Crystal Structure of Guanylyl-2′,5′-Cytidine Dihydrate: An Analogue of msDNA-RNA Junction in *Stigmatella aurantiaca*

R. KRISHNAN and T. P. SESHADRI*

Department of Physics, Indian Institute of Science, Bangalore 560012, India

SYNOPSIS

Sequence analysis of msDNA from bacterium such as *Stigmatella aurantiaca, Myxococcus xanthus* and *Escherichia coli B* revealed that the guanine residue of the single-stranded RNA is linked to the cytosine residue of the msDNA through a 2′-5′ instead of a conventional 3′-5′ phosphodiester bond. We have now obtained the crystal structure of the self-complementary dimer guanylyl-2′,5′-cytidine (G2′p5′C) that occurs at the msDNA–RNA junction. G2′p5′C crystallizes in the orthorhombic space group $P2_12_12_1$ with $a = 8.376(2)$, $b = 16.231(5)$, $c = 18.671(4)$. CuKα intensity data were collected on a diffractometer in the $\omega - 2\theta$ scan mode. The amount of 1699 out of 2354 reflections having $I \geq 3\sigma(F)$ were considered observed. The structure was solved by direct methods and refined by full-matrix least squares to a R factor of 0.054. The conformation of the guanine base about the glycosyl bond is syn ($\chi_1 = -54°$) and that of cytosine is anti ($\chi_2 = 156°$). The 5′ and 2′ and ribose moieties show C2′-*endo* and C3-*endo* mixed puckering just like in A2′p5′A, A2′p5′C, A2′p5′U, and dC3′p5′G. Charge neutralization in G2′p5′C is accomplished through protonation of the cytosine base. An important feature of G2′p5′C is the stacking of guanine on ribose 04′ of cytosine similar to that seen in other 2′-5′ dimers. G2′p5′C, unlike its 3′-5′ isomer, does not form a miniature double helix with the Watson–Crick base-pairing pattern. Comparison of G2′p5′C with A2′p5′C reveals that they are isostructural. A branched trinucleotide model for the msDNA–RNA junction has been postulated. © 1994 John Wiley & Sons, Inc.

INTRODUCTION

Abundance of 3′-5′ phosphodiester bonds in naturally occurring nucleic acids is well known. The 2′-5′ are second in importance only to 3′-5′ nucleic acids. A large excess of 2′-5′ oligomers had been reported to be produced in simulated prebiotic experiments. But the yield of 3′-5′ oligomers improved in the presence of some cations.[1] Experiments on nonenzymatic template directed synthesis performed with a view to understanding the chemical evolution showed preferential formation of 2′-5′ oligomers. A theoretical cyclic pathway model for the enrichment of 3′-5′ bonds from a prebiotic random copolymers containing a mixture of 2′-5′ and 3′-5′ bonds has been postulated.[2]

The 2′-5′ oligo As produced from ATPs by (interferon-induced) adenylate synthetase are known to be responsible for the antiviral action. It was concluded from an analysis of the experimental data on a number of phosphodiester isomers that the 2′-5′ phosphodiester bond of 2′-5′A is indispensible for its binding to the 2–5A dependent endonuclease (RNase-L, RNase-F). Further, it not only demonstrated that 2–5A was ten times more reactive than 3–5A in its binding to endonuclease or as an inhibitor of protein synthesis, but the substitution of even one 2′-5′ bond by a 3′-5′ link reduced the activity by one order of magnitude.[3]

Bacterium such as *Stigmatella aurantiaca, Myxococcus xanthus,* and *Escherichia coli* B, are known to contain msDNA.[4] Sequence analysis of this msDNA revealed that the guanine residue of a single-stranded RNA is linked to the cytosine residue of msDNA through a 2′-5′ instead of a conventional 3′-5′ phosphodiester bond.[4] Biochemical studies have shown that the substitution of the rG by C or A residue in the pre msDNA completely blocks the msDNA synthesis by reverse transcriptase, revealing the importance of the rG residue.[5]

The double-helical model of DNA has been widely

Biopolymers, Vol. 34, 1637–1646 (1994)

CCC 0006-3525/94/121637-10

* Author to whom correspondence should be addressed.

accepted. However, nonuniformity and local distortions in the DNA double helix have been envisaged for explaining the higher order organization of DNA and its interactions with proteins. Needless to say, evidences started emerging with the proliferation of 3′-5′ oligomer structures following the development of the triester method[6] synthesis of nucleic acid fragments. Besides several unusual features such as looped out,[7] unpaired,[8] and mismatched base pairs,[9] a novel tetraplex helical structure with hairpin loops has been reported for the oxytricha telomeric DNA.[10] In recent years, remarkable progress has been made in elucidating the anatomy of 3′-5′ oligomers. On the other hand, not much attention has been devoted to the stereochemistry of 2′-5′ nucleic acids. We have recently reported the crystal structures of cytidylyl-2′-5′-adenosine (C2′p5′A)[11] and adenylyl-2′,5′-adenosine (A2′p5′A).[12] The Pyr-Pur sequenced C2′p5′A forms a miniature parallel-stranded double helix whereas the Pur-Pur A2′p5′A shows a single-stranded helical conformation similar to its 3′-5′ analogue. The x-ray studies on guanylyl-2′,-5′-cytosine was taken in continuation of our earlier studies on 2′,5′-dinucleoside monophosphates. In addition, to determine whether it forms a miniature duplex with Watson–Crick base pairing like its 3′-5′ isomer[13], to compare its conformational features with the only other 2′-5′ self-complementary dimer (A2′p5′U)[14] reported so far and to unravel the effect of base sequence on the stereochemistry of 2′-5′ dimers.

