Proc. Nat. Acad. Sci. USA
Vol. 70, No. 3, pp. 849–853, March 1973

A Crystalline Fragment of the Double Helix: The Structure of the Dinucleoside Phosphate Guanylyl-3′,5′-Cytidine

(x-ray diffraction/complementary hydrogen bonding/dinucleoside phosphate/right-handed helix)

ROBERTA OGILVIE DAY*, NADRIAN C. SEEMAN, JOHN M. ROSENBERG, AND ALEXANDER RICH

Department of Biology, Massachusetts Institute of Technology, Cambridge, Mass. 02139

Contributed by Alexander Rich, January 15, 1973

ABSTRACT **The sodium salt of guanylyl-3′,5′-cytidine crystallizes in a monoclinic unit cell with one molecule in the asymmetric unit. Each molecule is related to another molecule by a 2-fold rotation axis which results in the formation of an antiparallel, right-handed double helix with complementary hydrogen bonding between the guanine and cytosine residues. The crystal is heavily hydrated with 36 water molecules in the unit cell. The geometry of this crystalline double helix is very similar to those which have been derived from studies of fiber x-ray diffraction patterns of double-stranded RNA, even though the latter do not yield data at atomic resolution.**

Double-helical nucleic acids play a central role in information storage and transfer processes in biological systems; nonetheless, we have very little data about their detailed atomic conformations. X-ray diffraction patterns of double-helical nucleic-acid fibers have provided generalized information about helical parameters, but since they rarely produce diffraction data beyond 3 Å in resolution (1–3), such patterns are incapable of providing detailed information about polynucleotides at the atomic level. Recently, we reported the solution of the crystal structure of the dinucleoside phosphate adenylyl-3′,5′-uridine (ApU), which crystallizes in the form of a right-handed, antiparallel double helix with Watson–Crick hydrogen bonding between the uracil and adenine bases (4). Here, we describe the structure of a second dinucleoside phosphate guanylyl-3′,5′-cytidine (GpC), which forms a similar structure in the solid state. These structures provide us with important information about the conformation of the ribose phosphate linkage that can be used in interpreting the molecular basis of the role of RNA in biological systems.

EXPERIMENTAL

A 5 mM solution of sodium GpC (Sigma) was prepared in 20 mM sodium cacodylate buffer at pH 6.2. 0.1 ml of this solution was mixed with 0.05 ml of a 10 mM 9-aminoacridine hydrochloride solution and 0.1 ml of 2-methyl-2,4-pentanediol was added. The solution was put into a depression in a pyrex spot plate, which was then placed in a sealed transparent container over a reservoir containing 75% aqueous 2-methyl-2,4-pentanediol. The container was stored at room temperature and small crystals were observed within a week. These eventually grew into clusters of large laths. Spectrophotometric analysis revealed that the crystals contained only GpC and no acridine. A crystal measuring about 0.6 mm × 0.2 mm × 0.1 mm was mounted for data collection in a thin-walled glass capillary in the presence of small droplets of mother liquor. This procedure was adopted because the crystals were found to be unstable when separated from their mother liquor and disintegrated visibly after 24 hr in the air. Three-dimensional x-ray intensity data were collected on a Syntex $P\bar{1}$ automated diffractometer by use of nickel-filtered CuKα radiation. The data were collected by use of a θ-2θ scanning mode at 10°C out to spacings of 0.89 Å. 2630 Reflections were recorded, and of these 2276 were considered to be observed in that they were more than 3 sigma above background.

The space group of GpC is C2 and the unit cell dimensions are $a = 21.460$, $b = 16.927$, $c = 9.332$ Å, $\beta = 90.54°$, $Z = 4$. Lorentz and polarization corrections were applied to the intensity data, but no absorption correction has been applied because of the low value of the linear-mass absorption coefficient. The phosphorus atom was located by resolution difference Patterson techniques (4). Approximate phosphate orientations were derived from Patterson superposition functions based on these phosphorus sites. Seven cycles of Fourier refinement revealed the entire GpC molecule, the sodium ion, and two water molecules. Six water molecules in general positions and two others lying on 2-fold axes were located in a series of Fourier difference syntheses, bringing the total to 36 water molecules per unit cell. The structure has been refined by full-matrix least squares using isotropic temperature factors and all of the observed data to a residual factor (R) of 11.1%.

