Molecular structure of deoxycytidyl-3'-methylphosphonate (R$_P$) 5'-deoxyguanidine, d[Cp(CH$_3$)G]. A neutral dinucleotide with Watson-Crick base pairing and a right handed helical twist

Fusen Han, William Watt, David J.Duchamp*, Larry Callahan[+,1], Ferenc J.Kézdy[§,1] and Kan Agarwal[1]

Physical and Analytical Chemistry Research, The Upjohn Company, Kalamazoo, MI 49001 and [1]Departments of Chemistry and Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637, USA

ABSTRACT

The crystal structure of d[Cp(CH$_3$)G] has been determined as part of a project to study the mechanism of the B→Z transition in DNA. The asymmetric unit contains two dinucleotides and the equivalent of 7.5 water molecules, partially disordered over 12 definable positions. The two symmetry-independent dinucleotides form a duplex with Watson-Crick base-pairing and a right-handed helical sense. Comparison with previously determined structures of the B and A conformation showed that this duplex is closer to B than to A but significantly different from B. It corresponds to a stretched out helix with a 4 Å rise per base pair and a helical twist of 32°. This structure may serve as a model for the bending of DNA in certain situations. The configuration at the methyl phosphonate is R$_P$, and a mechanism, based on this assignment, is presented for the B→Z transition in DNA.

INTRODUCTION

Self complementary oligonucleotides are useful model systems for study of nucleic acid conformations. Dinucleotide monophosphates are important models for understanding the intrinsic base pairing ability of individual base pairs in an environment where the interactions with nearest neighbors are minimized. In the crystal structures of the ribodinucleotide monophosphates r[GpC] and r[ApU] (1, 2) the dinucleotides have conformations related to polynucleotide conformation, namely Watson-Crick base pairing and a backbone configuration that could be extended to an A-RNA like conformation. Deoxyribonucleotides in the crystal, however, have exhibited conformations not usually associated with large nucleic acids.

The deoxyribodinucleotide d[CpG] (3) exhibits a novel self base pairing interaction in the crystal. The dinucleotide d[Ap(CH$_3$)T], which contains a methyl phosphonate in place of the phosphate, forms a non-helical infinite hydrogen-bonded network involving reverse Hoogsteen base pairing (4). One might have expected that exchanging the phosphate for a methyl phosphonate would favor duplex formation by eliminating charge repulsion. Despite the d[Ap(CH$_3$)T] results, we felt that d[Cp(CH$_3$)G] might exhibit Watson-Crick pairing, since C-G base pairing is stronger than A-T base pairing. This indeed turned out to be the case.

Certain polynucleotides undergo a salt induced conformational change from the right-handed B conformation to the left-handed Z conformation (5−7). Crystal structures of oligonucleotides in the B (8) and Z (9) conformation have revealed structural details of the B and Z conformations, but have shed little information on the nature of the transition itself. To study this transition, we synthesized a number of hexanucleotide analogs of d(CG)$_3$ and found that only one of the two possible diastereomers of d[Cp(CH$_3$)G] when incorporated into hexanucleotides undergoes the B→Z transition (10). The absolute stereochemistry of the phosphonate group in these dinucleotides is important to understanding this conformational transition. One of the stereoisomers was successfully crystallized, and its structure was determined by low temperature X-ray diffraction techniques. In this paper we present and discuss the results of this crystallographic study. The synthetic work, the study of conformational transition behavior, and a preliminary report of the crystallographic work has been published (10).

EXPERIMENTAL

Crystals, grown by slow evaporation from 0.2M triethyl-ammonium acetate (pH 7.2), are tetragonal, space group

---

* To whom correspondence should be addressed

+ Present address: Division of Cytokine Biology, Center for Biologics, Food and Drug Administration, Bethesda, MD 20892, USA

§ Present address: Biopolymer Chemistry Research, The Upjohn Company, Kalamazoo, MI 49001, USA

Table 1. Positional parameters and B_{eq} or B(*) of the dinucleotides. Estimated standard deviations are in parentheses.

