

---

# JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

---

(Registered in U. S. Patent Office) (Copyright, 1950, by the American Chemical Society)

VOLUME 72

JUNE 14, 1950

NUMBER 6

---

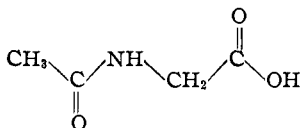
[CONTRIBUTION FROM THE GATES AND CRELLIN LABORATORIES OF CHEMISTRY, CALIFORNIA INSTITUTE OF TECHNOLOGY, No. 1330]

## The Crystal Structure of N-Acetylglycine<sup>1,2</sup>

BY GENE B. CARPENTER<sup>3</sup> AND JERRY DONOHUE

### Introduction

The determination of the crystal structures of individual amino acids and their derivatives is part of a long-range program of investigation of the constitution and configuration of proteins which is being carried out in these Laboratories. The structures of diketopiperazine,<sup>4</sup>  $\alpha$ -glycine,<sup>5</sup> DL-alanine,<sup>6</sup> and  $\beta$ -glycylglycine<sup>7</sup> have already been reported. The study of N-acetylglycine



reported here is a part of this program. This amino acid is of special interest in that it contains, in a simple environment, a peptide bond, which is the link that unites individual amino acids into proteins. Despite the great importance of this structural feature for the study of proteins, little detailed information about its geometry has been obtained. Much more information about peptide bonds in various environments is clearly desirable.

Furthermore, it is highly desirable that more knowledge be gained of the complex forces between peptides which determine the arrangement

of molecules in the crystal. Presumably, the same forces influence the types of folding realized in polypeptide chains in protein molecules. Among these forces are those resulting from hydrogen bond formation, electrostatic forces, van der Waals forces, and steric repulsion depending on the size and shape of the amino acid residues.

When preliminary examination of N-acetyl-glycine indicated that results of considerable accuracy might be obtained, it was decided to subject it to intensive study. This investigation is the first to employ complete three-dimensional data for the determination of a structure containing a peptide bond.

### Unit Cell and Space Group

N-Acetyl-glycine (Eastman white label) was recrystallized from water. Small portions of this purified material were used to grow numerous batches of crystals from aqueous solution. Optical goniometric examination of numerous crystals of acetyl-glycine showed them to be monoclinic with axial ratios  $a:b:c = 0.418:1:1.255$  and monoclinic angle  $\beta = 138^\circ 6'$ .

Preliminary measures of the axial lengths were obtained from layer-line spacings on rotation photographs taken about three axes. By means of these values, it was possible to index spots on oscillation photographs and thus to secure accurate lengths. The dimensions of the unit cell, based on the value  $\lambda = 1.542 \text{ \AA.}$  for Cu K $\alpha$ , are

$$a_0 = 4.86, b_0 = 11.54, c_0 = 14.63 \text{ \AA.}, \beta = 138^\circ 12'$$

with a maximum error of about 0.2%. These lengths lead to the axial ratios  $a:b:c = 0.421:1:1.267$  which are in sufficiently good agreement with the goniometric values.

(1) Presented in part before the Division of Physical and Inorganic Chemistry at the 115th Meeting of the American Chemical Society, San Francisco, California, March 30, 1949.

(2) Aided by a grant from the National Foundation for Infantile Paralysis.

(3) National Research Council Postdoctoral Fellow, 1947-1948. Present address: Department of Chemistry, Brown University, Providence 12, Rhode Island.

(4) R. B. Corey, *THIS JOURNAL*, **60**, 1598 (1938).

(5) G. Albrecht and R. B. Corey, *ibid.*, **61**, 1087 (1939).

(6) H. A. Levy and R. B. Corey, *ibid.*, **63**, 2095 (1941).

(7) E. W. Hughes and W. J. Moore, *ibid.*, **71**, 2618 (1949).