EXPERIMENTAL

Crystallization and Data Collection

G2′p5′C was purchased from Sigma Chemical Co., USA, and was used without further purification. The dinucleoside monophosphate crystals were grown by slow evaporation from aqueous solutions (pH 4.5) of the compound. Unit cell dimensions were calculated from the measurements made on the oscillation and Weissenberg photographs. Density measurements made using bromoform/alcohol mixtures indicated the presence of one G2′p5′C and two water molecules in the asymmetric unit. A crystal measuring 0.04 × 0.4 × 0.4 mm mounted on the tip of a glass fiber was used for data collection. Three-dimensional Ni-filtered CuKα intensity data were collected on a CAD-4 diffractometer in the $\omega - 2\theta$ scan mode upto a resolution of 0.85 Å ($\sin\theta/\lambda = 0.59$ Å^{-1}). The two reflections 2 2 2 and 2 2 3 monitored periodically during the data collection showed less than 4% variation in their intensities, indicating the crystal and instrument stability. The data was corrected for Lorentz and

Table I Crystal Data of Guanylyl-2′,5′-Cytidine · Dihydrate

Formula	$C_{19}H_{24}H_8O_{12}P \cdot 2H_2O$
Molecular weight	623.41
Space group	$P2_12_12_1$
Cell dimensions	a = 8.376(2) Å
	b = 16.231(5) Å
	c = 18.671(4) Å
Volume	2538.3 Å^3
	h 0 0 h odd absent
	0 k 0 k odd absent
Systematic absences	0 0 1 1 odd absent
Z	4
D_{calc}	1.63 mg/m^3
D_{meas}	1.62 mg/m^3
μ (CuKα)	19.2 cm^{-1}
Crystal size	0.02 × 0.4 × 0.4 mm
$\sin\theta_{max}/\lambda$	0.59 Å^{-1}
Data collection mode	$\omega - 2\theta$
Ranges of hkl	$0 \le h \le 10$
	$0 \le k \le 19$
	$0 \le 1 \le 22$
Total number of reflections	2354
Number of reflections with $I \ge 3\sigma(F)$	1699
R	5.4%
R_w	5.5%

Table II Final Atomic Coordinates with Estimated Standard Deviations (ESD's) in Parentheses and Equivalent Isotropic Values of the Anisotropic Thermal Parameters of non-H Atoms[a]

Atom	x	y	z	B_{eq}[a]
N9	−0.446(1)	−0.5060(4)	−0.0648(4)	2.1(2)
C8	−0.555(1)	−0.5682(6)	−0.0486(5)	2.7(2)
N7	0.585(1)	−0.6162(5)	−0.1021(4)	3.2(2)
C6	−0.480(1)	−0.6101(7)	−0.2299(5)	3.0(3)
O6	−0.546(1)	−0.6690(1)	−0.2623(8)	3.2(3)
C5	−0.495(1)	−0.5845(6)	−0.1577(4)	2.4(2)
C4	−0.406(1)	−0.5166(5)	−0.1350(5)	2.0(2)
N3	−0.305(1)	−0.4703(4)	−0.1754(4)	1.9(2)
C2	−0.294(1)	−0.4939(5)	−0.2424(4)	2.0(2)
N2	−0.194(1)	−0.4563(5)	−0.2897(4)	3.9(2)
N1	−0.380(1)	−0.5587(5)	−0.2696(4)	2.4(2)
C1'	0.381(1)	−0.4484(5)	−0.0139(5)	2.2(2)
C2'	−0.205(1)	−0.4588(5)	0.0021(5)	1.9(2)
C3'	−0.163(1)	−0.3739(6)	0.0308(5)	2.1(2)
C4'	−0.262(1)	−0.3170(6)	−0.0184(5)	2.2(2)
C5'	−0.178(1)	−0.2831(6)	−0.0809(5)	2.8(2)
O5'	−0.106(1)	−0.3459(4)	−0.1236(3)	3.6(2)
O4'	−0.397(1)	−0.3677(4)	−0.0413(3)	2.7(2)
O3'	−0.216(1)	−0.3690(5)	0.1024(4)	3.1(2)
O2'	−0.183(1)	−0.5240(3)	0.0522(3)	2.0(1)
P	−0.0244(3)	−0.5780(1)	0.0501(1)	1.90(5)
O1	−0.049(1)	−0.6423(4)	0.1061(3)	2.6(2)
O2	0.119(1)	−0.5233(4)	0.0562(3)	2.9(2)
O5'C	−0.019(1)	−0.6157(3)	−0.0277(3)	2.2(2)
O4'C	−0.119(1)	−0.6496(3)	−0.1723(3)	2.2(2)
O3'C	0.120(1)	−0.8162(4)	−0.1060(4)	3.1(2)
O2'C	0.006(1)	−0.8056(3)	−0.2439(3)	2.5(2)
C5'C	−0.125(2)	−0.6821(8)	−0.0433(7)	3.1(3)
C4'C	−0.089(1)	−0.7143(5)	−0.1200(5)	2.1(2)
C3'C	0.081(1)	−0.7387(5)	−0.1363(5)	1.9(2)
C2'C	0.085(1)	−0.7344(5)	−0.2176(5)	2.1(2)
C1'C	−0.017(1)	−0.6582(5)	−0.2312(5)	2.2(2)
C6C	0.106(1)	−0.5282(5)	−0.1843(5)	2.1(2)
C5C	0.194(1)	−0.4594(6)	−0.1919(5)	2.3(2)
C4C	0.264(1)	−0.4433(5)	−0.2594(4)	2.0(2)
N4C	0.359(1)	−0.3809(5)	−0.2719(5)	2.9(2)
N3C	0.233(1)	−0.4955(4)	−0.3144(4)	2.2(2)
C2C	0.137(1)	−0.5659(6)	−0.3066(4)	2.4(2)
O2C	0.110(1)	−0.6100(4)	−0.3585(3)	3.3(2)
N1C	0.078(1)	−0.5824(4)	−0.2399(3)	2.2(2)
OW1	−0.586(1)	−0.2790(3)	−0.1474(3)	4.6(3)
OW2	0.242(1)	−0.3864(6)	−0.0124(5)	6.2(3)

[a] $B_{eq} = 4/3 \sum_i \sum_j \beta_{ij} a_i a_j$.

polarization effects but not for absorption. Crystal data is presented in Table I.