FEATURES OF THE STRUCTURE

Solution of this structure was facilitated considerably by the recent solution of the dinucleoside phosphate ApU (4). ApU crystallizes in the form of a double helix in space group $P2_1$ with cell dimensions $a = 18.025$, $b = 17.501$, $c = 9.677$ Å, and $\beta = 99.94°$. In our initial attempts to crystallize GpC, a different monoclinic crystal form had been obtained that bore a striking similarity to ApU with respect to the lattice parameters ($P2_1$, $a = 18.6$ Å, $b = 16.7$ Å, $c = 9.6$ Å, $\beta = 97°$). Furthermore, the diffraction patterns of the $P2_1$ form of GpC strongly resembled the ApU pattern. From these similarities we inferred that the GpC molecule probably forms a double helix which is very similar to that observed for ApU. Antiparallel double-helical structures in general possess 2-fold rotation (dyad) axes which rotate one ribose phosphate chain into the other. A pseudo 2-fold axis is found in the crystal structure of ApU that relates the two antiparallel chains. In the C2 lattice of GpC, we deduced that this 2-fold rotation axis was coincident with a crystallographic 2-fold axis. It should be noted here that there are two types of 2-fold

* Present address: Department of Chemistry, University of Nebraska, Lincoln, Nebr. 68508.

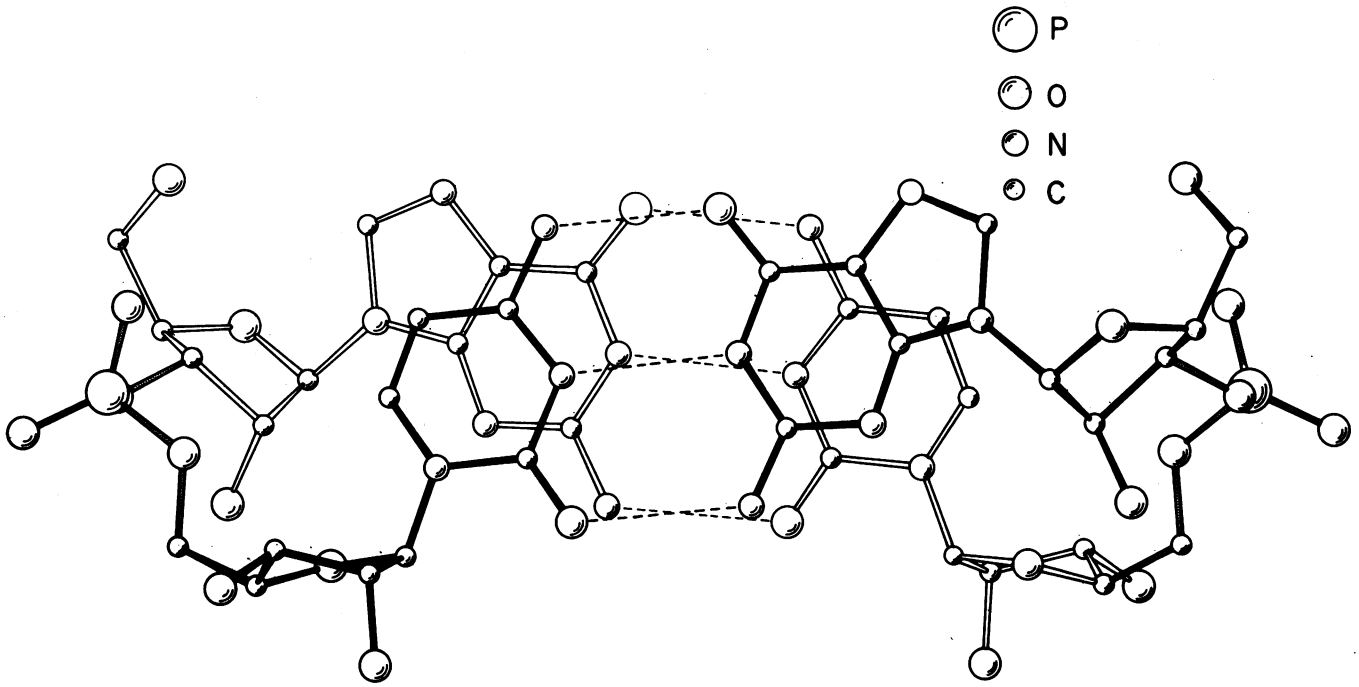

FIG. 1. A view of the GpC double-helical fragment approximately normal to the base planes. The shaded base-pair is nearest the reader.

axes found in double-helical polynucleotides. One of these lies in the plane of the base pairs while the other is located halfway between these planes. It is the latter 2-fold axis that relates the two molecules in the GpC structure as well as in ApU. Because of our surmise concerning the dyad axis, solution of the GpC structure used methods that were roughly similar to those used with ApU. However, as the asymmetric unit contained one GpC molecule rather than two, solution of the structure was correspondingly simpler.