The Laue symmetry of the crystals is  $C_{2h}2/m$  in agreement with their monoclinic aspect. Space group extinctions were ascertained from strongly exposed Weissenberg photographs. Reflections ( $h0l$ ) were absent for  $l$  odd and reflections ( $0k0$ ) were absent for  $k$  odd. This information is characteristic of the space group  $C_{2h}^5-P2_1/c$ , a common one for organic crystals. This space group is centrosymmetric and therefore the structure factors are all real. The general positions are four-fold:  $xyz$ ;  $\bar{x}\bar{y}\bar{z}$ ;  $\bar{x}, \frac{1}{2} + y, \frac{1}{2} - z$ ;  $x, \frac{1}{2} - y, \frac{1}{2} + z$ .

The density of several clear crystals was found to be 1.43 g./cc. There are then four (calculated 4.01) molecules per unit cell. Since the molecule of acetylglycine cannot contain a center of symmetry, the atoms are required to lie in general positions.

Marked cleavage parallel to (100) was observed. In addition, when a thin crystal resting on a (100) face on a microscope slide was crushed by lowering another slide onto it, it broke up into a bundle of fine needles with axes parallel to the  $c$ -axis.

#### Determination of Atomic Positions

**Estimation of Intensities.**—Complete sets of equi-inclination Weissenberg photographs about three axes were produced for intensity estimation. The three crystals used were roughly cylindrical in shape, small enough so that differences in absorption for different planes were not large enough to be significant. Nearly the entire range of reflections which can be observed with Cu  $K\alpha$  radiation was explored.

Intensities were estimated visually by the multiple film technique.<sup>8</sup> Many reflections occur on two or three different film sets, providing independent estimates of their relative intensities. The internal consistency of the data is good. Relative squared structure factor ( $F^2$ ) values were computed from the intensities by applying the usual Lorentz, polarization, and time factors. About 800 non-zero reflections were observed and measured out of a total of 1280 reciprocal lattice points included within the limiting sphere for copper radiation.

**Patterson Projection and Trial Structure.**—The values of  $F^2$  were placed approximately on the absolute scale by assuming that the geometric structure factor of the reflection (100) was 0.9. This assumption is based on the strong indications, which had accumulated during preliminary experiments, that the crystal has a layer structure: (a) The orders of reflection from (100) are exceptionally intense and the intensity decreases through four orders according to a normal decline; therefore, the atoms are concentrated in (100) planes. (b) The cleavage parallel to the (100) plane is good. Laue photographs exhibit non-Bragg brush-like blackening in the center

along a line perpendicular to (100), probably due to relatively large thermal motion perpendicular to the layers. These observations indicate that the molecules do not extend from one layer into another. (c) The perpendicular distance between (100) planes is only 3.24 Å., hence the layers are just one molecule thick. Therefore, the molecules must be nearly flat and cannot be tipped far out of the (100) plane. Consequently, the important features of the structure should be revealed by projections on (100), which can be calculated from ( $0kl$ ) data alone.

Accordingly, Patterson projections on (100) were prepared from the  $F^2$  values in the  $0kl$  zone. The first was prepared in the usual way; it gave little information and showed only slight detail in the region near the origin, illustrated on the left in Fig. 1. The second projection differed from the first in having the peak at the origin removed, the remaining peaks sharpened, and a modification function of the form  $e^{-a(\sin \theta)^2}$  applied to the  $F^2$  values. The sharpening was accomplished by dividing all  $F^2$  values by the square of the average atom form factor corresponding to the same spacing; the nitrogen form factor was taken as the average. The peak at the origin was removed by subtracting the average sharpened  $F^2$  value from the individual sharpened  $F^2$  values. These procedures, other than the application of the modification function, were suggested in Patterson's original paper.<sup>9</sup> The modification function is just an artificial temperature factor to insure convergence of the Fourier series. This second form, illustrated on the right in Fig. 1, gave more detail in the region near the origin. The dots represent the ultimate positions of projected interatomic vectors within one molecule.

The Patterson projection within about 3 Å. of the origin is most easily interpreted because peaks arise here chiefly from interatomic vectors within individual molecules. This region shows a roughly hexagonal pattern of peaks around the origin and about 1.5 Å. from it when the quadrant illustrated in Fig. 1 is repeated to fill in this region. This means that bonds are concentrated along three principal directions in the crystal. Examination of the expected configuration of acetylglycine revealed that there are also three bond directions in the molecule, about 120° apart. Since the bond lengths involved do not differ greatly, one molecule should indeed give rise to a hexagonal pattern of peaks around the origin. Therefore, the orientation of each molecule in the unit cell is restricted to one of three possibilities as a result of the Patterson projection. The orientations of the molecules are related, but it is fortuitous that the patterns from all four molecules in the unit cell nearly coincide.

