The Crystal Structure of Glycyl-L-Tryptophan Dihydrate*

By R. A. Pasternak

Gates and Crellin Laboratories of Chemistry, California Institute of Technology, Pasadena, California, U.S.A.

Glycyl-L-tryptophan dihydrate crystallizes in the monoclinic space group $P2_1$; $a_0 = 5.86$, $b_0 = 8.26$, $c_0 = 14.86$ Å, $\beta = 92.6°$. A trial structure was derived by the vector convergence method, and the atomic parameters were refined by three-dimensional Fourier and least-squares methods. The configuration and dimensions of the dipeptide chain are in close agreement with those found in other dipeptides. The plane of the peptide link and the plane of the carboxyl group are nearly perpendicular to each other, as in the molecules of glycylasparagine and glycyltyrosine. The structure is held together by a three-dimensional network of hydrogen bonds.

Introduction

In recent years, accurate information concerning the structure of many important amino acids and of the peptide group have been collected in this and other laboratories. The determination of the structure of the dipeptide glycyltryptophan has been undertaken in order to provide additional information concerning the dimensions and configuration of the peptide amide group and to obtain data for the important amino acid tryptophan.†

Experimental

Unit cell and space group

Glycyl-L-tryptophan dihydrate was obtained in the form of plate-like crystals from water–methylcellosolve solutions. On heating to about 100° C. the crystals disintegrated into a powder, indicating the presence of water of crystallization. Laue and Weissenberg photographs showed the crystals to be monoclinic with the unique b axis parallel to the crystal plate; the main face of the plate was chosen as (001). The unit-cell dimensions were determined from zero-layer Weissenberg photographs taken about all three crystallographic axes; the resulting values are

$$a_0 = 5.86\pm0.02, \quad b_0 = 8.26\pm0.03, \quad c = 14.86\pm0.05 \text{ Å},$$
$$\beta = 96.2\pm0.2° .$$

The density measured by flotation is 1.382 ± 0.007 g.cm.$^{-3}$. Assuming that the molecular unit is a dihydrate, the unit cell contains two molecules (calculated 1.99). A microanalytical water determination gave the

value 11.5%, which agrees fairly well with the value 12.1% calculated for the dihydrate.

The only systematic absences observed were the $(0k0)$ reflections with k odd, the range of observation in k being 0–10; the space group $C_2^2-P2_1$ was indicated, the centrosymmetric space group $C_{2h}^2-P2_1/m$ being excluded by the asymmetry of the molecule.

Intensity data

Although the crystal plates did not show pronounced cleavage, fragments of suitable dimensions (0.2–0.4 mm.) were obtained by cutting the plates with a razor blade. Samples were mounted along the three crystal axes, and equi-inclination Weissenberg photographs were taken with Cu $K\alpha$ radiation for all layer lines with inclination angles smaller than 30°. Over 1600 reflections, practically all those accessible with Cu $K\alpha$ radiation, were recorded.

The intensities were estimated visually from multiple films by comparison with intensity strips. Independent estimates were made by two observers; in general, they agreed within 10%. The intensities were corrected for Lorentz and polarization factors; no absorption corrections were applied. In the final selection of the intensity data equal weight was given to the photographs around the short a and b axes; data obtained from the c-axis photographs were used only for reflections outside the range of the two other sets of photographs. Finally, approximate temperature and scale factors were calculated by Wilson's method.

Trial structure

The asymmetric unit of the glycyltryptophan dihydrate structure

contains 21 atoms (excluding hydrogen). The 9 atoms of the ring system and the carbon atom attached to it might be expected to be coplanar, but the complexity of the unit remained so great that the conventional trial-and-error methods based on structure factors or on simple Patterson vector relationships did not look promising for deriving a trial structure. A three-dimensional Patterson function, sharpened and modified as described by Donohue & Trueblood (1952), was therefore prepared and an attempt was made to solve the structure by the vector convergence method.

The principle of the vector convergence, or superposition, method was discovered 16 years ago by Wrinch (1939). It was reformulated independently by Buerger (1950, 1951), Beevers & Robertson (1950), Clastre & Gay (1950), Garrido (1950), and McLachlan (1951), and various techniques for applying it to vector diagrams of real crystals were proposed by these and other authors (see Donohue & Trueblood, 1952; Shoemaker, Barieau, Donohue & Lu, 1953). A short description of the vector convergence method is included in Lipson & Cochran's textbook, *The Determination of Crystal Structures* (1953).

In the structure determination of glycyltryptophan a technique was tried which has some similarity to that used by Robertson & Beevers (1951) in the structure determination of sucrose. The same kind of approach has been recently described by Donohue & Bryden (1955).

For a successful application of the superposition method advance knowledge of the location of some atoms is in general required; these known atomic positions are used as the origins for the vector maps to be superposed. In most of the previous applications efforts were directed at finding the position of atoms in relation to the symmetry elements of the structure, in particular by making use of Harker relationships (Beevers & Robertson, 1950; Donohue & Trueblood, 1952; Pasternak, Katz & Corey, 1954). In this investigation, however, the symmetry of the structure was disregarded and no attempt was made to derive absolute atomic positions; instead, all the efforts were aimed at finding the *orientation* of a known group of atoms by studying the arrangement of peaks near the origin of the Patterson function.

