

## Structure of Guanylyl-3',5'-Cytidine Monophosphate. II. Description of the Molecular and Crystal Structure of the Calcium Derivative in Space Group $P2_1$

B. HINGERTY,\* E. SUBRAMANIAN,\*\* S. D. STELLMAN,\*\*\*  
S. B. BROYDE,† T. SATO,†† and R. LANGRIDGE, *Department  
of Biochemical Sciences, Frick Chemical Laboratory, Princeton  
University, Princeton, New Jersey 08540*

### Synopsis

The structural features of calcium guanosine-3',5'-cytidine monophosphate (GpC) have been elucidated by X-ray diffraction analysis. The molecule was crystallized in space group  $P2_1$  with cell constants of  $a = 21.224 \text{ \AA}$ ,  $b = 34.207 \text{ \AA}$ ,  $c = 9.327 \text{ \AA}$ , and  $\beta = 90.527^\circ$ ,  $Z = 8$ . The hydration of the crystal is 21% by weight with 72 water molecules in the unit cell. The four GpC molecules in the asymmetric unit occur as two Watson-Crick hydrogen-bonded dimers related by a pseudo-C face centering. Each dimer consists of two independent GpC molecules whose bases are hydrogen bonded to each other in the traditional Watson-Crick fashion. Each dimer possesses a pseudo twofold axis broken by a calcium ion and associated solvent.

The four molecules are conformationally similar to helical RNA, but are not identical to it or to each other. Instead, values of conformational angles reflect the intrinsic flexibility of the molecule within the range of basic helical conformations. All eight bases are *anti*, sugars are all C3'-*endo*, and the C4'-C5' bond rotations are *gauche-gauche*. The *R* factor is 12.6% for 2918 observed reflections at 1.2- $\text{\AA}$  resolution.

### INTRODUCTION

The dinucleoside phosphates have attracted considerable attention recently since they possess a representative backbone for polymeric nucleic acids, and as fragments of RNA may serve as models for larger structures such as tRNA.

We report here the essential features of the molecular and crystal structure of the  $\text{Ca}^{+2}$  salt of the dinucleoside phosphate, guanylyl-3',5'-cytidine

\* Present address: Max Planck Institute for Experimental Medicine, Gottingen, Federal Republic of Germany.

\*\* Present address: Laboratory of Molecular Biology, National Institutes of Health, Bethesda, Md. 20014.

\*\*\* Present address: American Health Foundation, 1370 Avenue of the Americas, New York, N. Y. 10019.

† To whom correspondence should be addressed at: School of Physics, Georgia Institute of Technology, Atlanta, Ga. 30332.

†† Present address: Shionogi Research Laboratory, Fukushima-ku, Osaka, Japan.

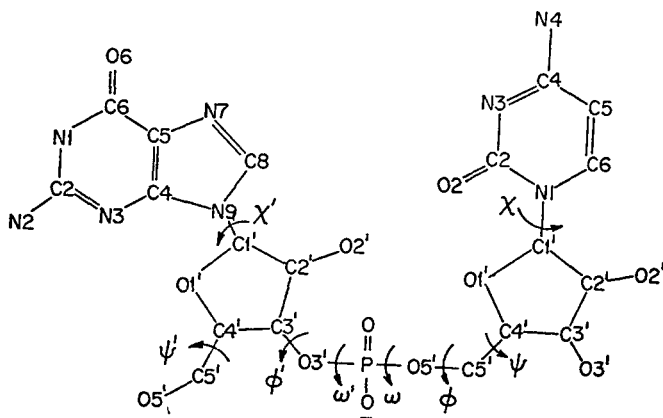


Fig. 1. Structure, numbering convention, and conformational angles for GpC.

monophosphate (GpC), shown in Figure 1. A structure for the  $\text{Na}^+$  salt of GpC was previously reported,<sup>1</sup> in which the molecule crystallized in space group  $C2$  with one molecule per asymmetric unit. In the present paper, the crystalline calcium salt occurred in space group  $P2_1$ , with four molecules per asymmetric unit. This crystal is not the same as the  $P2_1$  form of GpC, which was crystallized by Day et al.<sup>1</sup> but not solved. The structure of our modification of GpC is of particular interest in view of the large number of crystallographically independent molecules present in the unit cell, and because it is one of the largest crystallographic determinations made to date at atomic resolution, without use of isomorphous replacement. Of biochemical interest is the fact that this structure has been solved as the salt of a divalent metal, calcium. The effect of divalent cations on nucleic acid conformation in solution is well known,<sup>2-4</sup> and it is significant that we can now examine their effects on the structure of the solid state as well. Finally, in this crystal structure there are four independent observations of a helical RNA-like structure, whose ranges of conformational angles show GpC to be a rather flexible molecule within the limited helical range. The extent of this flexibility in the backbone may reflect the allowable conformations of RNA itself.

