Acta Cryst. (1951). **4**, 42

The Crystal Structure of Cysteylglycine–Sodium Iodide

By H. B. Dyer

Crystallographic Laboratory, Cavendish Laboratory, Cambridge, England

(*Received* 29 *April* 1950)

The structure of the dipeptide cysteylglycine has been studied by X-ray methods. The crystals are monoclinic, having $a = 11{\cdot}11$, $b = 5{\cdot}12$, $c = 16{\cdot}04$ A., $\beta = 90^\circ\ 57'$.

The space group is $A2$. The unit cell contains four molecules of cysteylglycine and two of sodium iodide. The molecules exist in an extended form, bound end to end by very strong hydrogen bonds, to form a series of chains. Adjacent chains are held together by weaker hydrogen bonds, by ionic bonds between oxygen atoms and sodium ions, and by similar bonds between hydrogen atoms and iodine ions.

Introduction

In recent years, the immense problem of protein structure investigation by X-ray methods has attracted the attention of numerous investigators. Perutz and his co-workers have made considerable progress in their attempts to determine the structures of certain crystalline globular proteins, notably horse methaemoglobin (Perutz, 1949), and horse myoglobin (Kendrew, 1950), while Astbury's work (see, for example, Astbury, 1947) on fibrous proteins is well known.

It has been established conclusively, by chemical means, that proteins are composed of polypeptide chains which themselves consist of a series of amino-acids, linked together by peptide bonds. It is also known that there are definite periodicities in the arrangement of the amino-acid residues in the peptide chain.

As the great complexity of protein molecules precludes the direct application of most of the conventional methods of X-ray analysis, an accurate knowledge of the structures of the simpler units, from which these molecules are built, is important. The amino-acids glycine (Albrecht & Corey, 1939), DL-alanine (Levy & Corey, 1941), the related compound diketopiperazine (Corey, 1938), and the dipeptide β-glycylglycine (Hughes & Moore, 1949), have already been analysed. The present paper describes the determination of the structure of the crystalline dipeptide, cysteylglycine, and preliminary work on certain tripeptides which form part of a programme for the investigation of progressively more complex units of the protein structure.

Experimental

The crystals of 2(cysteylglycine)NaI were very kindly supplied by Sir Charles Harington, of the National Institute for Medical Research. All the specimens examined were plates, elongated in the direction of the b axis, with (001) dominant. The crystal investigated by X-ray methods was about $1 \times 0{\cdot}3 \times 0{\cdot}1$ mm.

A modification of the 'θ-method' of Weisz, Cochran & Cole (1948), in which the setting of the crystal giving maximum reflexion intensity was determined by a Geiger-counter spectrometer, showed the crystal to be monoclinic, with unit cell dimensions

$$a = 11{\cdot}11,\quad b = 5{\cdot}12,\quad c = 16{\cdot}04\ \text{A.},$$
$$\pm 0{\cdot}2\,\%;\quad \beta = 90^\circ\ 57' \pm 2'.$$

The density was found to be 1·88 g.cm.$^{-3}$, so that each unit cell contains four molecules of

$$\mathrm{SH.CH_2.CH(NH_2).CO.NH.CH_2.COOH},$$

and two of NaI (calculated density 1·84 g.cm.$^{-3}$).

Reflexions (hkl) for which $(k+l)$ is odd, are absent; no other systematic absences were observed, hence the space group is $A2$. This space group has no centre of symmetry, though the b-axis projection is centred. The four cysteylglycine units were assumed to lie in general positions, with two sodium and two iodine atoms lying on diad axes.

Zero-layer X-ray reflexions about the a and b axes were recorded photographically, using a Weissenberg camera, and their intensities were measured by visual comparison with an intensity scale. After correcting for Lorentz and polarization factors, and for the shape of reflexions (Broomhead, 1948), the b-axis zero-layer data were used to calculate a two-dimensional Patterson synthesis (Fig. 1 (a)). The iodine atom lies at the origin of the unit cell, and therefore peaks arising from vectors between this and the lighter atoms in the molecule lie in the positions these atoms themselves occupy in the cell. As vector peaks other than those involving an iodine atom are of a smaller order of magnitude, the map gives, fairly accurately, the x and z parameters of the atoms in the cell.

A Fourier synthesis was now calculated, in which all the F's were taken to be positive; i.e. it was assumed that their signs were the same as those of the iodine contributions. Using the atomic parameters derived from this synthesis, structure factors, F_c, were calculated, and, in fact, only four were found to be negative. The Fourier synthesis was recalculated accordingly, and a further set of structure factors determined, these being scaled to fit observed F's by multiplying by a factor which varied smoothly with $\sin\theta$.

Attempts were made to correct this projection for errors arising out of the termination of the Fourier series, the effectiveness of two methods being compared. The first was the Booth 'back-correction' method (Booth, 1946). In the second, the series was extended beyond the limit imposed by the reflecting sphere after the manner suggested by Cruickshank (1949). To do this, the atomic scattering factor of each of the atoms was reduced to an 'observed' scale, by means of the same $\sin\theta$ factor as was used in relating the observed to the calculated structure factors. Graphs were drawn of the variation of these scattering factors, f_r, with $\sin\theta$, and at that point. This should effect an approximate correction for series-termination errors, since if