Obviously the benzene ring was the group to look for, and the interactions generated by its atoms were recognized in the Patterson diagram almost at first glance. In Fig. 1 the Patterson density in the plane through the origin which contains the benzene interactions is plotted. u' and w' are lines of intersection of this plane with the planes $w = 0$ and $v = 0$, respectively; they form angles of $31°$ and $-6°$ with the u and w axes, respectively. The peak at the origin has been eliminated. Contours are drawn at arbitrary intervals, the first contour being at zero; negative densities are an artifact caused by the modification of the Patterson coefficients. The vectors which were ascribed to the benzene ring are marked, interactions

of double weight by black dots and those of single weight by open circles; it is seen that all interactions of double weight fall near density maxima.
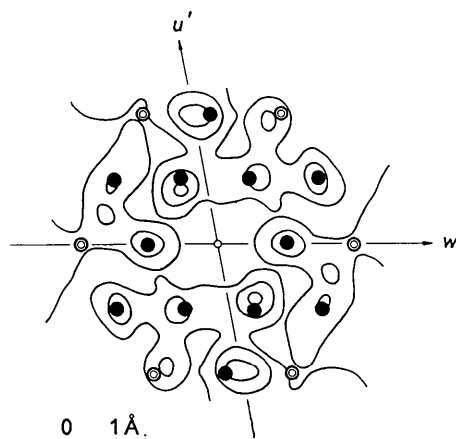

Fig. 1. Patterson section through the origin, parallel to the plane of the benzene ring. u' and w' are lines of intersection between this plane and the planes $w = 0$ and $v = 0$, respectively. Interactions originating from the atoms of the benzene ring are marked, those of double weight by black dots, those of single weight by double circles. Contours at arbitrary levels, first contour at 0.

The relative atomic positions defined by this arrangement of the benzene ring were used as centers for the sixfold superposition of the Patterson function; the technique used for constructing the overlaps has been described elsewhere (Donohue & Bryden, 1955). Since the benzene ring has a center of symmetry, the resulting set of overlaps also had a center of symmetry, coinciding with the center of the benzene ring. To eliminate, at least partly, the unwanted phantom set of overlaps, a seventh superposition was made using as origin the center of a good overlap apparently belonging to an atom of the pyrrole ring. In Fig. 2
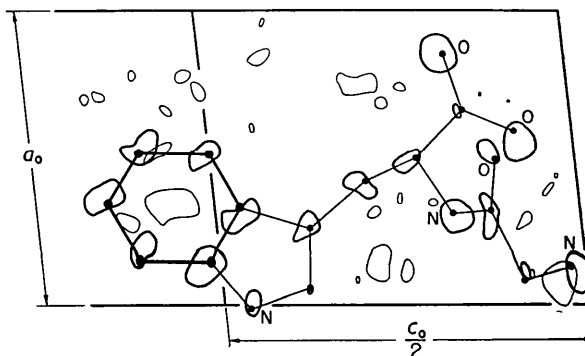

Fig. 2. A projection along y of all sevenfold overlaps between $y = 0$ and $y = \frac{1}{2}$. Those overlaps which define the trial structure shown in the drawing are marked with heavy contours. Unlabelled atoms are carbon atoms.

all the sevenfold overlaps obtained between $y = 0$ and $y = \frac{1}{2}$ are projected on the x, z plane; the origin

is chosen in an arbitrary way. Each closed contour represents the maximum spread of that overlap in a plane perpendicular to y; the overlaps extend, in general, by about the same amount along the y direction as along x or z.

The interpretation of the overlaps proved to be relatively easy, partly because the configuration of separate groups in the molecule—the indole ring system, the peptide amide link and the carboxyl group—was known approximately beforehand. The framework of the molecule was derived stepwise by selecting as probable atomic positions the centers of those overlaps which seemed to fit structurally. The correctness of each choice was tested and the position of the atom adjusted by checking systematically all interactions between this new atom and the atoms which had been located previously with some certainty. In this way reasonable positions for all the atoms of one molecule were found (except for the oxygen atoms of the two water molecules); as shown in Fig. 2, they lay all on sevenfold overlaps. However, some prominent overlaps are not utilized; they belong to the phantom overlaps which were not completely eliminated by the seventh superposition.

The known crystal symmetry was now taken into account for determining the position of the screw axis. There should be a second set of overlaps related by the screw axis to the set representing the first molecule, but it was not unexpected that this second set of overlaps was less complete than the first. Nevertheless, two prominent overlaps were recognized to belong to the oxygen atoms of the carboxyl group, and when the molecule was put into the unit cell in this way, most of the other atoms fell on or near overlaps. Finally, overlaps representing the two water molecules of the structure were found. The atomic parameters for all atoms of a trial structure were thus obtained; the structure factors for the $(h0l)$ zone were calculated, and the discrepancy factor of 0·32 left little doubt about the essential correctness of the trial structure.

The determination of the complete trial structure took approximately five weeks, starting with the three-dimensional Patterson plot. Undoubtedly the presence of a benzene ring in the structure enhanced the effectiveness of the technique, because the relative positions of the atoms in the benzene ring were found quite easily and with a fair degree of accuracy; a comparison with the final coordinates of the atoms revealed misplacements of only 0·08, 0·16, and 0·25 Å in respect to the center of the benzene ring. Undoubtedly, however, the same technique might be used to advantage whenever a structure contains a group of atoms of known configuration and preferably of a symmetry generating coinciding vectors. In the recent structure determination of leucyl-propyl-glycine (Leung & Marsh, 1955) this technique has contributed greatly to the derivation of the trial structure.

Refinement of the structure

The preliminary two-dimensional refinement of the structure was carried out by means of the centrosymmetric $(h0l)$ and the non-centrosymmetric $(0kl)$ electron-density projections. In the $(h0l)$ projection all but two atoms were fairly well resolved—a somewhat surprising fact considering the length of the axis of projection $(b = 8·26$ Å)—and refinement progressed rapidly. No further shifts of significance were indicated in the third $(h0l)$ projection, and an R factor of 0·22 was obtained.