Structure Solution and Refinement

This dinucleoside structure was solved by direct methods (SHELXS-86). All the molecular atoms were located unambiguously from the Fourier map computed with the phases of 668 reflections (FOM = 0.107). Full-matrix least-squares refinement with isotropic temperature factors for all the 42 nonhydrogen atoms dropped the R factor to 10%. Further refinement (with SHELXS-86) of the structure with

anisotropic thermal parameters reduced R to 7%. Twenty-one out of 24 of the molecular hydrogen atoms (except those attached to the ribose hydroxyl oxygens) were located from the difference Fourier map computed at this stage and were included in the calculations. However, on least-squares refinement the geometrical and the thermal parameters of the hydrogen atoms reached unacceptable values. Therefore, the hydrogen atoms were included only in the structure factor calculations but not in the refinement. The final R factor for 1699 observed reflections [$(I \geq 3\sigma(F)$] was 0.054. The sigma weighting scheme was used in the final cycles of refinement.

RESULTS AND DISCUSSION

Positional parameters are listed in Table II. Bond distances and angles are shown in Figure 1a and b, respectively. Backbone conformation is shown in Figure 2. Relevant torsion angles are listed in Tables III and IV. The molecular conformation of G2'p5'C is illustrated in Figure 3.

Conformation of the Guanine and the Cytosine Bases about the Glycosyl Bond

The guanine and cytosine bases are essentially planar. The conformation of the guanine base about the glycosyl bond is *syn* ($\chi_1 = 54°$) and that of cytosine is *anti* ($\chi_2 = -156°$; Figure 3). The guanines in dC3'p5'G,[15] G2'p5'G Ribonuclease T_1 complex,[16] left-handed Z-DNA,[17] and the tetraplex oxytricha telomeric DNA[10] are known to be in the *syn* conformation. This contrasts the *anti* conformation in G3'p5'C ($\chi_1 = -156°$). The *syn* conformation of the 5'-end guanine is stabilized by an intramolecular hydrogen bond N3···O5' [2.79(1) Å] like in A2'p5'A and G2'p5'G complexed with RNase T1. The stabilization of the *syn* conformation by the N3···O5' hydrogen bond is known to occur in several mononucleotides also.[18] In addition, the 2'-end cytosine conformation is stabilized by a C—H···O type of interaction[19] between the C6C and O5'C [C6C···O5'C = 3.41(1) Å] atoms (Figure 4). The bases do not show intramolecular stacking in spite of their parallel orientation as in A2'p5'A and A2'p5'C.[20]

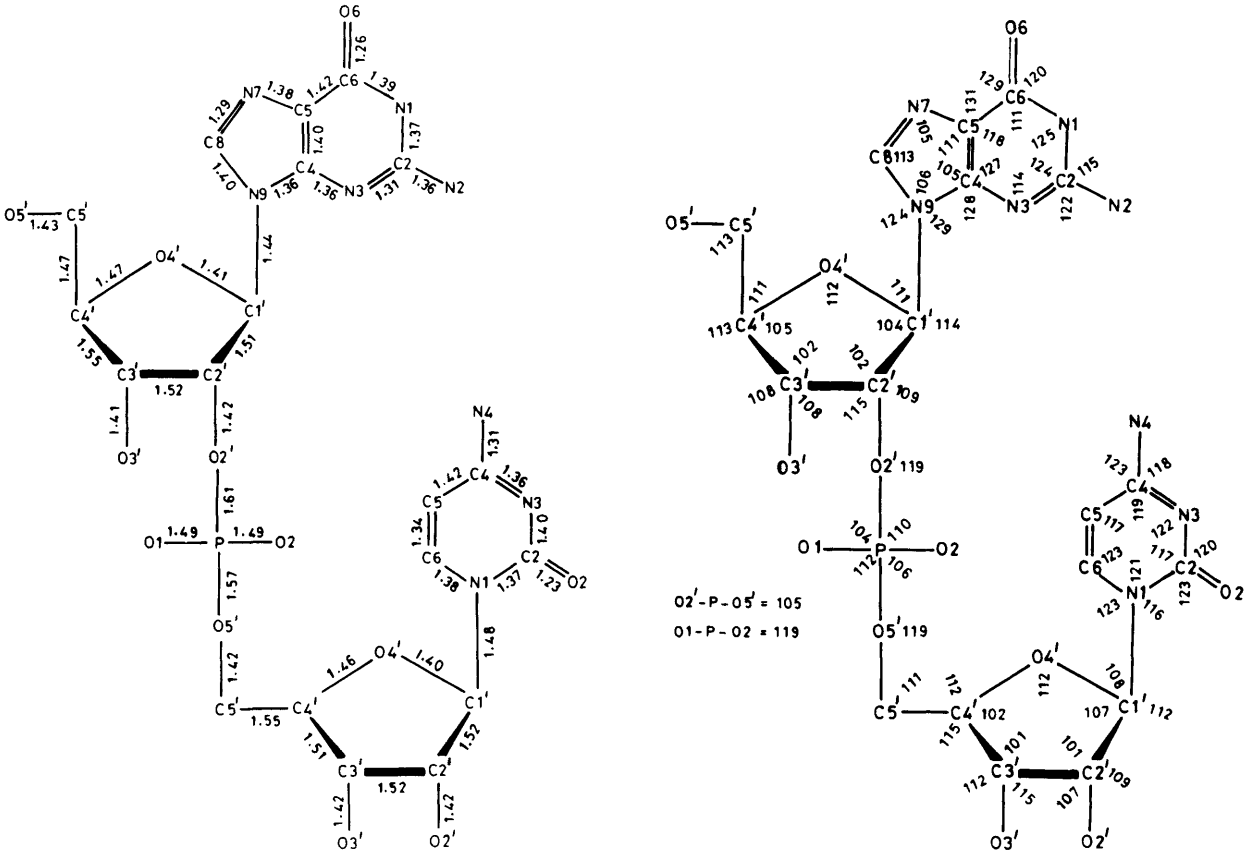

Figure 1. (Left) Bond lengths in G2'p5'C (average ESD = 0.02 Å). (Right) Bond angles in G2'p5'C (average ESD = 1°).