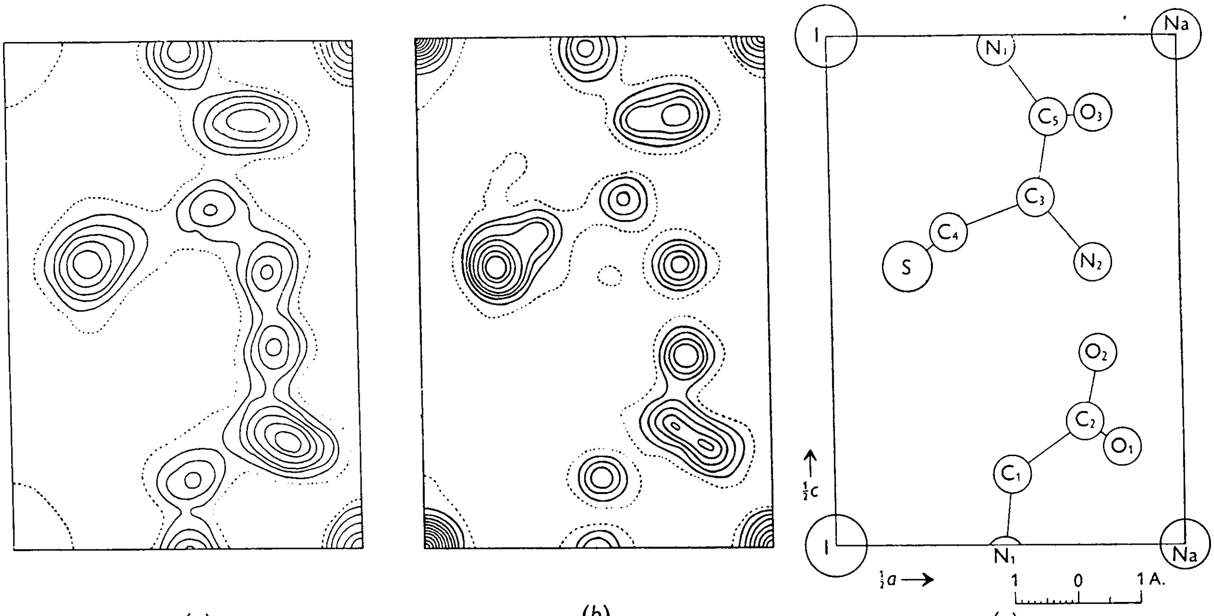
$$\rho' = \sum\sum_{0}^{N} (F_o - F_c) \cos 2\pi(hx/a + lz/c),$$

and

$$\rho_c = \sum\sum_{0}^{\infty} F_c \cos 2\pi(hx/a + lz/c),$$

the corrected density ρ is taken to be

$$\begin{aligned}\rho &= \sum\sum_{0}^{\infty} F_c \cos 2\pi(hx/a + lz/c) \\ &\quad + \sum\sum_{0}^{N} (F_o - F_c) \cos 2\pi(hx/a + lz/c) \\ &= \sum\sum_{0}^{N} F_o \cos 2\pi(hx/a + lz/c) \\ &\quad + \sum\sum_{N}^{\infty} F_c \cos 2\pi(hx/a + lz/c).\end{aligned}$$



(a) (b) (c)

Fig. 1. (a) Patterson projection on (010). (b) Electron-density map projected on (010). Contours at intervals of 2 e.A.$^{-2}$ for C, O and N; 4 e.A.$^{-2}$ for Na and S; 10 e.A.$^{-2}$ for I. First (broken) contour represents an electron-density of 2 e.A.$^{-2}$. (c) Key to (a) and (b).

extrapolated beyond $\sin\theta = 1$, to a point where f_r fell to 1 % of its value at $\sin\theta = 0$. Fourier transforms of each of these f_r curves were then calculated, and the spherically symmetrical distributions of electron density so obtained were transferred to a net similar to that used in representing the result of previous Fourier syntheses, the centres being made to coincide with the positions of the atomic centres as previously determined.

To this representation of electron density, ρ_c, was added, at each point in the net, the value of the Fourier synthesis

$$\rho' = \sum_{h}\sum_{l} (F_o - F_c) \cos 2\pi(hx/a + lz/c),$$

The electron-density map obtained by this method is shown in Fig. 1 (b).

A comparison was made between the atomic parameters determined by each of these methods for all the atoms which are completely resolved in projection, and in no case was the difference between corrected co-ordinates greater than 0·025 A. The mean difference was found to be 0·013 A.

A mean of the two results was taken as being nearest to the true value of each atomic co-ordinate, and a further set of structure factors was calculated. These are compared with the observed values in Table 4. The value of R ($= \sum | F_o - F_c | / \sum | F_o |$) was 0·13.

An a-axis Patterson projection was now calculated, but no immediate interpretation was possible. This was due, in part, to the inherent twofold ambiguity of this Patterson projection, for reasons of space-group symmetry, and, in part, to the fact that two molecules overlap. A possible structure was eventually postulated, however, which was in accord with the positions of the vector peaks, and with accepted bond lengths. A set of phase angles was determined on the basis of this structure, and a Fourier synthesis was calculated. However, only five atoms were completely resolved, and, even after successive refinements, the value of R had only fallen to 0·25. Accordingly, a further attempt was made to obtain, more accurately, the y co-ordinate of the atoms.

A full three-dimensional Fourier synthesis was not attempted, as the shape of the crystals was unsuited to the recording of three-dimensional data, and as the labour involved in calculating phase angles would have been prohibitive. Instead, a form of limited three-dimensional synthesis was decided upon. The method employed was that of Clews & Cochran (1949), with the difference that Patterson syntheses were used, a fact made possible by the close similarity of the Fourier and Patterson syntheses for this structure.

Oscillation and Weissenberg photographs were taken of the first and second layer lines about the b axis, and the intensities of the reflexions in the zero, first, and second layer lines were placed on the same scale. These were used as F^2's in the calculation of two Patterson syntheses of the form

$$P_K(x,z)=\sum_{\substack{h\\k=K}}\sum_{l} F^2_{(hkl)}\cos 2\pi(hx/a+lz/c)$$

for $K=1$ and $K=2$.

It may readily be shown that, if the height of a peak on the 'zero-layer-line' projection, $P_0(x,z)$, be taken as unity, there will be a peak in the same (x,z) position on the 'first-layer-line' projection, $P_1(x,z)$, of height proportional to $\cos 2\pi y_r/b$, where y_r is the y co-ordinate of the atom which gives rise to this peak through its vector to the iodine atom at the origin. Similarly, for the second layer line, the peak height will be proportional to $\cos 4\pi y_r/b$. The vector-density map of the function $P_1(x,z)$ is shown in Fig. 2, which should be compared with the map of $P_0(x,z)$ in Fig. 1 (a).

To determine the y co-ordinates of the atoms in the cell, the height of each peak on the first- and second-layer-line projections was measured. These values cannot be compared directly, for the reason that peak heights on higher-layer-line projections are, of necessity, less than on that corresponding to the zero layer, as the effective atomic scattering factors, and hence the coefficients $F^2_{(hkl)}$, decrease with increase in K. To reduce this effect the ratio

$$\alpha_K=\frac{\text{(height of peak)}}{\text{(height of origin peak of corresponding projection)}}$$

was determined, the denominator here being affected only by the reduction in scattering power at high angles. This does not effect a complete correction, as the atomic scattering factor of iodine, which is the main contributor to the origin peak, does not fall off with $\sin\theta$ in the same way as those of the lighter atoms. It is, however, the best available method.

