BIOPOLYMERS VOL. 12, 2731–2750 (1973)

Structure of Guanosine-3′,5′-cytidine Monophosphate. I. Semi-empirical Potential Energy Calculations and Model-Building

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

Synopsis

The conformation and packing scheme for guanosine-3′,5′-cytidine monophosphate, GpC, were computed by minimizing the classical potential energy. The lowest energy conformation of the isolated molecule had dihedral angles in the range of helical RNA's, and the sugar pucker was C3′ *endo*. This was used as the starting conformation in a packing search over orientation space, the dihedral angles being flexible in this step also. The packing search was restricted by constraints from our x-ray data, namely, (1) the dimensions of the monoclinic unit cell and its pseudo-C2 symmetry (the real space group is $P2_1$), (2) the location of the phosphorous atom, and (3) the orientation of the bases. In addition, a geometric function was devised to impose Watson-Crick base pairing. Thus, a trial structure could be sought without explicit inclusion of intermolecular potentials. An interactive computer graphics system was used for visualizing the calculated structures.

The packing searches yielded two lowest energy schemes in which the molecules had the same conformation (similar to double-helical RNA) but different orientations within the unit cell. One of these was refined by standard x-ray methods to a discrepancy index of 14.4% in the C2 pseudocell. This served as the starting structure for the subsequent refinement in the real $P2_1$ cell.[5]

INTRODUCTION

X-ray crystallographic methods have provided the bulk of our present knowledge of the three-dimensional structure of nucleosides and nucleotides, and from such studies it has been possible to deduce empirical rules governing regions of allowable conformations.[1–3] Elucidation of these structures by crystallographic techniques is often consuming; hence any method of predicting a good trial structure should greatly simplify the crystallographic analysis.

A common procedure for arriving at a trial structure (when other techniques have failed) has been to build models of the molecule under study

* Present address. Laboratory of Molecular Biology, National Institutes of Health, Bethesda, Maryland 20014.

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

and inspect the packing of the molecules in the unit cell, using information available from diffraction data and stereochemistry. This procedure can be greatly simplified by the use of interactive graphics display. However, for molecules possessing conformational flexibility, even with the use of display there would be a large number of trials necessary before arriving at a satisfactory trial structure that would explain the observed diffraction data. The number of trials can be greatly reduced if conformers corresponding to minimum potential energy are tried in the packing analysis.

In this paper we describe a systematic procedure used to determine a suitable trial conformation for guanosine-3′,5′-cytidine monophosphate (GpC), a dinucleoside phosphate for which X-ray diffraction data were available in our laboratory. Conventional crystallographic procedures failed to solve the structure in view of the fact that the diffraction data possesses pseudo-symmetry and also does not extend beyond a resolution of 1.2 Å. In the first stage of our analysis, the classical potential energy of a single GpC molecule was minimized with respect to its conformational angles. In the second stage, sets of conformational angles corresponding to the lowest energy minima were used as starting points in minimization searches of reasonable crystal packing schemes. A preliminary account of the first, intramolecular stage has been published previously.[4]

As will be shown below, minimization and search procedures were subject to constraints of two types: (a) chemical constraints, such as fixed bond lengths and bond angles, and potential energy parameters; and (b) experimental geometric constraints for this molecule deduced from the X-ray data. Choices for starting conformations for the *in vacuo* runs were based on observations of numerous related structures. An interactive computer graphics system was used to display calculated minimum-energy structures and examine possible packing schemes consistent with the observed unit cell parameters and crystallographic symmetry. Every model proposed was checked immediately against the X-ray data by computing the conventional crystallographic discrepancy index, R, and by direct visual inspection.

$$R = \sum|(|F_o| - |F_c|)|/\sum|F_o|$$

(Where F_o is the observed structure factor and F_c is the calculated structure factor on the basis of the model coordinates. All values of R reported here refer to 3-Å resolution.) One of the trial structures found by this method has satisfactorily explained the observed diffraction data and has been subjected to crystallographic refinement. The details of this crystallographic analysis will be presented elsewhere.[5]

Semi-empirical potential energy calculations have been widely used to predict conformations and/or crystal packing schemes for many classes of molecules. Olson and Flory[6] and Sasisekheran and co-workers[7,8] have calculated energies of mono- and polynucleotides, although no attempt was made to minimize the energy with respect to all the conformational angles simultaneously. Scheraga and co-workers[9,10] have calculated a complete

energy surface for a cyclic peptide, cyclo-(glyglyglypropro) under conditions of rigid bond lengths and bond angles (stage a), and also obtained the important minima in the potential energy surface when those parameters were allowed to vary (giving 141 independent variables) (stage b).[11] The thirteen minima found in stage a served as starting conformations for the minimizations of stage b. In calculations of a related nature, Bovey et al.[12] combined results of energy calculations with nmr coupling constants and other experimental data to obtain plausible structures (or groups of structures) for cyclic oligopeptides in solution.

The studies mentioned so far concern only intramolecular calculations, i.e., calculation of the conformational energy of a single molecule. Energy minimization methods have also been used in prediction of intermolecular packing schemes. Zugenmaier and Sarko[13] report minimization of the potential energy (repulsive only) of several different monosaccharides using the method of Williams,[14] to obtain crystal packing schemes for comparison with X-ray crystal diffraction analysis. In those calculations, the sugar molecule and its symmetry-related mates were moved within the fixed unit cell, and starting conformations were chosen by random translation and rotation of the rigid molecule. In a subsequent work on B-amylose, Zugenmaier and Sarko[15] permitted bond lengths and angles as well as the chain conformation to vary in calculating probable crystalline packing models. Ahmed et al.[16] minimized a combination of intra- and intermolecular potential functions for a series of organometallic compounds to obtain preferred conformation and crystal packing schemes for their molecules. In their case, one variable of each type was used. Coïro et al.[17] solved the crystal structure of *N,N'*-dicyclohexylurea, a molecule containing four internal degrees of freedom, by first minimizing the intramolecular potential energy of the molecule, and then packing the minimum-energy conformation as a rigid body within the experimental unit cell. Finally, Stellman et al.[18] combined X-ray diffraction data with crystal packing energy minimization to relate the structural and thermodynamic properties of poly-(1,4)-*trans*-butadiene single crystals.