## EXPERIMENTAL

4.9 mg of GpC (Miles Laboratories) were suspended in 2.0 ml of 10 mM tris buffer, pH 8.7, containing 5 mM  $\text{CaCl}_2$ . The vessel containing the suspension was sealed and placed in a water bath. The solid was dissolved at 50°C with stirring. The temperature of the bath was gradually decreased at the rate of 1 or 2°C/day, down to 22°C. The bath was placed against a cool surface, which supplied a temperature gradient. In this arrangement GpC began to crystallize at about 35°C. The crystals were unstable in the air when separated from the mother liquor. On slow drying, however, the crystals stayed intact without much loss of resolution, but were subject to a

small amount of shrinkage. A crystal shrunk in this way measuring  $0.03 \times 0.06 \times 0.45$  mm was used for data collection. X-ray intensity data were collected on a Hilger-Watts Y290 diffractometer using nickel-filtered  $\text{CuK}\alpha$  radiation. The data were recorded out to  $1.2 \text{ \AA}$  by use of  $\theta$ - $2\theta$  scanning. 4237 reflections were recorded, and of these 2918 exceeded background by one standard deviation, and were therefore considered to be observed. Only one standard deviation in intensity was used because of the presence of a high degree of pseudosymmetry causing a large number of systematically weak reflections.

The space group is  $P2_1$  based upon the systematic absence  $0k0$ ,  $k$  odd. Unit cell constants of  $a = 21.224 \text{ \AA}$ ,  $b = 34.207 \text{ \AA}$ ,  $c = 9.372 \text{ \AA}$ , and  $\beta = 90.527^\circ$ ,  $Z = 8$  were determined. The experimental density was  $1.512 \text{ g/cm}^3$  measured by a flotation gradient. This value indicates a high degree of hydration, with approximately 72 water molecules in the unit cell and four  $\text{Ca}^{+2}$  ions. The symmetry was sufficiently close to a doubled C2 cell, so that this pseudosymmetry could be used for obtaining the trial structure. The procedure employed for obtaining the conformation and packing scheme has been reported.<sup>5,6</sup>

## RESULTS

### Solution of the Structure

Since this structure was of such great complexity, containing 200 non-hydrogen atoms in the asymmetric unit, and since the methods employed in the analysis were somewhat unconventional, the detailed discussion of its solution will be dealt with elsewhere.<sup>7</sup> We can now report that the  $R$  factor is 12.6% for the 2918 observed reflections. In calculating this number all nonhydrogen atoms were given isotropic temperature factors, except the calcium and phosphorus atoms, which were anisotropic.

$$R = \frac{\sum |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum |F_{\text{obs}}|}$$

where the  $F_{\text{obs}}$  are the observed structure amplitudes and the  $F_{\text{calc}}$  the calculated structure factors based on the refined atomic coordinates.

### Description of the Structure

Within the asymmetric unit for four GpC molecules are similar but not identical. The molecules, numbered 1-4, obey the approximate relations given in Table I, which are exact in space group C2. Hence, the  $P2_1$  structure may be considered a deviation from C2 symmetry. The magnitude of the deviation is shown in Table II. Base pairing is between molecules related by an approximate twofold axis, which is not to be confused with the helix dyad axis, which is nearly perpendicular to it (Figure 2). Thus, molecule 1 is base paired to a unit cell-translated molecule 2, and similarly for molecules 3 and 4. One calcium of the asymmetric unit is associated with the phosphate group of molecules 1 and 2 (calcium 1), and the other in the same way with molecules 3 and 4 (calcium 2).

TABLE I  
Symmetry Relations Among the Four Molecules (Exact in C<sub>2</sub>, Approximate in P<sub>2</sub>)

Molecule				
1	$x$	$y$	$z$	
2	$-x$	$y$	$-z$	
3	$1/2 + x$	$1/4 + y$	$z$	
4	$1/2 - x$	$1/4 + y$	$-z$	

$1/4$  translation for  $y$  since in C<sub>2</sub> the cell is doubled.