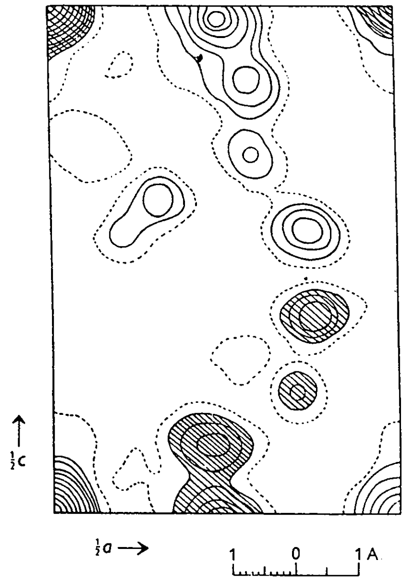


Fig. 2. The function $P_1(x,z)$. Areas for which this function is negative are shaded.

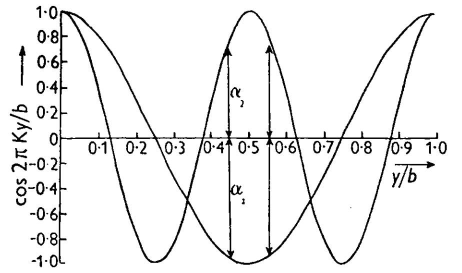


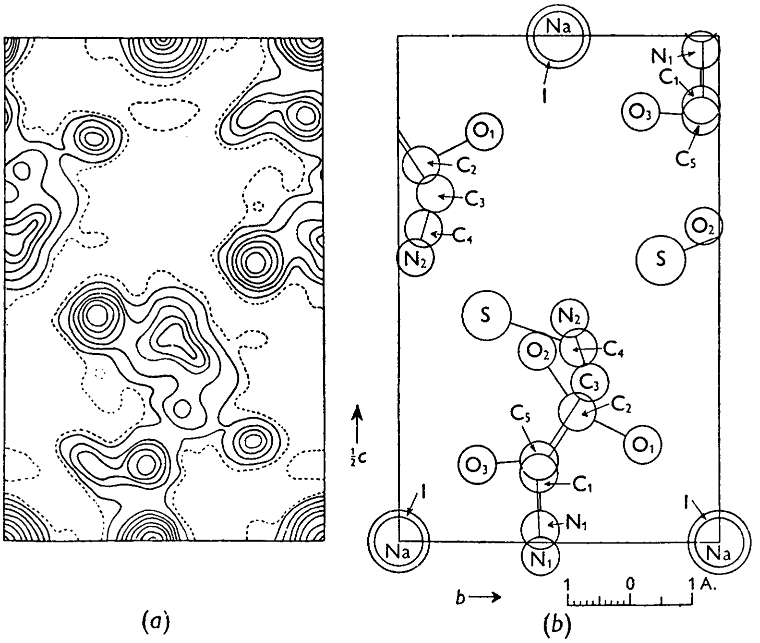
Fig. 3. Illustration of the method of determining atomic co-ordinates from the heights of peaks on the first- and second-layer-line Patterson projections. The case illustrated is that of atom O_2, which has $\alpha_1/\alpha_2=-1{\cdot}12/1{\cdot}00$.

The ratio of the value of α_K for a peak on the first-, to its value for this peak on the second-layer-line projection (α_1/α_2) was now determined. A graph was drawn (Fig. 3) on which the variation with y/b of the functions $\cos 2\pi y/b$ and $\cos 4\pi y/b$ was shown. For each peak a value of y/b was then found for which

$$\frac{\alpha_1}{\alpha_2}=\frac{\cos 2\pi y/b}{\cos 4\pi y/b}.$$

For each atom, however, there are two values of y for which the above equation is true, owing to the symmetry of the cosine function about $y/b=\frac{1}{2}$. For example, the atom O_2 has the ratio $\alpha_1/\alpha_2=-1{\cdot}12/1{\cdot}00$. As is shown in Fig. 3, this is compatible with its having a y co-ordinate 0·55 or 0·45.

Using the y co-ordinates obtained in this way, and with some help from the a-axis Patterson projection, a structure was now arrived at in which the y co-ordinates of the atoms were, in the case of some atoms, materially different from those previously postulated. A set of a-axis structure factors was calculated on the basis of the new atomic parameters, and the agreement between observed and calculated $F_{(0kl)}$'s was appreciably better. Successive a-axis Fourier projections were then calculated to determine, more accurately, the positions of those atoms which are completely resolved. The third of these is shown in Fig. 4 (a), from which it may be seen that only three atoms are completely, and three partially resolved.



(a) (b)

Fig. 4. (a) Electron-density map projected on (100). Contours at intervals of 2 e.A.$^{-2}$ for C, N and O; 4 e.A.$^{-2}$ for S; 10 e.A.$^{-2}$ for I. First (broken) contour represents an electron density of 2 e.A.$^{-2}$. (b) Key to (a).

Table 1. *Atomic parameters, expressed as fractions of cell edge*

Atom	x	y	z
C_1	0·255	0·943	0·569
C_2	0·362	1·070	0·623
C_3	0·293	1·113	0·343
C_4	0·167	1·077	0·308
C_5	0·315	0·943	0·420
O_1	0·415	1·269	0·596
O_2	0·382	0·950	0·688
O_3	0·379	0·755	0·424
N_1	0·247	0·940	0·487
N_2	0·375	1·053	0·279
S	0·109	0·797	0·277
Na	0·500	0·500	0·500
I	0·000	0·500	0·500

The positions of the resolved atoms were corrected for series-termination errors by the Booth 'back-correction' method, the positions of the others being taken as those arrived at by the 'higher-layer-line Patterson' method.

The final atomic parameters are shown in Table 1, and the final a-axis structure factors in Table 5. The R factor is 0·14.

Accuracy of results

In considering the accuracy with which atomic parameters have been determined, a distinction must be made between the y and the x and z parameters. The latter were obtained from a Fourier projection down the b axis, and are considered to be the more accurate, as this projection contains a centre of symmetry, and the atoms are generally well resolved from one another.

From a study of the agreement between observed and calculated $F_{(h0l)}$'s, and of the background undulations in the Fourier projection, it is thought that the probable error in x and z co-ordinates is less than 0·03 A.