These works are examples of what is being done in this active and expanding field. Recent reviews of conformational energy calculations have been made by Scheraga[19] and by Brant.[20]

METHOD

The PDP-10/LDS-1 interactive computer graphics system was used in this work.[21] A program was written which displays on the screen any dinucleoside ribophosphate, XpY, where X and Y may be adenine, guanine, cytosine, or uracil. The primary structure is constructed by the linked-atom algorithm of Scott and Scheraga,[22] with constant bond lengths and bond angles taken from small molecule studies. Figure 1 gives the structure, the numbering convention, and the conformational angles for GpC; Table I defines these angles, as per Sussman et al.[23] Dihedral angles are

TABLE I
Definition of Dihedral Angles for GpC

Angle[a]	Bonds
χ'	O1′–C1′–N9–C8
ψ'	C3′–C4′–C5′–O5′
ϕ'	P–O3′–C3′–C4′
ω'	O5′–P–O3′–C3′
ω	C5′–O5′–P–O3′
ϕ	C4′–C5′–O5′–P
ψ	C3′–C4′–C5′–O5′
χ	C6–N1–C1′–O1′

[a] All angles A–B–C–D are measured clockwise from A to D when viewed along B–C. A eclipsing D is 0°.[23]

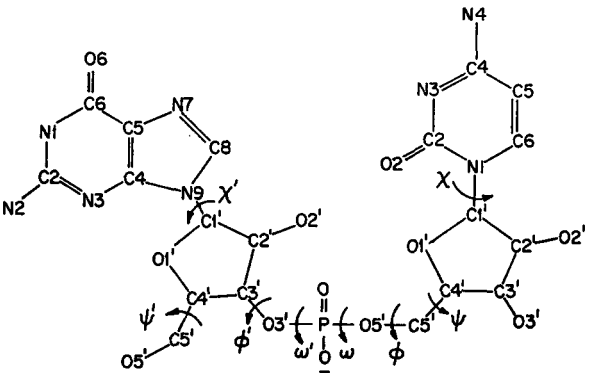


Fig. 1. Structure, numbering convention and conformational angles for GpC.

entered from knobs or from the teletype, allowing display of any geometrically possible conformation. Similarly, the molecule as a whole can be oriented within the cell and displayed. By displaying two or more adjacent cells, the graphics could be used to examine the mutual relationship among symmetry-related molecules.

Packing experiments were subjected to the following three constraints, deduced from the X-ray diffraction data:[5] (1) a monoclinic pseudo-unit cell with dimensions $a = 21.224$, $b = 17.104$, $c = 9.372$, $\beta = 90.527°$, and space group C2 (the true spacegroup is $P2_1$, with $b = 2 \times 17.104$); (2) the location of the phosphorus atom, obtained from the three-dimensional Patterson function; (3) the bases restricted to be parallel to the crystallographic y axis. The first constraint defines the symmetry of packing in the lattice. The second constraint fixes one point on the molecule within the cell. The third constraint reduces to two the number of independent Eulerian angles necessary to define the orientation of the GpC molecule as a rigid body.

ENERGY CALCULATION

As reported earlier,[4] the energy calculation was based on that of Scott and Scheraga for polypeptides,[24] using Eq. (1)

$$E = \sum_{i<j}\sum (a_{ij}r_{ij}^{-6} + b_{ij}r_{ij}^{-12}) + \sum_{i<j}\sum 332q_iq_jr_{ij}^{-1}\epsilon^{-1} + \sum_{k=1}^{8}\frac{V_{0,k}}{2}(1 + \cos 3\theta_k) \quad (1)$$

The first term represents the contribution to the energy, E, by nonbonded interactions, the second is electrostatic, and the third is torsional. The double sums extend pairwise over all interacting atoms i and j, where r_{ij} is the distance between the atoms, q_i is the charge on atom i, a_{ij} and b_{ij} are parameters in the Lennard-Jones potential, and ϵ is the dielectric constant. Pairs of atoms whose relative distances cannot change when only dihedral angles are varied, such as atoms in the rigid cytosine ring, were excluded from the sums. The single summation extends over all eight flexible dihedral angles, where θ_k is the kth dihedral angle and $V_{0,k}$ is the rotational barrier height for that rotation. The parameters a_{ij}, b_{ij}, as well as q_i, V_0, and ϵ were taken from Refs. 7 and 8, with a net charge of -1 on the phosphate group. A modified version of the algorithm of Powell,[25] obtained from the Courant Institute of New York University, was used to minimize the energy, using the dihedral angles, and later also the Eulerian orientation angles, as variable parameters. All minimizations were carried out to an accuracy of 1° in each angular parameter. No angle was permitted to vary by more than 100° at any given minimization step.

"IN VACUO" CALCULATION

The first stage of the calculations was a determination of the lowest minimum energy conformation for the isolated molecule, i.e., minimization of the energy, E, as given by Eq. (1). In this approximation the effects of solvent and of interaction with neighboring molecules is neglected, except that the dielectric constant is taken as 4.0.[8] Subsequent addition of terms to account for intermolecular interactions in the potential energy calculations significantly increases the number of variables, along with computation time. Therefore the starting conformation in the packing step should be as close as possible to the packed minimum for efficient convergence, and the lowest minimum for the isolated molecule was deemed to be a good starting point.

Choices of starting conformations for *in vacuo* runs were based on observed ranges for mono and polynucleotides.[1,3] In all, eighteen starting conformations were used, as follows: $\chi' = 15°$ (*anti*), $\psi' = 50°$, $\phi' = 225°$, $\omega' = 60°, 180°, 290°$, $\omega = 60°, 290°$, $\phi = 180°$, $\psi = 60°, 180°, 270°$, and $\chi = 15°$ (*anti*). The sugars were fixed at either C3′ *endo* or C2′ *endo*. In the latter case initial values for χ and χ' were 55°.

CRYSTAL PACKING

The next stage was to orient the molecule within the unit cell, subject to the constraints of the data mentioned above, and also including the effects of base pairing. A Watson-Crick type base pairing scheme was assumed, since guanine and cytosine will preferentially base pair in this way unless prevented for some steric or chemical reasons.[26] In UpA,[23] which was crystallized from acid solution, the N_1 of the adenine was protonated, precluding this kind of base pairing. Since, however, GpC was crystallized from basic solution, the N_1 of guanine was probably not protonated, so on chemical grounds it was reasonable to expect Watson-Crick type base pairing. This is indeed confirmed by the X-ray data where the strongest reflection of the data corresponds to a plane of spacing 3.4 Å; this together with an observed twofold pseudorotation axis strongly suggests a base pairing situation.