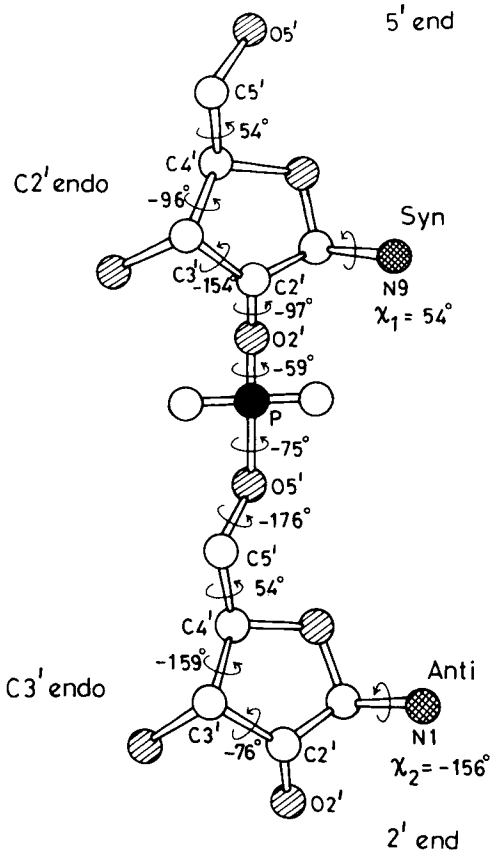

Figure 2. Sugar phosphate backbone conformation in G2'p5'C (average ESD in the torsion angle = 2°).

Conformations of the Ribose Moieties in G2'p5'C

The 5'- and 2'-end ribose moieties of G2'p5'C show C2'-*endo* and C3'-endo mixed puckering just like in A2'p5'A, A2'p5'C,[20] A2'p5'U,[14] and dC3'p5'G.[15] In contrast, the ribose moieties of G3'p5'C and A3'p5'U have C3'-*endo* puckering. The C2' atom of the 5'-end ribose and the C3' atom of the 2'-end ribose deviate from the best planes constituted by the remaining four atoms by 0.61 and 0.62 Å, respectively. The amplitude of puckering τ_m and the phase angle P of the pseudorotation parameters are 38° and 40° and 162° and 19° for the 5'- and 2'-end ribose sugars, respectively (Table III).

Phosphate Group

The phosphate group is in the staggered configuration. The bond distances and angles (Figure 1a and b) around the phosphorus atom indicate that it is negatively charged. The equality of the nonesterified bond lengths (P-O1, P-O2 = 1.49 Å) indicates that the charge could be distributed between the two oxygen atoms O1 and O2.

Charge Neutralization in G2'p5'C: Protonation of the Cytosine Base

As mentioned earlier, G2'p5'C was crystallized at acidic pH. It is conceivable that under these conditions the charge neutralization of the molecule could be accomplished through protonation of the cytosine base.[21] A strong peak of height 0.48 e/Å^3 for the hydrogen atom found at a distance of 1 Å from N3C in the difference Fourier map confirmed the protonation of the cytosine base. This was also borne out by the agreement between the cytosine bond distances and angles with the average values reported for protonated structures.[22]

Sugar–Phosphate Backbone: The Base–Ribose O4' Stacking of G2'p5'C resembles Z-DNA

The torsion angles about the phosphodiester bonds assume values of −75° (g^-) and −59° (g^-) characteristically seen in 3'-5' helical structures (Table IV). Both the 5'- and the 2'-end nucleosides have g^+ conformations about the C4'-C5' bond. An important feature of G2'p5'C is the stacking of guanine on ribose O4' of cytosine (Figure 5) similar to that seen in A2'p5'A(1), A2'p5'C, and A2'p5'U. Table IV lists the backbone torsion angles of the 2'-5' and 3'-5' of self-complementary sequence dimers. It can be seen that the 2'-5' and 3'-5' Pur-Pyr self-complementary structures essentially differ in the glycosyl bond, furanose ring, and C4'-C3' bond conformations.

G2'p5'C and A2'p5'U, Unlike Their 3'-5' Isomers, Do Not Form Miniature Duplexes

Both the 2'-5' self-complementary dimers, unlike their 3'-5' analogues, do not form miniature double helices in their crystal structures probably because

Table III Endocyclic Torsion Angles and Pseudorotation Parameters of the Furanose Rings

	5'-End	2'-End
θ_0	−37	38
θ_1	24	−39
θ_2	0	26
θ_3	−23	−1
θ_4	37	−24
	C2'-*endo*	C3'-*endo*
P	162°	19°
τ_m	38	40

Table IV Sugar–Phosphate Backbone Torsion Angles in Self-Complementary Dinucleoside Monophosphates and A2′p5′C

	Angle	G2′p5′C	A2′p5′C	G3′p5′C	A2′p5′U	A3′p5′U
O5′–C5′–C4′–C3′	γ_1	54	56	44	45	47
C5′–C4′–C3′–C2′	δ_1	−96	−95	82	−99	84
C4′–C3′–C2′–C1′	δ'_1	−154	−156	—	−155	—
C3′–C2′–O2′–P	ϵ_1	−97	−94	−149	−116	−147
C2′–O2′–P–O5′	ξ_1	−59	−57	−66	−127	−67
O2′–P–O5′–C5′	α_1	−75	−75	−75	−47	−72
P–O5′–C5′–C4′	β_1	−176	−179	−175	170	177
O5′–C5′–C4′–C3′	γ_2	54	56	51	59	57
C5′–C4′–C3′–C2′	δ_2	−159	−161	75	−157	74
C4′–C3′–C2′–C1′	δ'_2	−76	−76	—	−74	—
O4′–C1′–N9–C4	χ_1	54	54	−160	−128	−173
O4′–C1′–N1–C2	χ_2	−156	−163	−156	−169	−151
5′ Furanose	C2′-*endo*	C2′-*endo*	C3′-*endo*	C2′-*endo*	C3′-*endo*	C3′-*endo*
2′ or 3′ Furanose	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*	C3′-*endo*

of the protonation of the bases. It may be mentioned that the cytosine residue in the noncomplementary sequence A2′p5′C is also protonated but not the adenine residue. On the other hand, both the C and A residues are hemiprotonated in the C2′p5′A structure, which forms a miniature parallel-stranded double helix. So, the nonformation of duplex structure may not be entirely due to protonation but may also depend on other factors such as base sequence, strength of stacking interactions, phosphodiester links, etc. A right-handed single-stranded poly(G2′p5′C) helix has been generated by Parthasarathy et al.,[20] using GpCp as a repeating unit with 2.1 units/turn and 7.5 Å rise per unit. The conformation of this repeating unit is identical to the G2′p5′C as it has been modeled after the A2′p5′C structure (Table IV). Their observation that a stereochemially permissible double helix could not be built using A2′p5′C holds good for G2′p5′C as well. The nmr studies have shown that C2′p5′G, G2′p5′C,