Paper models of the four molecules were then moved about on a drawing of the  $x = 0$  section of the unit cell so as to try to satisfy the sym-

(8) J. M. Robertson, *J. Sci. Instr.*, **20**, 165 (1943).

(9) A. L. Patterson, *Z. Krist.*, **90**, 517 (1935).

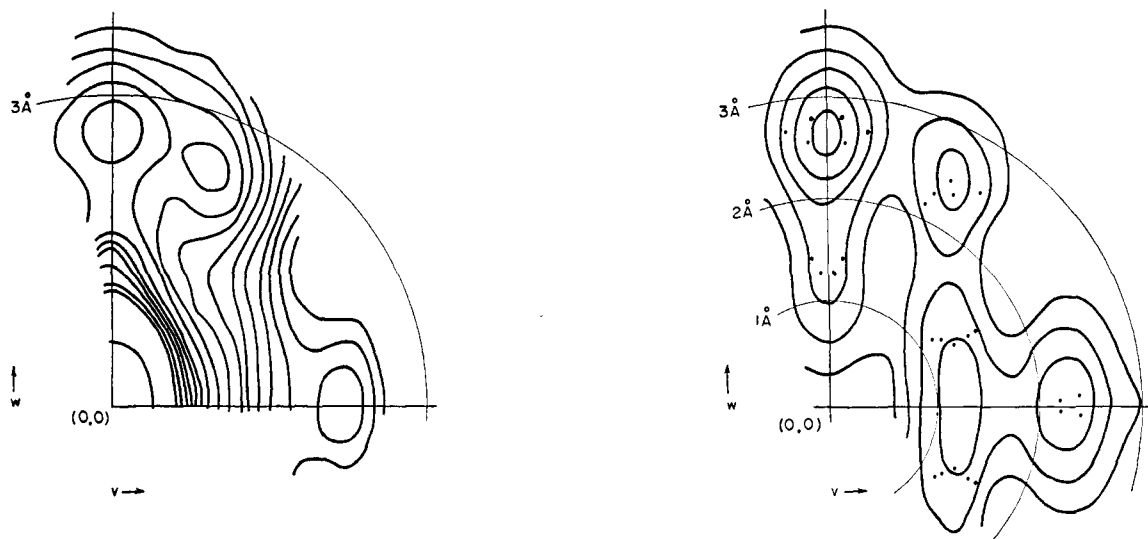


Fig. 1.—Region near the origin in Patterson projections on (100) for acetylglycine: left, unmodified Patterson projection; right, Patterson projection sharpened, with peak at origin removed, and with artificial temperature factor.

metry requirements of the space group while preserving one of the three possible orientations for the molecule chosen to specify the structure. The molecules were kept separated by distances determined by van der Waals atomic radii. It was thus possible to select one arrangement which, in addition, was consistent with the Patterson peaks more distant from the origin. This was taken as a trial structure for further examination.

Structure factors  $F_{0kl}$  were calculated for this trial structure; atom form factors from the International Tables<sup>10</sup> were modified by an estimated temperature factor. The calculated values were in encouraging agreement with the observed magnitudes except at large values of  $\sin \theta$ .

**Fourier Projections.**—A Fourier projection on (100) of the electron density in the unit cell was next calculated with 36 of the observed values of  $|F|_{0kl}$  to which the signs could be assigned with some confidence from the calculated values of  $F_{0kl}$ . The resulting projection exhibited a recognizable if poorly resolved picture of the molecules, from which improved atomic parameters were measured. From the improved parameters new structure factors were calculated; the agreement with observed magnitudes was appreciably better and more signs could be correctly assigned. The result of a series of such refinements is shown in Fig. 2; 104 reflections were used in forming this projection. The molecule is projected on a plane normal to the  $a$ -axis rather than onto (100) in order that the atoms may appear without distortion. A non-convergence correction like that discussed below for the three-dimensional syntheses was applied to the resulting parameters.