Thus, for GpC, base pairing can be treated as a purely geometric constraint, and can be simulated mathematically. By simultaneously minimizing the internal energy *and* the deviation from "correct" hydrogen bond geometry, we hoped to find at least one configuration in which intermolecular contacts (observable by interactive display) were also minimized. All other explicit intermolecular energy contributions were neglected. This approach greatly simplifies the calculations, and thereby cuts down on computer time (and real time) needed for minimization.

The standard Watson-Crick hydrogen bond configuration for a G-C pair is given in Figure 2a. A mathematical function H was devised which would be equal to zero when two selected molecules were oriented in this hydrogen bonding pattern. The function is non-negative and equals zero when the distance between the N4 on the cytosine of one molecule and O6 on the guanine of another was the value shown in Eq. (2) and the angle N4–H4–O6 was 180°; similar criteria were established for the cytosine N3–guanine N1 and the cytosine O2–guanine N2 interaction. An additional term, p, forces the bases to be parallel.

$$H = (d_1 - 2.9)^2 + (d_2 - 3.0)^2 + (d_3 - 2.9)^2 + 0.01(180 - \alpha_1)^2 + 0.01(180 - \alpha_2)^2 + 25p \quad (2)$$

where d_1, d_2, and d_3 are the hydrogen bond distances O2(C)...N2(G), N3(C)...N1(G), and N4(C)...O6(G) in Å, respectively, $\alpha_1 = \angle$N4–H4-(C)...O6(G) $\alpha_2 = \angle$O2(C)...H2–N2(G), with α_1 and α_2 expressed in degrees; $p = |\mathbf{c}_1 - \mathbf{c}_2|^2$ where $\mathbf{c}_1$ is a unit vector perpendicular to the cytosine plane, and $\mathbf{c}_2$ is a unit vector perpendicular to the guanine plane, and parallel to $\mathbf{c}_1$ when the bases are paired in the Watson-Crick scheme. This method of dealing with hydrogen bonding is very useful when the bonding partners can be deduced from chemical considerations. Semi-empirical functions[28] used to calculate hydrogen bond potentials are more complex mathematically, and more difficult to parameterize than the other compo-

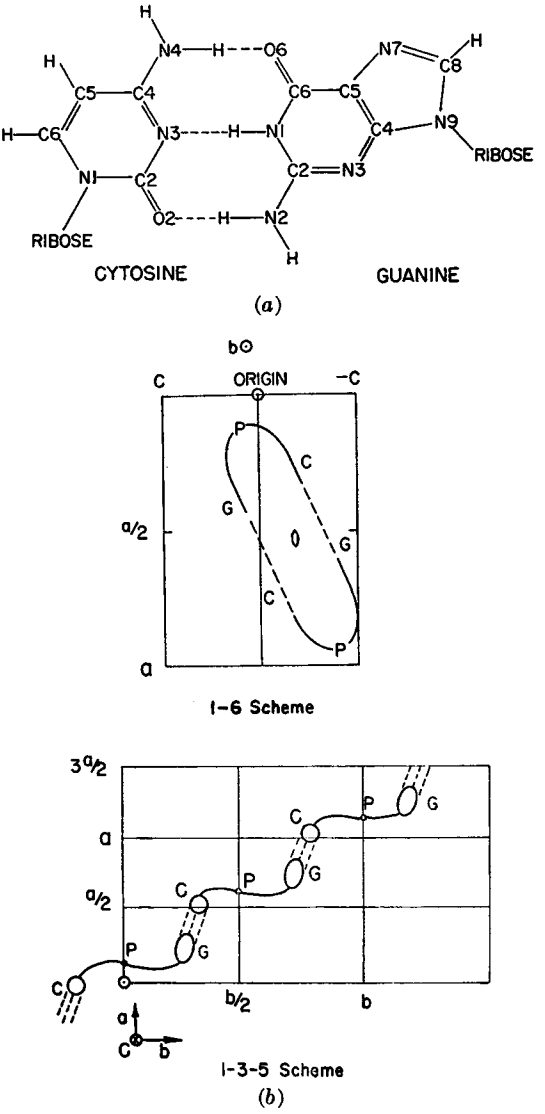


Fig. 2. (a) Watson-Crick hydrogen-bonding scheme for a guanine–cytosine pair. (b) Possible hydrogen bonding partners for GpC, molecules numbered as per Table II.

nents in Eq. (1) as well as less certain to lead to the expected base pairing scheme.

Three orientation parameters, corresponding to Eulerian rotations of the molecule as a rigid body, are needed to specify absolute orientations within the unit cell. Requiring the base planes to be parallel to the y axis as the X-ray data suggests, fixes one of those and leaves two to be varied in the minimizations.

Thus, the total function

$$F = E + H \tag{3}$$

was minimized with respect to the eight dihedral angles as well as the two remaining molecular orientations, giving in all ten variable parameters. By minimizing this sum, we locate minimum energy conformations which are also packed as Watson-Crick base pairs. In order for the minimization algorithm to act equally on E and H, it is necessary for these components to have values of the same order of magnitude at the minimum. To accomplish this, H, whose units are kcal/mole, was weighted by the factors given in Eq. (2).

CHOICE OF HYDROGEN BONDING PARTNERS

The choice of which guanine to allow to hydrogen bond to which cytosine depends on the spatial arrangement of neighboring molecules within the unit cell. The unit cell (in space group C2) has four molecules and the symmetry relationship among them is shown in Table II. The table also gives the coordinate transformations for two molecules in adjacent cells. When the starting conformation was helical, molecules 1 and 6 of Table II were linked to give dimers. For extended starting conformations, molecules 1, 3, and 5 were linked, giving a ribbon-like structure. They will be referred to as the 1–6 and the 1–3–5 bonding schemes, respectively. These hydrogen-bonded packing arrangements are illustrated in Figure 2b.

The local minima obtained are very much dependent on the starting conformation. Thus a major problem associated with this method is the choice of initial parameters. The local minimum energy conformation for the isolated molecule was generally used as the starting conformation. Packing was attempted for three different conformations, namely the lowest and the second lowest energies for C3′ *endo* sugar, and the lowest C2′ *endo*.