P	.2736(1)	.0717(1)	.7170(1)	3.3(1)	.3093(1)	.8396(1)	.5346(1)	3.3(1)
OP	.3049(3)	.0116(3)	.7402(2)	3.3(3)	.3470(3)	.9012(3)	.5179(2)	4.0(3)
CP	.1948(5)	.1034(5)	.7383(3)	3.5(4)	.2387(5)	.8097(5)	.5021(3)	3.6(4)
N_1C	.5002(3)	.2904(4)	.7202(2)	2.9(3)	.5341(4)	.6213(4)	.5501(2)	2.9(3)
C_2C	.4786(5)	.3563(5)	.7088(3)	3.6(4)	.5116(5)	.5530(6)	.5586(3)	4.2(4)
O_2C	.4164(3)	.3711(3)	.7085(2)	4.3(3)	.4479(4)	.5422(3)	.5561(2)	4.7(3)
N_3C	.5286(4)	.4040(4)	.6990(2)	3.6(3)	.5558(4)	.5027(4)	.5692(2)	3.3(3)
C_4C	.5961(5)	.3843(6)	.6966(2)	4.1(4)	.6241(6)	.5185(6)	.5728(3)	5.0(5)
N_4C	.6438(4)	.4333(4)	.6870(2)	3.6(3)	.6707(4)	.4679(4)	.5827(2)	3.7(3)
C_5C	.6170(5)	.3164(5)	.7041(3)	3.4(4)	.6492(5)	.5897(5)	.5655(3)	4.1(4)
C_6C	.5668(5)	.2711(5)	.7157(3)	4.0(4)	.6021(5)	.6373(5)	.5542(3)	3.9(4)
O_1,C	.4717(3)	.1905(3)	.7605(2)	3.9(3)	.5173(3)	.7298(3)	.5183(2)	3.8(3)
C_1,C	.4458(4)	.2405(4)	.7299(3)	3.0(4)	.4823(5)	.6714(5)	.5375(3)	3.6(4)
C_2,C	.4197(7)	.1979(7)	.6911(3)	6.3(6)	.4419(5)	.7041(5)	.5763(3)	3.3(4)
C_3,C	.3957(6)	.1316(5)	.7115(3)	4.6(5)	.4249(5)	.7754(5)	.5603(3)	3.9(4)
O_3,C	.3197(3)	.1402(3)	.7184(2)	4.0(3)	.3552(3)	.7724(3)	.5389(2)	3.5(3)
C_4,C	.4313(5)	.1282(5)	.7572(3)	4.3(5)	.4801(5)	.7927(6)	.5244(4)	5.3(6)
C_5,C	.4784(8)	.0669(9)	.7640(6)	9.8(10)*	.5209(10)	.8522(11)	.5243(6)	11.9(4)
O_5,C	.5229(8)	.0617(8)	.7298(8)	17.4(12)*	.5865(7)	.8505(7)	.5418(4)	14.0(3)
N_1G	.5064(4)	.3614(4)	.5826(2)	3.4(3)	.4878(4)	.5489(4)	.6896(2)	3.8(3)
C_2G	.4367(5)	.3471(5)	.5710(3)	4.0(4)	.4199(5)	.5662(5)	.6960(2)	3.7(4)
N_2G	.3984(4)	.4018(4)	.5593(3)	3.9(3)	.3756(4)	.5102(4)	.7037(2)	3.8(3)
N_3G	.4117(4)	.2840(4)	.5721(2)	3.1(3)	.3942(4)	.6293(4)	.6941(2)	3.6(3)
C_4G	.4580(5)	.2347(5)	.5837(3)	3.8(4)	.4445(5)	.6778(5)	.6868(2)	3.3(4)
C_5G	.5268(5)	.2446(5)	.5968(3)	3.5(4)	.5140(5)	.6668(5)	.6811(3)	3.6(4)
C_6G	.5552(5)	.3114(5)	.5951(2)	3.7(4)	.5411(5)	.6003(6)	.6821(2)	4.1(4)
O_6G	.6169(3)	.3317(3)	.6011(2)	3.6(3)	.6021(3)	.5773(3)	.6763(2)	4.1(3)
N_7G	.5571(4)	.1810(4)	.6072(2)	3.6(3)	.5490(4)	.7294(4)	.6740(2)	4.1(3)
C_8G	.5073(6)	.1371(6)	.5999(3)	5.0(5)	.4960(6)	.7746(5)	.6770(3)	4.2(4)
N_9G	.4462(4)	.1667(3)	.5862(2)	2.7(3)	.4328(4)	.7462(4)	.6833(2)	3.5(3)
O_1,G	.3302(3)	.1489(3)	.6059(2)	3.4(3)	.3245(3)	.7602(3)	.6520(2)	3.7(3)
C_1,G	.3824(5)	.1341(5)	.5742(3)	3.9(4)	.3681(5)	.7791(5)	.6883(3)	4.0(4)
C_2,G	.3831(6)	.0538(5)	.5725(3)	4.5(5)	.3668(5)	.8590(5)	.6866(3)	4.0(4)
C_3,G	.3052(5)	.0381(5)	.5745(3)	3.7(4)	.2892(5)	.8727(5)	.6768(3)	3.8(4)
O_3,G	.2765(3)	.0273(3)	.5315(2)	3.4(3)	.2506(3)	.8816(3)	.7167(2)	3.8(3)
C_4,G	.2719(5)	.1032(5)	.5941(2)	3.3(4)	.2635(5)	.8054(5)	.6537(3)	4.0(4)
C_5,G	.2279(5)	.0958(5)	.6348(3)	3.7(4)	.2357(5)	.8147(5)	.6080(3)	4.0(4)
O_5,G	.2644(3)	.0505(3)	.6671(2)	3.0(3)	.2834(3)	.8590(3)	.5825(2)	3.2(3)

Table 2. Positional parameters, B_{eq}, and population factors of the waters. Estimated standard deviations are in parentheses

	x	y	z	B_{eq}/B	Population	Alternate
O_{w1}	.7341 (6)	.6272 (5)	.6690 (2)	9.3 (5)	1.0	
O_{w2}	.7945 (5)	.4037 (9)	.6728 (4)	9.1 (7)	0.7	
O_{w3}	.6836 (5)	.7763 (4)	.4840 (3)	7.8 (5)	1.0	
O_{w4}	.7223 (6)	.2439 (6)	.6253 (4)	5.8 (6)	0.6	O_{w11}
O_{w5}	.6833 (8)	.1285 (7)	.6500 (4)	5.0 (7)	0.5	O_{w10}
O_{w6}	.8164 (5)	.4954 (7)	.6010 (4)	8.6 (6)	0.8	
O_{w7}	.3925 (8)	.8798 (8)	.4298 (4)	5.0 (7)	0.45	O_{w9}
O_{w8}	.5083 (6)	−.0373 (8)	.6317 (5)	7.6 (8)	0.6	
O_{w9}	.5577 (9)	−.0656 (7)	.7013 (4)	6.7 (7)	0.55	O_{w7}
O_{w10}	.6328 (7)	.1307 (8)	.6778 (4)	5.3 (7)	0.5	O_{w5}
O_{w11}	.7671 (10)	.2940 (13)	.6043 (5)	7.2 (11)	0.4	O_{w4}
O_{w12}	.6512 (12)	.7922 (11)	.6166 (12)	9.8 (16)	0.4	

$138°$ at $2°/min$ using $\theta-2\theta$ step scans with a scan width $> 3.4°$; 4087 reflections had intensities $>3\sigma$. Ten periodically monitored reflections showed no trend towards deterioration. $\sigma^2(I)$ was approximated by adding $(0.012\ I)^2$ to the standard deviations obtained from counting statistics, where the coefficient of I was calculated from the variations in intensities of the monitored reflections. Accurate cell parameters were determined by least squares fit of $CuK\alpha_1$ 2θ values $(\lambda(CuK\alpha_1) = 1.5402\ \text{Å})$ for 25 high 2θ reflections (11). Lorentz and polarization corrections appropriate for a monochromator with 50% perfect character were applied; no absorption correction was made.

