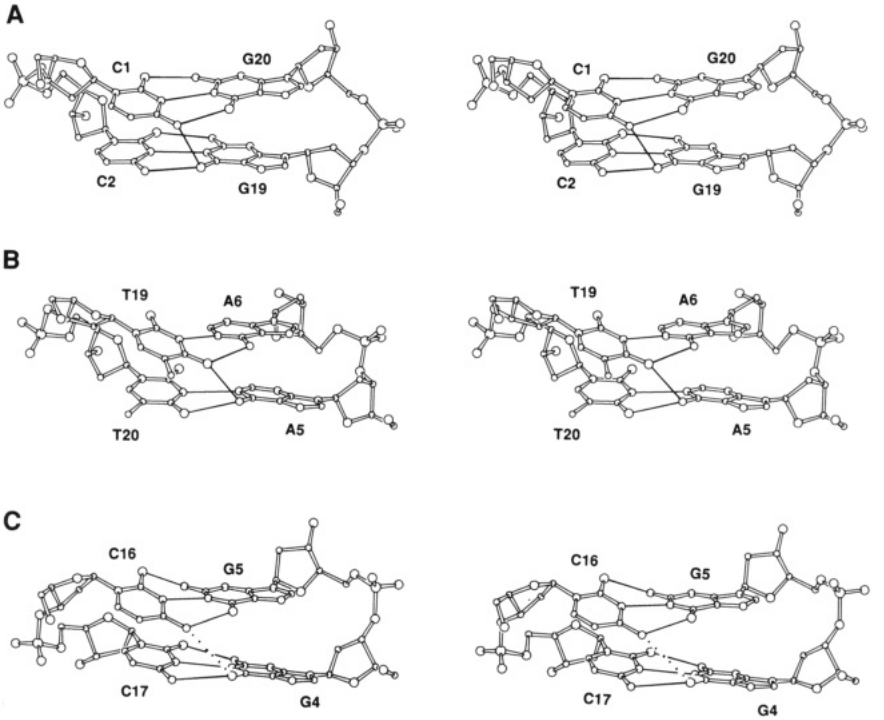


FIGURE 9: Three different types of three-centered hydrogen-bonded base pairs. (Top) GpG step in B-DNA from the G19pG20 of the d[CCAGGC(araC)TGG] structure (this work). (Middle) ApA step in B-DNA from the A5pA6 of the d(CGCAAATTTGCG) structure (Coll et al., 1987). (Bottom) GpG step in A-DNA from the G4pG5 of the d(ACCGGCCGGT) structure (Frederick et al., 1989).

cytosine base with a dihedral angle of 11.5°. Therefore, the O2′ is effectively surrounded by hydrophobic C–H hydrogen atoms, and it does not form direct intramolecular hydrogen bonds to other oxygen or nitrogen atoms in DNA. For example, the O2′ is 3.7 Å away from its own O5′ atom, a distance that is too long to be a hydrogen bond. In contrast, in the crystal structure of the free araC nucleoside, a strong hydrogen bond (2.65 Å) is found between the O2′ and O5′ oxygen atoms (Tougard & Lefebvre-Soubeyran, 1974). Instead, the O2′ in the decamer structure form hydrogen bonds to two first hydration shell water molecules, as seen in Figure 8. These water molecules bridge the O2′ to the O2P atoms of the phosphates in both 5′ and 3′ directions.

The presence of araC seems to affect the conformation of the associated base pairs. The araC7–G14 base pair has a higher buckle (−14°) in araC-DECA than the corresponding C7–G14 base pair (−9°) in HA-DECA. Similarly, the propeller twist of C6–G15 increases to −15°. Additionally, the G5–C6 step has a higher than normal helical twist angle of 44°. This combination of increased propellar twist and high helical twist angle brings the O2 oxygen atom of the C6 base closer to the O2 atom of the symmetry-related C16 base (the C6O2–C16O2 distance being 3.37 Å). A water molecule with strong electron density (with a temperature factor B of 25 Å^2, comparing to the averaged B of 36 Å^2 of all water molecules), located at the crystallographic 2-fold axis, forms strong hydrogen bonds (2.56 Å) to these two symmetry-related O2 oxygen atoms. This may stabilize the GpC dinucleotide step in its present conformation.