TABLE II
Possible Base-Paired Packing Schemes

Molecule no.	Fractional coordinates		
1	x	y	z
2	$\bar{x}$	y	$\bar{z}$
3	$1/2 + x$	$1/2 + y$	z
4	$1/2 - x$	$1/2 + y$	$\bar{z}$
5	$1 + x$	$1 + y$	z
6	$1 - x$	y	$-1 - z$
Scheme 1–3–5	G of 1 with C of 3		
	G of 3 with C of 5		
Scheme 1–6	G of 1 with C of 6		
	G of 6 with C of 1		

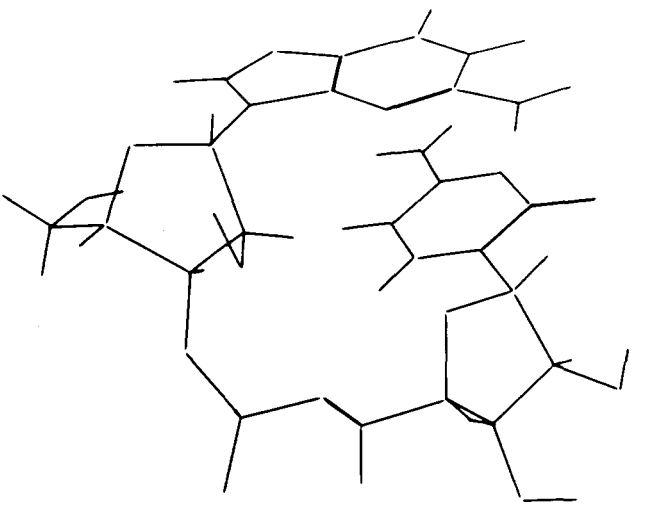
Note that molecules 1–4 are in the same unit cell.

Computer running time on the PDP-10 was about 1 hour for each ten-parameter trial when the system served no other user. For each starting conformation, an overall search of orientation space was made at 90° intervals of the two unconstrained Eulerian angles, for a total of 16 trials. The resulting conformations and packing schemes were examined on the graphics where those which were tangled could immediately be eliminated. This circumvented the need for a repulsive intermolecular potential in the energy calculation, which would otherwise have been essential, and thus permitted a substantial reduction in computational time.

RESULTS

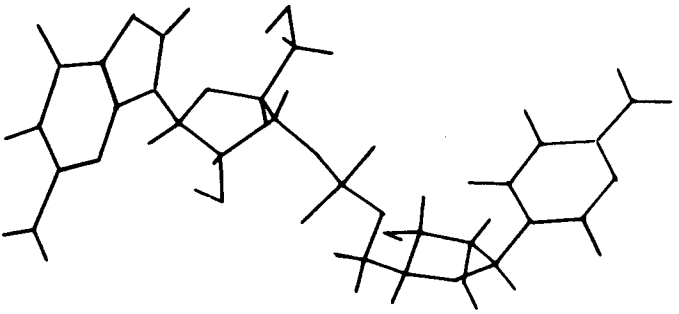
In Vacuo

For the isolated molecule with the sugar pucker at C3′ *endo,* seventeen different local minima were obtained from the initial eighteen conformations. Table III shows a selection of these. The lowest energy conformation, with $E = -39.0$ kcal/mole, resulted from two different sets of initial conditions, and had dihedral angles in the range of observed helical RNA's, the bases being almost parallel to one another. Very near it in energy was an extended conformation comparable to one deduced for aqueous solution from nmr data by Barry et al.[29] Of these seventeen local minima, eight were bent and nine were extended. With the exception of energy number 2, extended conformations were found to be of higher energy clustering at energies number 7, 10–14, 16, and 17. The bent conformations, other than the lowest energy form, had the bases tilted or even perpendicular to one another.

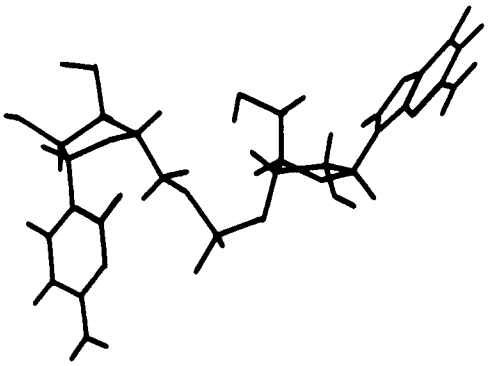


(*a*)

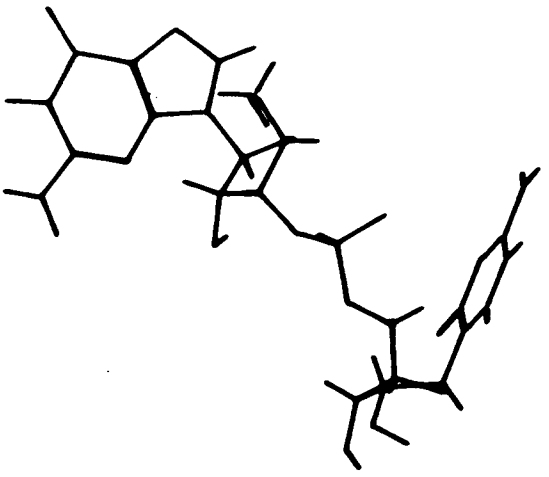
Fig. 3 (*continued*)



(*b*)



(*c*)



(*d*)

Fig. 3. (a) Lowest minimum energy conformation calculated for an isolated GpC molecule, sugar pucker C3′ *endo*. $E = -39.0$ kcal/mole. (b) Second lowest energy conformation, sugar pucker C3′ *endo*. $E = -38.1$ kcal/mole. (c) Lowest energy conformation obtained for isolated molecule with sugar pucker C2′ *endo*. $E = -16.7$ kcal/mole. (d) Lowest energy C2′ *endo* conformation found in this work. $E = -30.6$ kcal/mole.

TABLE III
Selected Minimum Energy Conformations Calculated for an Isolated GpC Molecule

Starting conformation—dihedral angles, degrees									Final conformation—dihedral angles, degrees								Energy, E, kcal/mole	Final order
χ'	ψ'	ϕ'	ω'	ω	ϕ	ψ	χ	Sugar pucker	χ'	ψ'	ϕ'	ω'	ω	ϕ	ψ	χ		
15	50	225	180	290	180	60	15	C3′ *endo*	2	62	205	296	279	183	56	25	−39.0	1
15	50	225	290	290	180	60	15		5	61	181	183	79	194	60	44	−38.1	2
15	50	225	60	60	180	60	15		−7	62	240	315	106	205	76	39	−37.3	3
15	50	225	290	60	180	60	15		4	61	180	23	207	161	311	32	−33.0	4
15	50	225	290	290	180	270	15		−3	61	190	129	238	100	350	39	−31.1	5
15	50	225	60	290	180	270	15		−2	62	183	112	91	263	141	44	−26.6	7
15	50	225	180	60	180	180	15		−2	62	191	344	247	178	160	29	−26.0	10
15	50	225	290	290	180	180	15		−2	61	201	46	57	210	302	39	−20.5	14
15	50	225	60	60	180	270	15		−2	62	189	192	295	171	303	38	−18.7	17
15	50	225	180	290	180	270	15											
55	50	225	290	60	180	180	55	C2′ *endo*	−15	294	299	332	281	178	171	153	−16.7	1
55	50	225	290	290	180	60	55		−23	20	214	296	264	253	45	155	+35.9	2
55	50	225	180	290	180	60	55		−20	50	293	185	160	273	−10	97	+66.7	18

