

## Crystal Structures of Two Complexes Containing Guanine and Cytosine Derivatives

By E. J. O'BRIEN\*

Laboratory of Molecular Biology, Hills Road, Cambridge, England  
and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, U.S.A.

(Received 16 November 1966)

The crystal structures of two complexes: (a) a 1:1 complex of 9-ethylguanine with 1-methylcytosine and (b) a 1:1 complex of 9-ethylguanine with 1-methyl-5-fluorocytosine are described. The crystals are isomorphous, space group  $P\bar{1}$ , and the second structure was solved from a knowledge of the first. Data for both structures were collected with a Weissenberg camera, and the intensities estimated visually. For the first structure, an analytical method was used to determine the plane of maximum density in the three-dimensional Patterson function, and the trial structure was found from inspection of this plane. The structure was refined in projection from inspection of successive difference-Fourier syntheses, and in three dimensions by a full-matrix least-squares method. Hydrogen atoms, except those in the alkyl groups, were located from a three-dimensional difference-Fourier synthesis. The final  $R$  value was 11.20%, unobserved intensities included, and the estimated standard deviation of bond lengths was 0.007 Å. The second structure was refined in three dimensions by the same least-squares method to an  $R$  value of 12.9% and an e.s.d. of 0.009 Å. Comparison of bond lengths and angles shows good agreement in the two structures. In both structures the asymmetric unit comprises the purine linked to the pyrimidine by three hydrogen bonds. This arrangement of hydrogen bonds is the same as that existing in the guanine-cytosine base-pair in the refined Watson-Crick model of deoxyribonucleic acid.

### Introduction

An important feature of the Watson-Crick structure of deoxyribonucleic acid (DNA) is the nature of the purine-pyrimidine base-pairing by hydrogen bonds. X-ray diffraction studies of DNA itself do not reveal the precise configuration of the base-pairs and it is necessary therefore to set up a model system. Purine-pyrimidine crystalline complexes provide such a system.

This paper gives a description of the X-ray diffraction study of the two complexes: (a) a 1:1 complex of 9-ethylguanine with 1-methylcytosine (GC), and (b) a 1:1 complex of 9-ethylguanine with 1-methyl-5-fluorocytosine (GFC). Preliminary accounts of these studies have been published (O'Brien, 1963, 1966). Other crystal structures of the guanine-cytosine type have been studied by Sobell, Tomita & Rich (1963) and Haschemeyer & Sobell (1964, 1965*b*). Crystal structures of the adenine-thymine type have been studied by Hoogsteen (1959, 1963), Mathews & Rich (1964*b*), Haschemeyer & Sobell (1963, 1965*a*) and Katz, Tomita & Rich (1965). The crystal structure of a complex of 9-ethyl-2-aminopurine with 1-methyl-5-fluorouracil has been reported by Sobell (1966). Arnott, Wilkins, Hamilton & Langridge (1965) have determined the overall size and shape of the base-pairs in DNA by Fourier synthesis studies with X-ray diffraction data from crystalline fibres of lithium DNA.

The investigation of the structure of GFC was carried out after that of GC and was undertaken to determine the possible effect of fluorine on the hydrogen-bonding properties of 1-methylcytosine. The two structures were found to be very similar and the fluorine atom was observed to have little effect on the geometry of the hydrogen bonding in the crystal (O'Brien, 1966).

### Unit-cell parameters

The unit-cell parameters (Table 1) were determined from measurements made on precession photographs taken with Cu  $K\alpha$  radiation ( $\lambda = 1.5418$  Å). The camera radius was calibrated from precession photographs taken with a single crystal of lead nitrate and Mo  $K\alpha$  radiation ( $\lambda = 0.7107$  Å). Measurements were made on photographs taken  $180^\circ$  apart around the spindle axis in order to eliminate crystal centring errors. Corrections for film shrinkage were applied.

The structure of 1-methylcytosine has been reported by Mathews & Rich (1964*a*). 9-Ethylguanine crystallizes in space group  $P4_12_12$  with unit cell:  $a = 10.92 \pm 0.03$ ,  $c = 29.34 \pm 0.10$  Å.

Table 1. Unit-cell parameters

GC	GFC
$a = 8.838 \pm 0.008$ Å	$a = 8.745 \pm 0.008$ Å
$b = 11.106 \pm 0.010$	$b = 11.227 \pm 0.010$
$c = 7.391 \pm 0.006$	$c = 7.513 \pm 0.009$
$\alpha = 107^\circ 49' \pm 5'$	$\alpha = 109^\circ 1' \pm 5'$
$\beta = 87^\circ 3' \pm 5'$	$\beta = 84^\circ 58' \pm 5'$
$\gamma = 91^\circ 27' \pm 5'$	$\gamma = 90^\circ 59' \pm 5'$
$V = 689.7 \pm 1.0$ Å <sup>3</sup>	$V = 694.7 \pm 1.0$ Å <sup>3</sup>
$D_m = 1.473 \pm 0.003$ g.cm <sup>-3</sup>	

\* Present address: Medical Research Council Biophysics Research Unit, King's College, 26-29 Drury Lane, London W. C. 2, England.

### Crystal preparation

The problem of preparing a mixed guanine-cytosine crystal was largely one of finding a suitable solvent since guanine and its derivatives are not very soluble in water. Waring & Katritzky (personal communication) found dimethyl sulphoxide to be satisfactory and prepared GC crystals by slowly evaporating an equimolar mixture of G (9-ethylguanine) and C (1-methylcytosine). It proved difficult, however, to find a crystal suitable for accurate data collection. With all the larger crystals, the reflections were smeared out or took the form of multiple spots. Several attempts were made to grow better crystals. The method most often adopted was to evaporate in a test tube at about 80°C a dimethyl sulphoxide solution containing a 1:1 mixture

of the constituents. The solution usually took about two to three weeks to evaporate to dryness, although crystals could be extracted from the sides of the tube before this time. In one batch there were a number of small crystals that gave reasonable X-ray photographs. Two of these were selected; their dimensions were 0.3 × 0.1 × 0.06 mm and 0.5 × 0.1 × 0.1 mm. The former was used for data taken around the *c* axis, the latter for data around the [1 $\bar{1}$ 0] axis. The crystals were elongated along [1 $\bar{1}$ 0].

Crystals of GFC were prepared in the same manner as were those of GC. The crystal formed with a different habit, however, with more numerous faces and elongated along the [100] axis. In addition, it proved easier to prepare good crystals than had been the case with GC. A crystal of dimensions 0.7 × 0.4 × 0.4 mm was selected.

### The GC structure solution

#### *Trial and error methods*

The predominant feature of the diffraction pattern was one very strong reflection, indexed 11 $\bar{2}$ , at a spacing of about 3.3 Å. The higher orders of this reflection were also strong, so that it was clear that the molecules must lie in one plane. The (11 $\bar{2}$ ) plane was drawn out and attempts were made to find reasonable positions for models of G and C. The presence of a centre of symmetry was at first assumed; it was later substantiated by statistical tests (Howells, Phillips & Rogers, 1950) made on two zones of reflections. The tests revealed not only that the crystal was centrosymmetric but also that the asymmetric unit was to a certain extent centrosymmetric (Lipson & Woolfson, 1952; Rogers & Wilson, 1953). Assuming that G would be linked to C by at least two hydrogen bonds, then the number of possible models for the asymmetric unit was limited to three; these are illustrated in Fig. 1. Model (a) is the one used by Watson and Crick in their model of DNA [with the option of there being either two or three hydrogen bonds (Crick & Watson, 1954)]. Model (b) was proposed by Donohue (1956) as an alternative base-pairing that could be accommodated in the DNA structure. There are no N-H...O bonds in model (c) and this makes it appear less likely, but there is the possibility of such bonds being formed with neighbouring molecules. All three models could be arranged in the (11 $\bar{2}$ ) plane in such a way that further hydrogen bonds were formed between adjacent asymmetric units. An attempt was made to determine the correct model by means of structure factor graphs for some of the strongest *hk*0 reflections. This proved unsuccessful and attempts at structure solution were discontinued until three-dimensional data had been collected.