The $(0kl)$ structure factors were calculated using z parameters taken from the last $(h0l)$ Fourier projection and y parameters from the trial structure; the agreement with the observed structure factors was quite satisfactory $(R = 0·27)$ for this stage of refinement. Unfortunately the $(0kl)$ Fourier projection refined only very slowly, probably because of bad overlaps between five atoms and interference between others; therefore, after two iterations this projection was abandoned with the R factor still 0·25.

In the next step all the $F(hkl)$ structure factors were calculated; they showed a fair agreement with the observed structure factors $(R = 0·25)$. The half-cell electron-density projection between $y = -\frac{1}{4}$ and
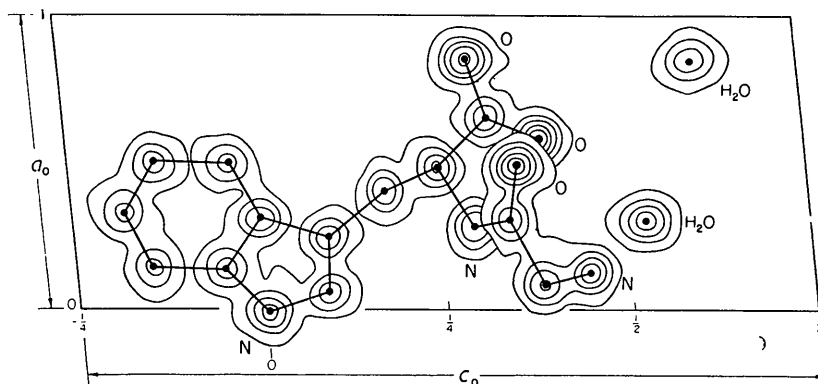

Fig. 3. Half-cell electron-density projection of the structure along y between the limits $y = -\frac{1}{4}$ and $y = +\frac{1}{4}$. Contours at about 1·5 e.Å^{-2} intervals; first contour at 1·5 e.Å^{-2}.

$y = \frac{1}{4}$ was then evaluated. No atomic center lies near these boundaries, and complete resolution of all atoms was obtained. In Fig. 3 the half-cell projection is shown; the drawing has been synthesized so as to represent a complete molecule. Small atomic shifts were indicated and the agreement between observed and calculated structure factors improved somewhat ($R = 0.22$).

It became obvious that only complete three-dimensional treatment of the data would lead to a significant improvement of the results, and three-dimensional refinement was initiated on the Datatron, the high-speed computer of the ElectroData Corporation in Pasadena. A great deal of time was spent in programming the computer for the evaluation of structure factors and for the refinement of atomic coordinates by the method of least squares. In its final form the program* proved to be of considerable effectiveness; a complete cycle of refinement, i.e. the calculation of 1650 structure factors, followed by the determination of the shifts in the 63 positional coordinates, required about five hours of machine time. Four such cycles of refinement were carried out. In the last refinement the average shifts in parameters was about 0.007 Å and the maximum shift was less than 0.02 Å, indicating convergence of the parameters. The R factor, however, was still 19%; furthermore, on calculation of bond lengths and angles, the benzene ring appeared to be badly distorted. It seemed likely that, because of great differences in temperature factors between the tightly hydrogen-bonded polar part and the loosely packed non-polar part of the structure, the least-squares refinement did not converge correctly. In order to check this point and to locate the hydrogen atoms experimentally, a three-dimensional difference Fourier synthesis was calculated. Some shifts in atomic parameters were clearly indicated, the maximum shift being 0.02 Å, and they were incorporated into the atomic coordinates. The final set of parameters is given in Table 1. These shifts, however, did not eliminate the distortion in the benzene ring. It is not thought that this distortion is real, because the electron density around carbon atom C_3, the apparent locus of distortion, showed disturbances not explainable in terms of shifts and anisotropic vibrations; they might arise from accidental accumulations of anisotropic temperature effects of other atoms and falsify the position of C_3.

The carbon atoms of the benzene ring not shared by the pyrrole ring showed fairly large anisotropic temperature vibrations, as did the oxygen atoms of the carboxyl group and, especially pronouncedly, the water oxygen atoms. Although the introduction of anisotropic temperature factors for all atoms probably

---

* The program is designed for the refinement of crystal structures belonging to the non-centrosymmetric point groups 2 and m. Lavine & Rollett (1956) have recently described two programs for the refinement of crystal structures belonging to the point groups $2/m$ and 222, respectively.

Table 1. *Atomic coordinates for one molecule*

	x	y	z
C_1	0.4938	−0.0003	0.9618
C_2	0.4922	0.0197	0.8675
C_3	0.3196	0.0968	0.8169
C_4	0.1406	0.1698	0.8560
C_5	0.1418	0.1522	0.9496
C_6	0.3124	0.0691	0.0031
C_7	0.2548	0.0768	0.0933
C_8	0.0562	0.1574	0.0925
C_9	0.3946	−0.0016	0.1751
C_{10}	0.4774	0.1131	0.2504
C_{11}	0.6498	0.0287	0.3208
C_{12}	0.3018	0.3149	0.3433
C_{13}	0.0744	0.3700	0.3772
N_1	−0.0060	0.2095	0.0014
N_2	0.2828	0.1829	0.2919
N_3	0.1116	0.4936	0.4497
O_1	0.8528	0.0140	0.3018
O_2	0.5786	−0.0431	0.3894
O_3	0.4786	0.3939	0.3603
O_4	0.1602	0.7985	0.3925
O_5	0.6898	0.6197	0.4760

would have improved appreciably the agreement between observed and calculated structure factors, corrections to the structure factors were applied only for the anisotropy of the oxygen atoms O_4 and O_5 of the two water molecules. The temperature anisotropy of these two atoms was determined from the difference Fourier by a method similar to that described by Hamilton (1955). For both atoms, the anisotropy can be expressed approximately by elongated ellipsoids of rotation of identical dimensions but of different orientation. The temperature factor along the minor axes is $B_1 = 3.1$ Å^2, which is identical with the average temperature factor of the structure, and the temperature factor along the major axis is $B_2 = 5.6$ Å^2. For both atoms the major axis lies approximately in a plane parallel to (100); for oxygen atom O_4, however, the angle between the y direction and this axis is