The y co-ordinates, however, are less accurately known, and may be in error by as much as 0·05 A. The probable error in bond lengths is estimated as 0·05 A., and the corresponding error in bond angle as about 3°.

Structure of molecule

A diagram of the cysteylglycine molecule, as seen in projection down the b axis, is shown in Fig. 5. There are shown, in addition to the heavier atoms, three hydrogen atoms whose probable positions were decided upon by consideration of the van der Waals contacts which they make with other atoms. The only other hydrogen atoms about which anything is known are those associated with the atoms N_2 and O_2, and which form hydrogen bonds between these atoms and the respective oxygen and nitrogen atoms in adjacent molecules. There is, however, no certain method of determining where, within any hydrogen bond, one of these hydrogen atoms may lie.

Values obtained for the interatomic bond lengths and bond angles are listed in Tables 2 and 3, where they are compared with values obtained for glycine (Albrecht & Corey, 1939), DL-alanine (Levy & Corey, 1941), diketopiperazine (Corey, 1938), and β-glycylglycine (Hughes & Moore, 1949).

Though the molecule as a whole is by no means plane, the carbon atom C_1 is found to be coplanar with the atoms of the carboxyl radicle, a fact in accord with the structures of oxalic acid dihydrate (Dunitz & Robertson, 1947), glycine, DL-alanine and β-glycylglycine.

The C–O distances of 1·23 and 1·28 A. are in good agreement with those found in glycine, DL-alanine and β-glycylglycine. The bond angles about the carbon atom C_2 also agree very well with those of β-glycylglycine, there being similar asymmetry of the carboxyl radicle about the C_1–C_2 bond. The C_5–O_3 bond length of 1·21 A. is a little less than those found in the carboxyl radicle, a fact which may possibly be ascribed to its being more nearly a pure double bond.

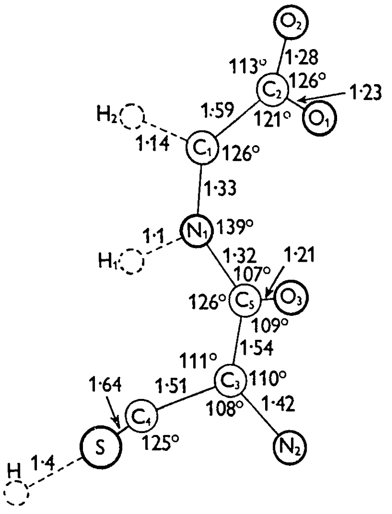


Fig. 5. The cysteylglycine molecule projected on (010), showing interatomic bond lengths and bond angles, and postulated hydrogen positions.

Table 2. *Comparison of bond lengths in cysteylglycine and in related compounds*

Bond	Cysteyl-glycine	Glycyl-glycine	Glycine	Alanine	Diketo-piperazine
C_1–C_2	1·59	1·53	1·52	1·54	—
C_3–C_4	1·51	—	—	—	—
C_3–C_5	1·54	1·53	—	1·54	—
C_1–N_1	1·33	1·48	—	—	—
C_5–N_1	1·32	1·29	—	—	1·33
C_3–N_2	1·42	1·51	1·39	1·42	1·41
C_5–O_3	1·21	1·23	—	—	1·25
C_2–O_1	1·23	1·21	1·25	1·23	—
C_2–O_2	1·28	1·27	1·27	1·25	—
C_4–S	1·64	—	—	—	—

Table 3. *Comparison of bond angles in cysteylglycine and in β-glycylglycine*

Bond angle	Cysteyl glycine	Glycyl glycine
O_2–C_2–O_1	126°	124°
C_1–C_2–O_2	113°	112°
C_1–C_2–O_1	121°	123°
N_1–C_1–C_2	126°	110°
C_1–N_1–C_5	139°	122°
N_1–C_5–C_3	126°	114°
N_1–C_5–O_3	107°	125°
C_3–C_5–O_3	109°	121°
C_5–C_3–N_2	110°	110°
S–C_4–C_3	125°	—
C_5–C_3–C_4	111°	—
C_4–C_3–N_2	108°	—

The C–C distances of 1·51, 1·54 and 1·59 A. all agree fairly well with accepted values for a single bond of this type, while the bond angles about the carbon atom C_3 are each sufficiently close to 109° to suggest a tetrahedral arrangement.

The C_3–N_2 bond length of 1·42 A. agrees well with values obtained for bonds of similar type in glycine, alanine and diketopiperazine. Hughes & Moore, however, record a value of 1·51 A. for the corresponding bond in β-glycylglycine, and suggest that the shortness of the bond found in glycine, etc., is due to experimental error.

The C–N distances found in the peptide link are appreciably shorter than this. This might be expected of the C_5–N_1 bond, as it may have some double-bond character. The value of 1·32 A. obtained agrees favourably with similar bonds found in diketopiperazine and in β-glycylglycine. There appears, however, to be no obvious explanation of the shortening of the N_1–C_1 bond, other than that of experimental error.

The C_4–S bond length of 1·64 A. is rather less than the value obtained by the addition of the Pauling (1945) covalent radii of carbon and sulphur, and again there seems to be no obvious explanation.

A further anomaly is the extended form of the molecule in the direction of the peptide chain; i.e. the bond angles C_3–C_5–N_1, C_5–N_1–C_1 and N_1–C_1–C_2 are all larger than corresponding angles found in β-glycylglycine. The C_5–N_1–C_1 bond angle is particularly large, being 139°. This cannot be ascribed entirely to experimental error, as shifts of even 0·1 A. in the y co-ordinates of any of these atoms produce very small changes in this angle; and, as all these atoms have y co-ordinates which are very nearly equal to $\frac{1}{2}b$, their being incorrectly placed as a result of the ambiguity of the 'higher-layer-line' Patterson syntheses would not involve large changes in their y co-ordinate.

Intermolecular bonding

The spatial relationships of the molecules within the crystals, and the bonds formed between atoms of adjacent molecules, are illustrated in Figs. 6, 7 and 8.

The cysteylglycine molecules are linked together by hydrogen bonds to form chains in the direction of their greatest length. Each hydrogen bond links the terminal nitrogen, N_2, of one molecule to an oxygen atom, O_2, of the carboxyl group of the next. The N–O distance, 2·55 A., is rather shorter than previously determined hydrogen bonds of this type, though Broomhead (unpublished), in the structure of guanine.HCl, H_2O, records an N–H–O distance of 2·62 A.