That GpC forms a double-helical complex with itself can be seen in Fig. 1, in which we are looking at the helix in a direction perpendicular to the planes of the base pairs. Note that the guanine and cytosine residues are connected by three hydrogen bonds in a manner similar to that which is believed to exist in double-helical polynucleotides. It can be seen from this figure that there is a considerable amount of overlap between the purine and pyrimidine rings in this double-helical fragment. This is true despite the clearly visible rotational relationship of the lower base pair relative to the upper one. This rotation of the base pairs is due to the helical twist, which is imposed on the molecule by the conformation of the backbone.

The guanine is bound to the cytosine by three hydrogen bonds (see Fig. 3 for the numbering scheme): cytosine N4 to guanine O6 (2.90 Å), guanine N1 to cytosine N3 (2.90 Å), and guanine N2 to cytosine O2 (2.86 Å). The distance between the glycosidic carbon atoms C1′ of guanosine and C1′ of cytidine is 10.67 Å. This type of hydrogen bonding interaction between guanine and cytosine residues has been seen in other intermolecular complexes of these bases, and the geometry of the present base pair is similar to that which has been reported (5–8). The planar guanine and cytosine rings are almost coplanar. Their least-squares planes intersect at a dihedral angle of 11.6°, which is similar to that which has been reported for intermolecular complexes of guanine and cytosine derivatives.

A view of the helical complex perpendicular to the helix axis is shown in Fig. 2. For comparative purposes we include the same views of the ApU structure and the structure proposed for double-helical RNA (9) with 11 base pairs per turn (RNA-11). The right-handed nature of the helical rotation is evident by studying the orientation of adjacent ribose residues along the polynucleotide chains. The similarity of the three structures is apparent.

In the crystal lattice of GpC the double-helical fragments are stacked with the bases of neighboring molecules in parallel orientations. The guanine residues of adjacent helical molecules overlap each other, while the cytosine residue is in close contact with O1′ of the adjacent guanosine ribose. A somewhat similar stacking was observed in the ApU structure (4). In this manner the stacked double-helical fragments are oriented in the form of long rod-like arrays which run through the crystal. These rods have base pairs in the center and the ribose phosphate groups situated on the outside. Above and below the rods there is a connected network of water molecules so that in the direction parallel to the rod array, there are continuous channels of water. These channels branch periodically into large pockets containing additional water molecules. It is likely that this hydration structure accounts for the instability of the crystals when they are separated from their mother liquor.

The arrays of stacked double-helical fragments are connected to each other through a series of electrostatic interactions involving the sodium ions that are coordinated with the phosphate groups as shown in Fig. 3. Here it can be seen that every phosphate group is connected to two sodium ions each of which has octahedral coordination with additional water molecules. One face of this octahedral coordination shell is