A partial trial solution, (60 atoms), obtained using MULTAN80 (12), was extended using successive Fourier syntheses. Most hydrogens of the dinucleotides were found in difference maps close to positions generated using planar or tetrahedral geometry. Generated positions were used for all dinucleotide hydrogens, except hydroxyl hydrogens, which were excluded from the calculations. The water of solvation is partially disordered. A chemically reasonable model was developed, involving 12 water positions, 9 of which are occupied at any one time. A computer search showed no holes big enough to accomodate an additional water molecule. The two $5'$ end atoms, $O_{5'}$ and $C_{5'}$, of each dinucleotide are partially disordered.

Atomic coordinates and thermal parameters for nonhydrogen atoms, and an overall scale factor were refined by least squares. Anisotropic thermal parameters were used for all nonhydrogen atoms, except the $5'$ end atoms of one dinucleotide were held isotropic. Seven population factors for the disordered solvent were set manually by observing the behavior of the temperature factors of the water oxygens. Hydrogen parameters for the dinucleotide molecules, except hydroxyl hydrogens, were included in the calculations but were not refined. The function minimized was $\Sigma w\ (F_o^2-F_c^2)^2$, where $w = (\sigma(F_o^2))^{-2}$. Since they are most affected by deficiencies in the model used to represent the disorder, reflections with $\sin^2\theta/\lambda^2 < 0.05$ were excluded from the refinement, but not the agreement index. Atomic form factors for non-hydrogen atoms were from Doyle & Turner (13), and, for hydrogen, from Stewart, Davidson & Simpson (14). In the final refinement cycle, the largest shifts were about 1σ, except for the atoms in the disordered $5'$ end which had larger shifts $(<6\sigma)$. The final agreement index, R, was 0.109 for those 4087 reflections with $F_o^2>3\sigma(F_o^2)$; the final standard deviation of fit was 4.1. The CRYM system of computer programs was used (15). Table 1 contains final atomic coordinates of the dinucleotides along with the isotropic equivalent temperature factors calculated by

$B_{eq}=4/3\ (a^2B_{11} + b^2\ B_{22} + c^2\ B_{33} + ab\ \cos\gamma\ B_{12} + ac\ \cos\beta\ B_{13} + bc\ \cos\alpha\ B_{23})$

Table 2 gives the final positional, B_{eq}, and population parameters for the waters. Some of the waters were paired up, so that the population of one was constrained to be 1.0 minus the population of the other; paired waters are indicated as alternates in Table 2.

RESULTS

The asymmetric unit consists of two symmetry independent dinucleotides, base-paired in an anti-parallel way to form a short segment of double helix. This is the smallest (2-base pair) example of a deoxyoligonucleotide with a right-handed helical conformation (in the absence of an intercalator) to be determined by crystallographic methods. Having only two base pairs is a disadvantage for extrapolating results to DNA because the effect of the continuing helix is not present. On the other hand, the crystallographic asymmetric unit is small enough and the diffracting power of the crystals strong enough, so that high resolution 'small molecule' crystallographic methods could be used, producing much more precise atomic positions than the larger oligonucleotide examples of a B-DNA type conformation (8). Also atomic positions were refined without bond distance or angle constraints, producing independently determined atomic positions.

Bases

Bond distances (in Angstrom units) and bond angles (in degrees) for the two cytosine and the two guanine bases are shown in Figure 1. In each case the first number corresponds to dinucleotide 1, the second to dinucleotide 2. The average estimated deviations of bond lengths is 0.013 Å, and of bond angles is $1.0°$. These values compare well with guanine and cytosine bases in other high resolution crystal structures (1, 3, 16), except for the $C_4 - N_9$ distances of guanine which, in both dinucleotides, is about 0.04 Å shorter than the corresponding distance (1.377 Å) in the proposed standard for nucleic acid base residues (16).

Watson-Crick type base pairing is present in both base pairs. The hydrogen bond lengths in the base pairs, shown in Table 3, are within the normally expected range for hydrogen bonds of this type. Deviations from least-squares planes through the ring atoms of the bases are shown in Table 4. In both cytosines, O_2 and N_4 deviate significantly from the plane, and in the guanosines, O_6 shows a large deviation. These deviations may
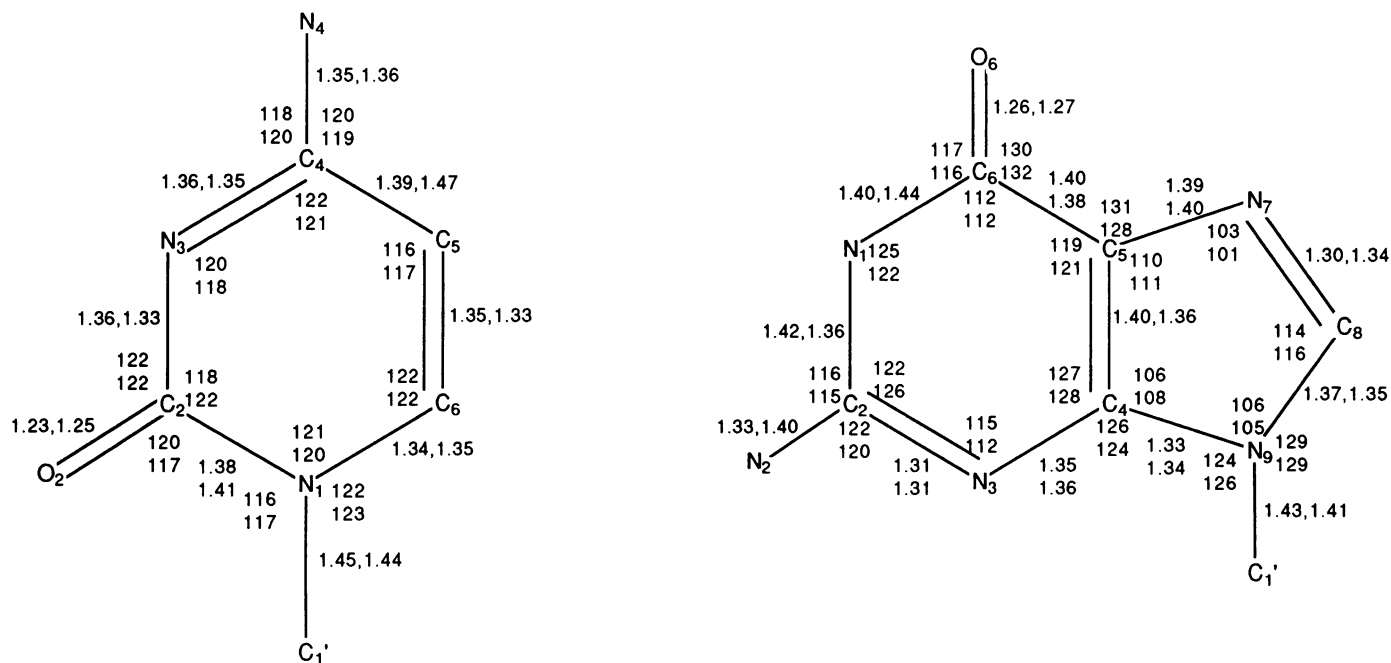
Figure 1. Bond distances(Å) and bond angles (°) for cytosine and guanine bases.