#### *Data collection*

For the collection of data a Weissenberg camera was used with nickel-filtered copper radiation. The camera was of the double-crystal type, limited to a maximum

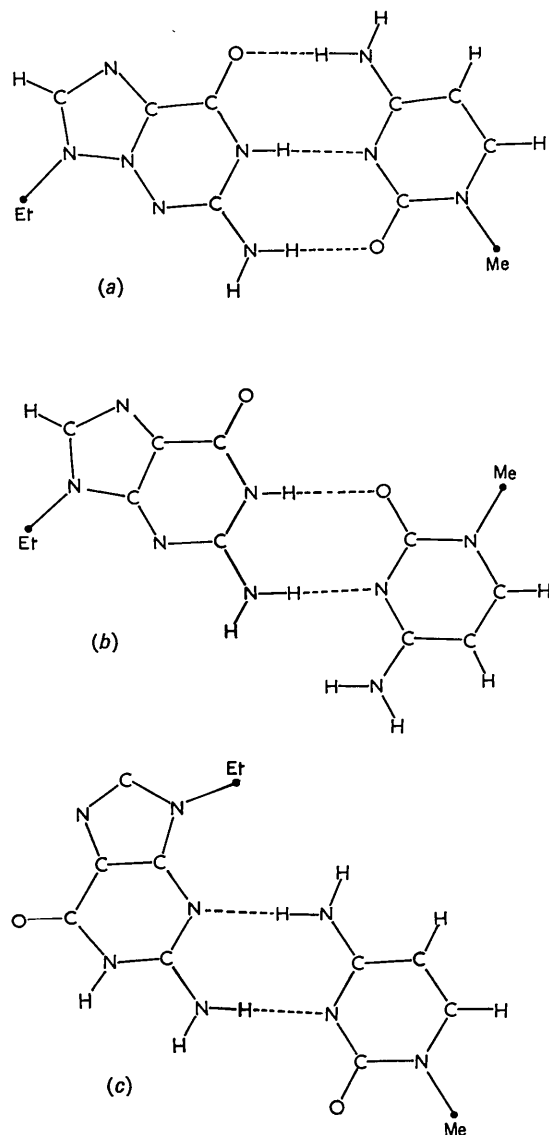


Fig. 1. Three possible models for the asymmetric unit.

inclination of  $20^\circ$  and a maximum scan of  $195^\circ$ . Levels  $hk0$  to  $hk3$ , and  $hhl$  to  $h, \bar{h}+6, l$  were recorded. On each upper level two settings were used, covering  $360^\circ$  rotation of the crystal, in order to include the entire reciprocal lattice for that level. Intensities were measured by visual comparison of the spots with a calibrated intensity scale. Empirical corrections were made for  $\alpha_1\alpha_2$  splitting. The Phillips correction (Phillips, 1956) was used to compensate for spot shape extension on upper-level photographs; only extended spots were measured. Since the highest layer line recorded was at an equi-inclination angle of less than  $20^\circ$ , the spot shape correction was not severe for more than a relatively small number of reflections. A further correction was necessary, for the  $c$ -axis data, in order to compensate for the change of spot shape caused by the non-uniform shape of the crystal (the  $c$  axis was nearly perpendicular to the long direction of the crystal, and the crystal was too small to cut to shape). The use of a correction of this type has been discussed by Broomhead (1948). Lorentz-polarization corrections were applied. The  $c$ -axis and the  $[1\bar{1}0]$ -axis data were merged by a least-squares method (Rollett & Sparks, 1960). The data were then sharpened by the function:

$$M(S) = \left(\frac{1}{f}\right)^2 \exp\left[\frac{-\pi^2}{7.25} S^2\right] \quad (\text{Lipson \& Cochran, 1957})$$

where  $S = 2 \sin \theta / \lambda$  and  $f =$  average unitary scattering factor for the atoms in the structure. 1976 independent reflections were recorded out of a possible total of about 3000 for the copper sphere.

#### 'Best plane' of the Patterson function

The method of Tollin & Cochran (1964) was used to find the plane of maximum density in the three-dimensional Patterson function. This method integrates the Patterson function over a circular disc of chosen radius, centred at the origin. The disc assumes different orientations in a stepwise manner and the value of the integral is calculated for each orientation. Angular coordinates,  $\theta$  and  $\varphi$ , of the normal to the disc form a grid in a Sanson-Flamsteed equi-area projection of a spherical quadrant, and the values for the integral are plotted on this grid (Fig. 2).

The program was set up for  $P2$  symmetry, but could be used with little modification because two of the GC cell angles are close to  $90^\circ$  ( $\beta^* = 92.6$ ,  $\gamma^* = 89.4$ ). The results showed one large peak with low background; this gave the best plane as within one degree of the  $(11\bar{2})$  plane, neglecting the errors due to the non-orthogonality of the axes. The program was run first for a disc of radius  $5 \text{ \AA}$ , which is about the size of a single base, and then for a disc of radius  $10 \text{ \AA}$ , which is about the size of a base-pair. For the  $10 \text{ \AA}$  rotation the peak was more elongated; this suggested that the molecules in the base-pair were inclined to each other at a small angle. The results of the  $10 \text{ \AA}$  rotation, which was only carried out over a limited range of  $\theta$ , are shown in Fig. 2.

#### Patterson solution

The data were reindexed as follows:

$$H = -h - k - l, \quad K = 2h + l, \quad L = l.$$

This changes  $11\bar{2}$  into  $00\bar{2}$  while preserving a right-handed set of axes; the cell now becomes  $A$ -centred. The  $a$  axis and  $b$  axis in the new cell are the vectors  $[\bar{1}\bar{1}\bar{1}]$  and  $[201]$  in the old cell. Sections in the three-dimensional, sharpened, Patterson synthesis were calculated parallel to the  $(001)$  plane of the new cell.

The basal plane of the Patterson function (Fig. 3) shows hexagonal arrangements of peaks around the origin, the first set of six corresponding to nearest-neighbour vectors ( $1.4 \text{ \AA}$ ), and the second corresponding to next-to-nearest neighbour vectors ( $2.4 \text{ \AA}$ ). Two of the peaks of the second set are considerably higher than the others. The peaks are higher on account of parallel contributions from the oxygen atoms and amino groups on the six-membered rings of both guanine and cytosine. Consequently, two possible orientations for the six-membered rings can immediately be determined. There are also two strong  $2.8 \text{ \AA}$  peaks; these suggest the presence of parallel hydrogen bonds.

At a distance of  $7.5 \text{ \AA}$  from the origin is a very large peak (height  $0.4 \times$  height of origin peak), and around this is an arrangement of peaks similar to that around the origin. Hence there is a pseudo-repeat distance of  $7.5 \text{ \AA}$ , that is, adjacent base-pairs are nearly parallel to each other. With the added requirement of the presence of a centre of symmetry between adjacent base-pairs, it is now possible to draw the structure shown in full lines in the lower half of Fig. 3. The position of the five-membered ring of guanine still remains to be determined, but the limitations imposed by the centres of symmetry lead to only one stereochemically acceptable position; this position is indicated in Fig. 3 by dashed lines. Now the possibility arises of additional hydrogen bonds being formed, between N(3) of guanine and N(2) of a neighbouring guanine and between N(7) of guanine and N(4) of a neighbouring cytosine (Fig. 5); the model therefore looks very satisfactory. It remains only to fix the second carbon atom

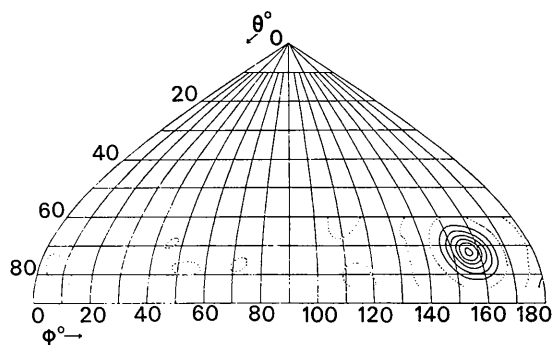


Fig. 2. Results for the 'best-plane' rotation program, plotted on a Sanson-Flamsteed equi-area projection. Contours are at equal, arbitrary intervals, with the zero contour dashed.

of the ethyl group of guanine, and this was revealed when the first difference Fourier synthesis was calculated.