Table 2. *Coordinates of the hydrogen atoms in one molecule*

	x	y	z
$H_1(C_1)$	0.620	−0.061	0.000
$H_2(C_2)$	0.618	−0.028	0.830
$H_3(C_3)$	0.314	0.105	0.747
$H_4(C_4)$	0.000	0.231	0.817
$H_5(C_5)$	−0.040	0.204	0.143
$H_6(N_1)$	−0.146	0.267	0.987
$H_7(C_9)$	0.540	−0.053	0.155
$H_8(C_9)$	0.286	−0.086	0.207
$H_9(C_{10})$	0.550	0.217	0.225
$H_{10}(N_2)$	0.134	0.124	0.280
$H_{11}(C_{13})$	−0.020	0.277	0.397
$H_{12}(C_{13})$	−0.044	0.412	0.327
$H_{13}(N_3)$	0.134	0.605	0.428
$H_{14}(N_3)$	0.220	0.483	0.505
$H_{15}(N_3)$	−0.034	0.517	0.475
$H_{16}(O_4)$	0.066	0.870	0.358
$H_{17}(O_4)$	0.300	0.850	0.383
$H_{18}(O_5)$	0.610	0.550	0.433
$H_{19}(O_5)$	0.634	0.570	0.530

Table 3. *Observed and calculated structure factors*

The calculations are based on McWeeny scattering curves. The *F*'s are multiplied by ten.

0kℓ

k	ℓ	F_o	F_c
1	1	270	290
1	2	254	214
1	3	356	396
1	4	431	501
1	5	83	81
1	6	122	107
1	7	215	222
1	8	161	185
1	9	298	307
1	10	28	21
1	11	74	74
1	12	103	131
1	13	136	128
1	14	88	97
1	15	32	36
1	16		24
1	17		10
1	18		3
2	0	453	682
2	1	447	559
2	2	84	106
2	3	309	355
2	4	437	498
2	5	342	360
2	6	91	87
2	7	233	207
2	8	161	168
2	9	40	49
2	10	72	57
2	11	21	15
2	12	84	81
2	13	32	29
2	14	91	74
2	15	30	35
2	16	48	37
2	17	28	21
2	18	18	22
3	1	215	193
3	2	199	196
3	3	47	45
3	4	209	216
3	5	312	268
3	6	149	107
3	7	170	187
3	8	90	63
3	9	30	18
3	10	72	50
3	11	30	38
3	12	70	67
3	13	30	18
3	14	51	67
3	15	26	9
3	16		25
3	17		23
3	18	26	27
4	0	26	11
4	1	65	93
4	2	39	23
4	3	64	44
4	4	63	32
4	5	24	39
4	6	101	78
4	7	188	208
4	8	113	112
4	9	88	65
4	10	53	54
4	11	60	66
4	12	77	70
4	13	34	38
4	14	37	38
4	15	18	20
4	16	10	17
5	1	149	161
5	2	195	208
5	3	65	46
5	4	18	33
5	5	163	138
5	6	106	89
5	7	146	142
5	8	57	42
5	9	107	96
5	10	102	128
5	11	72	94
5	12	32	37
5	13	24	13
5	14	30	28
5	15		9
5	16	36	48
6	0	208	152
6	1	80	88
6	2	24	45
6	3	109	108
6	4	47	37
6	5	81	93
6	6	70	102
6	7	90	108
6	8	92	91
6	9	69	70
6	10	75	53
6	11	86	65
6	12	75	64
6	13	42	23
6	14	60	69
6	15	15	21
7	1	21	13
7	2	21	24
7	3	141	150
7	4	57	51
7	5	61	64
7	6	69	85
7	7	92	121
7	8	45	46
7	9	110	103
7	10	15	5
7	11	51	37
7	12	21	28
7	13	40	37
7	14	18	
8	0	48	45
8	1	36	28
8	2	34	31
8	3	51	56
8	4	40	64
8	5	77	71
8	6	32	7
8	7	21	24
8	8	43	33
8	9	34	21
8	10	65	64
8	11	47	45
8	12	34	49
9	1		9
9	2	32	37
9	3	53	60
9	4	18	19
9	5	21	15
9	6		11
9	7		3
9	8	24	26
10	0	55	59
10	1	24	20
10	2	48	23
10	3		3
10	4	15	29
10	5	15	22
10	6	15	35