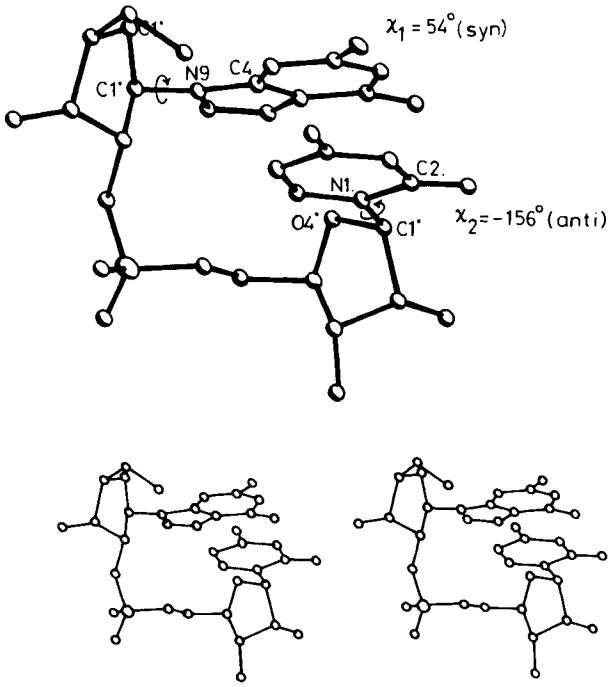

Figure 3. Mono and stereo diagrams of G2′p5′C depicting the *syn* and *anti* conformation of G and C residues, respectively.

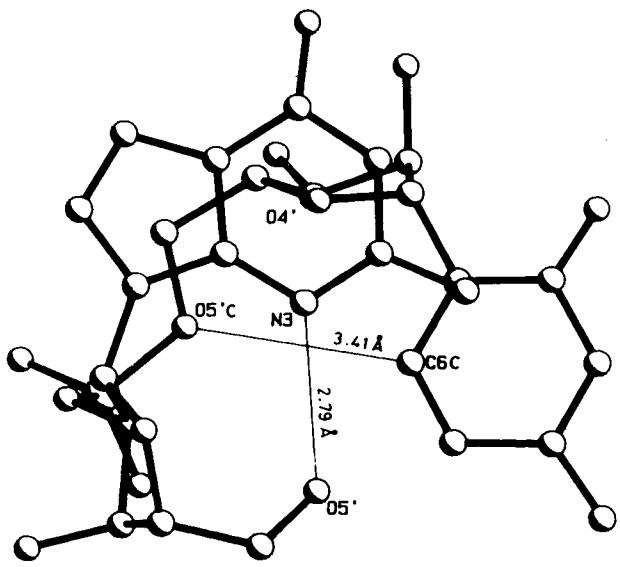

Figure 4. Molecular conformation of G2′p5′C showing the N3···O5′ (guanine) and C6C···O5′C (cytosine) intramolecular hydrogen bonds.

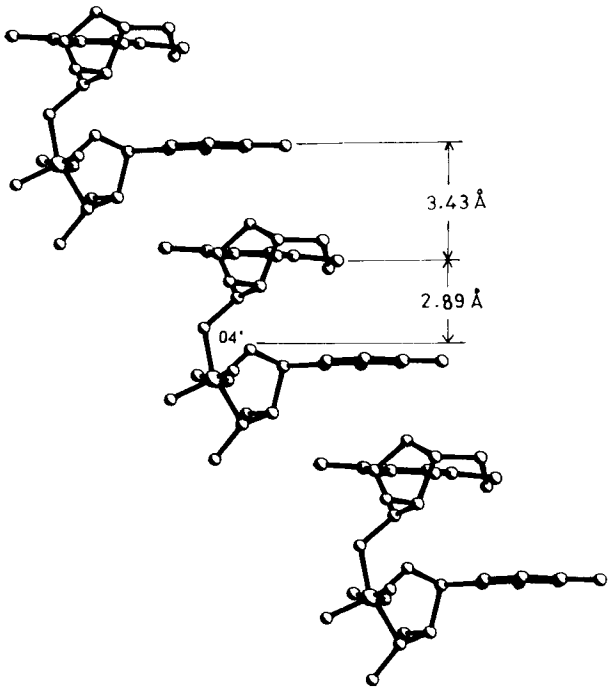

Figure 5. Intermolecular stacking of guanine bases along the *a* axis at intervals of 3.43 Å and base-ribose O4' stacking. The stacking has the O4'–base–base · · · O4'–base–base . . . sequence.

and their mixtures do not form miniature duplexes, unlike their 3'-5' counterparts. All of them indicate that the proximity of the phosphate group to the helix axis in 2'-5' structures must be responsible for the destacking and disruption of the base-pair formation by deflecting the Watson–Crick hydrogen-bonding sites away from each other (Figure 6a and b).