In the trial structure the dimensions of the base-pair as given by Spencer (1959*a, b*) were used.

### Refinement

The structure was first refined in projection. The *c*-axis projection was selected for refinement, in order to give the best resolution of the molecule. The trial structure showed little overlap in this projection. A two-dimensional refinement program, written by Wells (1961) for the Edsac II computer, scales observed to calculated structure amplitudes and calculates Fourier and difference Fourier syntheses. The program was used to calculate successive difference-electron-density maps, and on each map refined coordinates were obtained by inspection. In this way, the *R* value for the

projection was reduced from 50% to 20% ( $R = \frac{\sum |F_o| - |F_c|}{\sum |F_o|}$ ).

A three-dimensional refinement program by Wells (1961) was then used. The program employed the method of differential synthesis (Booth, 1946). The *z* coordinates were calculated assuming that the molecules were exactly in the  $(11\bar{2})$  plane. The first three-dimensional *R* value was 39.9%; this was reduced to 19.3% after several cycles of refinement. Positional parameters and an overall scale factor, or positional parameters and an overall temperature factor were refined. During refinement conditions on certain classes of reflections were imposed: those reflections for which  $|F_o - F_c|$  was very large, or for which  $|F_o|$  was large but  $|F_c|$  small, were excluded while the new parameters were calculated.

At this stage of the refinement three-dimensional Fourier and difference Fourier syntheses were calculated with an Edsac II program written by Ward (1962). A diagram showing the  $(11\bar{2})$  sections of these syntheses was given in the preliminary publication (O'Brien, 1963). All the hydrogen atoms in or close to the  $(11\bar{2})$  plane were located from the difference synthesis. An attempt was made to locate the hydrogen atoms attached to the alkyl groups, but the positive regions where hydrogen atoms would be expected were in this case not resolved from each other or from the positive regions associated with the anisotropic thermal motion of the adjacent carbon atoms. The  $\text{CH}_3$  groups of both guanine and cytosine are free to rotate, and disorder of the hydrogen atoms in these groups would therefore be expected in the crystal. The  $\text{CH}_2$  group of guanine is not free to rotate, however, and the probable positions of the two hydrogen atoms in this group were determined by model building.

The refinement was continued with an IBM 7090 computer and the full-matrix least-squares program of Busing, Martin & Levy (1962). Atomic scattering factors were taken from *International Tables for X-ray Crystallography* (1962). The function minimized was  $\sum w(|F_o| - |F_c|)^2$ ; *w*, the weighting factor, was given by the expression:  $1/w = a + |F| + b|F|^2$  (Cruickshank, Pilling, Bujosa, Lovell & Truter, 1961) where *a* and *b* are constants:  $a \approx 2|F_{\min}|$ ,  $b \approx 1/|F_{\max}|$ . Unobserved reflections were given the value  $|F_{\min}|/2$  (Hamilton, 1955), with weight  $(2/F_{\min})^2$ . A single scale factor was included as an adjustable parameter. No correction was made for extinction.

Two refinement cycles, in which the temperature factors of all atoms excluding hydrogen atoms were varied first isotropically and then anisotropically, reduced *R* to 13.4%. A three-dimensional difference Fourier synthesis was again calculated, this time with the program of Sly, Shoemaker & Van den Hende (1962). The  $(11\bar{2})$  section is illustrated in Fig. 4. Negative regions, which were evident in the previous difference-electron-density map and were associated with the anisotropic motion of the atoms, have for the most part disappeared. In addition, the hydrogen atom peaks are higher, the peak

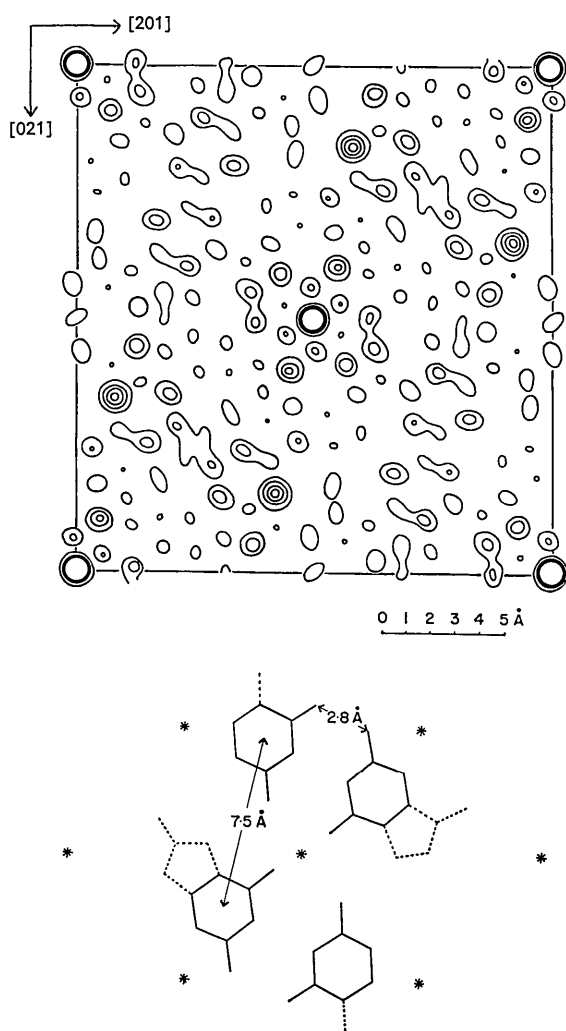


Fig. 3. The upper half of the figure shows the  $(11\bar{2})$  section of the Patterson synthesis. Contours are at equal, arbitrary intervals, commencing at the second contour. The lower half of the figure shows the trial structure.

height in each case being above  $0.6 \text{ e.}\text{\AA}^{-3}$ . There are other positive regions, which cannot be identified as hydrogen atoms, but in no case is the peak height of such a region as high as  $0.6 \text{ e.}\text{\AA}^{-3}$ .