h0ℓ

h	ℓ	F_o	F_c
0	2	172	176
0	3	403	441
0	4	268	-234
0	5	123	114
0	6	278	286
0	7	134	-97
0	8	120	130
0	9	57	97
0	10	72	94
0	11	121	105
0	12		-42
0	13	78	104
0	14	37	-40
0	15	26	-6
0	16	50	34
0	17		15
0	18		-6
1	1	202	-181
1	2	47	-38
1	3	381	449
1	4	112	-92
1	5	18	13
1	6	80	-61
1	7	149	-173
1	8	287	313
1	9	15	-31
1	10	106	92
1	11	116	131
1	12	46	65
1	13	135	-122
1	14	61	81
1	15	24	-4
1	16	57	-38
1	17		26
1	18	18	13
2	0	131	121
2	1	160	144
2	2	343	371
2	3	227	-214
2	4	148	111
2	5	158	136
2	6	137	-107
2	7	109	118
2	8	258	243
2	9	18	18
2	10	53	-34
2	11	53	-64
2	12	24	
2	13	30	-56
2	14	63	-18
2	15	55	37
2	16	24	-22
2	17	21	-7
3	0	70	-52
3	1	130	90
3	2	53	54
3	3	55	-60
3	4	322	318
3	5	18	-29
3	6	163	144
3	7	18	-45
3	8	28	45
3	9	65	46
3	10	165	174
3	11	135	136
3	12	57	56
3	13		3
3	14		-10
3	15	57	55
3	16		11
4	0		-18
4	1	53	73
4	2	67	36
4	3		-14
4	4	72	-67
4	5	18	-14
4	6	32	
4	7	24	16
4	8	100	70
4	9	74	66
4	10	24	-18
4	11		20
4	12	46	23
4	13		4
4	14		3
5	0	163	-166
5	1	97	-78
5	2	32	-38
5	3	21	9
5	4	142	-113
5	5	72	65
5	6	52	13
5	7	18	-16
5	8	24	23
5	9	18	-5
5	10	62	64
5	11		5
5	12	45	32
6	0	92	86
6	1	51	32
6	2	84	-68
6	3	32	19
6	4	86	65
6	5		16
6	6	55	61
6	7	30	12
6	8	21	-8
6	9	36	22
7	0	51	38
7	1		-5
7	2	39	41
7	3	34	15
7	4		-9
-1	1	143	139
-1	2	169	131
-1	3	210	218
-1	4	114	79
-1	5	148	-108
-1	6	55	-18
-1	7	67	-74
-1	8	199	-134
-1	9	65	33
-1	10	45	62
-1	11	18	9
-1	12	120	111
-1	13	70	85
-1	14	99	108
-1	15	48	32
-1	16	34	7
-1	17	60	-57
-1	18	55	48
-2	1	143	151
-2	2	238	-251
-2	3	347	418
-2	4	373	-409
-2	5	341	337
-2	6	15	35
-2	7	180	-172
-2	8	120	103
-2	9	34	15
-2	10	144	-169
-2	11	72	96
-2	12	43	-44
-2	13	61	59
-2	14	21	16
-2	15	60	-60
-2	16	110	108
-2	17	46	-35
-2	18	46	-33
-3	1	21	-11
-3	2	101	94
-3	3		-31
-3	4	43	6
-3	5	67	47
-3	6	221	221
-3	7	24	32
-3	8	125	-128
-3	9	167	173
-3	10	90	78
-3	11	46	86
-3	12	157	196
-3	13	28	34
-3	14	24	29
-3	15		8
-3	16		-15
-3	17	50	52
-4	1	185	167
-4	2		-5
-4	3	108	108
-4	4	162	172
-4	5	47	-42
-4	6	30	-12
-4	7	95	-92
-4	8	177	152
-4	9		42
-4	10		24
-4	11	83	80
-4	12		29
-4	13	110	-95
-4	14	88	82
-4	15	15	12
-5	1	147	138
-5	2	24	32
-5	3	57	-16
-5	4	99	99
-5	5	57	67
-5	6	72	-83
-5	7	36	54
-5	8	53	-40
-5	9	69	-43
-5	10	53	53
-5	11	21	-21
-5	12	30	44
-5	13	24	23
-5	14		11
-5	15		18
-6	1		29
-6	2	74	97
-6	3	96	76
-6	4	46	25
-6	5		7
-6	6	88	76
-6	7	36	52
-6	8	32	47
-6	9		8
-6	10	55	-38
-6	11		-15
-6	12		-23
-7	1		-12
-7	2	24	-27
-7	3	21	9
-7	4	43	44
-7	5	24	-5
-7	6	58	50
-7	7		-16

60° and for O_5 130°. The direction of maximum vibration for O_4 is about perpendicular to the directions of the three hydrogen bonds in which it takes part, and for O_5 it is approximately perpendicular to two strong hydrogen bonds and forms an angle of about 30° with the third weak hydrogen bond.

Finally, the contributions of the hydrogen atoms to the structure factors were introduced. The coordinates used for the hydrogen atoms are given in Table 2. Bond distance of approximately 1·05 Å for the C–H, of 1·00 Å for the N–H, and of 0·95 Å for the O–H bonds were taken. In the case of the amino group and the water molecules, the N–H and O–H bonds were pointed as nearly as possible along the directions of the hydrogen bonds. The positions of most of the hydrogen atoms were also indicated by the presence of diffuse electron-density peaks in the difference Fourier.

A list of the observed and calculated $(0kl)$ and $(h0l)$ structure factors is given in Table 3.* The few very strong reflections are observed weaker than calculated, probably because of extinction effects, but otherwise the overall agreement is satisfactory, and the reliability factor for all reflections, excluding those too weak to be observed, is 0·165. Characteristically, the agreement for the centrosymmetric $(h0l)$ data $(R = 0·173)$ is worse than for the non-centrosymmetric $(0kl)$ data $(R = 0·155)$. The average standard error calculated on the basis of the least-squares results is 0·014 Å in a bond length and 1·0° in an angle of about 110°.

Discussion of the structure

Configuration of the molecule

The intramolecular bond lengths and bond angles are listed in Table 4; they are also shown in Fig. 4, which gives schematic views of the molecule along the b axis. One might somewhat arbitrarily divide the molecule into three parts: the indole ring with its β-carbon atom, the peptide group with the amino nitrogen atom, and the carboxyl group with its α-carbon atom.

Indole ring.—The atoms of the indole ring system with the methylene carbon attached to it were found to be coplanar. The median plane through all these atoms was calculated by least-squares methods; the average deviation of the atoms from this plane is 0·010 Å, and the maximum deviations are +0·022 and −0·027 Å, for the atoms C_8 and N_1, respectively. Carbon atom C_3 was not included in the calculation

of the median plane, because its coordinates, as has been discussed in the previous section, were considered less reliable. It lies apparently 0·047 Å out of the plane.