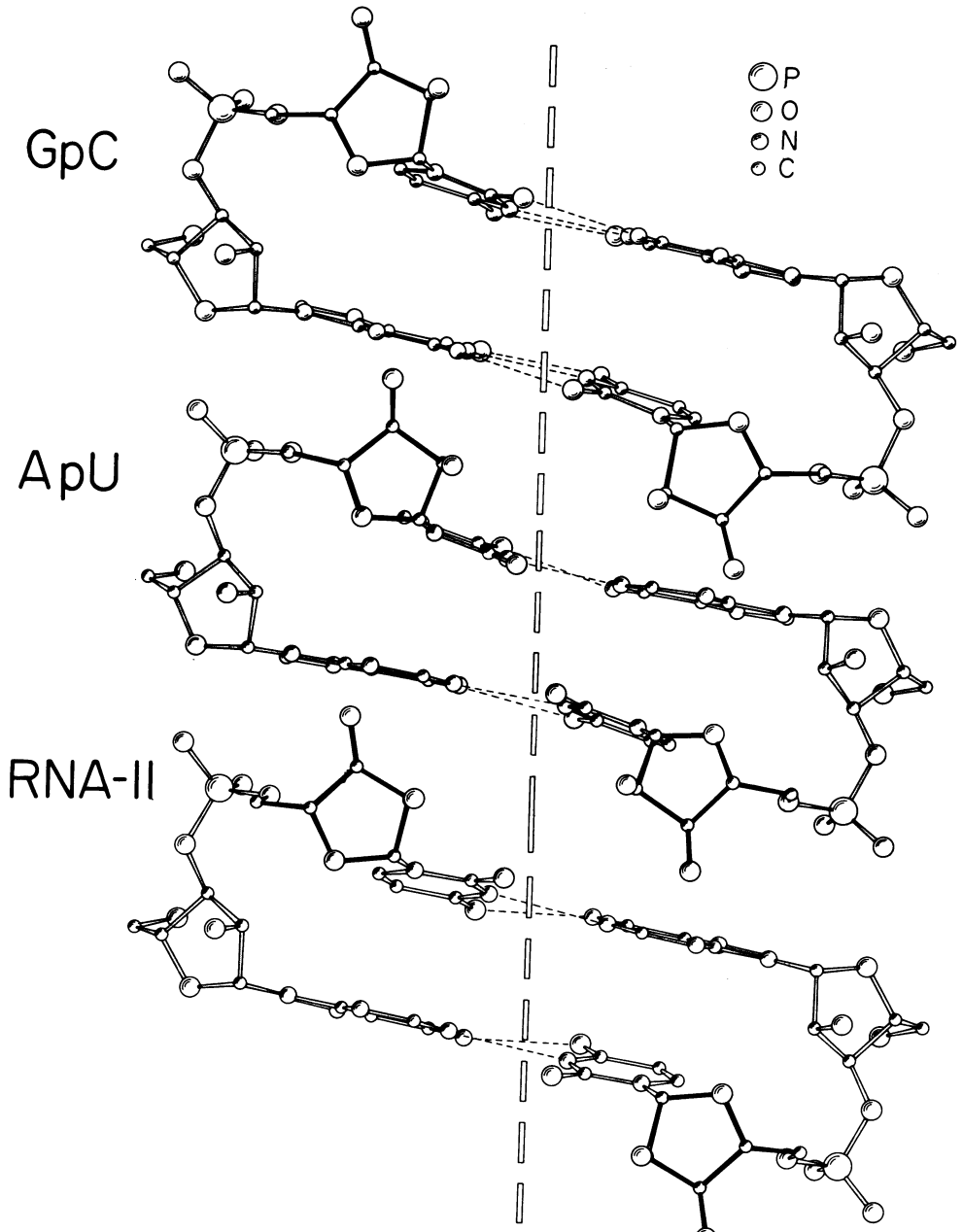

FIG. 2. View of ApU, GpC, and RNA-11 approximately normal to the vertical helix axis. The similarities between the structures are quite visible, as is the overall right-handed helical relationship between the successive residues on each of the chains.

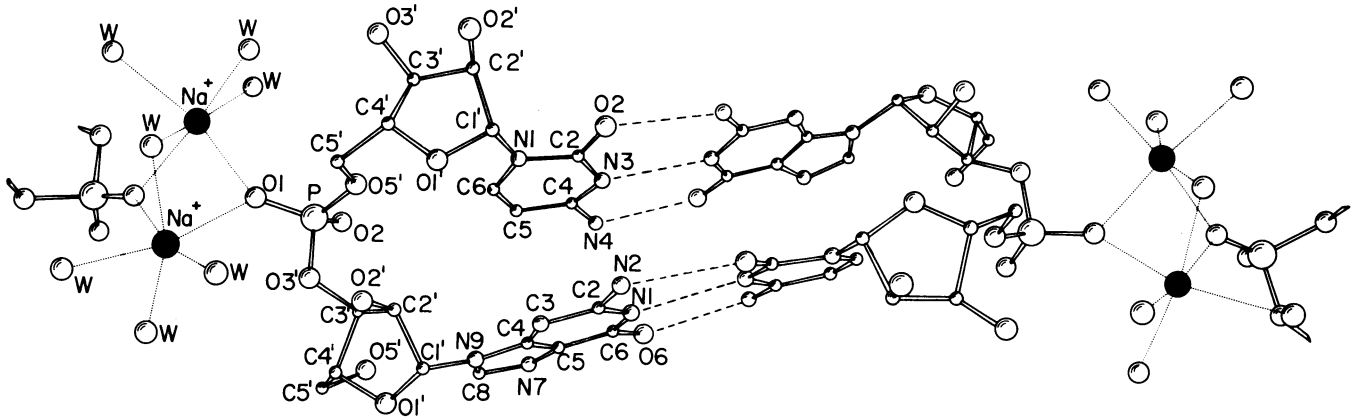

FIG. 3. A perspective view of the GpC double-helical fragment showing the molecular numbering scheme and the face-sharing octahedral coordination of sodium ions.