The next cycle refined anisotropic temperature factors of all atoms again, but this time ten hydrogen atoms were included with fixed parameters. The hydrogen atoms were given positions obtained from theoretical predictions and not from the difference maps, but the discrepancy between the predicted and observed positions was in most cases not significant. Only the six hydrogen atoms of the methyl groups were still excluded from the refinement. In the next cycle the positional parameters of all the atoms, including the ten hydrogen atoms, were refined. The  $R$  value then stood at 11.3%. An attempt was made to refine the positional parameters and isotropic temperature factors of the hydrogen atoms while keeping the non-hydrogen atoms fixed. Errors in the hydrogen thermal parameters were large, so that most of the shifts which occurred after one cycle were not significant. For the two hydrogen atoms G-H(9A), and G-H(9B), however, the temperature factors increased to become several units greater than the temperature factor of the adjacent carbon atom, G-C(9) (see Fig. 5 for the numbering of the atoms in G and C). All of the other hydrogen temperature factors decreased, and for one atom, G-H(8), the change was about  $2\sigma$  and the temperature factor became negative. The negative temperature factor implied that the scattering factor being used for hydrogen was incorrect. Negative values for hydrogen temperature factors have been observed by

Hirschfeld, Sandler & Schmidt (1963). Jensen & Sundaralingam (1964) derive a linear relationship between the temperature factor,  $B'(H)$ , of a hydrogen atom and the temperature factor,  $B(\text{non-H})$ , of the atom to which the hydrogen is bonded. They show that if  $B(\text{non-H})$  is below 2.5, then  $B'(H)$  is likely to be negative. As  $B(\text{non-H})$  increases above 2.5 then  $B'(H)$  also increases. For large values of  $B(\text{non-H})$  (above 4)  $B'(H)$  is considerably greater than  $B(\text{non-H})$ . The negative value of  $B$  for G-H(8) (for G-C(8),  $B \approx 2.8$ ), and the large values of  $B$  for G-H(9A) and G-H(9B) (for G-C(9),  $B \approx 3.5$ ) are in agreement with these results. The modified hydrogen scattering factors of Stewart, Davidson & Simpson (1965) were not available at the time of this refinement.

A final refinement cycle varied positional and thermal parameters of most of the non-hydrogen atoms. 170 parameters were varied, nearly the maximum possible for the program. The shifts in all the positional parameters, and in nearly all of the thermal parameters, were less than the corresponding standard deviations. The refinement was therefore terminated. The  $R$  value was 11.20%, unobserved reflections included.

#### The GFC structure solution

The structure was first examined in projection. A GFC-GC difference Fourier synthesis, calculated with coefficients prepared from the  $hk0$  structure factors of GC and the  $hk0$  structure amplitudes of GFC, revealed a prominent peak in the expected position for the fluorine atom but was otherwise relatively featureless.

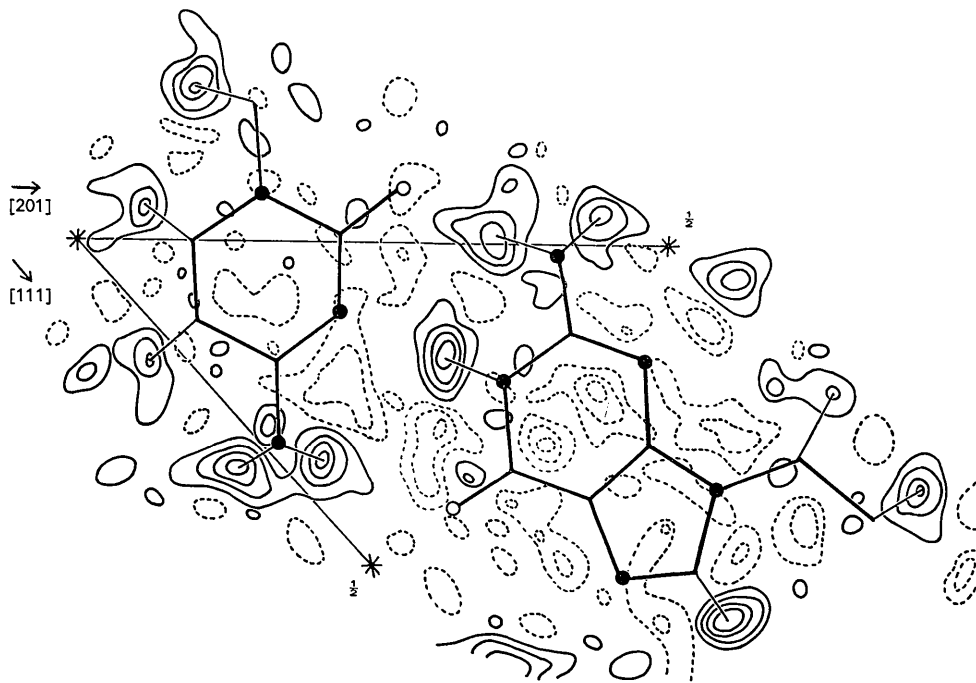


Fig. 4. The  $(11\bar{2})$  section of the difference Fourier synthesis. Contours are at intervals of  $0.2 \text{ e.}\text{\AA}^{-3}$ , with negative contours dashed and the zero contour omitted.

Table 2(a). Positional and thermal parameters of non-hydrogen atoms in GC

All values are multiplied by  $10^4$ . Average estimated standard errors are indicated at the head of each column. The temperature factor is defined by the expression:

$$T = \exp(-\beta_{11}h^2 - \beta_{22}k^2 - \beta_{33}l^2 - 2\beta_{12}hk - 2\beta_{13}hl - 2\beta_{23}kl).$$

	$x/a$ $\sigma=4$	$y/b$ $\sigma=4$	$z/c$ $\sigma=6$	$\beta_{11}$ $\sigma=5$	$\beta_{22}$ $\sigma=3$	$\beta_{33}$ $\sigma=10$	$\beta_{12}$ $\sigma=4$	$\beta_{13}$ $\sigma=6$	$\beta_{23}$ $\sigma=5$
1-Methylcytosine									
N(1)	3075	-718	1187	76	43	189	-4	-32	22
C(1)	3031	-2102	401	116	39	245	-12	-41	8
C(2)	4382	-135	2075	73	46	165	8	-36	26
O(2)	5428	-826	2167	97	42	295	11	-71	17
N(3)	4460	1138	2721	82	38	192	2	-46	16
C(4)	3284	1839	2564	86	45	160	11	-12	26
N(4)	3421	3099	3201	114	42	269	14	-53	24
C(5)	1913	1268	1779	88	62	251	12	-27	39
C(6)	1867	-2	1107	73	69	261	2	-56	30
9-Ethylguanine									
N(1)	7257	2165	4583	80	32	215	10	-42	20
N(2)	8084	126	4049	106	26	296	9	-85	7
C(2)	8350	1361	4828	69	36	182	5	-44	18
N(3)	9602	1767	5754	70	31	194	5	-34	15
C(4)	9695	3048	6385	75	33	158	0	-23	18
C(5)	8650	3925	6249	81	31	210	6	-42	17
C(6)	7325	3488	5250	82	33	201	9	-30	26
O(6)	6301	4111	4914	104	40	354	14	-83	38
N(7)	9172	5150	7118	91	32	239	-2	-46	13
C(8)	10478	4991	7745	86	29	228	1	-49	4
N(9)	10867	3743	7368	78	31	194	-1	-47	12
C(9)	12240	3212	7902	90	53	254	-2	-70	26
C(10)	13435	4195	8663	94	69	309	-17	-83	9

Table 2(b). Positional parameters of hydrogen atoms in GC

All values are multiplied by  $10^3$

1-Methylcytosine			9-Ethylguanine				
$x/a$ $\sigma=7$	$y/b$ $\sigma=6$	$z/c$ $\sigma=9$	$x/a$ $\sigma=7$	$y/b$ $\sigma=6$	$z/c$ $\sigma=9$		
H(4A)	278	358	309	H(1)	625	177	402
H(4B)	427	345	388	H(2A)	703	-23	344
H(5)	107	189	166	H(2B)	873	-41	413
H(6)	98	-55	72	H(9A)	1213	255	873
				H(9B)	1259	240	660
				H(8)	1122	569	853

This result indicated that in a three-dimensional analysis the known structure of GC could be used as a trial structure for GFC.

Levels  $0kl$  to  $8kl$  inclusive were recorded with an integrating Weissenberg camera and copper  $K\alpha$  radiation. The camera had a maximum rotation angle of about  $230^\circ$  so that most of the reciprocal lattice of an upper level could be recorded on one set of films. Intensities were measured by visual comparison with a calibrated intensity scale. Lorentz-polarization and absorption corrections were applied.