The molecules of glycine, alanine and glycylglycine are all believed to exist in the form of *Zwitter*-ions. If the cysteylglycine molecules assume this form, i.e. $\overset{+}{\mathrm{N}}\mathrm{H_3.CH}\ldots\mathrm{CH_2.CO}\overset{-}{\mathrm{O}}$, the terminal nitrogen might be expected to form three hydrogen bonds. In fact, one other hydrogen bond is formed linking two adjacent

Table 4. *Comparison of observed and calculated values of* $F_{(h0l)}$

$h0l$	$F_{\text{obs.}}$	$F_{\text{calc.}}$
002	19·5	24·0
004	36·6	39·8
006	17·6	22·2
008	20·1	20·4
0,0,10	40·1	35·0
0,0,12	22·3	20·5
0,0,14	23·0	20·5
0,0,16	7·8	6·1
0,0,18	7·6	7·6
0,0,20	4·2	1·3
100	9·6	5·7
102	25·1	19·8
104	32·4	34·3
106	48·5	43·2
108	22·6	19·6
1,0,10	16·2	18·5
1,0,12	4·2	1·4
1,0,14	21·7	19·1
1,0,16	5·1	5·9
1,0,18	6·0	6·4
1,0,20	3·5	3·3
200	17·4	19·5
202	40·0	37·6
204	12·2	13·8
206	28·9	31·6
208	23·1	20·4
2,0,10	34·9	31·8
2,0,12	20·3	16·3
2,0,14	4·2	8·7
2,0,16	0	3·9
2,0,18	6·6	7·2
2,0,20	2·3	3·6
300	44·6	40·3
302	30·8	25·4
304	8·1	9·3
306	28·9	31·5
308	4·4	1·1
3,0,10	14·8	19·3
3,0,12	−4·3	−1·0
3,0,14	10·3	10·3
3,0,16	4·2	5·2
3,0,18	4·0	3·7
400	6·8	2·5
402	61·6	68·6
404	35·5	30·9
406	40·0	40·7
408	20·7	20·6
4,0,10	9·2	11·4
4,0,12	18·5	15·4
4,0,14	16·0	14·9
4,0,16	9·0	8·8
500	21·5	22·7
502	19·9	21·2
504	5·8	6·3
506	6·5	10·5
508	20·2	20·7
5,0,10	19·9	19·8
5,0,12	14·1	14·9
5,0,14	4·1	1·6
5,0,16	7·3	7·1
600	15·2	13·4
602	13·9	17·4
604	30·3	29·3
606	24·1	23·3
608	15·8	17·3
6,0,10	6·3	8·0
6,0,12	7·9	10·1
6,0,14	4·0	4·7
700	14·7	15·6
702	20·4	18·5
704	14·3	15·1
706	5·5	6·3
708	11·5	14·6
7,0,10	−7·1	−2·2
7,0,12	13·6	13·2
7,0,14	3·3	1·5
800	32·1	33·0
802	11·7	12·5
804	23·9	22·2
806	6·8	8·5
808	17·8	19·2
8,0,10	7·6	7·8
8,0,12	6·1	6·8
8,0,14	5·5	6·3
900	9·3	9·6
902	10·2	6·3
904	14·4	15·4
906	5·1	5·4
908	8·5	9·1
9,0,10	5·8	5·9
9,0,12	6·3	4·9
9,0,14	5·5	6·2
10,0,0	16·0	17·2
10,0,2	5·7	5·2
10,0,4	9·0	10·1
10,0,6	6·8	7·3
10,0,8	6·2	6·4
10,0,10	5·5	7·2
10,0,12	3·6	4·6
11,0,0	10·3	9·1
11,0,2	6·3	6·4
11,0,4	6·2	7·9
11,0,6	5·8	7·1
11,0,8	0	1·2
11,0,10	4·2	4·6
11,0,12	−4·2	−2·2
12,0,0	4·1	5·6
12,0,2	8·3	9·7
12,0,4	0	2·9
12,0,6	8·3	8·4
12,0,8	0	3·4
12,0,10	4·5	6·6
13,0,0	3·8	3·7
13,0,2	4·6	4·8
13,0,4	1·5	1·6
13,0,6	4·4	4·9
14,0,0	0	1·4
14,0,2	5·4	6·3
$\bar{1}02$	6·2	5·9
$\bar{1}04$	54·7	52·4
$\bar{1}06$	11·5	9·2
$\bar{1}08$	26·2	31·0
$\bar{1},0,10$	4·5	5·4
$\bar{1},0,12$	14·8	14·5
$\bar{1},0,14$	0	3·8
$\bar{1},0,16$	7·9	3·1
$\bar{1},0,18$	5·7	6·7
$\bar{1},0,20$	6·2	2·0
$\bar{2}02$	10·5	14·3
$\bar{2}04$	49·9	55·4
$\bar{2}06$	22·0	19·4
$\bar{2}08$	36·1	39·8
$\bar{2},0,10$	14·6	13·9
$\bar{2},0,12$	4·4	10·0
$\bar{2},0,14$	14·8	11·8
$\bar{2},0,16$	10·1	9·9
$\bar{2},0,18$	6·2	6·2
$\bar{2},0,20$	6·5	4·9
$\bar{3}02$	36·9	42·3
$\bar{3}04$	11·9	12·4
$\bar{3}06$	4·5	2·5
$\bar{3}08$	25·8	29·1
$\bar{3},0,10$	8·5	7·4
$\bar{3},0,12$	31·4	28·6
$\bar{3},0,14$	11·8	8·4
$\bar{3},0,16$	7·4	8·4
$\bar{3},0,18$	5·5	2·9
$\bar{4}02$	52·1	50·3
$\bar{4}04$	28·8	27·8
$\bar{4}06$	42·0	42·4
$\bar{4}08$	29·8	34·0
$\bar{4},0,10$	14·3	13·6
$\bar{4},0,12$	25·3	23·7
$\bar{4},0,14$	5·0	3·8
$\bar{4},0,16$	10·2	9·9
$\bar{4},0,18$	7·0	4·7
$\bar{5}02$	33·2	33·6
$\bar{5}04$	18·9	14·4
$\bar{5}06$	24·8	27·5
$\bar{5}08$	−3·6	−1·6
$\bar{5},0,10$	13·4	15·2
$\bar{5},0,12$	4·5	5·4
$\bar{5},0,14$	4·9	5·1
$\bar{5},0,16$	9·2	9·3
$\bar{5},0,18$	3·5	1·0
$\bar{6}02$	34·6	32·8
$\bar{6}04$	6·5	7·2
$\bar{6}06$	25·1	30·3
$\bar{6}08$	16·4	14·5
$\bar{6},0,10$	16·0	15·9
$\bar{6},0,12$	11·3	10·7
$\bar{6},0,14$	6·4	7·7
$\bar{6},0,16$	7·4	6·5
$\bar{6},0,18$	3·8	3·2
$\bar{7}02$	14·5	12·9
$\bar{7}04$	10·2	9·3
$\bar{7}06$	19·4	20·2
$\bar{7}08$	5·7	5·0
$\bar{7},0,10$	18·3	20·0
$\bar{7},0,12$	5·7	4·4
$\bar{7},0,14$	9·6	8·8
$\bar{7},0,16$	4·3	2·2
$\bar{8}02$	15·2	15·4
$\bar{8}04$	25·1	26·1
$\bar{8}06$	17·6	17·1
$\bar{8}08$	8·6	7·2
$\bar{8},0,10$	13·5	15·1
$\bar{8},0,12$	5·0	5·2
$\bar{8},0,14$	11·1	10·8
$\bar{8},0,16$	6·2	4·4
$\bar{9}02$	11·1	10·3
$\bar{9}04$	13·4	14·7
$\bar{9}06$	5·4	4·6
$\bar{9}08$	13·0	13·7
$\bar{9},0,10$	4·4	4·5
$\bar{9},0,12$	4·1	2·2
$\bar{9},0,14$	4·2	4·4
$\overline{10},0,2$	4·9	6·8
$\overline{10},0,4$	14·2	13·4
$\overline{10},0,6$	7·4	6·9
$\overline{10},0,8$	8·4	8·2
$\overline{10},0,10$	6·3	5·5
$\overline{10},0,12$	6·7	6·7
$\overline{11},0,2$	0	0·4
$\overline{11},0,4$	8·1	7·5
$\overline{11},0,6$	4·2	4·1
$\overline{11},0,8$	8·3	9·4
$\overline{11},0,10$	3·8	2·6
$\overline{11},0,12$	6·5	6·7
$\overline{12},0,2$	11·0	10·7
$\overline{12},0,4$	9·4	10·2
$\overline{12},0,6$	4·5	3·8
$\overline{12},0,8$	6·5	6·5
$\overline{12},0,10$	4·3	3·9
$\overline{13},0,2$	4·2	4·7
$\overline{13},0,4$	3·9	4·0
$\overline{13},0,6$	3·7	3·3
$\overline{13},0,8$	2·0	2·2
$\overline{14},0,2$	3·3	3·8
$\overline{14},0,4$	3·1	1·0