The increased propeller twist may also have the effect of further stabilizing the three-centered hydrogen bond involving the N4 of C6 residue. The distances of N4 of C6 to the O6 of G15 and O6 of G14 are 2.83 and 3.32 Å, respectively. This three-centered hydrogen bond in the major groove is also seen in the G19–G20 steps (Figure 9). The N4 of C1 is 2.88 and 3.28 Å away from the O6 of G20 and G19 residues, respectively.

The three-centered (bifurcated) hydrogen bond in the major groove of the DNA helix was initially discovered in the $(dA)_n$–$(dT)_n$ stretch of B-DNA dodecamers, as in d(CGCAAATTTGCG) (Coll et al., 1987) and d(CGCAAAAAAGCG)–d(CGCTTTTTTGCG) (Nelson et al., 1987). In those cases, the three-centered hydrogen bonds occur between the N6 of the adenine and O4 of two thymines from the complementary strand (Figure 9). It has been suggested that the three-centered hydrogen bonds in the A_n–T_n sequence may be related to the narrow minor groove in the same nucleotide stretch.

Interestingly, we have noticed that the G–C base pairs in the A-DNA structure of d(ACCGGCCGGT) have high propeller twist angles. For example, the G5–C16 and C2–G19 base pairs have high propeller twist angles of 22.4° and 17.2°, respectively (Frederick et al., 1989). One may wonder whether the three-centered hydrogen bonds also occur in this molecule. We found that in this structure the N4 of C16 residue is 2.93 and 3.56 Å from the O6 of the G5 and G4 bases, respectively. Therefore, the unique base pair stacking interaction in A-DNA has caused the adjacent guanine bases to slide slightly against each other (Figure 9C). This movement increases the distance between G4O6 and C16N4 beyond the reasonable hydrogen-bond range. It is likely that the three-centered hydrogen bond does not exist in A-DNA or RNA structures, despite the possibility of having high propeller twist angle. Figure 9 compares these three type of purine–purine steps in which bifurcated hydrogen-bonding interactions exist in the GG and AA steps in B-DNA (Figure 9A,B), but not in the GG step in A-DNA structure (Figure 9C).

The lack of direct intramolecular hydrogen bonds involving O2′ in B-DNA is in sharp contrast to that observed previously in the Z structure of d(CG)(araC)d(GCG) (Teng et al., 1989) and in the $(araC–dG)_3$ structures. Both of these have the intramolecular hydrogen bonds between the O2′ hydroxyl group of the arabinose and the NH_2 of guanine in the (dG)-p(araC) steps. In the mode of these intramolecular hydrogen-bonding interactions in the Z helix, the O2′ of araC(n+1) receives a hydrogen bond from N2 of G(n), and it simultaneously donates a hydrogen bond to its own O1P. In Z-DNA, the O2′ hydroxyl groups from the araC residues fill in the cleft surrounded by the (dG)p(araC) portion of the helix. This cleft is normally occupied by one or two well-defined water molecules in normal Z-DNA helix. Therefore, the O2′ hydroxyl groups perform the role of those water molecules and may stabilize the Z-DNA conformation.

Discussion

DNA conformation is polymorphic, depending on the nucleotide sequence as well as many extrinsic factors such as ions and pH (Rich et al., 1984). This point of sequence-dependent DNA conformation is nicely illustrated by the structure of two DNA decamers having identical base compositions, namely, d(CCAGGCCTGG) and d(ACCGGCCGGT). Despite the same core GGCC tetramer sequence in the center of the two decamers, they adopt different conformations. The decamer d(CCAGGCCTGG) (Heinemann & Alings, 1989) and its araC-modified d[CCAGGC(araC)TGG] (this work) adopt the B-DNA conformation. However, the decamer d-(ACCGGCCGGT), in which the A and T nucleotides move from the internal positions to the terminal positions, adopts a rather typical A-DNA conformation (Frederick et al., 1989). The crystallization conditions for these three decamers do not differ significantly. What might be the reason for this observation? We and others have suggested that the GG (and CC) sequence may have a preferred stacking interaction that favors the A-DNA conformation (Wang et al., 1982); McCall et al., 1984). We infer that the decamer d(ACCGGCCGGT), with a longer segment of contiguous GG (and CC) sequence, has a higher propensity to adopt the A-DNA conformation. In comparison, in the decamer d(CCAGGCCTGG), the A–T base pairs are sandwiched between CC and GG sequences, thereby disrupting the repeating CC/GG sequences. For this reason, this decamer adopts the B-DNA conformation.