The integration of the reflections did not fully compensate the difference in spot shape on upper and lower halves of the films. For a reflection appearing on both halves of the film the intensity on the lower half was up to 50% larger (in a few cases, however, spots measured on the upper half were stronger). A correction was made for this effect by treating the upper and lower halves of the film as separate quantities; a scale factor was found relating the two halves from the values of

those intensities measured on both halves. This procedure was carried out for all the upper levels. The  $Hkl$  and  $hk0$  data were then scaled together.

The refined GC parameters were used as a starting point for the refinement of GFC. For the fluorine atom  $x$  and  $y$  coordinates were taken from the GFC-GC difference Fourier synthesis and the  $z$  coordinate was calculated assuming the atom to lie in the  $(11\bar{2})$  plane. The refinement program, weighting scheme and treatment of unobserved reflections were the same as used in the least-squares refinement of GC. Two cycles were run: the first in which the positional parameters of the non-hydrogen atoms varied, and the second in which the same positional parameters varied in conjunction with the nine scale factors (one for each  $Hkl$  layer). The  $R$  value fell from 34% to 19%. A cycle in which the anisotropic temperature factors varied reduced  $R$  to 15.6%. A cycle of varying positional parameters with scale factors improved this to 14.2%. The  $R$  value (including unobserved reflections) was brought to 12.9% by two cycles which repeated the sequence of parameters varied in the previous two cycles, with the addition, in the last cycle, that the positional parameters of the hydrogen atoms were also allowed to vary. In the last cycle nearly all the parameter shifts were less than the corresponding standard deviations, and the refinement was therefore discontinued.

## Results and discussion

Tables 2 and 3 give the final positional and thermal parameters for GC and GFC. At the head of each

Table 3(a). Positional and thermal parameters of non-hydrogen atoms in GFC

All values are multiplied by  $10^4$ . Average estimated standard errors are indicated at the head of each column. The temperature factor is defined by the expression:

$$T = \exp(-\beta_{11}h^2 - \beta_{22}k^2 - \beta_{33}l^2 - 2\beta_{12}hk - 2\beta_{13}hl - 2\beta_{23}kl).$$

	$x/a$ $\sigma=6$	$y/b$ $\sigma=5$	$z/c$ $\sigma=8$	$\beta_{11}$ $\sigma=8$	$\beta_{22}$ $\sigma=5$	$\beta_{33}$ $\sigma=13$	$\beta_{12}$ $\sigma=5$	$\beta_{13}$ $\sigma=8$	$\beta_{23}$ $\sigma=6$
1-Methyl-5-fluorocytosine									
N(1)	3106	-778	1212	64	61	231	-14	-53	26
C(1)	3113	-2168	462	108	54	277	-23	-60	19
C(2)	4406	-191	2072	64	54	202	-3	-47	36
O(2)	5477	-858	2186	77	63	320	7	-78	21
N(3)	4445	1090	2699	61	60	213	-11	-48	30
C(4)	3289	1797	2541	45	59	170	-4	-23	25
N(4)	3347	3043	3141	90	54	249	1	-56	19
F(5)	771	1870	1523	70	86	315	7	-76	45
C(5)	1945	1171	1692	53	75	198	6	-40	32
C(6)	1913	-87	1061	63	73	221	-12	-57	32
9-Ethylguanine									
N(1)	7256	2112	4493	63	46	213	6	-59	23
N(2)	8164	102	3949	90	47	324	1	-95	23
C(2)	8367	1339	4733	52	58	185	1	-35	20
N(3)	9608	1771	5654	58	47	206	4	-46	23
C(4)	9639	3053	6315	56	47	179	1	-31	20
C(5)	8590	3888	6153	58	48	218	-4	-39	27
C(6)	7266	3436	5164	52	51	208	10	-23	29
O(6)	6242	4035	4818	79	53	337	7	-77	44
N(7)	9046	5124	7013	79	51	229	0	-26	30
C(8)	10363	4987	7662	73	44	253	6	-48	26
N(9)	10801	3748	7288	55	52	226	-4	-47	23
C(9)	12143	3226	7831	72	78	256	-2	-71	34
C(10)	13341	4241	8581	78	82	330	-10	-72	38

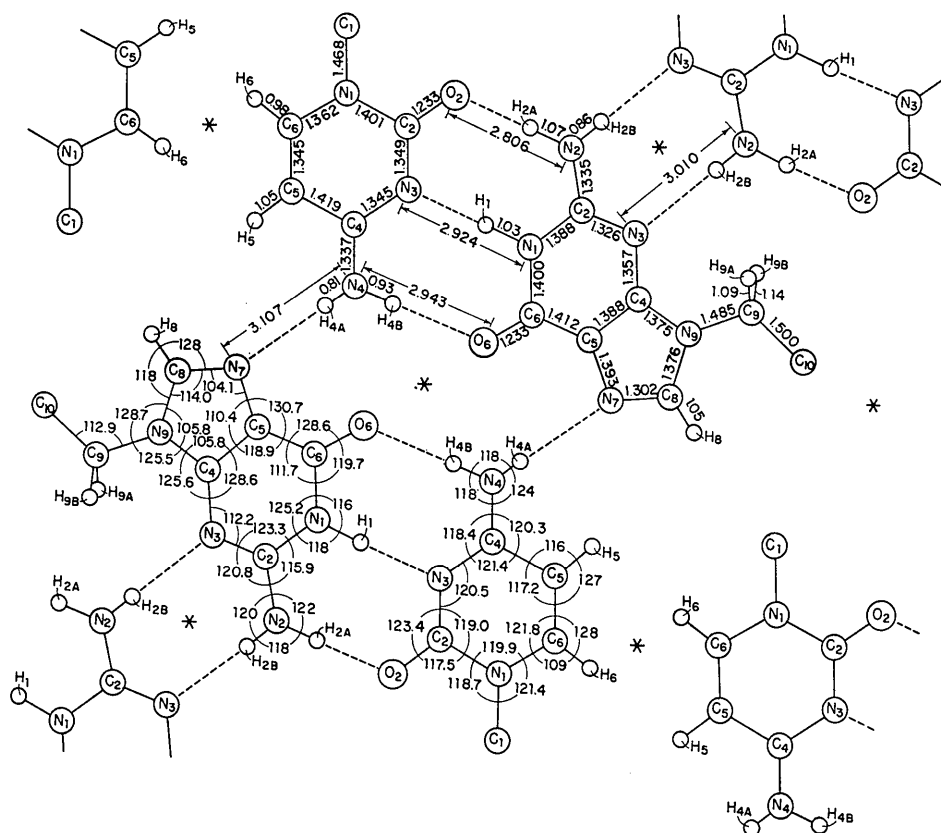


Fig. 5. Bond lengths and angles in GC.

Table 3(b). *Positional parameters of hydrogen atoms in GFC*

All values are multiplied by  $10^3$ .

1-Methyl-5-fluorocytosine				9-Ethylguanine			
	$x/a$	$y/b$	$z/c$		$x/a$	$y/b$	$z/c$
	$\sigma=9$	$\sigma=7$	$\sigma=11$		$\sigma=9$	$\sigma=7$	$\sigma=11$
H(4A)	260	349	292	H(1)	638	178	398
H(4B)	416	343	369	H(2A)	733	-18	350
H(6)	91	-47	41	H(2B)	874	-43	413
				H(8)	1105	563	815
				H(9A)	1210	253	855
				H(9B)	1260	260	655

column is listed the average standard error; this is the mean of the errors given in the output of the least-squares refinement program. Taking the mean is satisfactory because, for any given type of parameter, the error for each atom is approximately the same. It should be noted, however, that the errors for the carbon atoms of the alkyl groups are 20–30% higher than are those for the other non-hydrogen atoms.