Table 5. *Comparison of observed and calculated values of* $F_{(0kl)}$

$0kl$	$F_{\text{obs.}}$	$F_{\text{calc.}}$	$0kl$	$F_{\text{obs.}}$	$F_{\text{calc.}}$	$0kl$	$F_{\text{obs.}}$	$F_{\text{calc.}}$
002	19·5	19·9	020	31·0	27·2	042	15·8	18·9
004	36·2	33·2	022	30·2	29·0	044	14·8	16·0
006	15·5	17·7	024	25·7	25·2	046	7·7	8·3
008	18·7	16·7	026	28·8	35·6	048	10·5	11·8
0,0,10	38·1	30·6	028	14·8	17·5	0,4,10	8·9	13·4
0,0,12	25·1	23·2	0,2,10	16·4	17·7	0,4,12	10·5	11·0
0,0,14	24·7	21·0	0,2,12	14·5	14·2	0,4,14	8·9	9·2
0,0,16	10·5	10·3	0,2,14	13·8	10·0	051	8·9	13·3
0,0,18	11·0	10·6	0,2,16	17·6	14·9	053	8·4	11·1
0,0,20	2·0	1·3	0,2,18	6·3	4·9	055	5·5	7·1
011	5·5	8·2	031	18·7	18·0	057	11·4	13·6
013	45·5	43·8	033	17·9	20·2	059	5·5	5·9
015	13·8	16·9	035	20·3	21·1	0,5,11	11·8	11·1
017	30·5	34·0	037	13·4	16·4	060	6·3	7·2
019	17·6	17·8	039	9·5	18·0	062	5·5	6·6
0,1,11	21·7	17·9	0,3,11	8·4	11·1	064	6·3	10·9
0,1,13	17·6	13·6	0,3,13	11·8	13·8	066	4·5	7·5
0,1,15	11·0	9·2	0,3,15	7·1	6·3	068	6·3	7·3
0,1,17	7·7	8·1	0,3,17	7·7	7·2			
0,1,19	9·5	9·0	040	25·1	23·1			

molecules, across the chain, and making an angle with the 2·55 A. bond roughly equal to the tetrahedral angle. The length of this bond is 3·03 A. The third hydrogen attached to the terminal nitrogen may possibly be directed toward the oxygen atom, O_2, of the carboxyl group of the molecule which, in the next unit cell, lies directly above the one to which the 2·55 A. bond is made. If this is the case, no hydrogen bond of the normal type exists, as the N–O distance is too great, being 3·42 A. On this assumption, however, the nitrogen atom makes four tetrahedrally disposed bonds. Thus, while in glycylglycine the terminal nitrogen makes three hydrogen bonds of roughly equal length to neighbouring oxygen atoms, in cysteylglycine one very strong bond is formed, together with one which is rather weaker, and a third which approximates more nearly to a van der Waals bond.

A further set of cross links between the chains is provided by the sodium ions, about each of which are grouped four oxygen atoms in approximately tetrahedral arrangement. Each sodium ion is bonded electrostatically to the oxygen atoms, O_1, of a pair of molecules related by a diad axis passing through the soium ion; and to the oxygen atoms, O_3, of a further pair of molecules related in a similar manner.

Both the Na^+–O_1 distances of 2·16 A. and the Na^+–O_3 distances of 2·23 A. appear to be slightly less than the value predicted by the sum of the ionic radii of sodium and oxygen, quoted by Pauling (1945). However, this whole arrangement is unusual, as the sodium ion shows here a co-ordination number of four, as compared with the usual one of six.