Of the eighteen local minima found for C2′ *endo* sugars, ten were extended. The lowest energy form was extended with $E = -16.7$ kcal/mole, which was 2.0 kcal/mole above the highest energy C3′ *endo* found. Other C2′ *endo* minimum energy conformations were of much higher energy, all being positive (see Table III). When orientation parameters were included in the minimization (see below), a number of conformations whose energies lie between -4.2 and -30.6 kcal/mole occurred. In no case, however, was a C2′ *endo* conformation found whose energy was as low as -39.0 kcal/mole. Some of the local minima listed in Table III have unusual values of some torsion angles. As has been pointed out by the referee, this may well reflect the sensitivity of semi-empirical potential energy calculations to minor variations in the choice of parameters, especially van der Waals' radii. Figure 3 shows the two lowest C3′ *endo* as well as the lowest C2′ *endo* conformations.

Packing Schemes

Since the lowest energy C2′ *endo* form was extended, it was decided to try packing it in the ribbon-like rather than the dimeric scheme at first. The dihedral angles at the local minimum for the isolated molecule were used as starting parameters. From the sixteen orientation trials, seven packed conformations were found in which the function H was less than 5, indicating that the base pairing alignment was reasonably good. Each of these cases also had negative energies. However, four of these proved to be tangled. The remaining conformations when compared to the X-ray data had discrepancy indexes indicating no better than random fit. The other packed conformations either had energies that were positive or H functions that were greater than 5, and were not explored further. No duplicate minima were obtained for the sixteen trials. Table IV gives dihedral angles and orientations for all calculated packing schemes that had negative energies, H functions of less than 5, and no tangling. Figures 4–8 show some of these.

The second lowest C3′ *endo* conformation for the isolated molecule was also extended, so the ribbon-like scheme was used in packing trials, with the dihedral angles at the local minimum as initial parameters. In this case ψ' was set at 61°, a value from which it had been found to deviate by less than a degree in all seventeen local minima for the isolated molecule. Thus only nine parameters were varied. Each of the sixteen trials resulted in a different local packing minimum. Only two of these had both negative energies and an H function of less than 5 and one of them proved to be tangled. The other, listed in Table IV and shown in Figure 7 had an energy of -15.7 kcal/mole and H of 4.0, indicating a less satisfactory energy and alignment than for other packed structures tabulated.

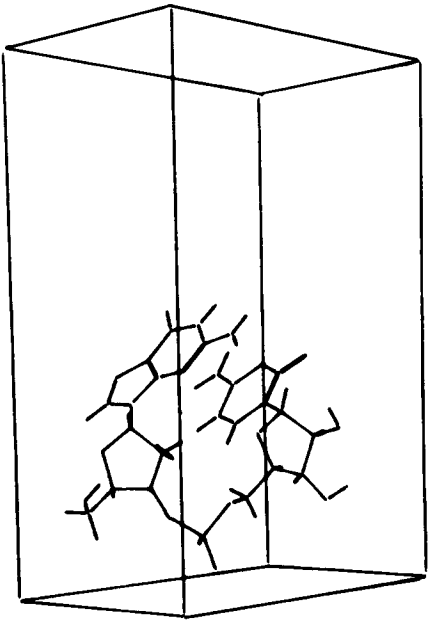
The lowest energy C3′ *endo* conformation for the isolated molecule was bent, and we therefore attempted to pack it as a dimer, using the dihedral angles that produced this local minimum as initial parameters. The sixteen trial conformations resulted in ten minima with an energy near

TABLE IV
Results of Flexible Packing Calculations

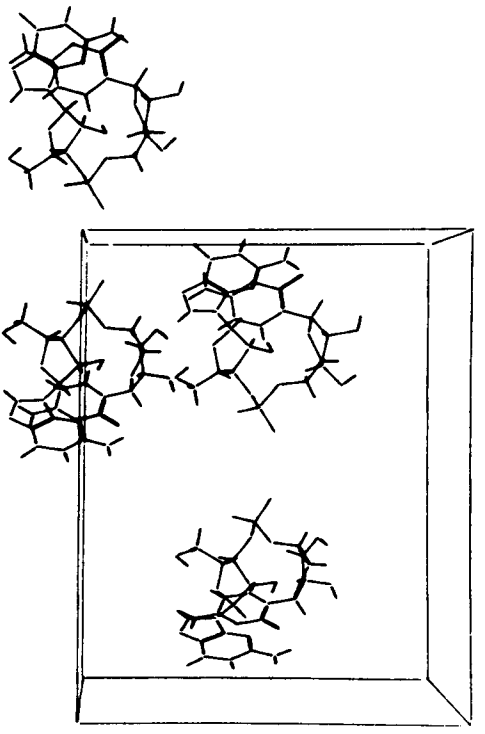
Starting conformation												Final conformation												
Dihedral angles, degrees								Orientation parameters,[a] degrees		Sugar pucker	Bonding scheme	Dihedral angles, degrees								Orientation parameters,[a] degrees		Energy, E,[b] kcal/mole	H	R, %
χ'	ψ'	ϕ'	ω'	ω	ϕ	ψ	χ	θ	Ξ			χ'	ψ'	ϕ'	ω'	ω	ϕ	ψ	χ	θ	Ξ			
2	62	205	296	279	183	56	25	90	0	C3′-*endo*	1–6	−1	62	200	296	271	194	57	21	223	280	−38.0	0.3	39.3
								180	270							same								
								270	0							same								
								270	270							same								
								180	0							same								
								0	90			−1	62	203	294	277	192	52	21	8	100	−38.0	0.4	59.0
								90	90							same								
								180	180							same								
								270	90							same								
								270	180							same								
								0	180			−13	62	245	263	295	152	53	13	9	101	−35.6	0.4	60.1
55	50	225	290	290	180	60	55	0	270	C2′-*endo*	1–6	−4	295	284	208	202	177	−11	−5	12	93	−30.6	22.5	74.3
								90	0			147	295	271	68	247	230	290	−5	17	100	−23.3	0.2	63.7
								270	270			−22	293	232	344	303	248	−9	99	211	281	−4.2	1.7	66.6
5	61	181	183	79	194	60	44	180	0	C3′-*endo*	1–3–5	269	61	155	114	119	69	25	97	163	246	−15.7	4.0	62.7
−15	294	299	332	282	178	171	153	90	90	C2′-*endo*	1–3–5	−12	295	210	217	200	224	42	136	175	249	−22.2	0.4	57.9
								90	180			−5	294	197	250	322	180	289	154	163	240	−20.7	1.7	79.2
								180	270			−19	297	220	177	29	100	286	162	161	246	−18.5	0.8	75.8

[a] Defined as follows: The molecule is first translated so that C4 of the cytosine coincides with the origin and C5 lies along the y axis. Then θ is the angle of counterclockwise rigid rotation of the molecule about a vector from C4 to C5, measured from the positive x axis. Ξ corresponds to a subsequent rotation about a vector perpendicular to the cytosine ring, passing through C4.