Table 3. Watson-Crick hydrogen bond lengths in base pairs

H-Bond (C-G)	Base Pair (C1-G2)	Base Pair (C2-G1)	r[GpC] (1)
O_2... $H-N_2$	2.80 (1) Å	2.87 (1) Å	2.86 Å
N_3... $H-N_1$	2.91 (1)	2.91 (1)	2.95
N_4-H... O_6	2.91 (1)	2.88 (1)	2.91

Table 4. Deviations from planes through the bases. Atoms used to calculate the plane are indicated by *

Atom	Cytosine 1 Deviation	Cytosine 2 Deviation	Atom	Guanine 1 Deviation	Guanine 2 Deviation
N_1*	0.04 Å	−0.01 Å	N_1*	−0.01 Å	−0.01 Å
C_2*	−0.04	0.01	C_2*	0.00	0.02
O_2	−0.14	0.06	N_2	−0.01	0.03
N_3*	0.01	0.00	N_3*	0.02	0.00
C_4*	0.02	−0.01	C_4*	0.00	0.00
N_4	0.08	−0.05	C_5*	0.03	0.00
C_5*	−0.02	0.01	O_6	−0.14	−0.04
C_6*	−0.01	0.00	C_6*	−0.02	0.00
$C_{1'}$	−0.03	−0.04	N_7*	0.02	0.00
			C_8*	−0.02	0.02
			N_9*	−0.02	−0.02
			$C_{1'}$	−0.10	0.01

be due to the base stacking interactions, since the two base pairs are not parallel, but pinch down on each other in the region of the Watson-Crick hydrogen-bonding.

Sugar and methylphosphonate dimensions

Distances and angles for the deoxyriboses and the methylphosphonates are shown in Figure 2. Standard deviations are as above (0.013Å and 1°), except those involving $C_{5'}$ and $O_{5'}$ which are 0.025 Å and 1.5°, and those involving P which

are 0.008 Å and 0.5°. In both cytidines, $C_{5'}$ and its hydroxyl group are partially disordered. Their temperature factors (Table 1) are much higher than those of other atoms in the dinucleotides. This disorder is also evident in the $C_{4'}$ − $C_{5'}$ and $C_{5'}$ − $O_{5'}$ bonds and in the angles associated with these atoms. All other bonds agree with previous small molecule crystal structure results, except the glycosyl bonds are about 0.03 − 0.04 Å shorter than the values given by Saenger (17).

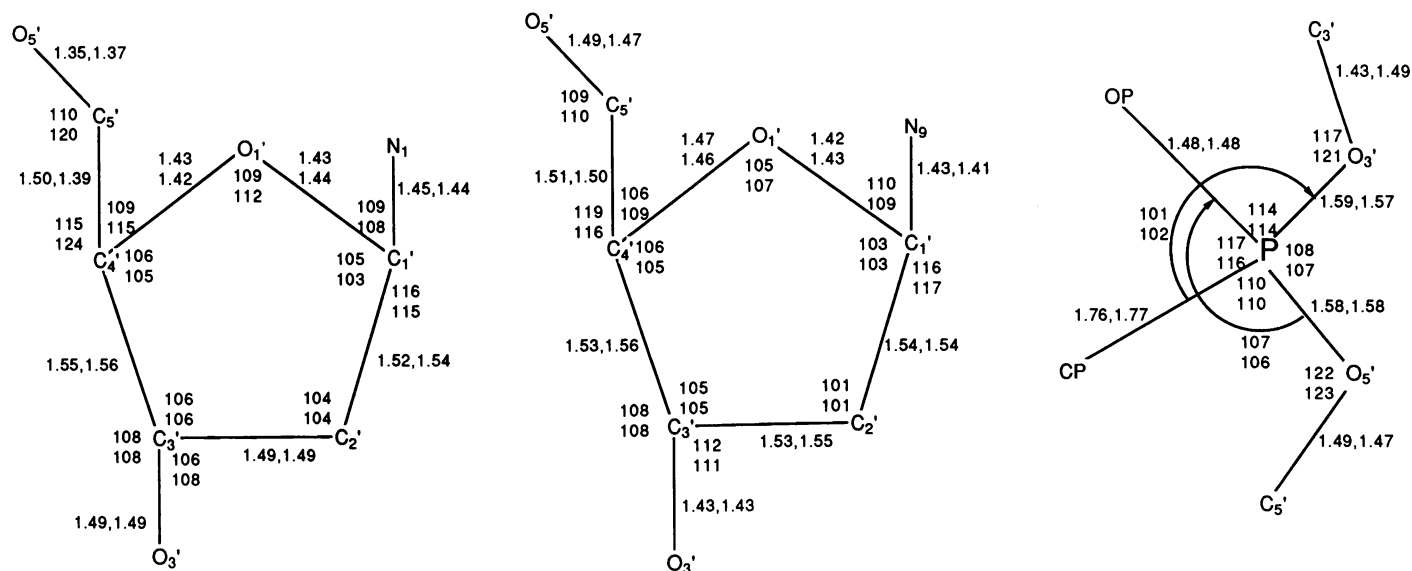The methylphosphonate bond angles (see Figure 2) differ

Figure 2. Bond distances(Å) and bond angles(°) for the cytosine deoxyriboses, guanine deoxyriboses and methylphosphonates.