TABLE II  
Fractional Coordinates of Phosphorous Atoms Observed in GpC, and Expected Values From C<sub>2</sub> Pseudosymmetry Relations of Table I

Molecule	Observed			Expected			Positional Deviation (Å)
	$x$	$y$	$z$	$x$	$y$	$z$	
1	0.1106	-0.0080	-0.1247				
2	0.8638	0.0017	0.1039	0.8894	-0.0080	0.1247	0.66
3	0.6242	0.2522	-0.1322	0.6106	0.2420	-0.1247	0.46
4	0.3807	0.2447	0.1089	0.3894	0.2420	0.1247	0.25

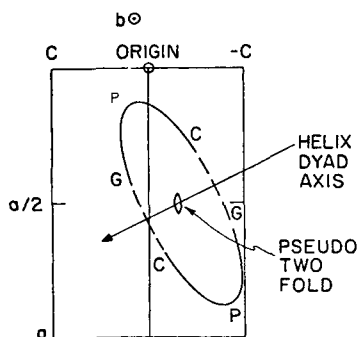


Fig. 2. Watson-Crick base paired dimer showing the difference between the helix dyad axis, which is in the direction of the arrow, and the crystallographic pseudo-twofold axis, which is perpendicular to the plane of the paper; G = guanine, C = cytosine, P = phosphorus.

Hence, there is one  $\text{Ca}^{+2}$  ion for every two GpC molecules in the crystal and therefore two in the asymmetric unit. The  $\text{Ca}^{+2}$  ion is unsymmetrically located since it is not on the pseudo-twofold axis (Figure 3). The GpC molecules are paired by six Watson-Crick hydrogen bonds per dimer (Figure 4), and the dimeric molecules pack so as to maximize the overlap of six-membered aromatic rings.

Of interest is the phosphorous-phosphorous distance in the dimers as this gives some measure of the relation of the base paired GpC's to each other. We have two different observations of this quantity, which are nearly identical, namely 17.66 and 17.65 Å. Another informative quantity is the interglycosidic carbon-carbon distance (C1'-C1'), since it gives a

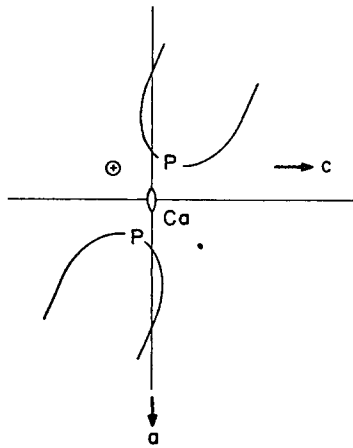
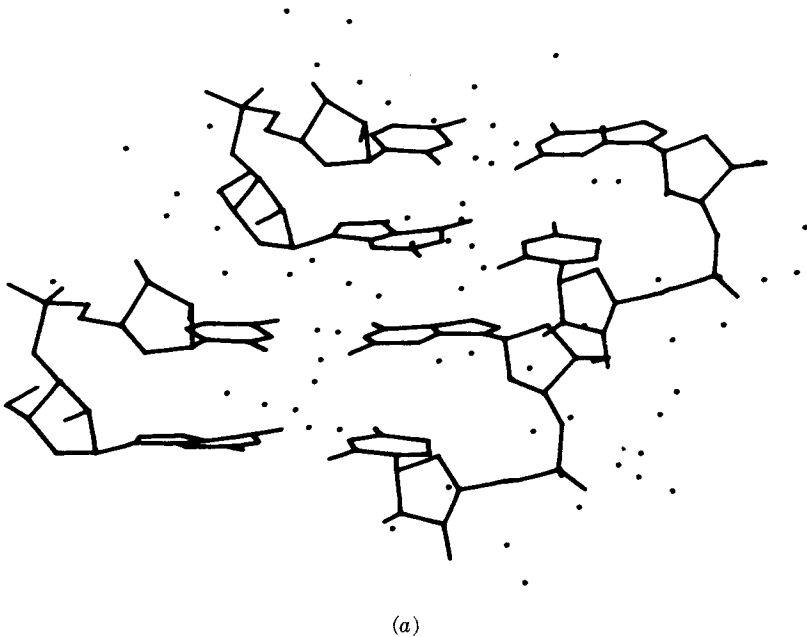


Fig. 3. Schematic representation of a calcium position where  $\oplus$  denotes the spot of the vacant calcium site on the other side of the pseudo-twofold axis (denoted by  $\phi$ ); P = phosphorus.