The bonds within the benzene ring, excluding the ones involving carbon atom C_3, have a length of 1·39–

Table 4. *Intramolecular bond distances and bond angles*

C_1–C_2	1·411 Å	C_6–C_1–C_2	117·4°
C_2–C_3	1·353	C_1–C_2–C_3	122·1
C_3–C_4	1·390	C_2–C_3–C_4	121·5
C_4–C_5	1·399	C_3–C_4–C_5	116·4
C_5–C_6	1·388	C_4–C_5–C_6	123·1
C_6–C_1	1·405	C_5–C_6–C_1	118·8
C_6–C_7	1·418	C_5–C_6–C_7	106·9
C_7–C_8	1·338	C_6–C_7–C_8	108·0
C_8–N_1	1·430	C_7–C_8–N_1	106·9
N_1–C_5	1·307	C_8–N_1–C_5	109·3
		N_1–C_5–C_6	108·8
C_7–C_9	1·534	C_4–C_5–N_1	127·9
C_9–C_{10}	1·506	C_1–C_6–C_7	134·3
C_{10}–C_{11}	1·540		
C_{12}–C_{13}	1·541	C_6–C_7–C_9	124·7
		C_8–C_7–C_9	127·3
C_{10}–N_2	1·470	C_7–C_9–C_{10}	115·1
N_2–C_{12}	1·329	C_9–C_{10}–C_{11}	110·8
C_{13}–N_3	1·483	C_9–C_{10}–N_2	110·8
		N_2–C_{10}–C_{11}	111·9
C_{11}–O_1	1·256	C_{10}–C_{11}–O_1	117·3
C_{11}–O_2	1·286	C_{10}–C_{11}–O_2	120·2
C_{12}–O_3	1·226	O_1–C_{11}–O_2	122·0
		C_{10}–N_2–C_{12}	122·8
		N_2–C_{12}–O_3	124·8
		N_2–C_{12}–C_{13}	114·1
		C_{13}–C_{12}–O_3	121·0
		C_{12}–C_{13}–N_3	112·1
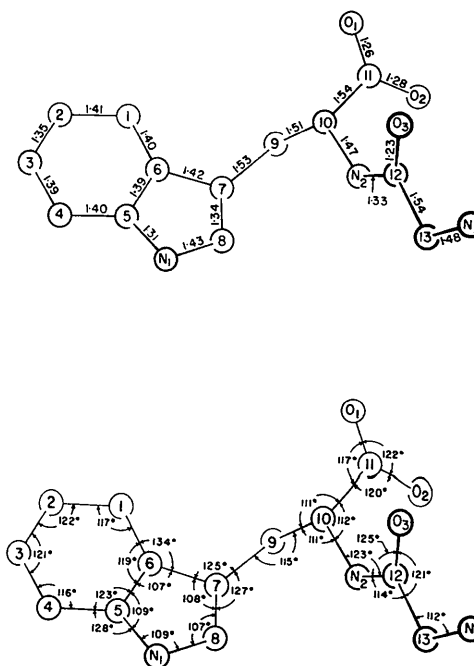
Fig. 4. Schematic drawings of the molecule viewed along the b axis of the crystal showing bond lengths (Ångström units) and angles. Atoms labelled with numbers only are carbon atoms.

1·41 Å; it is doubtful if they are significantly longer than in the isolated benzene ring. The bond angles in the ring do not deviate significantly from 120°. In the pyrrole ring the bond angles and bond lengths are approximately the same as reported for pyrrole (Schomaker & Pauling, 1939) except for the bond $C_5–N_1$, which was found to be unexpectedly short (1·31 Å). In view of this surprising result, an accurate crystal-structure determination of indole or β-methyl indole would be of great interest.

Peptide group.—The configuration and dimensions of the peptide group were found to be in excellent agreement with those observed in accurate structure determinations of other dipeptides (see review by Pauling & Corey, 1954). In particular, the five atoms associated with the group are again coplanar, the maximum deviation from the median plane being 0·020 Å and the average deviation 0·010 Å. Carbon atom C_9 is located near the plane, as is amino nitrogen atom N_3, the $C_{13}–N_3$ bond forming an angle of only 15° with the plane. It is interesting to notice that in all dipeptides which have been investigated so far the $C–NH_3$ bond forms a small angle with the plane of the peptide group and is approximately *cis* to the $C=O$ bond (α- and β-glycylglycine, diglycylcystine, glycyltyrosine (Smits & Wiebenga, 1953), glycylasparagine). Apparently the oxygen and the nitrogen atoms attract each other by electrostatic forces. This attraction is also indicated by the configuration around the carboxyl end of most amino acids and of peptides; the amino nitrogen and the peptide amide nitrogen atoms, respectively, are near the plane of the carboxyl group, and approach closely to one oxygen atom.

Carboxyl group.—The planarity of the carboxyl group with its α-carbon atom has been established in many investigations; in glycyltryptophan the carboxyl group was also found to be very nearly planar. The average deviation of an atom from the median plane is 0·025 Å; the maximum deviation of 0·05 Å shown by the central atom C_{11} seems to be significant.
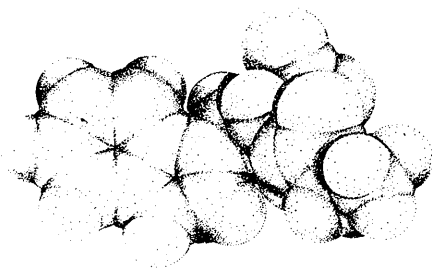
Fig. 5. Packing drawing of one molecule; view along *b* axis.