Table 5. Endocyclic torsion angles and pseudorotation angle of deoxyriboses

Torsion Angle	Designation	Dinucleotide 1		Dinucleotide 2	
		Cytosine	Guanine	Cytosine	Guanine
$C_{4'}$-$O_{1'}$-$C_{1'}$-$C_{2'}$	ν_0	−32.6	−45.7	−29.0	−44.1
$O_{1'}$-$C_{1'}$-$C_{2'}$-$C_{3'}$	ν_1	31.5	40.9	31.7	41.9
$C_{1'}$-$C_{2'}$-$C_{3'}$-$C_{4'}$	ν_2	−19.1	−20.8	−23.8	−24.9
$C_{2'}$-$C_{3'}$-$C_{4'}$-$C_{1'}$	ν_3	0.5	−5.1	7.4	0.2
$C_{3'}$-$C_{4'}$-$O_{1'}$-$C_{1'}$	ν_4	20.2	31.8	14.1	27.4
Pseudorotation	P	125	117	137	124

somewhat between d[Cp(CH₃)G] and d[Ap(CH₃)T] (4). The largest bond angle about the phosphorus in each case is between the methyl and the nonesterified oxygen. In d[Cp(CH₃)G] the P − O₅′ and the P − O₃′ bonds are both 1.58 Å, near the value expected for esterified phosphate oxygens; the short P − O₅′ found in d[Ap(CH₃)T] is not observed here.

Endocyclic torsion angles of the deoxyribose rings are given in Table 5, along with the pseudorotation angle (18). The conformation of all four sugar rings is clearly $C_{1'}$-*exo*. The observed torsion angles fit theoretical curves (19) for furanose torsion angles: cytosine 1 and guanosine 2 match exactly $C_{1'}$-*exo*; guanosine 1 is slightly on the $O_{1'}$-*endo* side; and cytosine 2 is slightly on the $C_{2'}$-*endo* side. The $C_{1'}$-*exo* conformation is one of the four most common sugar pucker conformations (8), and is assumed by 11 of the 24 deoxyriboses in the d[CGCGAATTCGCG] dodecamer crystal structure (20).

Methyl phosphonate configuration and B→Z transition mechanism

In our previous communication (10), we reported that only one of two diastereomeric hexanucleotide analogs which contain a methylphosphonate moiety in place of the first 5′ phosphate group in d(CGCGCG) underwent the salt induced transition to Z-DNA. Since methylphosphonate linkages are chiral and uncharged with one of the nonesterified phosphate oxygens being replaced by a methyl group, the absence of the Z-DNA transition could be due to the lack of a hydrogen bond involving one of these

oxygens. The lack of the Z-DNA transition is due to a kinetic effect, since we could not determine any essential structural destabilization of Z-DNA or stabilization of B-DNA that would result from the replacement of the first 5′ phosphate with a methylphosphonate. Determination of the stereocheemistry of the methylphosphonate was essential for understanding the role of this oxygen in the B→Z transition. The absolute configuration is Rₚ for the dinucleotide the structure of which has been analyzed, and hence for the hexanucleotide that does not undergo the B→Z transition. In our previous communication, the methyl phosphonate configuration was mislabelled throughout. Therefore the dinucleotide which eluted first in the HPLC elution profile shown in Figure 2 of our previous work (10) has the Rₚ configuration.

How can this configuration account for the kinetic control of the B→Z transition? We simulated the B→Z transition using CPK models of the dinucleotide d(CpG). During this simulation we rotated the bases inward from the N₂-amino group of guanine. As the bases became unstacked, the N₂-amino group of guanine rotated past the prochiral Sₚ (prochiral with respect to methyl substitution) oxygen of guanosine's 5′ phosphate group, and formed a hydrogen bonding interaction with the prochiral Rₚ phosphate oxygen. This hydrogen bonding interaction subsequently broke as the bases restacked in the Z-conformation. This simulation cannot rule out the possibility of a water molecule being involved in this hydrogen bond interaction; however, it does indicate that the absence of this hydrogen bond would

kinetically prevent the R_p hexamer from assuming the Z-conformation. The prochiral R_p phosphate oxygen can provide an interaction that lowers the barrier to the B → Z transition, and at the same time, prevents the continued rotation of guanosine's 2-amino group past the 5' phosphate.

This mechanism provides an explanation for the lack of Z-DNA formation in the deoxyoligomer $d(GpC)_4$ (21) and larger oligomers $D(GpC)_{n>4}$ [Roger Jones, personal communication]. In these oligonucleotides the first guanosine lacks a 5' phosphate, and therefore, since nucleation of the B → Z transition occurs at the ends (22), these oligonucleotides are kinetically prevented from forming Z-DNA. This mechanism is consistent with the B → Z transition mechanism previously mentioned (10), including the role of dehydration in accelerating the rate of this transition.
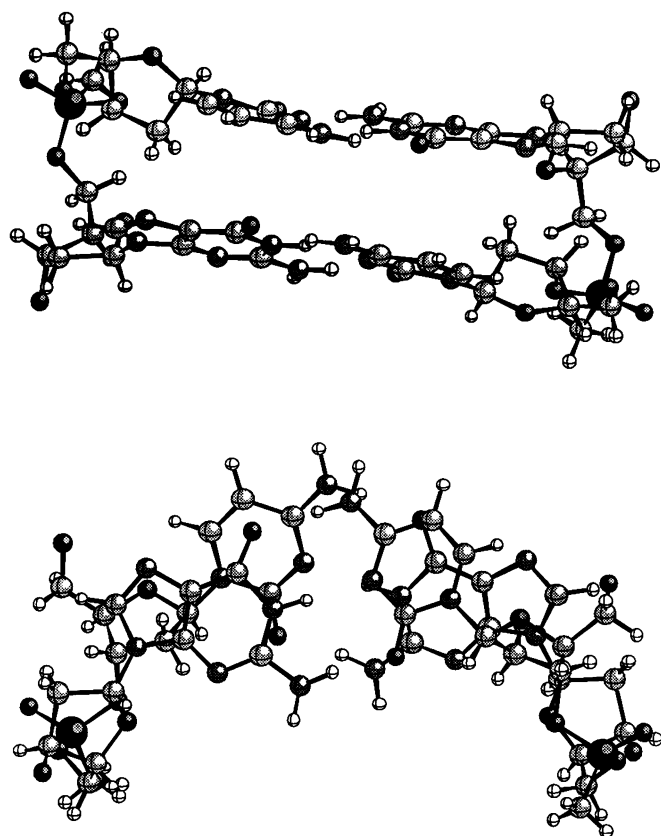

Figure 3. Drawings of the duplex viewed from the parallel direction to the base pairs (top) and perpendicular direction to the base pairs (bottom).