[b] E is the potential energy of a single molecule having the tabulated conformational angles without consideration of intermolecular forces.

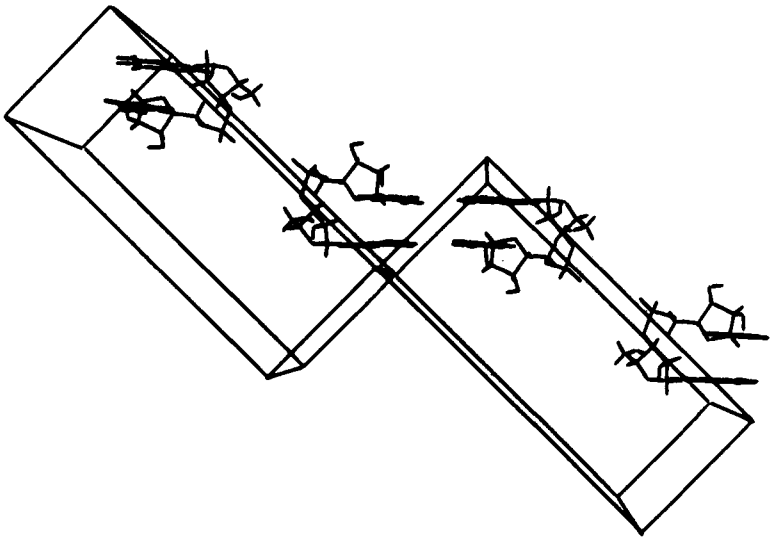


(*a*)

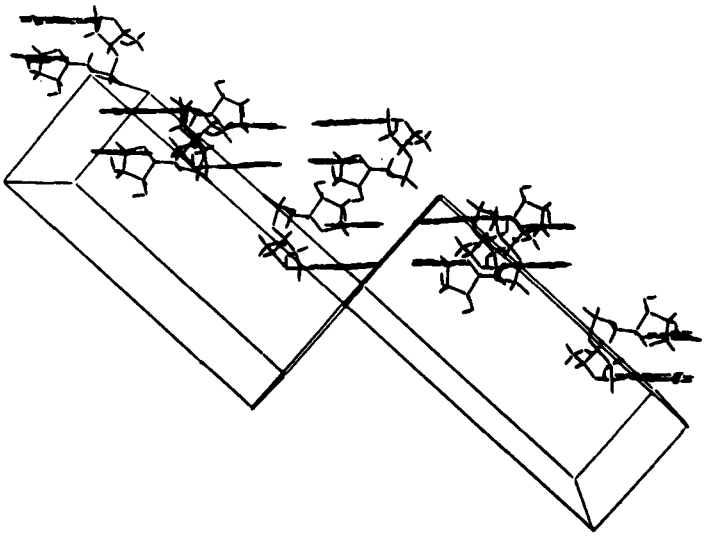


(*b*)

Fig. 4 (*continued*)



(c)



(d)

Fig. 4. Lowest energy packing scheme calculated for GpC, sugar pucker C3′ *endo*, denoted form A. This trial structure was successfully refined against our X-ray data. $E = -38.0$ kcal/mole. (a) Molecule 1 (see Table II) in unit cell. (b) Entire unit cell. (c) Two contiguous unit cells viewed along crystallographic b axis. For clarity, only two of the four molecules in the unit cell are shown. (d) Same view as (c) with all four molecules in the unit cell shown.

−38.0 kcal/mole, a value of $H = 0.3$, and similar sets of dihedral angles, again in the helical RNA range. These ten similar minima occurred in two distinct groups of five each, differing in their orientation parameters; that is, two types of packing are possible for the same conformation. Figures 4 and 5 show how these molecules are packed. The discrepancy

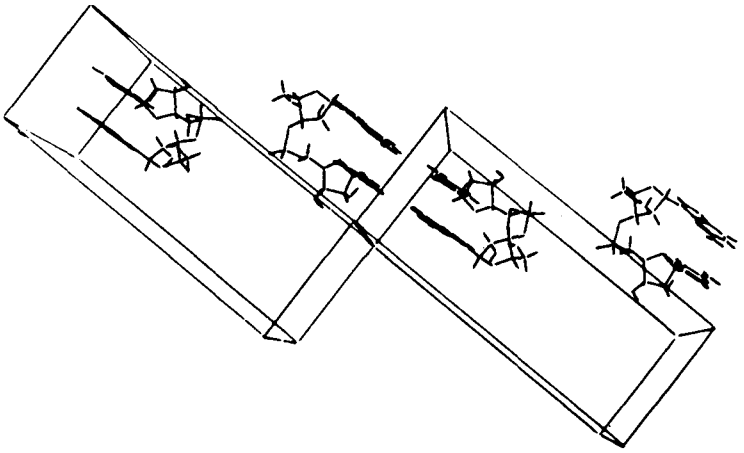


Fig. 5. Lowest energy packing scheme calculated for GpC, sugar pucker C3′ *endo*, denoted form B. This trial structure could not be refined. $E = -38.0$ kcal/mole. Same view as Fig. 4c.