Fig. 5 shows bond lengths and angles for GC, and the arrangement of the base-pairs in the  $(11\bar{2})$  plane. Besides the three Watson–Crick type hydrogen bonds, two others are formed, linking adjacent base-pairs across centres of symmetry. The positions of the hydrogen atoms, as taken from the least-squares refinement, show the hydrogen bonds all to be highly linear, the

maximum deviation from linearity being less than  $3^\circ$  (standard deviation is  $4^\circ$ ).

Fig. 6 represents the GC structure as viewed along the  $a$  axis; we can see in this figure that the two molecules are slightly inclined to each other. This inclination was predicted from the results of the Patterson rotation program.

Fig. 7 shows two layers of the GC structure, viewed in a direction normal to the  $(11\bar{2})$  plane. There is little overlap in this view. The vertical separation of the layers is  $3.3 \text{ \AA}$ .

Table 6 gives the covalent bond lengths for GC, GFC, 1-methylcytosine (Mathews & Rich, 1964a and personal communication) and guanine hydrochloride dihydrate (Iball & Wilson, 1963). In addition, in Table 6, a comparison between GC and GFC is made by forming the difference,  $\Delta$ , for each corresponding bond; the term  $\Delta/\sigma$  ( $\Delta$ ) then tests whether or not the difference is significant. In GC and GFC the ethylguanine should, of course, have identical bond lengths. The  $\Delta$  for C(4)–C(5) of ethylguanine is at the level of possible significance (Cruickshank & Robertson, 1953). This indicates that the errors in one or both of the structures may have been underestimated. Comparison of bond lengths in 1-methylcytosine with the corresponding lengths in GC and GFC shows good agreement. Similarly, bond lengths in guanine hydrochloride dihydrate

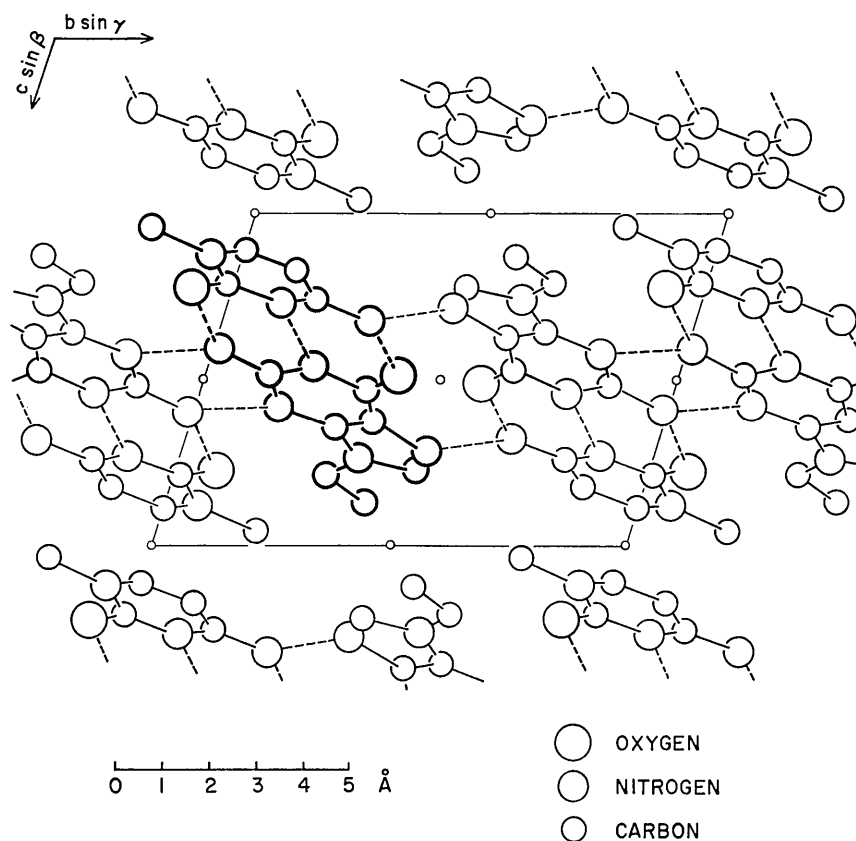


Fig. 6. The GC structure viewed along the  $a$  axis. One base-pair is shown in heavier lines.



Table 4. Observed and calculated structure factors for GC

The columns contain, from left to right, the values of h, k, l0|Fo|, and 10Fe. Unobserved reflections are listed as 1/2 . 10|Fmin|, and are marked with asterisks.

Table with 11 columns: L=0, 7, 6, 37, -36, -2, -10, 6\*, -8, 3, -3, 96, -72, 7, 3, 4A, -34, -3, 12, 42, -31, 3, -5, 17, -4, -6, 2, 222, 209. The table contains observed and calculated structure factors for various reflections, with some marked as unobserved with asterisks.







Table 6. Comparison of covalent bond lengths in GC, GFC, 1-methylcytosine (Mathews &amp; Rich, 1964a and personal communication) and guanine hydrochloride dihydrate (Jball &amp; Wilson, 1963)

All values are in Å [except  $\Delta/\sigma(\Delta)$ ]

	GC $\sigma=0.007$	GFC $\sigma=0.009$	$\Delta$ $\sigma=0.012$	$\Delta/\sigma(\Delta)$	1-Methylcytosine $\sigma=0.004$
1-Methylcytosine					
1-Methyl-5-fluorocytosine					
N(1)-C(2)	1.401	1.403	0.002	0.2	1.400
C(2)-N(3)	1.349	1.360	0.011	0.9	1.359
N(3)-C(4)	1.345	1.334	0.011	0.9	1.338
C(4)-C(5)	1.419	1.446	0.027	2.3	1.436
C(5)-C(6)	1.345	1.336	0.009	0.8	1.343
C(6)-N(1)	1.362	1.345	0.017	1.4	1.373
C(2)-O(2)	1.233	1.235	0.002	0.2	1.240
C(4)-N(4)	1.337	1.324	0.013	1.1	1.346
N(1)-C(1)	1.468	1.477	0.009	0.8	1.469
C(5)-F(5)		1.342			
9-Ethylguanine					
N(1)-C(2)	1.388	1.375	0.013	1.1	1.375
C(2)-N(3)	1.326	1.336	0.010	0.8	1.319
N(3)-C(4)	1.357	1.362	0.005	0.4	1.341
C(4)-C(5)	1.388	1.363	0.025	2.1	1.373
C(5)-C(6)	1.412	1.427	0.015	1.3	1.416
C(6)-N(1)	1.400	1.405	0.005	0.4	1.391
C(2)-N(2)	1.335	1.336	0.001	0.1	1.336
C(6)-O(6)	1.233	1.223	0.010	0.8	1.233
C(5)-N(7)	1.393	1.395	0.002	0.2	1.379
N(7)-C(8)	1.302	1.320	0.018	1.5	1.328
C(8)-N(9)	1.376	1.381	0.005	0.4	1.334
C(4)-N(9)	1.375	1.381	0.006	0.5	1.376*
N(9)-C(9)	1.485	1.467	0.018	1.5	
C(9)-C(10)	1.500	1.535	0.035	2.9	

\* This figure (Wilson, personal communication) is more recent than the published value of 1.397.

agree well with corresponding bond lengths in GC and GFC, especially GC. The agreement is much better, however, for the six-membered ring of guanine than for the five-membered ring. Guanine hydrochloride dihydrate is protonated at N(7) (Wilson, personal communication) so that the N(7)-C(8) bond in this compound has less double-bond character than its counterpart in 9-ethylguanine; hence the observed differences can be explained.