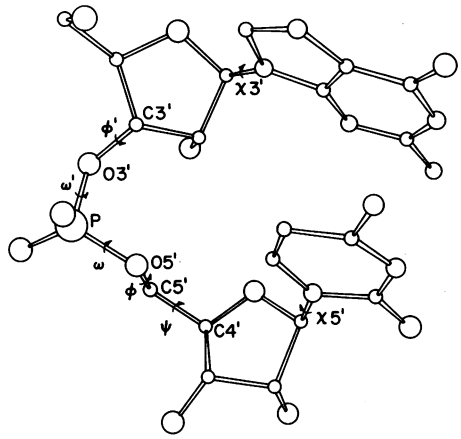

FIG. 4. A representation of the GpC molecule defining the backbone and glycosidic torsion angles. The conventions used are those of ref. 12.

shared between the two sodium ions. The position of the sodium ions in this structure is in sharp contrast to the coordination that is seen in ApU, in which one sodium ion is coordinated to the uracil residues in the minor groove of the helix.

It is interesting that the crystal lattices of both ApU and GpC are heavily hydrated (ApU has 26 water molecules in the unit cell). Apart from the stacking interactions, the molecules are largely shielded from their neighbors by the widely dispersed water structure in both crystals. Thus, these molecules may be fairly accurate models of a double-helical polyribonucleotide existing in an aqueous environment. This suggests that the molecular conformations are determined largely by internal constraints within the ribose phosphate backbone, rather than by crystal packing forces. Indeed, there are only two hydrogen bonds between GpC molecules that are not involved in the base pairing. One of these connects N4 of cytosine with O3′ of a cytidine related to it by a 2-fold screw axis. Its length is 2.95 Å. The other hydrogen bond involves O1′ of cytosine, which is 2.91 Å away from O5′ of guanosine in a molecule also related by a 2-fold screw axis.

The discovery of crystalline fragments of a double helix is important for two reasons. First, it illustrates the important interactions of nucleotides which have an intrinsic specificity for forming complementary hydrogen bonds with each other: guanine to cytosine and adenine to uracil. This specificity can be seen in solution studies and can also be studied in the solid state, as in the present investigation. Secondly, these crystalline fragments allow us to obtain precise information about the detailed geometry of the ribose-phosphate backbone (10). As mentioned above, most of our information on double helices is obtained from studies of fiber-diffraction patterns, which usually do not extend to resolutions greater than 3 Å. This has led to considerable discussion of the actual nature of the hydrogen bonds which exist in double-helical nucleic acids. The present studies provide high-resolution data about the hydrogen bonding and conformation of double-helical nucleotides. Therefore, they provide us with a set of basic data which can be of great value in understanding polynucleotide conformation. The data include bond lengths, bond angles, and the torsion angles which define the conformation of the polynucleotide backbone.

TABLE 1. *Bond lengths (Å) in ribose phosphate linkages*

	C4′–C5′	C5′–C5′	05′–P	P–03′	03′–C3′	Ref.
GpC	1.52	1.45	1.60	1.62	1.44	This work
ApU1	1.53	1.47	1.59	1.60	1.42	4
ApU2	1.49	1.49	1.60	1.61	1.44	4
UpA1	1.51	1.41	1.59	1.62	1.43	12
UpA2	1.51	1.44	1.60	1.62	1.39	12
RNA-11	1.51	1.45	1.60	1.60	1.42	9
RNA-10	1.52	1.43	1.60	1.60	1.42	16

TABLE 2. *Bond angles (°) in ribose phosphate linkages*

	C4′–C5′–05′	C5′–05′–P	05′–P–03′	P–03′–C3′
GpC	109	121	103	119
ApU1	107	119	105	121
ApU2	108	118	103	120
UpA1	111	121	105	118
UpA2	110	124	101	119
RNA-11	110	121	101	123
RNA-10	110	121	102	121