rough measure of the distance between the base pairs. For this there are the four observations, 10.59, 10.42, 10.53, and 10.66 Å. The dihedral angles between the mean-square planes of the guanine and cytosine rings are given in Table III and indicate a significantly smaller angle ( $6.9^\circ$  average) than the  $11.6^\circ$  for Na(GpC).<sup>1</sup> Our latest estimates at this stage of



(a)  
Fig. 4 (continued)

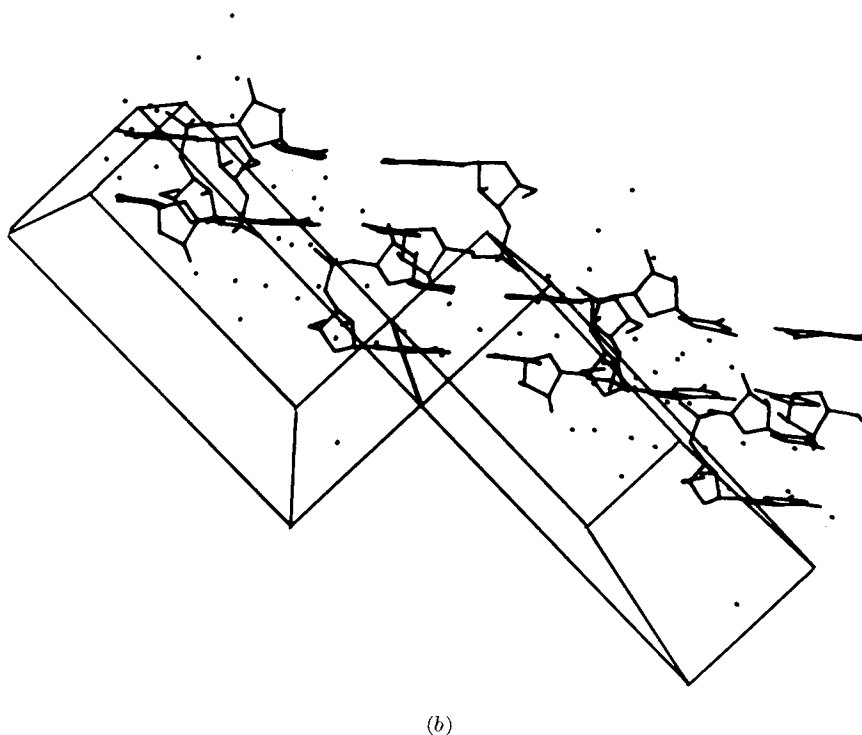


Fig. 4. (a) Contents of the asymmetric unit of GpC, showing the two Watson-Crick dimer pairs. Locations of water molecules are represented by dots. (b) View down  $b$  axis of two adjacent asymmetric units of GpC, showing dimeric Watson-Crick base pairing and solvation scheme.

refinement of the standard deviation are  $0.1 \text{ \AA}$  for bond lengths and  $6^\circ$  for bond angles. Within this limit our results are consistent with those values reported in the literature.<sup>8-11</sup>

### Conformational Angles

The values of dihedral angles of rotation about single bonds have become the signature by which different forms of the same molecule are distinguished from each other. In Table III we show the values of eight conformational angles (defined according to the convention of the first paper in this series<sup>6</sup>) for each of the four molecules of the asymmetric unit. First it is noted that all nucleosides are in the *anti* conformation. Although all of the observations can be classified as helical RNA-like, there are significant differences in some angles. These differences are not attributable to the uncertainty in the coordinates, which produces an estimated error in any dihedral angle of at most  $7^\circ$  at this stage. If all four observations are averaged, the estimated standard deviation is reduced to  $4^\circ$ . Of the seven angles reported by Day et al. for Na(GpC) only  $\omega$ ,  $\psi$ , and  $\chi$  fall within the range of our observations for those angles. The differences may be related

TABLE III  
Conformational Angles Observed for Four Reported GpC Molecules, Sodium Salt of GpC, and Two Recent Models of RNA

Molecule	Ref.	$\chi'$	$\psi'$	$\phi'$	$\omega'$	$\omega$	$\phi$	$\psi$	$\chi$	Ribose Puckering	Interplanar Spacing	Deviation of GC Planes from Parallelism (deg)
GpC-1		12	57	227	287	294	175	45	38	G C3'endo-C2'exo C C3'endo-C2'endo	3.58	8.5
GpC-2		-2	49	219	287	295	171	56	14	G C3'endo C C3'endo-C2'endo	3.73	7.2
GpC-3	This work	3	49	222	291	289	165	59	28	G C3'endo-C2'endo C C3'endo	3.66	6.5
GpC-4		3	49	217	290	279	179	58	32	G C3'endo-C2'exo C C3'endo-C2'exo	3.62	5.4
Na(GpC)	1	13		209	291	284	186	51	25	C3'endo	3.55	11.6
RNA-11	13			209	286	298	180	48	12	C3'endo	3.45	—
RNA-12	13			193	300	295	193	44	17	C3'endo	3.47	—

The G or C notation in the ribose column shows to which nucleoside moiety the ribose belongs.

to the different solvation scheme for the sodium and calcium salts (see below).