Hydrogen Bonding and Molecular Packing

Table V lists the hydrogen-bond distances. As in A2'p5'A and A2'p5'C, there are intermolecular stacking of bases. They are stacked along the *a* axis at 3.43 Å, as shown in Figure 5. There are no base–base hydrogen-bonding interactions. Instead, the Watson–Crick sites in both cytosine and guanine bases are involved in hydrogen bonds with the phosphate group, ribose hydroxyls, and water molecules (Figure 7a). A pair of hydrogen bonds, N4C · · · O1 and N3C · · · O2, links the protonated cytosine base to the negatively charged phosphate group (Figure 7b). Similar base phosphate interactions have been observed in the crystal structures of ADP, ADP · Rb, and ADP · K.[23] Guanine base forms two hydrogen bonds N2 · · · O2'C and N1 · · · O3' with the ribose hydroxyl oxygens. The third Watson–Crick site in both the bases interact with water molecules forming O2 · · · OW2 and O6 · · · OW1 hydrogen bonds (Figure 5). Similar interactions involving the Watson–Crick sites of guanine base have been reported in the structure of G2'p5'G complexed with RNase T1. The N1 and N2 atoms of the guanine base form hydrogen bonds with oxygens of Glu 46 and the O6 atom with a nitrogen atom of Asn 44 of the enzyme, suggesting that the intermolecular interactions observed in the present structure could be of relevance in understanding the stereochemistry of RNase–inhibitor complexes. In addition, the water molecules OW1 and OW2 are involved in hydrogen bonds with the phosphate oxygen atom O2 and ribose hydroxyl oxygen O3', respectively.

G2'p5'C and A2'p5'C Are Isostructural

The stereochemical features of G2'p5'C and A2'p5'C are in spectacular agreement. Figure 8 and Table IV illustrate the degree of agreement between them. Both molecules can be superposed with an rms deviation of less than 0.20 Å in their atomic positions. In addition, the G2'p5'C and A2'p5'C crystal lattices can also be superposed with very little deviation.

Table V Hydrogen Bonding in G2'p5'C[a]

A	B	Distance (Å)	Symmetry	Cell Translation of Atom B		
		Intramolecular				
N3	O5'	2.79(1)	1	0	0	0
O5'C	C6C	3.41(1)	1	0	0	0
		Intermolecular				
O1	N4C	2.80(1)	2	0	−1	0
O2	N3C	2.74(1)	2	0	−1	0
N2	O2'C	2.97(1)	4	0	0	−1
N1	O3'	2.78(1)	2	−1	−1	−1
O5'	O2'C	2.69(1)	4	0	0	−1
O1	O3'C	2.86(1)	3	−1	−1	0
		G2'p5'C and water molecules				
O6	OW1	2.69(1)	4	−1	−1	−1
O3'	OW1	2.76(1)	3	0	−1	0
N4C	OW1	2.89(1)	1	1	0	0
O2	OW2	2.76(1)	1	0	0	0
O2C	OW2	3.13(1)	2	0	−1	0

[a] Symmetry codes: (1) x, y, z; (2) $1/2 - x, -y, 1/2 + z$; (3) $1/2 + x, 1/2 - y, -z$; (4) $-x, 1/2 + y, 1/2 - z$. A and B do not always symbolize donors and acceptors respectively.

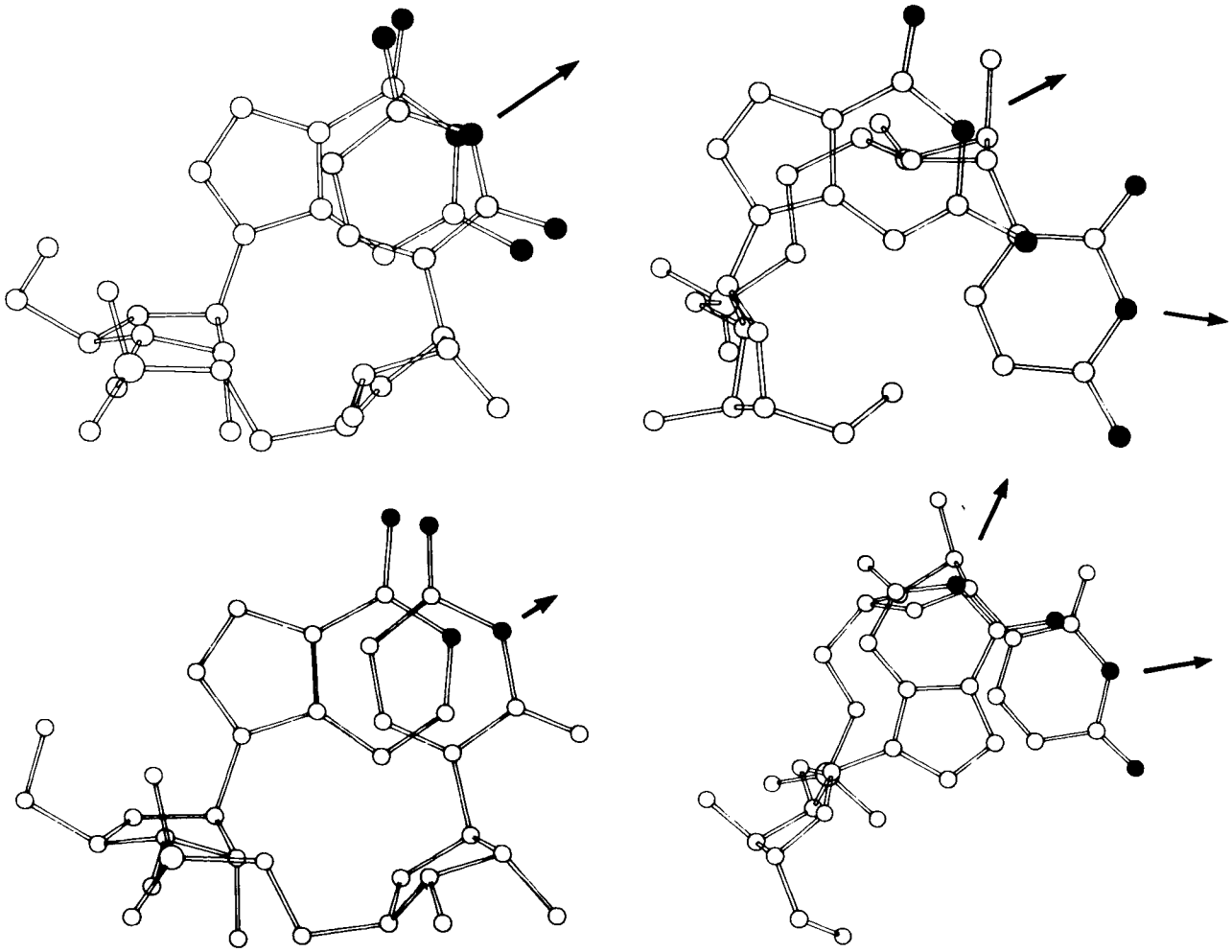

Figure 6. (Top) G3′p5′C and G2′p5′C and (Bottom) A3′p5′U and A2′p5′U viewed perpendicular to their 5′-end bases illustrating the conformational differences between them. The Watson–Crick hydrogen-bonding sites of the Pur-Pyr bases in 3′-5′ point in the same direction, unlike in the 2′-5′ dimers.