In GpC, both ribose residues are in the 3′-endo conformation, which was also noted in the ApU structure, as well as in the protonated dinucleoside phosphate uridylyl-3′,5′-adenosine (UpA) (11–13). This conformation has frequently been observed in ribonucleosides and in mononucleotides (14, 15). However, there is limited information available about the conformation of the phosphate linkage connecting the ribose groups (16). For comparative purposes we list in Tables 1, 2, and 3 the relevant information pertaining to the structure of the phosphodiester linkage in the three naturally occurring dinucleoside phosphates whose structures have been determined, together with those used in models of polynucleotide structure, RNA-10 (16) and RNA-11 (9). Table 1 lists the bond lengths in the phosphodiester linkage. The bond lengths of the three dinucleoside phosphate structures are similar to those that have been used in building the helices of RNA-10 and RNA-11. Table 2 lists the bond angles, while Table 3 lists the important torsion angles which define the stable conformations.

Some interesting features are found in the bond angles. It can be seen, for example, that all three polynucleotides show angles around the phosphate group that are somewhat

TABLE 3. *Torsion angles (°) in the ribose phosphate linkages*

	χ3′	χ5′	ψ	φ	ω	ω′	φ′
GpC	13(G)	25(C)	51	186	284	291	209
ApU1	7(A)	29(U)	57	177	289	294	213
ApU2	2(A)	30(U)	57	169	296	285	220
UpA1	12(U)	37(A)	55	203	82	81	206
UpA2	19(U)	44(A)	54	192	271	164	224
RNA-11	14(A)	15(U)	49	186	294	294	202
RNA-10	9(A)	9(U)	84	180	261	287	212

The torsion angles are shown in Fig. 4 and defined in ref. 12. References for the various structures are listed in Table 1.

smaller than the tetrahedral angle. The angles around the oxygen atoms O3′ and O5′ are not tetrahedral, but are very close to 120° in all structures, implying sp^2 hybridization for these atoms, while the angle around C5′ is close to the tetrahedral angle. The orientation of the base around the glycosidic bond is described by the angle χ as defined in Fig. 4. It can be seen that the χ values fall in two distinct classes, those at the 3′ end of the dinucleoside phosphate being appreciably smaller than those at the 5′ end. This discrepancy is undoubtedly associated with the fact that the environments are different at the two ends of the molecules. Inspection of the torsion angles in the phosphodiester linkage readily reveals which angles are important in determining helical conformation. The protonated UpA has two independent molecules in the crystal lattice and neither of them generates a potentially helical structure. The phosphate-ester angles ω and ω' differ markedly in UpA from those found in the helical forms ApU, GpC, and the RNA models. It should be noted that there is little variation in the angles ψ, ϕ, and ϕ' among the different dinucleoside phosphates. This suggests that there is a range of preferred conformations in the nucleoside unit, as pointed out previously (11). The material presented in Tables 1–3 provides us with realistic parameters for interpreting the details of polynucleotide structures, not only for double-helical RNA but also for more complex molecules such as transfer RNA (17).

DISCUSSION

We have been struck by the apparent analogies between the crystal structures of ApU and GpC. As pointed out above, an important feature of both structures is the high degree of hydration, which may allow the molecules to assume a natural conformation in the crystal lattice. However, there is another element of similarity that should be noted. The base stacking seen in the ApU structure is very similar to that shown in Fig. 1 for the GpC structure. It should be noted, however, that a markedly different kind of base stacking would be observed if the sequence of bases in the molecule was reversed and the backbone maintained the same conformation. Helical structures of the inverted sequences UpA and CpG would have sharply different types of base stacking. It will be an interesting exercise for the reader to interchange the two base paris seen in Fig. 1 so that the more heavily shaded base pair in front is switched with the lighter base pair behind. When these two are exchanged, the reader will find that there is no longer an overlap of purines and pyrimidines, but rather the purine residues overlap purine residues only and the pyrimidine residues do not overlap at all! Thus, it may be that the purine–pyrimidine sequence permits similar modes of base stacking in ApU and GpC, which in turn contributes to the similarity of the two crystal structures. In this regard it will be interesting to examine the crystal structure of the $P2_1$ form of GpC since we anticipate that it would have the same type of base pairing and base stacking as that observed in the present structure.