In addition to the nucleotide sequences, many chemical modifications of DNA molecules can also have profound effects on their stability and reactivity. Some of these modifications have been exploited to design useful probes for studying the structure and function of DNA. For example, the replacement of the phosphodiester linkage with a thiophosphate or methylphosphate group alters the nuclease susceptibility of the nucleic acids (Matsukura et al., 1987; Smith et al., 1986). In fact, one can take advantage of the similarity in chemical structure between nucleoside analogues and natural nucleosides to use the former compounds as drugs. A recent timely example of utilizing these similarities involves the nucleoside analogue AZT, which is used to treat the AIDS disease caused by the HIV virus (Mitsuya et al., 1990). The molecular mechanism of AZT is associated with its ability to function as a chain terminator.

Another example is the clinically important anticancer drug araC, which is studied here. What is the molecular mechanism of the anticancer activity of araC? An important aspect may be related to its misincorporation into nucleic acid. Interestingly, araC has been shown to incorporate into DNA, but not RNA (Keating et al., 1982; Kufe & Spriggs, 1985; Mikita & Beardsley, 1988). Why does differential incorporation occur? Our structural analysis suggests that there may be profound differences in the conformational consequences due to the insertion of araC into various nucleic acid structures, including B-DNA, A-DNA, RNA, and DNA–RNA hybrid double helices. In the present araC-DECA structure, we note that in the decamer B-DNA there are close contacts between the O2′ hydroxyl group of araC and other atoms in DNA, causing subtle but significant structural distortions in the double helix.

These distortions may have important repercussions affecting the properties of DNA. They may alter the stability of the DNA double helix. To date, there have been no systematic studies of inserting araC into oligonucleotides and measuring their melting temperatures to assess the stability of the oligomer duplex. A recent NMR study of the DNA octamer d(CG)(araC)d(TAGCG) incorporating an araC showed that araC seemed to allow the octamer to form a hairpin structure readily under a very low salt conditions (Pieters et al., 1990). However, that study did not include a control experiment using the parent sequence of d-(CGCTAGCG). Therefore, it is not possible to compare the stability of the two octamers with and without araC.

Another possibility is that the extra O2′ hydroxyl group provides direct steric hindrance to proteins that interact with DNA. In Figure 5 it can be seen that the hydrophilic O2′ hydroxyl is tucked in a hydrophobic cavity. Consequently, DNA-binding proteins that have amino acid side chains interacting with this hydrophobic area would not bind to the araC-substituted DNA nearly as well. In addition, as mentioned earlier, the conformation of the nucleotides in the complementary strand (opposite to the araC strand) is more distorted. This will also significantly affect the binding of proteins (like polymerases). This interpretation is consistent with the observation made by Mikita and Beardsley (1988) who pointed out that DNA polymerase (Klenow fragment) is retarded at the araC site in DNA.

It should be noted that the extent of the close contact between the O2′ hydroxyl group and its neighboring nucleotide may be sequence-dependent. In the highly overwound C2pA3 step (ω = 52°), there would be little contact between the atoms of the A3 base on the 3′-side and the O2′ hydroxyl group, if the C2 residue is replaced by an araC residue. One would expect only small conformational change in this case. However, if the C2 residue is replaced by an rC residue in the CpA step, the O2′ hydroxyl group of rC may form a hydrogen bond

to the N7 site in adenine A3. This may be tested by solving the structure of the decamer d(C)r(C)d(AGGCCTGG). It is clear that the structural influence by a modified nucleoside residue depends on the exact starting conformation of the individual dinucleotide step. For example, we have recently determined the crystal structure of the covalent adduct between daunorubicin and an araC-containing hexamer, d[CG-(araC)GCG], cross-linked by formaldehyde (unpublished results). The structure contains an araCpG step in which the O2′ is in contact with the C8 H8 of the 3′-guanine base but exhibits little structural perturbation to the overall drug–DNA complex (Wang et al., 1991).