The spatial relationships between oxygen atoms and sodium ions, and the nitrogen to oxygen hydrogen bonds of length 2·55 A., are shown in Figs. 6 and 7, while the hydrogen bonds of length 3·03 A. are shown in Fig. 6 only.

The iodine ions apparently play an equally important, though less obvious, part in holding the structure together. If the hydrogen atom associated with the sulphur atom, S, be assumed to lie somewhere on the surface of a sphere of radius 1·4 A., described about the sulphur atom (this distance corresponds to the sum of the Pauling covalent radii of sulphur and hydrogen), a position can be found at which it is equidistant from two iodine ions. The H–I distance is here 4·25 A. In this position the hydrogen atom of one molecule makes van der Waals contacts with both the corresponding

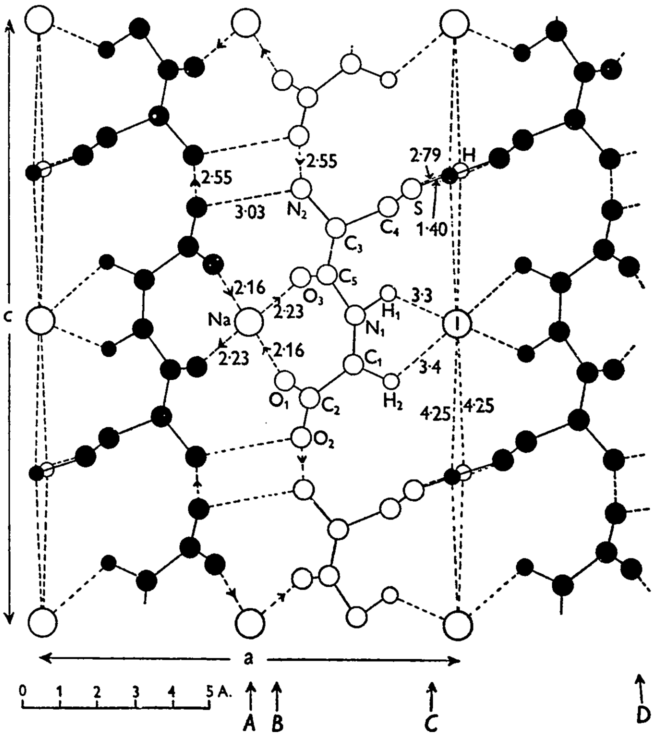


Fig. 6. Molecules in unit cell projected on (010). Intermolecular bonds represented by broken lines. Black circles represent atoms of molecules related to those represented by open circles by the operation of diad axes through the iodine and sodium ions.

hydrogen and sulphur atoms of an adjacent molecule. The H–H distance is approximately 2·5 A., and the S–H distance 2·8 A. This arrangement is illustrated in Figs. 6 and 8.

If, for the atoms C_1 and N_1, hydrogen atoms are taken to be at distances from the atomic centres equal to the sum of the appropriate covalent radii, these hydrogen atoms can once again be placed in positions equidistant from two iodine ions. The H–I distances in this case are 3·4 A. for the C_1 hydrogen, and 3·3 A. for the N_1 hydrogen, and agree well with the theoretical value of 3·36 A. for such a van der Waals contact. Further, the two hydrogen atoms are situated at a distance of 2·4 A. apart, a distance approximately equal to the sum of their van der Waals radii. These probable hydrogen positions are shown in Figs. 6 and 8.

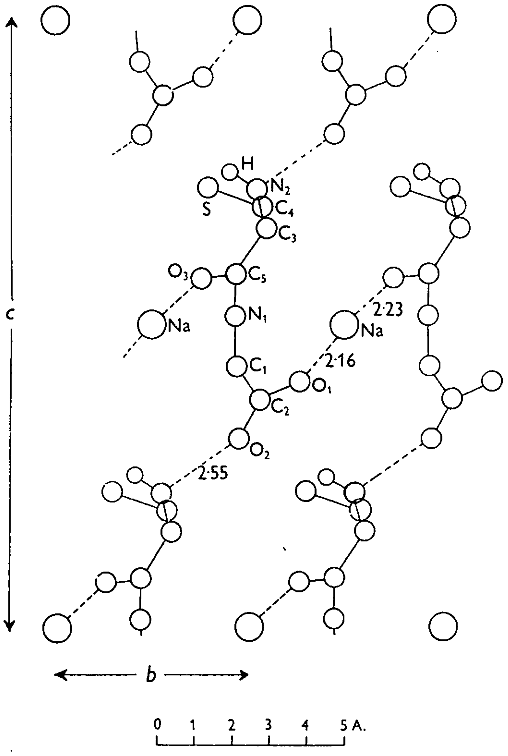


Fig. 7. Section lying between points *A* and *C* on Fig. 6, projected on (100).

It thus appears that cysteylglycine molecules are bound together, end to end, by very strong hydrogen bonds, to form a series of chains. Adjacent chains are held together by weaker hydrogen bonds; by ionic bonds between oxygen atoms and sodium ions; by similar bonds between hydrogen atoms and iodine ions; and by van der Waals contacts between neighbouring SH groups.

Further programme

An X-ray investigation of the structures of the isomorphous compounds triglycylethyl ester hydrochloride, and triglycylethyl ester hydrobromide has been started. Both compounds crystallize in the orthorhombic system, the space group being *Pcab*. The two unit cells have almost identical dimensions, approximately

$$a = 8{\cdot}4,\quad b = 8{\cdot}9,\quad c = 30{\cdot}8\text{ A}.$$

Patterson and Fourier projections along the *a* axis have been prepared and interpreted.

Preliminary work has also been begun on crystals of glycylglycylglycine, the unit-cell dimensions of which have been found to be the same as those observed by Bernal (1931).

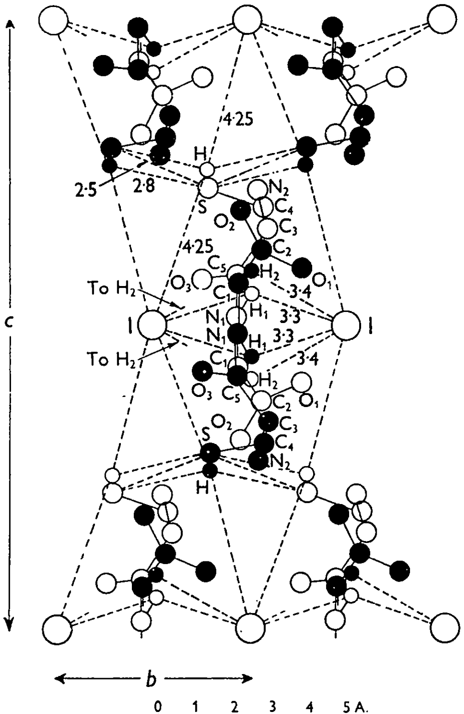


Fig. 8. Section lying between points *B* and *D* on Fig. 6, projected on (100). Black and open circles correspond to atoms similarly represented in Fig. 6.