It is our expectation that crystals of helical oligonucleotides will be found not only with dinucleoside phosphates but also with longer oligonucleotides. It is anticipated that the detailed molecular analysis of this type of molecule will eventually provide us with a body of information which will enable us to understand finer details of polynucleotide-chain folding.

This research was supported by grants from the National Institutes of Health, the National Science Foundation, and the American Cancer Society. R.O.D. and N.C.S. are postdoctoral fellows of the Damon Runyon Foundation, J.M.R. is a predoctoral trainee of the NIH. We thank Drs. Jung Ja Park Kim, Don Hatfield, and Victor Day for useful discussions. We are grateful for the skillful technical assistance rendered by John Genova.

1. Fuller, W., Wilkins, M. H. F., Wilson, H. R. & Hamilton, L. D. (1965) "The molecular configuration of deoxyribonucleic acid IV. X-ray diffraction study of the A form," *J. Mol. Biol.* **12**, 60–75.
2. Langridge, R. & Gomatos, P. J. (1963) "The structure of RNA," *Science* **141**, 694–698.
3. Tomita, K. & Rich, A. (1964) "X-ray diffraction investigations of complementary RNA," *Nature* **201**, 1160–1163.
4. Rosenberg, J. M., Seeman, N. C., Kim, J. J. P., Suddath, F. L., Nicholas, H. B. & Rich, A. (1973) "The double helix at atomic resolution," *Nature*, in press.
5. Sobell, H. M., Tomita, K. & Rich, A. (1963) "The crystal structure of an intermolecular complex containing a guanine and a cytosine derivative," *Proc. Nat. Acad. Sci. USA* **49**, 885–892.
6. Haschemeyer, A. E. V. & Sobell, H. M. (1965), "The crystal structure of a hydrogen bonded complex of deoxyguanosine and 5-bromodeoxycytidine," *Acta. Crystallogr.* **19**, 125–130.
7. O'Brien, E. J. (1966) "Structure of a purine–pyrimidine crystalline complex: 9-Ethylguanine with 1-methyl-5-fluorocytosine," *J. Mol. Biol.* **22**, 377–379.
8. O'Brien, E. J. (1967) "Crystal structures of two complexes containing guanine and cytosine derivatives," *Acta Crystallogr.* **23**, 92–106.
9. Arnott, S., Hukins, D. W. L & Dover, S. D. (1972) "Optimised parameters for RNA double helices," *Biochem. Biophys. Res. Commun.* **48**, 1392–1399, Model A.
10. Kim, S. H., Berman, H. M., Seeman, N. C. & Newton, M. D. (1973)" Seven basic conformations of nucleic acid structural units," *Acta. Crystallogr.*, in press.
11. Seeman, N. C., Sussman, J. L., Berman, H. M. & Kim, S. H. (1971) "Nucleic acid conformation: Crystal structure of a naturally occurring dinucleoside phosphate (UpA)," *Nature New Biol.* **233**, 90–92.
12. Sussman, J. L., Seeman, N. C., Kim, S. H. & Berman, H. M. (1972) "Crystal structure of a naturally occurring dinucleoside phosphate: Uridylyl-3′,5′-adenosine phosphate. Model for RNA chain folding", *J. Mol. Biol.* **66**, 403–422.
13. Rubin, J., Brennan, T. & Sundaralingam, M. (1972) "Crystal and molecular structure of a naturally occurring dinucleoside monophosphate," *Biochemistry* **11**, 3112–3128.
14. Sundaralingam, M. (1969) "Stereochemistry of nucleic acids and their constituents," *Biopolymers* **7**, 821–860.
15. Voet, D. & Rich, A. (1970) "The crystal structure of purines, pyrimidines and their intermolecular complexes," *Progr. Nucl. Acid Res. Mol. Biol.* **10**, 183–265.
16. Arnott, S., Dover, S. D. & Wonacott, A. J. (1969) "Least-squares refinement of the crystal and molecular structures of DNA and RNA from x-ray data and standard bond lengths and angles," *Acta Crystallogr.* **B25**, 2192–2206, Model B.
17. Kim, S. H., Quigley, G. J., Suddath, F. L., MacPherson, A., Sneden, D., Kim, J. J. P., Weinzierl, J. & Rich, A. (1973) "The crystal structure of yeast phenylalanyl tRNA: The folding of the chain," *Science*, **179**, 285.