### Sugar Pucker

While all eight ribose groups of the asymmetric unit are C3'-*endo*, they differ substantially in the contribution of other components (Table III). Three of the sugars exhibit a C2'-*endo* component, three others have a C2'-*exo* component, while the two remaining sugars are a purely C3'-*endo* envelope conformation, with the C2' atom in the plane. Of the 40 in-ring sugar conformational angles, all but two are within our 6° error limit for previously observed ranges for these angles.<sup>12</sup>

### Intramolecular Hydrogen Bonds

The C8 of guanine is in close contact with both O1' and O5' of the adjoining sugar. The distances range from 2.95 to 3.39 Å and from 2.54 to 2.84 Å for C8...O5' and C8...O1', respectively, indicating possible hydrogen bonds between these two sets of atoms.<sup>11</sup>

### Intermolecular Hydrogen Bonds

In addition to participating in six Watson-Crick hydrogen bonds listed in Table IV, each GpC molecule exhibits two other close contacts that are probably hydrogen bonds. These are O1'(C)...O5'(G) and N4(C)...O3'(C). The molecules that participate in these bonds are given in Table V.

TABLE IV  
Watson-Crick Hydrogen-Bond Distances for GpC Dimers. Parentheses Show Hydrogen-Bonded Atom of Dimeric Partner

Molecule	O6(-N4)	N1(-N3)	N2(-O2)
1	2.85	2.98	2.77
2	2.88	2.96	2.85
3	3.01	2.97	2.94
4	2.81	2.98	2.88

Distances given in angstroms.

TABLE V  
Additional Close Interatomic Contacts Between Atoms of Different Molecules

Atom	Molecule	Atom	Molecule	Distance (Å)
O1'(C)	1	O5'(G)	4	3.02
O1'(C)	2	O5'(G)	3	3.06
O1'(C)	3	O5'(G)	1	3.13
O1'(C)	4	O5'(G)	2	3.16
N4(C)	1	O3'(C)	3	3.03
N4(C)	2	O3'(C)	4	2.92
N4(C)	3	O3'(C)	2	2.87
N4(C)	4	O3'(C)	1	2.93



### Location of Cations and Solvent

The two divalent cations are an important stabilizing factor in the crystal. Two of the six octahedral sites in the coordination sphere of each calcium are occupied by oxygens of phosphate groups from two GpC molecules, each of which belongs to a different dimer, while the remaining four sites are occupied by water. Figure 5 shows this symmetry relationship and shows how two of the waters in the  $\text{Ca}^{+2}$  sphere are also hydrogen bonded to the guanosine groups of molecules 2 and 4. These hydrogen bonds may be responsible for the unequal values of  $\chi$  among the four molecules.

One of the most important aspects of the structure is the solvation scheme. Its elucidation was fairly difficult in the higher symmetry C2 space group since it yielded 14 sites for nine waters, showing that many sites were partially occupied. The  $\text{Ca}^{+2}$  ion could be located at this stage but with only half occupancy. When the refinement continued in the real space group P2<sub>1</sub>, the importance of the  $\text{Ca}^{+2}$  ion became clear. It was present on only one side of the pseudo-twofold axis, strictly breaking the symmetry (Figure 2). The waters that were present near the vacant calcium site rearranged themselves in order to fill in the hole. This further broke the symmetry and explains why we observed several waters with only half occupancy in the space group C2.

As can be seen from Table II the  $\text{Ca}^{+2}$  ion appears to have perturbed the molecules also. One phosphate group has moved up in  $y$  while the other has moved down. This could be due to the size of the  $\text{Ca}^{+2}$  ion, which is larger than the  $\text{Na}^+$ , and also to the fact that the  $\text{Ca}^{+2}$  is unsymmetrically located, affecting each molecule differently. The high charge on  $\text{Ca}^{+2}$  compared to  $\text{Na}^+$  may also be a contributing factor.