In contrast to the relatively small conformational perturbation in B-DNA induced by araC, the perturbation is expected to be significantly larger for the RNA double helix. When an araC is put into the RNA double helix, many severe close contacts between O2′ and other atoms (e.g., C6 of pyrimidine or C8 of purine, O4′ and O5′ of the 3′ residue) arise. Therefore, it is possible that due to these steric clashes, araC cannot be inserted into the resultant DNA–RNA hybrid (which has an RNA-11 helix conformation) during transcription (Mikita & Beardsley, 1988). This explains well the observation that araC is not incorporated into RNA molecules, presumably because RNA polymerase would not be able to insert an araC into the elongating RNA chain due to this steric hindrance.

In comparison, no such steric clashes were detected in araC-(Z-DNA). In fact, araC can be incorporated into a Z-DNA structure with minimum conformational distortion. The insertion of araC into alternating dC–dG sequence may actually facilitate the B to Z conversion by providing a strong intramolecular O2′–N2 hydrogen bond, thereby stabilizing the guanine in syn conformation. Whether there is any biological implication associated with araC-(Z-DNA) remains to be explored as discussed before (Teng et al., 1989).

In conclusion, we show that araC can be readily inserted into a B-DNA double helix. The structure of the araC-containing decamer conclusively demonstrates that the incorporation of araC into the B-DNA structure induces some conformational perturbations in the vicinity of the araC site that may hinder the binding interactions of relevant proteins such as polymerases. The long term goal of this research is to increase our repertoire of information concerning DNA oligomers incorporated with anticancer nucleosides including araC and 5-fluorouridine (Wang, 1987; Coll et al., 1989). By extending our structural studies to DNA oligomers containing other nucleoside analogues, we hope to enhance our understanding of the structure–function relationship of those anticancer nucleosides.

References

Coll, M., Frederick, C. A., Wang, A. H.-J., & Rich, A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8385–8389.

Coll, M., Saal, D., Frederick, C. A., Aymami, J., Rich, A., & Wang, A. H.-J. (1989) *Nucleic Acids Res. 17*, 911–923.

de Vroom, E., Roelen, H. C. P. F., Saris, C. P., Budding, T. N. W., van der Marel, G. A., & van Boom, J. H. (1988) *Nucleic Acids Res. 16*, 2987–3003.

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekman, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C.-S., Wang, A. H.-J., & Zhurkin, V. B. (1989) *Nucleic Acids Res. 17*, 1797–1803.

DiGabriele, A. D., Sanderson, M. R., & Steitz, T. A. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 1816–1820.

Drew, H. R., & Dickerson, R. E. (1981) *J. Mol. Biol. 151*, 535–556.

Fratini, A. V., Kopka, M. L., Drew, H. R., & Dickerson, R. E. (1981) *J. Biol. Chem. 257*, 14686–14707.

Frederick, C. A., Teng, M.-K., Quigley, G. J., Coll, M., van der Marel, G. A., van Boom, J. H., Rich, A., & Wang, A. H.-J. (1989) *Eur. J. Biochem. 181*, 295–307.

Gessner, R. G., Frederick, C. A., Quigley, G. J., Rich, A., & Wang, A. H.-J. (1989) *J. Biol. Chem. 264*, 7921–7935.

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

Heinemann, U., & Alings, C. (1989) *J. Mol. Biol. 210*, 369–381.

Hendrickson, W. A., & Konnert, J. (1979) in *Biomolecular Structure, Conformation, Function and Evolution* (Srinvasan, R., Ed.) pp 43–57, Pergamon, Oxford.

Keating, M. J., McCredie, K. B., Bodey, G. P., Smith, T. L., Gehan, E., & Freieich, E. J. (1982) *J. Am. Med. Assoc. 248*, 2481–2486.