The two C–O bond lengths and the two C–C–O angles are identical within the accuracy of this investigation. The average length of the C–O bonds (1·27 Å) is somewhat longer than usually found; it might be explained by the fact that each oxygen atom takes part in two good hydrogen bonds. In addition, oxygen

atom O_2 participates in a third weak hydrogen bond (see next section).

The configuration of the molecule as a whole can be best described in reference to the three planes defined by the indole ring, the peptide, and the carboxyl groups, respectively; they are easily recognized in the scale drawing of the molecule, Fig. 5. The plane of the indole ring forms a dihedral angle of 73° with the plane of the peptide group and an angle of 30° with that of the carboxyl group. The planes of the peptide and of the carboxyl groups, which have atom C_{10} in common, are nearly at right angles (88°) and intersect approximately along the $C_{10}–N_2$ bond. Virtually the same spatial relationship between carboxyl and peptide groups has been found in glycyltyrosine (Smits & Wiebenga, 1953) and glycylasparagine.

The environment of the molecule

In Fig. 6 views of the structure along the *b* and *a* axes are given.

The molecules are arranged in the crystal as is characteristic for a structure containing polar and non-polar groups; a separation into layers of non-polar groups and layers of polar groups occurs. These layers, which are parallel to the *a, b* plane, can be easily recognized in Fig. 6. A closely similar packing is encountered in the crystal of glycyltyrosine (Smits & Wiebenga, 1953).

The indole ring systems are approximately parallel to the *z* axis and form a dihedral angle of 32° with the *a, c* plane. The centers of the indole rings which are related by screw axes form approximately a face-centered array. The weakly basic nitrogen atom of the pyrrole ring does not take part in hydrogen-bond formation, and the layer is held together by van der Waals forces only. This arrangement of the indole groups is very similar to that found in the crystal structure of naphthalene (Abrahams, Robertson & White, 1949), the naphthalene molecules forming puckered layers parallel to the *a, b* plane also.

The polar part of the molecules consists of the planar peptide group with the amino nitrogen atom and of the

Table 5. *Hydrogen bond lengths and angles*

From atom y in molecule M	to atom y	in molecule	Distance N $\cdots$ O or O $\cdots$ O (Å)	Angle O $\cdots$ N–C or O $\cdots$ O $\cdots$ O (°)
N_2	O_1	M'	2·89	123·4* 112·6†
N_3	O_2	A	2·86	124·4
N_3	O_4	M	2·68	115·4
N_3	O_5	M'	2·75	113·6
O_4	O_1	M'_1	2·77 ⎱	102·6
O_4	O_2	M_1	2·78 ⎰	
O_5	O_2	A	2·99 ⎱	83·1
O_5	O_3	M	2·74 ⎰	

* Angle $O_1 \cdots N_2–C_{10}$.   † Angle $O_1 \cdots N_2–C_{12}$.
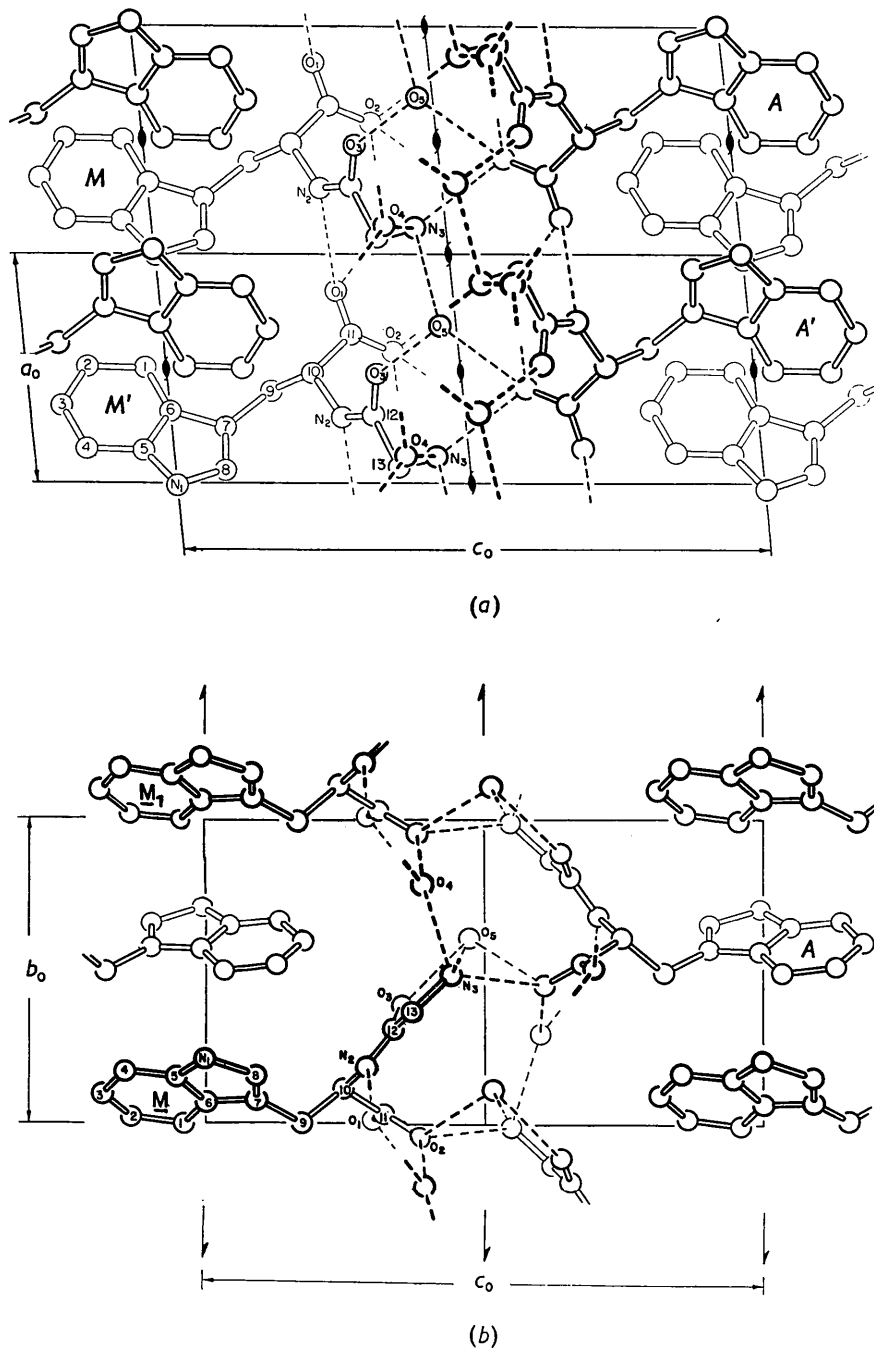
(a)