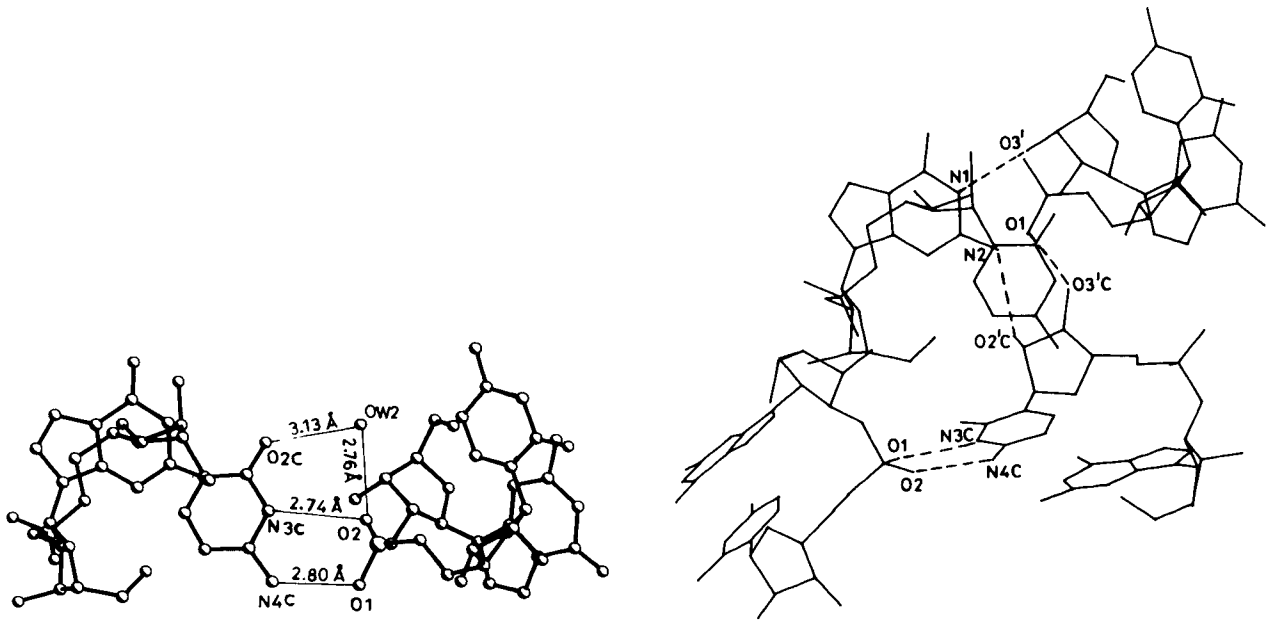

Figure 7. Intermolecular hydrogen-bonding pattern involving Watson–Crick sites in G2′p5′C.

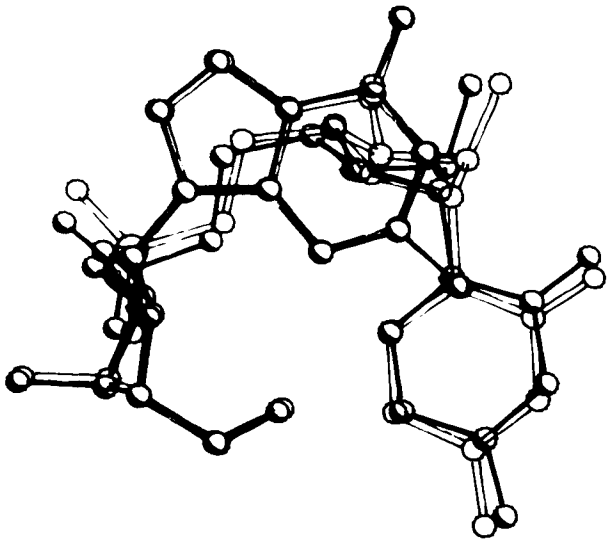

Figure 8. A superposition of G2'p5'C (dark) on A2'p5'C (thin) viewed perpendicular to their bases.

Their unit cell dimensions, contents, and volumes are very nearly the same (Table VI). Further, both of them have identical structure and intermolecular interactions (Table VI and Figure 8). These observations suggest that G2'p5'C and A2'p5'C are isostructural.

A Model of the Branched Trinuleotide $G^{2'p5'C}_{3'p5'C}$: Visualization of msDNA–RNA Junction

Recently, Remaud et al.[25] have carried out nmr studies on branched tri-, tetra-, penta-, and heptanucleotides. They concluded that the stability of the branched structure depends on the length of the oligomers. Further, in the longer fragments the conformation of the branch point resembles the A-RNA structure. Moreover, the conformation of the branched trimer largely resembles the conformation of the 2'-5' rather than the corresponding 3'-5' dimers. We have utilized this property to postulate a branched trinucleotide model for the msDNA–RNA junction as no crystal structure is available at the moment. The branched trimer model was generated through retaining the 2'-5' unit as in the G2'p5'C crystal structure and attaching a cytidine-5'-monophosphate to the 3'-hydroxyl oxygen of the 5'-end ribose moiety. A stereopicture of the branched trimer is shown in Figure 9. The conformation of the cytosine base about the glycosyl bond is *anti* and the ribore puckering is C3'-*endo*. The stereochemistry of the model has been checked on an interactive graphics system. As discussed above both G2'p5'C and A2'p5'C are structurally similar. Conceivably, the stereochemistry of the trimers with either G or A at the branch point could also be very nearly the same. Therefore, it is plausible the termination of the msDNA synthesis by reverse transcriptase observed on substitution of G by A could be arising due to perturbation in the RNA structure away from the branch point.