Of the 36 water molecules in the asymmetric unit, 32 were found with full occupancy while four were disordered, with two sites for each position. These disordered sites can be explained by the fact that they are located in positions that are not near the molecule. Thus, they are not strongly bound in the structure and may therefore display partial site occupancy.

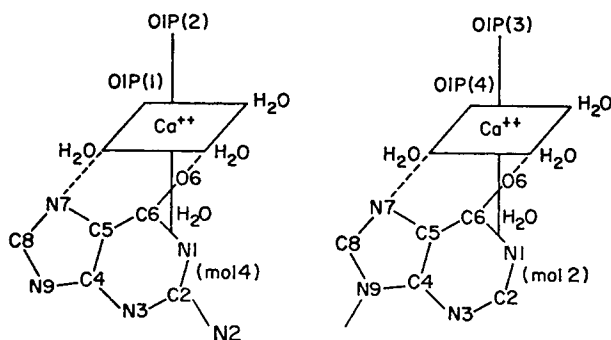


Fig. 5. Schematic representation of calcium coordination system showing the octahedral coordination scheme and the indirect hydrogen bonding to the guanine rings of molecules 2 and 4.

The GpC pair is surrounded by water on all sides, which minimizes intermolecular interactions. As pointed out by Day et al.<sup>1</sup> this seemingly allows the molecule to take on its natural or *in vivo* state. This must be the reason why our conformational and packing calculations, which ignored such interactions except for Watson-Crick base pairing, yielded the solution.<sup>5,6</sup>

In summary then, the  $\text{Ca}(\text{GpC})_2$  and  $\text{Na}(\text{GpC})$  have significantly different crystal structures, mainly reflected in the modified solvation scheme for the cations. The molecules are in turn perturbed by the new solvation and the result is four crystallographically independent observations of the GpC molecule in the asymmetric unit. These molecules are now no longer identical as in  $\text{Na}(\text{GpC})$ , and this shows the extent to which the backbone angles can be modified while still preserving the helical character of the structure.

Our present work on this structure is directed towards further refinement by correcting the data for absorption, secondary extinction, and anisotropic thermal parameters for the remainder of the nonhydrogen atoms. A final attempt is being made to find the hydrogen atoms. All of this is being done with the hope of seeing how far a large structure such as this can be refined at a resolution of 1.2 Å.

The authors thank Edwin Heilweil and Russell Massey for the photographic work. They also express their appreciation to Prof. J. R. Fresco for his critical comments. The molecular model building was performed at the Princeton University Computer Graphics Laboratory, supported by NIH Grant RR-00578; other support was received from NIH Grant GM-16539 and NSF Grant GB-28021.

### References

1. Day, R. O., Seeman, N. C., Rosenberg, J. M. & Rich, A. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 849-853.
2. Lindahl, T., Adams, A. & Fresco, J. R. (1966) *Proc. Nat. Acad. Sci. USA* **55**, 941-948.
3. Adams, A., Lindahl, T. & Fresco, J. R. (1967) *Proc. Nat. Acad. Sci. U. S.* **57**, 1684-1691.
4. Danchin, A. (1972) *Biopolymers*, **11**, 1317-1333.
5. Stellman, S. D., Hingerty, B., Broyde, S. B., Subramanian, E., Sato, T. & Langridge, R. (1973) *Macromolecules* **6**, 652-653.
6. Stellman, S. D., Hingerty, B., Broyde, S. B., Subramanian, E., Sato, T. & Langridge, R. (1973) *Biopolymers* **12**, 2731-2750.
7. Hingerty, B. et al. (in preparation).
8. Rosenberg, J. M., Seeman, N. C., Day, R. O. & Rich, A. (1973) *Amer. Crystallogr. Assoc. Summer Meeting* **1**, Ser. 2, 130.
9. Arnott, S. (1970) *Progr. Biophys. Mol. Biol.* **21**, 265-319.
10. Sundaralingam, M. (1969) *Biopolymers* **7**, 821-860.
11. Sussman, J., Seeman, N. C., Kim, S.-H. & Berman, H. J. (1972) *J. Mol. Biol.* **66**, 403-421.
12. Altona, C. & Sundaralingam, M. (1972) *J. Amer. Chem. Soc.* **94**, 8205-8212.
13. Arnott, S., Hukins, D. & Dover, S. (1972) *Biochem. Biophys. Res. Commun.* **48**, 1392-1399.

Received April 9, 1974

Accepted May 7, 1974