(b)

Fig. 6. (a) A projection of the structure along the b axis of the crystal. Hydrogen atoms are omitted.

(b) A projection of the structure along the a axis of the crystal.

planar carboxyl group; the planes of both are approximately parallel to a, as shown in Fig. 6(b). In the two sets of molecules the polar ends protrude on opposite sides of the interlocking indole rings, and tie in with the polar ends extending from the next non-polar layer. The structure is held together by a network of hydrogen bonds, the water molecules being of major importance for the coherence of the structure. All the

hydrogen bonds feasible are formed; the hydrogen-bond distances and angles are given in Table 5. The two equivalent molecules are labelled M and A; molecules displaced by the unit vector b are labelled M_1 and A_1, while those displaced by $-a$, are labelled M' and A'.

Of the eight hydrogen bonds only two do not involve the water molecules. The peptide nitrogen atom

is hydrogen bonded to carboxyl oxygen atom O_1 in the adjacent molecule M'. The oxygen atom is located near the plane of the peptide link, the N $\cdots$ O vector pointing only 11° out of the plane; furthermore, the angles formed by this vector with the C–N bonds are near 120°. Thus, the direction of the N $\cdots$ O vector makes it very likely that the N–H bond is also coplanar with the peptide link.

The amino nitrogen atom undoubtedly is present as the charged NH_3^+ group. Three hydrogen bonds are formed by it, one with carboxyl oxygen atom O_2 in molecule A, which is of main importance in tying the structure together along the c direction, and two with water oxygen atoms; one of these hydrogen bonds is unusually short (2·68 Å). The hydrogen bonds and the C–N bond are arranged approximately tetrahedrally around the nitrogen atom.

The water molecules use both of their hydrogen atoms for the formation of hydrogen bonds. Water oxygen atom O_4 acts as a bridge along a, joining the carboxyl oxygen atoms O_1 and O_2 of adjacent molecules M and M'. Of the two hydrogen bonds involving the hydrogen atoms of water oxygen atom O_5, one is rather long and probably contributes little to the coherence of the structure.

No unusually short contacts between non-bonded atoms are found in the structure; the packing as a whole is fairly loose, especially at the junction between polar and non-polar layers.

References

ABRAHAMS, S. C., ROBERTSON, J. M. & WHITE, J. G. (1949). *Acta Cryst.* **2**, 238.

BEEVERS, C. A. & ROBERTSON, J. H. (1950). *Acta Cryst.* **3**, 164.

BEEVERS, C. A., McDONALD, T. R. R., ROBERTSON, J. H. & STERN, F. (1952). *Acta Cryst.* **5**, 689.

BUERGER, M. J. (1950). *Acta Cryst.* **3**, 87.

BUERGER, M. J. (1951). *Acta Cryst.* **4**, 531.

CLASTRE, J. & GAY, R. (1950). *C. R. Acad. Sci., Paris*, **230**, 1876.

DONOHUE, J. & TRUEBLOOD, K. N. (1952). *Acta Cryst.* **5**, 414.

DONOHUE, J. & BRYDEN, J. H. (1955). *Acta Cryst.* **8**, 314.

GARRIDO, J. (1950). *C. R. Acad., Sci., Paris*, **230**, 1878; **231**, 297.

HAMILTON, W. C. (1955). *Acta Cryst.* **8**, 199.

LAVINE, L. R. & ROLLETT, J. S. (1956). *Acta Cryst.* **9**, 269.

LEUNG, Y. C. & MARSH, R. E. (1955). Private communication.

LIPSON, H. & COCHRAN, W. (1953). *The Determination of Crystal Structures*, pp. 163–169, 186–189. London: Bell.

McLACHLAN, D., JR. (1951). *Proc. Nat. Acad. Sci., Wash.* **37**, 115.

PASTERNAK, R. A., KATZ, L. & COREY, R. B. (1954). *Acta Cryst.* **7**, 225.

PAULING, L. & COREY, R. B. (1954). *Fortschr. Chem. org. Naturstoffe*, **11**, 181.

ROBERTSON, J. H. & BEEVERS, C. A. (1951). *Acta Cryst.* **4**, 270.

SCHOMAKER, V. & PAULING, L. (1939). *J. Amer. Chem. Soc.* **61**, 1769.

SHOEMAKER, D. P., BARIEAU, R. E., DONOHUE, J. & LU, CHIA-SI (1953). *Acta Cryst.* **6**, 241.

SMITS, D. W. & WIEBENGA, E. H. (1953). *Acta Cryst.* **6**, 531.

WRINCH, D. M. (1939). *Phil. Mag.* (7), **27**, 98.