(10) "International Tables for the Determination of Crystal Structures", Gebrüder Borntraeger, Berlin, 1935, Vol. II, p. 571.

At this point the observed and calculated ( $0kl$ ) structure factors were in good agreement. The average discrepancy  $\sum_{0kl} \frac{|F|_{\text{obs.}} - |F|_{\text{calcd.}}}{\sum_{0kl} |F|_{\text{obs.}}}$ , often suggested as a criterion of agreement, had a value of 14%. This value is computed from those reflections only which were observed. The final parameters from the ( $0kl$ ) data, together with the

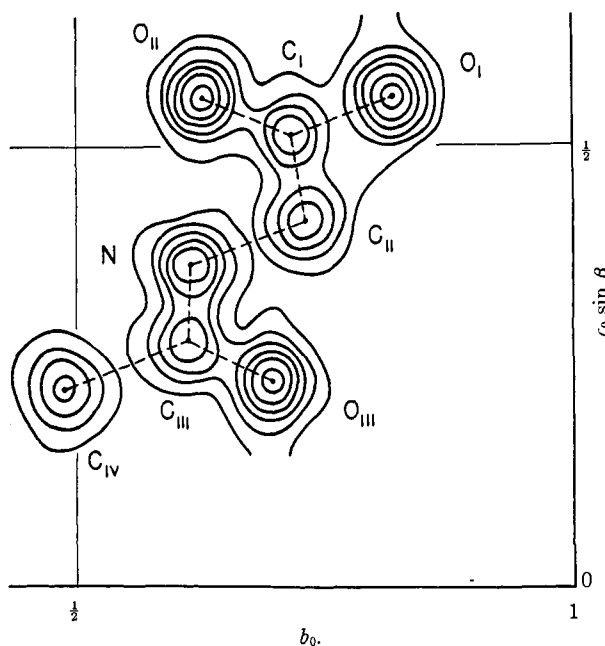


Fig. 2.—Fourier projection along the  $a$ -axis of one molecule of acetylglycine.

parameters for the trial structure with which the refinement began, are compared with the final parameters from the ( $hkl$ ) data in Table II.