Kufe, D. W., & Spriggs, D. R. (1985) *Semin. Oncol. 12*, 34–48.

Lauer, S. J., Pinkel, D., Buchanan, G. R., Sartain, P., Cornet, J. M., Krance, R., Borella, L. D., Caspaer, J. T., Kun, L. E., Hoffman, R. G., & Camitta, B. M. (1987) *Cancer 60*, 2366–2371.

Matsukura, M. K., Shinozuka, K., Zon, G., Mitsuya, H., Reitz, M., Cohen, J., & Broder, S. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 7706–7710.

Mikita, T., & Beardsley, G. P. (1988) *Biochemistry 27*, 4698–4705.

Mitsuya, H., Yarchoan, R., & Broder, S. (1990) *Science 249*, 1533–1544.

Nelson, H. C. M., Finch, J. T., Luisi, B. F., & Klug, A. (1987) *Nature 330*, 221–226.

Pieters, J. M., de Vroom, E., van der Marel, G. A., van Boom, J. H., Koning, T. M., Kaptein, R., & Altona, C. (1990) *Biochemistry 29*, 788–799.

Prive, G., Yanagi, K., & Dickerson, R. E. (1991) *J. Mol. Biol. 217*, 177–199.

Rabinovich, D., Haran, T. E., Eisenstein, M., & Shakked, Z. (1988) *J. Mol. Biol. 200*, 151–161.

Rich, A., Nordheim, A., & Wang, A. H.-J. (1984) *Annu. Rev. Biochem. 53*, 791–846.

Smith, C. C., Aurelian, L., Reddy, M., Miller, P. S., & Ts'o, P. O. P. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2787–2791.

Teng, M.-K., Liaw, Y.-C., van der Marel, G. A., van Boom, J. H., & Wang, A. H.-J. (1989) *Biochemistry 28*, 4923–4928.

Tougard, P., & Lefebvre-Soubeyran, O. (1974) *Acta Crystallogr. B30*, 86–89.

Wang, A. H.-J. (1987) in *Nucleic Acids and Molecular Biology* (Eckstein, F., & Lilley, D. M., Eds.) Vol. 1, pp 32–54, Springer, Berlin.

Wang, A. H.-J., & Gao, Y.-G. (1990) *Methods 1*, 91–99.

Wang, A. H.-J., & Teng, M.-K. (1990) in *Crystallographic and Modeling Methods in Molecular Design* (Bugg, C. E., & Ealick, S. E., Eds.) pp 123–150, Springer-Verlag, New York.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. A., & Rich, A. (1979) *Nature 282*, 680–686.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., van der Marel, G. A., van Boom, J. H., & Rich, A. (1981) *Science 211*, 171–176.

Wang, A. H.-J., Fujii, S., van Boom, J. H., & Rich, A. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 3968–3972.
Wang, A. H.-J., Gao, Y.-G., Liaw, Y.-C., & Li, Y.-K. (1991) *Biochemistry 30*, 3812–3815.
Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 184*, 119–145.
Yanagi, K., Prive, G., & Dickerson, R. E. (1991) *J. Mol. Biol. 217*, 201–214.

Characterization of Conformational Features of DNA Heteroduplexes Containing Aldehydic Abasic Sites†

Jane M. Withka, Joyce A. Wilde, and Philip H. Bolton*
Department of Chemistry, Wesleyan University, Middletown, Connecticut 06459
Abhijit Mazumder and John A. Gerlt
Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742
Received May 16, 1991; Revised Manuscript Received July 17, 1991

ABSTRACT: The DNA duplexes shown below, with D indicating deoxyribose aldehyde abasic sites and numbering from 5′ to 3′, have been investigated by NMR. The ^{31}P and ^{31}P–^{1}H correlation data indicate