Overall duplex

Figure 3 shows the duplex viewed parallel and perpendicular to the axis of a helix relating the two base pairs. The distance between the $C_{1'}$ atoms of the deoxyriboses are 10.56 and 10.58 Å, respectively for base pair 1 (cytosine$_1$ − guanine$_2$) and base pair 2 (cytosine$_2$ − guanine$_1$). The corresponding distance is significantly longer, 10.67 Å, in r[GpC] (1), perhaps due to the repulsion between the negatively charged phosphates; d[Cp(CH$_3$)G], which is neutral, lacks this repulsion.

The overall conformation of the duplex is characterized by seven torsion angles, six, α, β, γ, δ, ϵ, and ζ, defined the backbone between phosphorus atoms, and the seventh, χ, describes the rotation about the glycosyl bond between the deoxyribose and the base. Table 6 gives these angles, along with the average values from the B-DNA dodecamer d[CGCGAATTCGCG] (19) and A-DNA average values from the A-DNA octamer d[GGTATACC] (20). [Cp(CH$_3$)G] is clearly closer to B-DNA than to A-DNA, since the A-DNA and B-DNA structures differ primarily at the δ angle, where this duplex agrees well with B-DNA.

Helical parameters were calculated by using the relationships among the vectors between the two $C_{1'}$ atoms and between the attached base nitrogens of one base pair and the corresponding vectors of the other base pair. These calculations give a helical twist of 32° and a rise per base pair of 4.0 Å, yielding 11.25 base pairs and 45 Å per turn of the helix. These may be compared to the values of 36° and 3.3 Å (10 base pairs and 33 Å per turn) for B-DNA (19), and 32.3° and 2.87 Å (11.2 base pairs and 32 Å per turn) for A-RNA (23).

A helix built using these parameters is shown in Figure 4, along with a similar number of base pairs of B-DNA built according to fiber data (Chandrasekaran, R. and Arnott, S., private communication). No attempt was made to model the ends of the duplex to bond with the adjacent duplexes in the helix. The helix is stretched out compared to B-DNA and is probably more flexible and less stable than the B-DNA helix.

Water structure and intermolecular contacts

A schematic illustration of the network of interduplex hydrogen bonds and hydrogen bonds to water is shown in Figure 5; three waters which alternate with other waters were eliminated from the diagram. Arrows in the schematic show probable directions of donation of hydrogens inferred from study of the structure. One of the hydrogen bonds, $N_4,C_2 → O_6G_1$ is an intraduplex Watson-Crick hydrogen bond. Four hydrogen bonds, involving the $O_{3'}$ hydroxyl group of the guanosines, between symmetry related duplexes, $N_2G_1 → O_3,G_2$, $O_3,G_2 → OP_1$, $N_2G_2 → O_3,G_1$, and $O_3,G_1 → OP_2$, could influence the conformation of the guanosine sugars. None of the stabilizing water bridges found

Table 6. Torsion angles (degrees) describing the overall duplex conformation

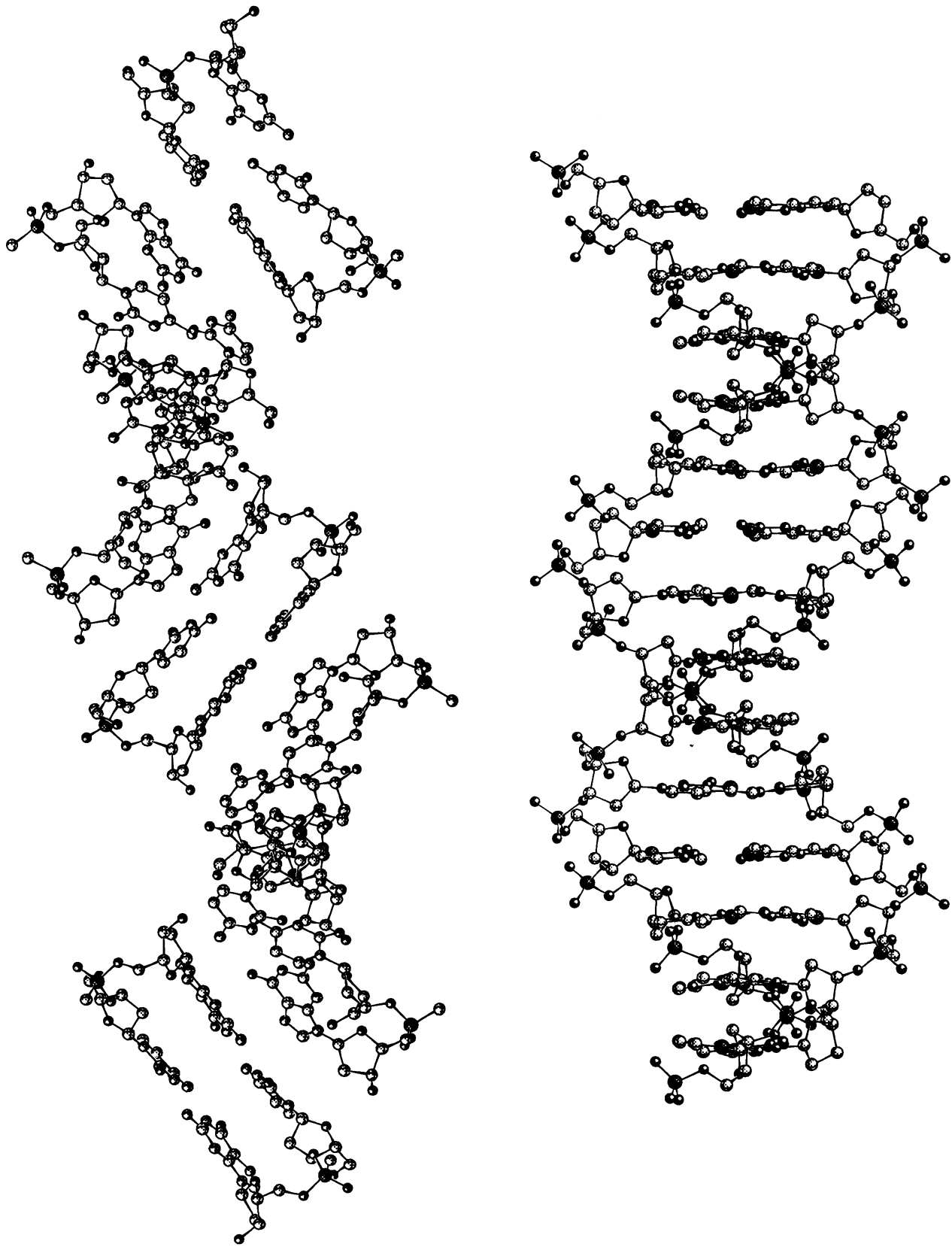| Angle	Designation	Dinucleotide 1	Dinucleotide 2	B-DNA	A-DNA
$O_{1'}$-$C_{1'}$-N_1-C_2	χ	−151.4	−163.3	−117	−160
$O_{5'}$-$C_{5'}$-$C_{4'}$-$C_{3'}$	γ	45.5	46.3	54	52
$C_{5'}$-$C_{4'}$-$C_{3'}$-$O_{3'}$	δ	116.1	121.4	123	88
$C_{4'}$-$C_{3'}$-$O_{3'}$-P	ϵ	−96.4	−89.1	−169	−152
$C_{3'}$-$O_{3'}$-P- $O_{5'}$	ζ	−77.0	−70.3	−96	−78
$O_{3'}$-P -$O_{5'}$-$C_{5'}$	α	−63.0	−65.7	−63	−62
P -$O_{5'}$-$C_{5'}$-$C_{4'}$	β	131.0	131.9	171	173
$O_{5'}$-$C_{5'}$-$C_{4'}$-$C_{3'}$	γ	50.2	95.7	54	52
$C_{5'}$-$C_{4'}$-$C_{3'}$-$O_{3'}$	δ	128.0	117.3	123	88
$O_{1'}$-$C_{1'}$-N_9-C_4	χ	−72.3	−68.7	−117	−160