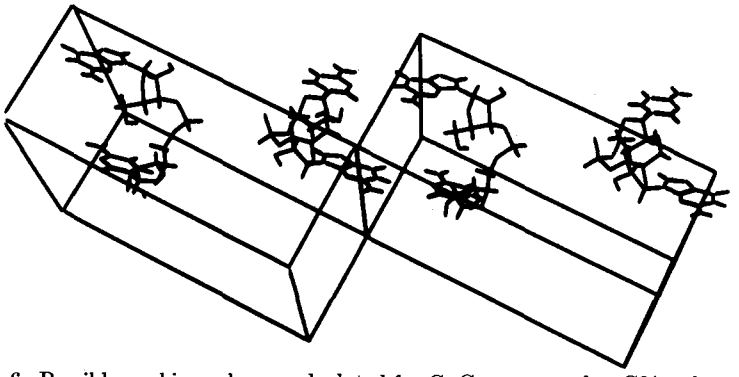


Fig. 6. Possible packing scheme calculated for GpC, sugar pucker C2′ *endo*. $E = -23.3$ kcal/mole. View is similar to Fig. 4c.

indexes calculated for the strong reflections within a 3-Å shell for the two schemes were 39.3% for orientation parameters (defined in Table IV) $\theta = 223°$ and $\Xi = 280°$ (form A), and 59.0% for parameters $\theta = 8°$ and $\Xi = 100°$ (form B). Note that these two packing schemes are "flipped over" versions of each other. That is, form A can be converted to form B by rotations of 215° and 180° in θ and Ξ, respectively. Immediate attempts to refine these structures showed that form A could be refined successfully against our X-ray data while form B could not. Of the six remaining packed conformations, one other had dihedral angles like the helical RNA's, with an energy of −35.6 kcal/mole, a value of 0.4 for H, and orientation parameters close to form B. Of the other conformations, all but one had a positive energy; however, all were poorly aligned with respect to base pairing.

For completeness another sixteen-trial run was made, to see if a dimeric, helical RNA type structure with sugars C2′ *endo* could be packed in our

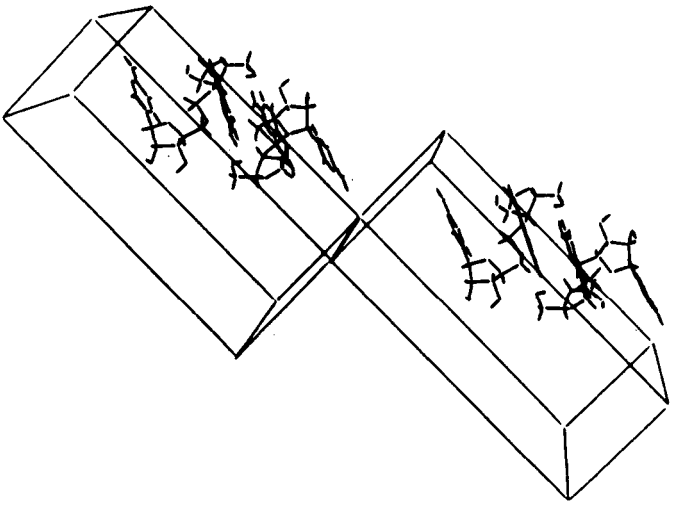


Fig. 7. Possible packing scheme calculated for GpC, sugar pucker C3′ *endo*. E = −15.7 kcal/mole. View similar to Fig. 4c.

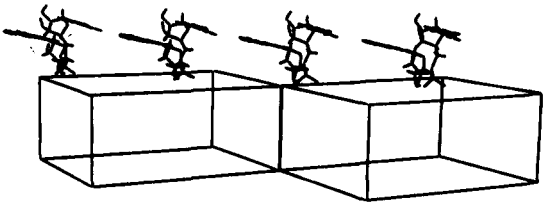


Fig. 8. Possible packing scheme calculated for GpC, sugar pucker C2′ *endo*. E = −20.7 kcal/mole. View similar to Fig. 4c.

cell. All ten parameters were allowed to vary, but molecule 6 instead of molecules 3 and 5 was used as the hydrogen-bonding partner. Two packed conformations were found with both negative energies and a good base pairing alignment. On visual examination with the graphics both proved to be packed nicely but gave discrepancy indexes no better than that given by a random distribution of atoms. A conformation whose energy was −30.6 kcal/mole (see Table IV and Fig. 3d), the lowest obtained for a C2′ *endo* conformation, was poorly aligned, with H = 22.5. The minimization algorithm did not find these low-energy conformations for the isolated molecule, although these energy regions became accessible when our packing constraints were applied. This is probably due to our not including the *syn* region ($\chi \simeq 180°$) as starting conformations for the cytosine base. This range is observed more frequently for C2′ *endo* sugars than for C3′ *endo* in nucleosides, although it has not been found at all for 5′-nucleotides.[2]

Using the graphics, an attempt was made to pack an earlier rigid body model of a dinucleotide having the elevenfold RNA dihedral angles[1] into our cell. This yielded an orientation close to that of the energy mini-

mization result but would not refine against the X-ray data. It was found that only the packed minimum energy result would refine and that differences from this conformation of as little as 15° in the internal dihedral angles were enough to inhibit X-ray refinement.

Refinement of Calculated Trial Structure

Isotropic refinement of form A in the pseudocell yielded a structure (R = 14.4%) conformationally similar to that of Day et al.[30] Further refinement of the structure in the real ($P2_1$) space group,[5] using form A as a starting point, is near completion and shows one Ca^{++} for every two GpC's in the crystal, with no Cl^- ions detected (our GpC was crystallized from $CaCl_2$ solution at basic pH, where the phosphate is singly ionized). This Ca^{++} is found to be at a position not on the pseudo-2-fold axis between the phosphate groups, thus breaking the C2 symmetry and causing our crystal to be $P2_1$. Unit cell dimensions and physical parameters for the molecule refined in space group C2 are given in Table V, and are compared with those of Day et al.,[30] whose GpC was crystallized as the sodium salt and was rigorously C2. Also shown are values for recent models of eleven- and twelvefold RNA.[31]

TABLE Va
Conformational Angles of GpC and Helical RNA's

		Dihedral angles, degrees							
	Ref.	χ'	ϕ'	ω'	ω	ϕ	ψ	χ	R, %
GpC	this work	17	230	287	291	170	57	37	14.4
GpC	30	13	209	291	284	186	51	25	11.1
RNA-11	31	a	209	286	298	180	48	12	38
RNA-12	31	a	193	300	295	193	44	17	27

[a] Same as χ.

TABLE Vb
Experimental Unit Cell Data for GpC

Monoclinic space group	This work (real cell), $P2_1$	This work (pseudo cell), C2	Day et al.,[30] C2
a	21.224 Å	21.224 Å	21.460 Å
b	34.207 Å	17.104 Å	16.927 Å
c	9.372 Å	9.372 Å	9.332 Å
β	90.527°	90.527°	90.54°
Z	8	4	4

DISCUSSION

The approach employed here to calculate the conformation and packing scheme for GpC has led to a trial structure that satisfactorily explains our X-ray crystallographic data. Two distinct computational steps were involved. First, the lowest energy conformation was computed for the

isolated molecule. This then served as a starting conformation in the subsequent packing minimization. In the latter case, both the energy and a geometric function which constrained the molecule to Watson-Crick base pairing were minimized simultaneously, with eight dihedral angles and two orientation parameters used as variables. The search over orientation space was limited by constraints from the X-ray data, as previously noted.