I am grateful to Prof. Sir Lawrence Bragg, Dr W. H. Taylor and Dr W. Cochran for their constant encouragement and guidance, and to Sir Charles Harington, who provided the crystals which were used in this determination. I am also indebted to the University of South Africa for the award of a Croll Scholarship, during the tenure of which this work was carried out.

References

Albrecht, G. & Corey, R. B. (1939). *J. Amer. Chem. Soc.* **61**, 1087.

Astbury, W. T. (1947). *Proc. Roy. Soc.* B, **134**, 303.

Bernal, J. D. (1931). *Z. Krystallogr.* **78**, 363.

Booth, A. D. (1946). *Proc. Roy. Soc.* A, **188**, 77.

Broomhead, J. M. (1948). *Acta Cryst.* **1**, 324.

Clews, C. J. B. & Cochran, W. (1949). *Acta Cryst.* **2**, 46.

COREY, R. B. (1938). *J. Amer. Chem. Soc.* **60**, 1598.
CRUICKSHANK, D. W. J. (1949). *Acta Cryst.* **2**, 65.
DUNITZ, J. D. & ROBERTSON, J. M. (1947). *J. Chem. Soc.* **142**, 1947.
HUGHES, E. W. & MOORE, W. J. (1949). *J. Amer. Chem. Soc.* **71**, 2618.
KENDREW, J. C. (1950). In the Press.
LEVY, H. A. & COREY, R. B. (1941). *J. Amer. Chem. Soc.* **63**, 2095.
PAULING, L. (1945). *The Nature of the Chemical Bond.* Correll: University Press.
PERUTZ, M. F. (1949). *Proc. Roy. Soc.* A, **195**, 474.
WEISZ, O., COCHRAN, W. & COLE, W. F. (1948). *Acta Cryst.* **1**, 83.

Acta Cryst. (1951). **4**, 50

The Configurations of the Diastereomeric 1-Chloromercuri-2-Methoxycyclohexanes

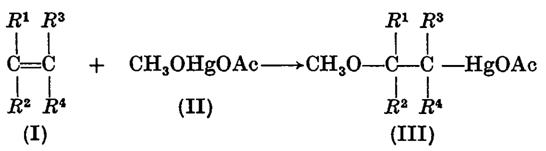
BY A. G. BROOK AND GEORGE F WRIGHT

Department of Chemistry, University of Toronto, Canada

(*Received* 6 *April* 1950)

Structure determinations of the α- and β-1-chloromercuri-2-methoxycyclohexanes have shown, contrary to earlier opinions based on chemical evidence, that the α form has the diastereomeric configuration known as *trans* while the β form has the *cis* configuration. The cyclohexane ring in each exists as the puckered or crown form. The structure determination was facilitated by the fact that the chloro-, bromo- and iodomercurials of each diastereoisomer form isostructural crystalline series from which an initial estimate could be made for the relative positions of the halogen and mercury atoms. A new interpretation of the chemical evidence, based on cycloco-ordination, can be obtained from these structure determinations.

Methoxymercuration may be defined as the addition to a carbon–carbon double bond of the elements of methoxy and acetoxymercuri:



$$\underset{R^2\ \ R^4}{\overset{R^1\ \ R^3}{C{=}C}} + CH_3OHgOAc \longrightarrow CH_3O{-}\underset{R^2\ \ R^4}{\overset{R^1\ \ R^3}{C{-}C}}{-}HgOAc$$

(I) (II) (III)

The addend (II) evidently originates from a solution of mercuric acetate in methanol. The reaction is of interest as one means of evaluating the detailed mechanism whereby an alkene is converted to an alkane because of certain characteristics. First, the reaction can easily be followed kinetically (Romeyn & Wright, 1947; Connor & Wright, 1946). Secondly, if the alkane (I) is capable of geometric isomerism (either R^1, R^3 or R^2, R^4 are different), then each geoisomer reacts to give a specific uncontaminated diastereoisomer, and the *cis* geoisomer (R^2, R^4 = hydrogen) always reacts more rapidly than the *trans* form (Wright, 1935; Brown & Wright, 1940). Finally, methoxymercuration is accelerated by small amounts of electron acceptors and is inhibited by traces of electron donors (Birks & Wright, 1940). In spite of these favorable diagnostic features, the application of methoxymercuration to studies of reaction mechanism has been limited by the fact that the spatial configuration of the methoxymercurials (III) cannot reliably be dmonstrated by chemical means.

Romeyn & Wright (1947) prepared the two possible diastereoisomers of 1-chloromercuri-2-methoxycyclohexane. Since cyclohexane can exist in but one geoisomeric form, only one of these diastereoisomers (α or *dd*, *ll*) can be found when the cycloalkene is treated with mercuric acetate in methanol and the product converted by aqueous sodium chloride from the acetoxymercurial to the chloromercurial. The β (*dl*, *ld*) diastereomer forms the greater part of a mixture produced when the α isomer is subjected to equilibration conditions by introduction of catalysts which produce free radicals. Since the α isomer was less stable and more easily soluble in alkali than the β form, Romeyn & Wright considered that the added groups were closer together in the α than in the β form. This opinion, which specifies that the addition does not occur with a single Walden inversion, will be shown in the present report to be in error.

The positions of the methoxy and mercuri groups in these two diastereomers are fixed relative to the cyclohexane ring, since free rotation cannot occur owing to the cyclic structure. Anionic exchange of acetoxy by chloro, bromo and iodo ions can be accomplished simply and without configurational change to yield well-crystallized compounds. An attempt has been made, therefore, to determine the spatial configurations of these methoxyhalomercurials by X-ray diffraction methods.

The three α (*dd*, *ll*) halomercurials of methoxycyclohexane were found to exist as an isostructural crystalline series, the unit cell of which was triclinic. In consequence rotation and Weissenberg moving-film photographs had to be taken about each of the three