Figure 4. Drawings of the helix built from dinucleotide (left) and B-DNA helix (right).
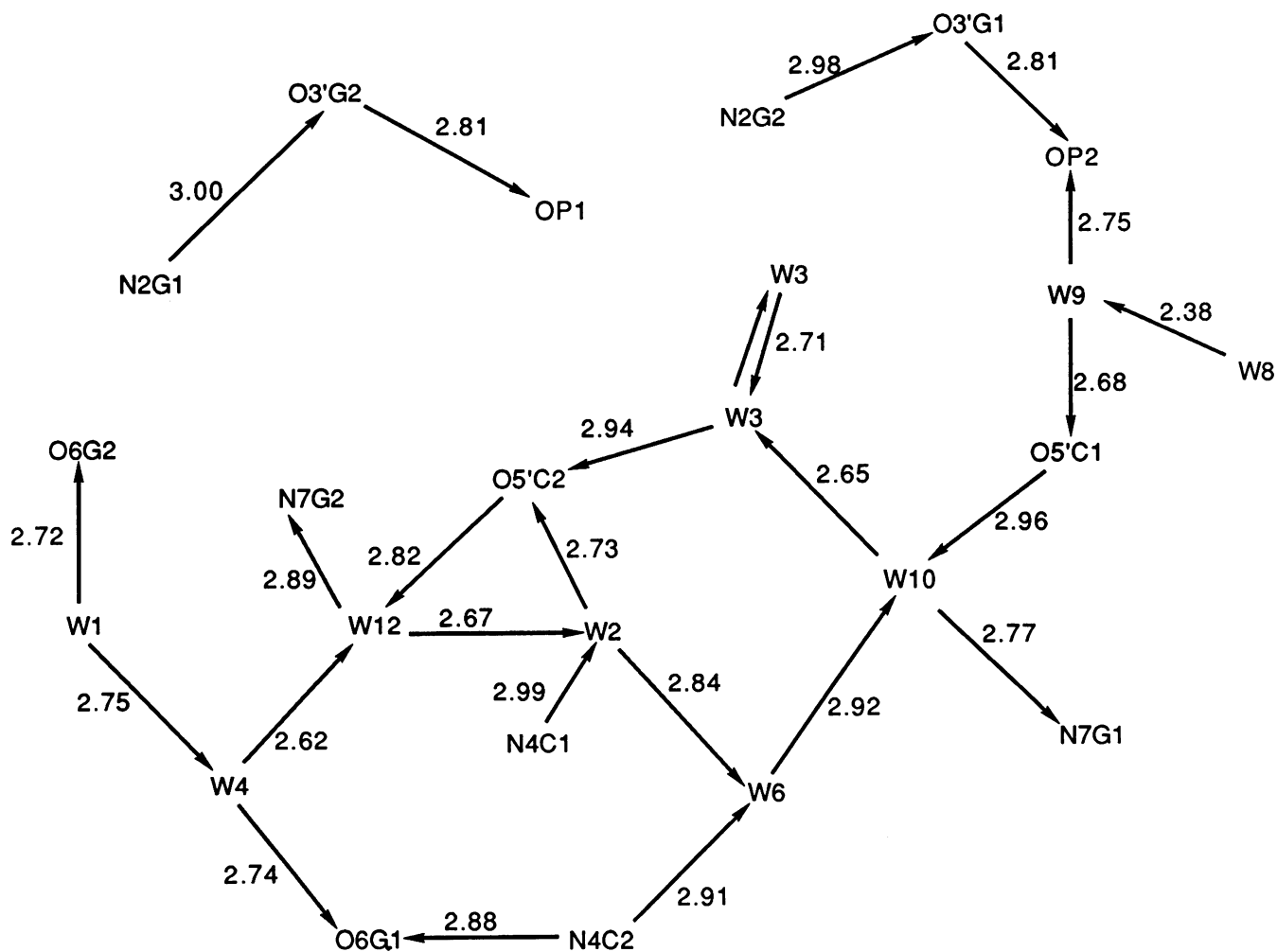
Figure 5. Schematic representation of the hydrogen bonding network.

in the B-DNA dodecamer d[CGCGAATTCGCG] (24) occur here, however, rings of hydrogen bonds found in many hydrated crystal structures are present.