The lowest energy conformation for the isolated molecule, calculated without reference to the experimental X-ray data, is quite close to the conformation observed in crystalline GpC. While many other local minima were also obtained, they were all of higher energy, and proved not to be consistent with the X-ray data when attempts were made to pack those that appeared promising. The inferences derived from the X-ray data, as well as a good guess on the type of base pairing from chemical knowledge, made it possible to pack the molecule without calculating either an intermolecular lattice potential or a hydrogen bond potential. Instead, the molecular orientation was "pinned down" within these restrictions. Our method differs in this respect with the work of others[13-18] who used intermolecular potential functions in their calculations of packing schemes. Further, in packing the molecule, the dihedral angles as well as the orientation parameters were varied simultaneously. Thus, here the molecule was not packed as a rigid body, but was *flexible*, with the bias that the initial values for the dihedral angles were the lowest energy values from the isolated molecule calculation. It was not necessary to vary the bond lengths and bond angles. It is remarkable that the orientation search did not lead to many false local minima. Rather, the lowest energy packed conformation resulted from ten out of sixteen trials in the search over orientation space; five of the ten were identically oriented and consistent with the X-ray data (form A) while the other five packed in an alternate fashion which was not consistent with our data (form B). It should also be noted that the energy of the GpC molecule changed by just 1.0 kcal/mole when the *in vacuo* conformation was packed into the unit cell.

The computer graphics was in constant use in our search for the solution. To visualize the conformations of the various minima from the computed dihedral angles would have otherwise necessitated the difficult task of building a large number of actual models by hand. The graphics also proved to be a feasible method for visually determining a packing scheme, although in this case the molecule had to be treated as a rigid body. Furthermore, the calculated packing schemes could readily be examined to determine if the base pairing lineup was satisfactory or if intermolecular tangling had occurred. The packed minimum energy conformations could also be examined visually to ascertain if planes of densest packing did indeed correspond to strong diffraction maxima. In particular, an orientation in which the bases were parallel to the 4, 0, 2 plane was immediately believed to be the right trial conformation, and this turned out to be the case.

Thus, the minimization method led to the correct packed structure by making use of both chemical and experimental information, where either alone would not have been sufficient.

The authors thank Mr. Edwin Heilweil for the photographic work. The molecular modeling was performed at the Princeton University Computer Graphics Laboratory, supported by NIH Grant RR-00578; other support was received from NIH GM-16539 and NSF GB-28021.

References

1. M. Sundaralingam, *Biopolymers*, **7**, 821 (1969).
2. N. Yathindra and M. Sundaralingam, *Biopolymers*, **12**, 297 (1973).
3. S. Arnott, *Prog. Biophys. Mol. Biol.*, **21**, 265 (1970).
4. S. Stellman, B. Hingerty, S. Broyde, E. Subramanian, T. Sato, and R. Langridge, *Macromol.*, **6**, 652 (1973).
5. B. Hingerty et al., to be published.
6. W. K. Olson and P. J. Flory, *Biopolymers*, **11**, 25 (1972).
7. V. Renugopalakrishnan, A. V. Lakshminarayan, and V. Sasisekheran, *Biopolymers*, **10**, 1159 (1967).
8. A. V. Lakshminarayan and V. Sasisekheran, *Biopolymers*, **8**, 475(1969); *Ibid.*, **8**, 489(1969).
9. N. Go and H. A. Scheraga, *Macromol.*, **3**, 188 (1970).
10. N. Go, P. N. Lewis, and H. A. Scheraga, *Macromol.*, **3**, 628 (1970).
11. G. Niu, N. Go, and H. A. Scheraga, *Macromol.*, **6**, 91 (1973).
12. F. Bovey, A. Brewster, D. Patel, A. Tonelli, and D. Torchia, *Acct. Chem. Res.*, **5**, 192 (1972).
13. P. Zugenmaier and A. Sarko, *Acta Cryst.*, **B28**, 3158 (1972).
14. D. Williams, *Acta Cryst.*, **A25**, 464 (1969).
15. P. Zugenmaier and A. Sarko, *Biopolymers*, **12**, 435 (1973).
16. N. A. Ahmed, A. I. Kitaigorodsky, and K. V. Mirskaya, *Acta Cryst.*, **B27**, 867 (1971).
17. V. M. Coiro, P. Giacomello, and E. Giglio, *Acta Cryst.*, **B27**, 2112 (1971).
18. J. Stellman, A. E. Woodward, and S. Stellman, *Macromol.*, **6**, 330 (1973).
19. H. A. Scheraga, *Chem. Rev.*, **71**, 195 (1971).
20. D. Brant, *Ann. Rev. Biophys. Bioeng.*, **1**, 369 (1972).
21. R. Langridge et al., to be published.
22. R. Scott and H. A. Scheraga, *J. Chem. Phys.*, **44**, 3054 (1966).
23. J. Sussman, N. Seeman, S.-H. Kim, and H. Berman, *J. Mol. Biol.*, **66**, 403 (1972).
24. R. Scott and H. A. Scheraga, *J. Chem. Phys.*, **45**, 2091 (1966).
25. M. Powell, *Computer J.*, **7**, 155 (1964).
26. D. Voet and A. Rich, *Prog. Nucleic Acid Res. Mol. Biol.*, **10**, 183 (1970).
27. J. Rubin, T. Brennan, and M. Sundaralingam, *Biochem.*, **11**, 3112 (1972).
28. R. McGuire, F. Momany, and H. A. Scheraga, *J. Phys. Chem.*, **76**, 375 (1972).
29. C. Barry, J. Glasser, A. North, R. Williams, and A. Xavier, *Biochem. Biophys. Acta*, **262**, 101 (1972).
30. R. Day, N. Seeman, J. Rosenberg, and A. Rich, *Proc. Natl. Acad. Sci. (U.S.)*, **70**, 849 (1973). The phosphorus position of our crystal was confirmed with Dr. N. Seeman of this group.
31. S. Arnott, D. Hukins, and S. Dover, *Biochem. Biophys. Res. Comm.*, **48**, 1392 (1972).

Received May 18, 1973
Revised August 27, 1973