Table VI Crystal Data of G2'p5'C and A2'p5'C

	GpC	ApC
Formula	$C_{19}H_{24}N_8O_{12}P \cdot 2H_2O$	$C_{19}H_{24}N_8O_{11}P \cdot 2H_2O$
a =	8.376 Å	8.631 Å
b =	16.231 Å	18.099 Å
c =	18.671 Å	16.101 Å
Space group	$P2_12_12_1$	$P2_12_12_1$
v =	2538.3 Å^3	2515.2 Å^3

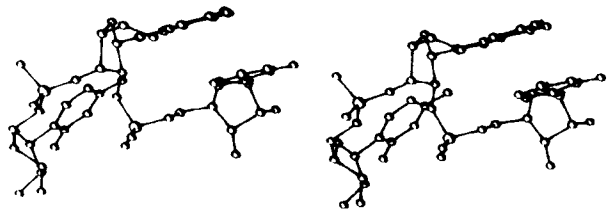

Figure 9. A stereo picture of the $G^{2'p5'C}_{3'p5'C}$ branched trinucleotide model of the msDNA–RNA junction.

CONCLUSION

This paper describes the crystal structure of G2'p5'C. Both the 2'-5' self-complementary dimers A2'p5'U and G2'p5'C do not form miniature duplexes, unlike their respective 3'-5' isomers. A noteworthy feature of G2'p5'C is the stacking of guanine on ribose 04' of cytosine. Suprisingly, this self-complementary G2'p5'C and the noncomplementary A2'p5'C dimers have identical stereochemical features. We envision from the astoundingly similar stereochemical features of G2'p5'C, A2'pC, A2'p5'U, and A2'p5'A that the 2'-5' dimers with the Pur-Pyr sequence overwhelmingly appears to prefer the *syn,anti* conformation about the glycoxyl bond, C2'-*endo*, C3'-*endo* mixed ribose puckering, and base–ribose 04' stacking pattern. This hypothesis, although attractive, must await adjudication by 2'-5' oligomeric structures. A branched trinucleotide model for msDNA–RNA junction has been postulated.

We thank the CSIR for financial support and DBT for computations.

REFERENCES

1. Joyce, C. F. (1987) *Cold Spring Harbor Symp. Quant. Biol.,* **LII,** 41–51.
2. Usher, D. A. (1977) *Science,* **196,** 311–312.
3. Lesiak, K., Imai, J., Floyd-Smith, G. & Torrence, P. F. (1983) *J. Biol. Chem.* **258,** 13082–13088.
4. Furuichi, T., Dhundale, A., Inouye, M. & Inouye, S. (1987) *Cell* **48,** 47–58.
5. Hsu, M.-Y., Inouye, S. & Inouye, M. (1989) *J. Biol. Chem.* **264,** 6214–6219.
6. Kennard, O. & Hunter, N. N. (1989) *Quart. Rev. Biophys.* **22,** 327–329.
7. Joshua-Tor, L., Rabinovich, D., Hope, H., Frolow, F., Appella, E. & Sussman, J. (1988) *Nature* (London) **334,** 82–84.
8. Miller, M., Harrison, R. W., Wlodawer, A., Appella, E. & Sussman, J. L. (1988) *Nature* (London) **334,** 85–86.
9. Kennard, O. (1985) *J. Biomol. Struct. Dynam.* **3,** 205–226.
10. Kang, C. H., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. (1992) *Nature* (London) **356,** 126–131.
11. Krishnan, R., Seshadri, T. P. & Viswamitra, M. A. (1991) *Nucleic Acids Res.* **19,** 379–384.
12. Krishnan, R. & Seshadri, T. P. (1993) *J. Biomol. Struct. Dynam.,* **10,** 727–745.
13. Rosenberg, J. M., Seeman, N. C., Day, R. O. & Rich, A. (1976) *J. Mol. Biol.* **104,** 145–167.
14. Shefter, E., Malcolm, B., Sparks, R. A. & Trueblood, K. N. (1969) *Acta Cryst. B* **25,** 895–909.
15. Ramakrishnan, B. & Viswamitra, M. A. (1988) *J. Biomol. Struct. Dynam.* **6,** 511–523.
16. Koepke, J., Maslowska, M., Heinemann, U. & Saenger, W. (1989) *J. Mol. Biol.* **206,** 475–488.
17. Wang, A. H. J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. & Rich, A. (1979) *Nature* (London) **282,** 680–686.
18. Rao, S. T. & Sundaralingam, M. (1970) *J. Am. Chem. Soc.* **92,** 4963–4970.
19. Taylor, R. & Kennard, O. (1982) *J. Am. Chem. Soc.* **104,** 5063–5070.
20. Parthasarathy, R., Malik, M. & Fridey, S. M. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 7292–7296.
21. Adams, L. P. A., Knowler, J. T. & Leader, D. P. (1986) *The Biochemistry of Nucleic Acids,* Chapman & Hall, London, chap. 2. p. 7.
22. Taylor, R. & Kennard, O. (1982a) *J. Mol. Struct.* **78,** 1–28.
23. Viswamitra, M. A., Katti, S. K. & Hosur, M. V. (1981) in *Structural Studies on Molecules of Biological Interest,* Dodson, G., Glusker, J. P. & Sayre, D., Eds., Clarendon Press, Oxford, pp. 154–165.
24. James, M. & Crabbe, C. (1989) *Desk Top Molecular Modeller (DTMM),* version 1.2, Oxford University Press, Oxford.
25. Remaud, G., Vial, J. M., Balgobin, N., Koole, K. H., Strom, A., Drake, A. F., Zhou, X-X., Glemarec, C. & Chattopadhyaya, J. (1990) in *Structure and Methods: DNA and RNA,* Sarma, R. H. & Sarma, M. H., Eds. Adenine Press, New York, pp. 319–337.

Received January 31, 1994
Accepted June 7, 1994