Table 7. Comparison of covalent bond angles in GC and GFC

	GC $\sigma=0.4^\circ$	GFC $\sigma=0.5^\circ$	$\Delta$ $\sigma=0.6^\circ$	$\Delta/\sigma(\Delta)$
1-Methylcytosine, 1-Methyl-5-fluorocytosine				
C(1)-N(1)-C(2)	118.7°	117.7°	1.0°	1.6
C(1)-N(1)-C(6)	121.4	121.7	0.3	0.5
C(2)-N(1)-C(6)	119.9	120.6	0.7	1.1
N(1)-C(2)-O(2)	117.5	118.7	1.2	1.9
N(1)-C(2)-N(3)	119.0	118.4	0.6	1.0
O(2)-C(2)-N(3)	123.4	122.9	0.5	0.8
C(2)-N(3)-C(4)	120.5	122.2	1.7	2.7
N(3)-C(4)-N(4)	118.4	121.6	3.2	5.1
N(3)-C(4)-C(5)	121.4	118.4	3.0	4.8
N(4)-C(4)-C(5)	120.3	120.0	0.3	0.5
C(4)-C(5)-C(6)	117.2	119.3	2.1	3.4
N(1)-C(6)-C(5)	121.8	121.0	0.8	1.3
C(4)-C(5)-F(5)		119.1		
C(6)-C(5)-F(5)		121.5		

Table 7 (cont.)

	GC $\sigma=0.4^\circ$	GFC $\sigma=0.5^\circ$	$\Delta$ $\sigma=0.6^\circ$	$\Delta/\sigma(\Delta)$
9-Ethylguanine				
C(2)-N(1)-C(6)	125.2°	125.3°	0.1°	0.2
N(1)-C(2)-N(2)	115.9	116.2	0.3	0.5
N(1)-C(2)-N(3)	123.3	123.3	0.0	0.0
N(2)-C(2)-N(3)	120.8	120.5	0.3	0.5
C(2)-N(3)-C(4)	112.2	112.3	0.1	0.2
N(3)-C(4)-C(5)	128.6	128.3	0.3	0.5
N(3)-C(4)-N(9)	125.6	124.5	1.1	1.8
C(5)-C(4)-N(9)	105.8	107.1	1.3	2.1
C(4)-C(5)-C(6)	118.9	119.8	0.9	1.4
C(4)-C(5)-N(7)	110.4	110.8	0.4	0.6
C(6)-C(5)-N(7)	130.7	129.4	1.3	2.1
N(1)-C(6)-C(5)	111.7	111.0	0.7	1.1
N(1)-C(6)-O(6)	119.7	120.0	0.3	0.5
C(5)-C(6)-O(6)	128.6	129.1	0.5	0.8
C(5)-N(7)-C(8)	104.1	103.5	0.6	1.0
N(7)-C(8)-N(9)	114.0	114.0	0.0	0.0
C(4)-N(9)-C(8)	105.8	104.6	1.2	1.9
C(4)-N(9)-C(9)	125.5	125.5	0.0	0.0
C(8)-N(9)-C(9)	128.7	129.8	1.1	1.8
N(9)-C(9)-C(10)	112.9	111.6	1.3	2.1

Table 7 gives the covalent bond angles in GC and GFC. Significant differences occur in the region of C(4) and C(5) of cytosine.

The best plane, in the least-squares sense, was calculated for the ring atoms of each base. A program, written by Haschemeyer (personal communication),

and based on the method of Blow (1960), was used. Table 8 gives the distances of the atoms from the least-squares planes. For the cytosine ring, the r.m.s. deviation of the atoms in the plane is 0.017 Å for GC, 0.007 Å for GFC. The extra-ring atoms, C(1) and N(4), are significantly out-of-plane. The r.m.s. deviation for the guanine plane is 0.011 Å for GC, 0.007 Å for GFC; O(6) and C(10) are significantly out-of-plane. The angle

between the guanine and cytosine planes is 6.5° for GC, 5.4° for GFC. In GC, the (11 $\bar{2}$ ) plane makes an angle of 4.0° with the cytosine plane, 2.6° with the guanine plane, and 0.3° with the plane through the ring atoms of both molecules. The equation for the plane through the ring atoms of both molecules is:

$$0.374X + 0.309Y - 0.875Z + 0.040 = 0$$

Table 8. Distances of atoms from least-squares planes calculated through the ring atoms (marked with asterisks) of each base

1-Methylcytosine, 1-Methyl-5-fluorocytosine			9-Ethylguanine		
	Distance from plane (Å)			Distance from plane (Å)	
	GC	GFC		GC	GFC
*N(1)	0.024	0.010	*N(1)	0.009	0.006
*C(2)	-0.020	-0.008	*C(2)	0.001	0.003
*N(3)	-0.004	-0.001	*N(3)	-0.015	-0.005
*C(4)	0.022	0.008	*C(4)	0.005	-0.008
*C(5)	-0.016	-0.006	*C(5)	-0.019	-0.012
*C(6)	-0.006	-0.002	*C(6)	0.010	0.004
			*N(7)	-0.011	-0.004
			*C(8)	0.009	0.008
			*N(9)	0.011	0.007
C(1)	0.079	0.064	N(2)	0.027	0.037
O(2)	-0.034	0.002	O(6)	0.047	0.052
N(4)	0.056	0.028	C(9)	0.002	-0.028
F(5)		0.001	C(10)	0.250	0.250

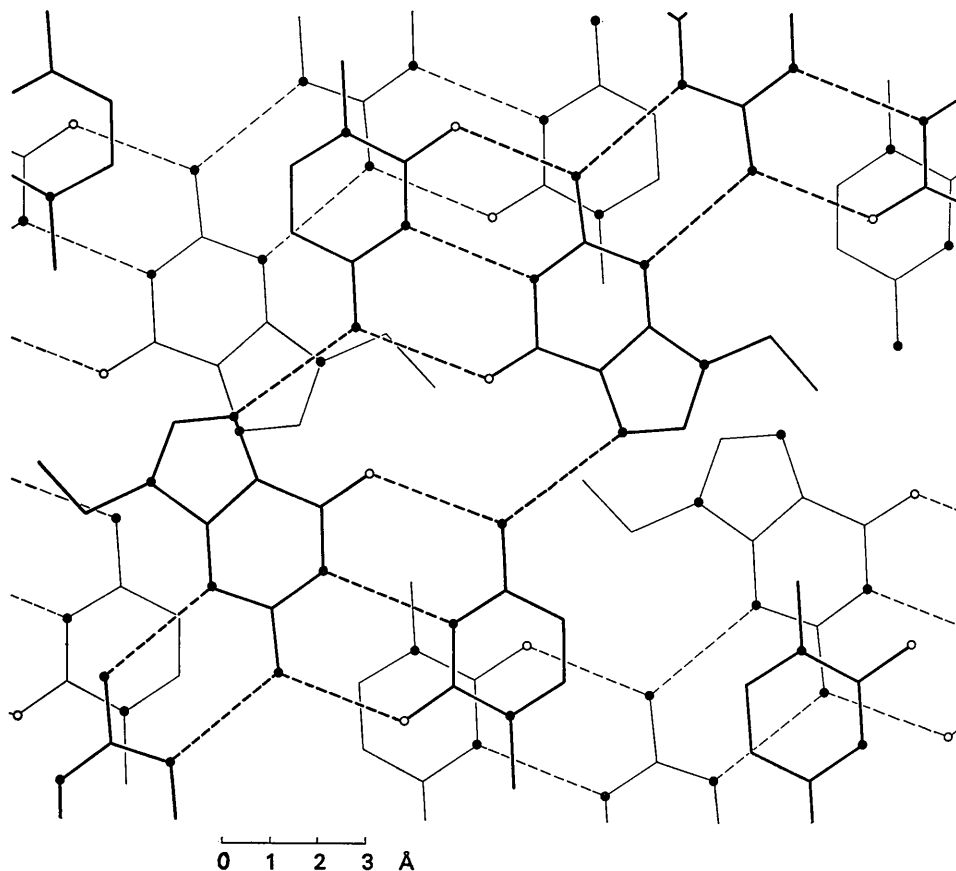


Fig. 7. Two layers of the GC structure, viewed in a direction normal to the (11 $\bar{2}$ ) plane. Hydrogen bonds are shown, but hydrogen atoms are omitted.