Structural and biological relevance

Several features of this structure have structural and biological implications. This dinucleotide crystallized as a right-handed helix in which Watson-Crick base pairing was maintained, although the helical diameter is reduced significantly and the helical rise increased significantly from both B-DNA and A-DNA. This structure therefore is indicative of conformational flexibility that can occur within double-stranded DNA. A rise of 4 Å probably prevents this conformation from being maintained over regions of a DNA double helix longer than several base pairs, since stacking interactions would be reduced. Thus, a dinucleotide unit of this type could occur for short segments within a DNA helix, providing an interesting, easily accessible, distortion of the overall helical structure. Preliminary model building studies indicate that this dinucleotide unit can be readily placed within an idealized B-DNA helix without significant distortion of the backbone. The placement of this unit within the B-DNA helix leads to a 15° to 20° bend of the helix. Kinking of DNA has been previously postulated to occur by bending the dinucleotide toward the minor

groove (25), whereas our structure is bent toward the major groove of the DNA.

Kinking of DNA in chromatin has been popular as the explanation for the bending of DNA around a nucleosome core particle (26), although smooth bending of DNA could also account for this phenomenon. Histones and many other DNA binding proteins have regions of high positive charge density which are involved in interacting with one side of the DNA helix. The neutralization of charge has been postulated to be a driving force which causes the DNA to wrap around the histone core particle (27). It is tempting to speculate that the nonionic nature of our dinucleotide allows it to serve as a model system for the bending of DNA within chromatin. Since the bending of the dinucleotide makes the major groove more accessible for interactions, it is likely that a structure of this type might be found in the interaction of proteins with the major groove of the DNA.

In conclusion, this dinucleotide structure represents a model for short sections of polynucleotide structures. Its non-ionic nature makes it a valid model for regions of DNA where the negative charge density is reduced. Short regions of DNA may assume this bent conformation, perhaps under the influence of charge neutralizing cations, providing an easily accessible distortion which allows greater access to the major groove.

ACKNOWLEDGEMENTS

REFERENCES

1. Rosenberg,J.M., Seeman,N.C., Day,R.O. and Rich,A. (1976) *J. Mol. Biol.* **104**, 145−167.
2. Seeman,N.C., Rosenberg,J.M., Suddath,F.L., Kim,J.J.P. and Rich,A. (1976) *J. Mol. Biol.* **104**, 109−144.
3. Cruse,W.B.T., Egert,E., Kennard,O., Sala,G.B., Salisbury,S.A. and Viswamitra,M.A. (1983) *Biochemistry* **22**, 1833−1839.
4. Chacko,K.K., Linder,K., Saenger,W. and Miller,P.S. (1983) *Nucl. Acids Res.* **11**, 2801−2814.
5. Pohl,F.M. and Jovin,T.M. (1972) *J. Mol. Biol.* **67**, 375−396.
6. Vorlickova,M., Kypr,J., Stokrova,S. and Sponar,J. (1982) *Nucl. Acids Res.* **10**, 1071−1091.
7. Gaffney,B.L., Marky,L.A. and Jones,R.A. (1982) *Nucl. Acids Res.* **10**, 4351−4361.
8. Dickerson,R.E., Drew,H.R., Conner,B.N., Wing,R.M., Fratini,A.V. and Dopka,M.L. (1982) *Science* **216**, 475−485.
9. Wang,A.H.J., Quigley,G.J., Kolpak,F.J., Crawford,J.L., van Boom,J.H., van der Marel,G. and Rich,A. (1979) *Nature (London)* **282**, 680−686.
10. Callahan,L., Han,Fu-Son, Watt,W., Duchamp,D., Kézdy,F.J. and Agarwal,K. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1617−1621.
11. Duchamp,D.J. (1977) *ACS Symp. Ser.* **No. 46**, 98−121.
12. Main,P., Fiske,S.J., Hull,S.E., Lessinger,L., Germain,G., Declercq,J.P. and Woolfson,M.M. (1980) *MULTAN80. A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data.* Univs. of York, England, and Louvain, Belgium.
13. Doyle,P.A. and Turner,P.S. (1968) *Acta Cryst.* **A24**, 390−397.
14. Stewart,R.F., Davidson,E.R. and Simpson,W.T. (1965) *J. Chem. Phys.* **43**, 3175−3187.
15. Duchamp,D.J. (1984) *CRYM, a System of Crystallographic Programs.* The Upjohn Company, Kalamazoo, MI.
16. Taylor,R. and Kennard,O. (1982) *J. Am. Chem. Soc.* **104**, 3209−3212.
17. Saenger,W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, p. 70.
18. Altona,C. and Sundaralingam,M. (1972) *J. Am. Chem. Soc.* **100**, 8205−8212.
19. Levitt,M. and Warshel,A. (1978) *J. Am. Chem. Soc.* **100**, 2607−2613.
20. Dickerson,R.E. and Drew,H.R. (1981) *J. Mol. Biol.* **149**, 761−786.
21. Cosstick,R. and Eckstein,F. (1985) *Biochemistry* **24**, 3630−3638.
22. Pohl,F.M. and Jovin,T.M. (1972) *J. Mol. Biol.* **67**, 375.
23. Shakked,Z., Rabinovich,D., Kennard,O., Cruse,W.B.T., Salisbury,S.A. and Viswamitra,M.A. (1983) *J. Mol. Biol.* **166**, 183−201.
24. Kopka,M.L., Fratini,A.V., Drew,H.R. and Dickerson,R.E. (1983) *J. Mol. Biol.* **163**, 129−146.
25. Crick,F.H.C. and Klug,A. (1975) *Nature* **255**, 530−533.
26. Mirzabekov,A.D. and Rich,A. (1979) *Proc. Natl. Acad. Sci. USA* **75**, 4185−4188.
27. Taylor,R. and Kennard,O. (1982) *J. Am. Chem. Soc.* **104**, 443−444.