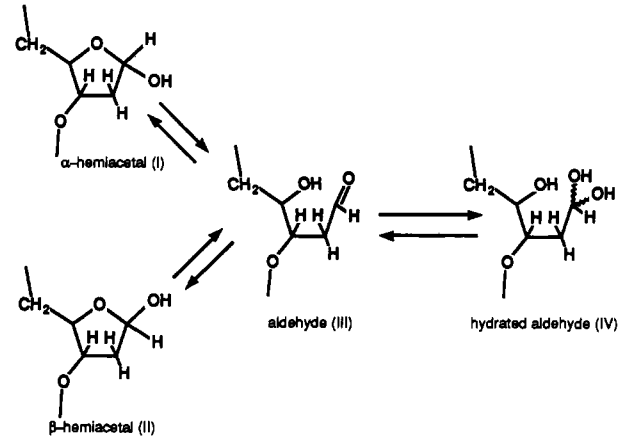
$d(C_1\ G_2\ C_3\ A_4\ G_5\ D_6\ C_7\ A_8\ G_9\ C_{10}C_{11})$
$d(G_{22}C_{21}G_{20}T_{19}C_{18}A_{17}G_{16}T_{15}C_{14}G_{13}G_{12})$

$d(C_1\ G_2\ C_3\ A_4\ G_5\ D_6\ C_7\ A_8\ G_9\ C_{10}C_{11})$
$d(G_{22}C_{21}G_{20}T_{19}C_{18}G_{17}G_{16}T_{15}C_{14}G_{13}G_{12})$

that the backbones of these duplex DNAs are regular. One- and two-dimensional ^{1}H NMR data indicate that the duplexes are right-handed and B-form. Conformational changes due to the presence of the abasic site extend to the two base pairs adjacent to the lesion site with the local conformation of the DNA being dependent on whether the abasic site is in the α or β configuration. The aromatic base of residue A17 in the position opposite the abasic site is predominantly stacked in the helix as is G17 in the analogous sample. Imino lifetimes of the AT base pairs are much longer in samples with an abasic site than in those containing a Watson–Crick base pair. The conformational and dynamical properties of the duplex DNAs containing the naturally occurring aldehyde abasic site are different from those of duplex DNAs containing a variety of analogues of the abasic site.

The in vivo repair of chemical damage to the bases in DNA as well as the misincorporation of bases that can occur during replication is frequently initiated by the hydrolysis of the N-glycosyl bond to yield an abasic site (Lindahl, 1982; Friedberg, 1985; Loeb & Preston, 1987). For example, the spontaneous deamination of cytosine to uracil occurs at a genetically significant rate, and the resulting uracil is removed by uracil-DNA glycosylase to produce an abasic site. The abasic site is an equilibrium mixture of α- (I) and β- (II) hemiacetals (2-deoxy-D-*erythro*-pentofuranoses), aldehyde (III), and hydrated aldehyde (IV), as depicted in Scheme I. The hemiacetal forms predominate, with about 1% aldehyde being present (Manoharan et al., 1988a; Wilde et al., 1989). The strand cleavage at the 3′-side of the abasic site catalyzed by UV endonuclease V of bacteriophage T4 or endonuclease III of *Escherichia coli* occurs via a syn β-elimination reaction (Manoharan et al., 1988b; Mazumder et al., 1989, 1991). The hydroxide-catalyzed reaction proceeds via an anti β-elimination reaction (Mazumder et al., 1991).

Scheme I



We have previously examined duplex heptamers containing aldehyde abasic sites (Manoharan et al., 1988a,b; Mazumder et al., 1989; Wilde et al., 1989), but these do not have sufficient thermal stability for detailed characterization by NMR. Thus, samples of the DNA heteroduplex

$d(C_1\ G_2\ C_3\ A_4\ G_5\ D_6\ C_7\ A_8\ G_9\ C_{10}C_{11})$
$d(G_{22}C_{21}G_{20}T_{19}C_{18}A_{17}G_{16}T_{15}C_{14}G_{13}G_{12})$

numbering from 5′ to 3′, have been prepared by methods

† This research was supported, in part, by Grant NP-750 from the American Cancer Society (P.H.B.) and by the Bristol-Myers Squibb Corp. via participation in a State of Connecticut Cooperative High Technology Research and Development Grant (P.H.B.). J.M.W. is a recipient of a traineeship in molecular biophysics via NIH 1T32 GM-08271 (P.H.B.).

0006-2960/91/0430-9931$02.50/0