Table 9(a). Comparison of hydrogen bond distances in GC and GFC

Values are in Å [except  $\Delta/\sigma(\Delta)$ ].

	GC $\sigma=0.007$	GFC $\sigma=0.009$	$\Delta$ $\sigma=0.012$	$\Delta/\sigma(\Delta)$
Guanine to cytosine				
N(2)---O(2)	2.806	2.823	0.017	1.4
N(1)---N(3)	2.924	2.942	0.018	1.5
O(6)---N(4)	2.943	2.964	0.021	1.8
N(7)---N(4)	3.107	2.995	0.112	9.3
Guanine to guanine				
N(2)---N(3)	3.010	2.993	0.017	1.4

Table 9(b). Comparison of angles involving hydrogen bonds in GC and GFC

	GC $\sigma=0.4^\circ$	GFC $\sigma=0.5^\circ$	$\Delta$ $\sigma=0.6^\circ$	$\Delta/\sigma(\Delta)$
Cytosine to guanine				
C(4)-N(4)---O(6)	116.4	113.4	3.0	5.0
C(4)-N(4)---N(7)	121.5	127.9	6.4	10.7
Guanine to cytosine				
C(6)-N(1)---N(3)	114.3	113.0	1.3	2.2
C(2)-N(1)---N(3)	120.4	121.6	1.2	2.0
C(2)-N(2)---O(2)	122.9	121.5	1.4	2.3
Guanine to guanine				
C(2)-N(2)---N(3)	119.9	121.1	1.2	2.0

where  $X$ ,  $Y$  and  $Z$  are orthogonal coordinates in Å, with  $X$  along  $\mathbf{a}$ ,  $Y$  in the  $ab$  plane and  $Z$  along  $\mathbf{c}^*$ .

Table 9 compares hydrogen bond lengths and angles in GC and GFC. Only one hydrogen bond has significantly different lengths and makes significantly different angles in the two structures. A discussion of the hydrogen bond lengths in GC and GFC, and a comparison with the corresponding hydrogen bond lengths in the 1:1 crystalline complex of 9-ethylguanine with 1-methyl-5-bromocytosine (Sobell, Tomita & Rich, 1963) have been given in the previous communication (O'Brien, 1966).

The author would like to thank Dr F.H.C. Crick, Dr M.F. Perutz and Professor A. Rich for continued support and encouragement during this work; Drs A. Katritzky and A.J. Waring for providing the first samples of GC crystals; Miss Sheila Chapman and Mrs Joan Blows for assistance with data processing; the University of Cambridge Mathematical Laboratory and the Computation Center at the Massachusetts Institute of Technology for computer time. The work was supported by grants from the Rockefeller Foundation, the American Cancer Society and the National Institutes of Health (U.S. Public Health Service), and in part by a Public Health Service fellowship (no. 7-F2-CA-25, 632-01) from the National Cancer Institute.

#### References

ARNOTT, S., WILKINS, M. H. F., HAMILTON, L. D. & LANGRIDGE, R. (1965). *J. Mol. Biol.* **11**, 391.  
BLOW, D. M. (1960). *Acta Cryst.* **13**, 168.

BOOTH, A. D. (1946). *Trans. Faraday Soc.* **42**, 444.  
BROOMHEAD, J. M. (1948). *Acta Cryst.* **1**, 324.  
BUSING, W. R., MARTIN, K. O. & LEVY, H. A. (1962). *ORFLS, A Fortran Crystallographic Least-Squares Program*. Oak Ridge National Laboratory, Oak Ridge, Tennessee.  
CRICK, F. H. C. & WATSON, J. D. (1954). *Proc. Roy. Soc. A*, **223**, 80.  
CRUICKSHANK, D. W. J. & ROBERTSON, A. P. (1953). *Acta Cryst.* **6**, 698.  
CRUICKSHANK, D. W. J., PILLING, D. E., BUJOSA, A., LOVELL, F. M. & TRUTER, M. R. (1961). In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, p.45. Oxford: Pergamon Press.  
DONOHUE, J. (1956). *Proc. Nat. Acad. Sci. Wash.* **42**, 60.  
HAMILTON, W. C. (1955). *Acta Cryst.* **8**, 185.  
HASCHEMEYER, A. E. V. & SOBELL, H. M. (1963). *Proc. Nat. Acad. Sci. Wash.* **50**, 872.  
HASCHEMEYER, A. E. V. & SOBELL, H. M. (1964). *Nature, Lond.* **202**, 969.  
HASCHEMEYER, A. E. V. & SOBELL, H. M. (1965a). *Acta Cryst.* **18**, 525.  
HASCHEMEYER, A. E. V. & SOBELL, H. M. (1965b). *Acta Cryst.* **19**, 125.  
HIRSCHFELD, F. L., SANDLER, S. & SCHMIDT, G. M. J. (1963). *J. Chem. Soc.* p.2108.  
HOOGSTEEN, K. (1959). *Acta Cryst.* **12**, 822.  
HOOGSTEEN, K. (1963). *Acta Cryst.* **16**, 907.  
HOWELLS, E. R., PHILLIPS, D. C. & ROGERS, D. (1950). *Acta Cryst.* **3**, 210.  
IBALL, J. & WILSON, H. R. (1963). *Nature, Lond.* **198**, 1193.  
*International Tables for X-ray Crystallography* (1962). Vol. III, p.202. Birmingham: Kynoch Press.  
JENSEN, L. H. & SUNDARALINGAM, M. (1964). *Science*, **145**, 1185.  
KATZ, L., TOMITA, K. & RICH, A. (1965). *J. Mol. Biol.* **13**, 340.  
LIPSON, H. & COCHRAN, W. (1957). *The Determination of Crystal Structures*, 2nd edition, page 172: London: Bell.  
LIPSON, H. & WOOLFSON, M. M. (1952). *Acta Cryst.* **5**, 680.  
MATHEWS, F. S. & RICH, A. (1964a). *Nature, Lond.* **201**, 179.  
MATHEWS, F. S. & RICH, A. (1964b). *J. Mol. Biol.* **8**, 89.  
O'BRIEN, E. J. (1963). *J. Mol. Biol.* **7**, 107.  
O'BRIEN, E. J. (1966). *J. Mol. Biol.* **22**, 377.  
PHILLIPS, D. C. (1956). *Acta Cryst.* **9**, 819.  
ROGERS, D. & WILSON, A. J. C. (1953). *Acta Cryst.* **6**, 439.  
ROLLETT, J. S. & SPARKS, J. A. (1960). *Acta Cryst.* **13**, 273.  
SLY, W. G., SHOEMAKER, D. P. & VAN DEN HENDE, J. H. (1962). *ERFR2, Two- and Three-Dimensional Crystallographic Fourier Summation Program for the IBM 7090 Computer*. Massachusetts Institute of Technology, Esso Research and Engineering Company.  
SOBELL, H. M., TOMITA, K. & RICH, A. (1963). *Proc. Nat. Acad. Sci. Wash.* **49**, 885.  
SOBELL, H. M. (1966). *J. Mol. Biol.* **18**, 1.  
SPENCER, M. (1959a). *Acta Cryst.* **12**, 59.  
SPENCER, M. (1959b). *Acta Cryst.* **12**, 66.  
STEWART, R. F., DAVIDSON, E. R. & SIMPSON, W. T. (1965). *J. Chem. Phys.* **42**, 3175.  
TOLLIN, P. & COCHRAN, W. (1964). *Acta Cryst.* **17**, 1322.  
WARD, J. (1962). Personal communication.  
WELLS, M. (1961). In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, page 102. Oxford: Pergamon Press.