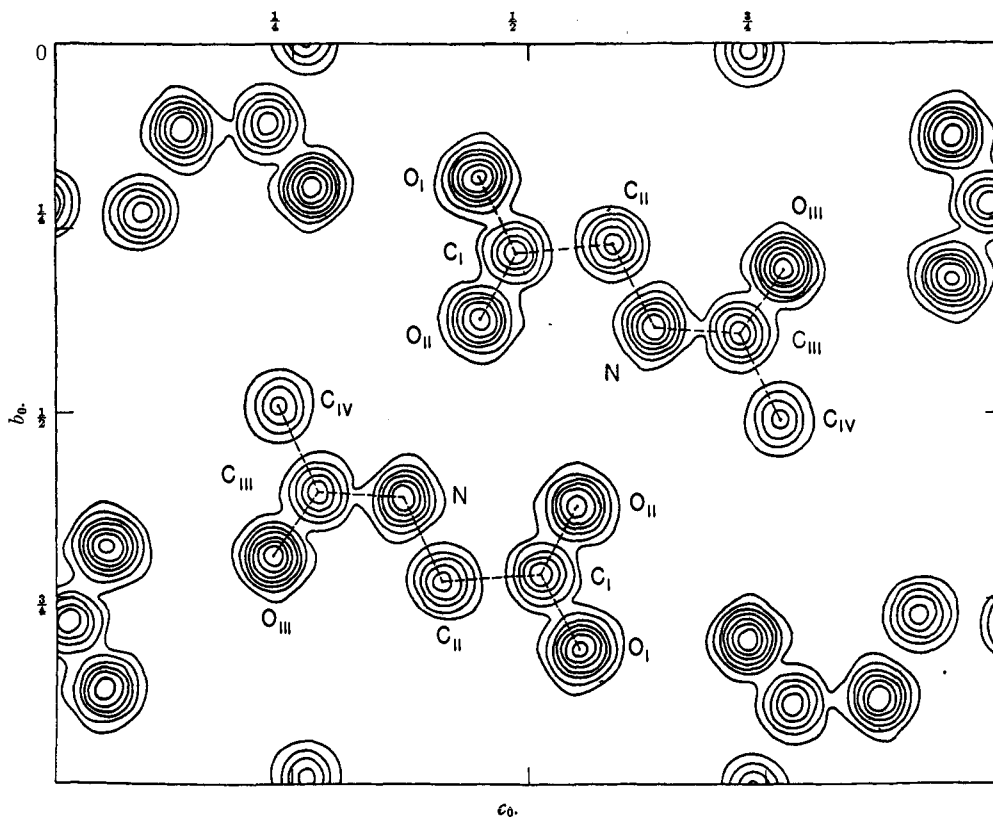


Fig. 3.—Acetylglycine electron density at  $x = 0$  from second three-dimensional Fourier synthesis.

**Plane-parallel Approximation and Three-dimensional Fourier Syntheses.**—The shape of the molecule in projection verified the expectation that the atoms are not far out of the (100) plane. This fact permits the use of a simple approximation for calculating the structure factor  $F_{hkl}$ . Assuming that carbon, nitrogen and oxygen atoms have proportional atom form factors, the structure factor for the space group  $P2_1/c$  is

$$F_{hkl} = 4\hat{f}_{hkl} \sum_i Z_i \cos 2\pi(hx_i + lz_i) \cos 2\pi ky_i$$

for  $k + l$  even

$$= -4\hat{f}_{hkl} \sum_i Z_i \sin 2\pi(hx_i + lz_i) \sin 2\pi ky_i$$

for  $k + l$  odd

where  $\hat{f}_{hkl}$  is the average atom form factor per electron,  $Z_i$  is the number of electrons, and  $x_i$ ,  $y_i$ ,  $z_i$  are the parameters of atom  $i$ . The introduction of the approximation  $x_i = 0$  simplifies these expressions to

$$F_{hkl} \cong 4\hat{f}_{hkl} \sum_i Z_i \cos 2\pi ky_i \cos 2\pi lz_i \text{ for } k + l \text{ even}$$

$$\cong -4\hat{f}_{hkl} \sum_i Z_i \sin 2\pi ky_i \sin 2\pi lz_i \text{ for } k + l \text{ odd}$$

Therefore

$$F_{hkl}/\hat{f}_{hkl} \cong F_{0kl}/\hat{f}_{0kl}$$

To test this approximation, the observed structure factor magnitudes were tabulated with  $h$  varying across each row but  $k$  and  $l$  remaining

fixed. The magnitudes across each row decreased smoothly as  $\sin \theta$  increased, for all but a few of the weaker reflections. Therefore, the approximation was verified and, in particular, the sign of  $F_{0kl}$  was correct also for  $F_{hkl}$  for the stronger reflections at least. As the signs of the former were known from the projection calculations, it was possible to proceed immediately to the computation of a three-dimensional Fourier synthesis; 740 of the 800 observed reflections were assigned signs and were used in this summation.

The  $x$  parameters measured from the first three-dimensional Fourier synthesis indicated that all atoms (excluding hydrogen) were less than 0.2 Å. from the plane (100). The structure was then carried through a series of Fourier refinements. The result of the second three-dimensional Fourier summation for the section  $x = 0$  is shown in Fig. 3. All other parallel sections are very similar to this one. The third three-dimensional synthesis using observed structure factor magnitudes differs only very slightly from the second; it includes all but six small structure factors for which the signs are still uncertain.

**Locating Atomic Centers.**—The Fourier summations were carried out on International Business Machine equipment using punched cards.<sup>11</sup> The results are obtained in the form of

(11) V. Schomaker, unpublished work.

printed values proportional to the electron density at a network of points throughout the cell. For acetylglucine these numbers were entered on layers with constant  $x$  values and lines were sketched through points of equal electron density. The resulting diagrams, such as Fig. 3, give important information about the shape of the atoms and the degree of resolution but do not provide an accurate method for locating atomic centers. To accomplish this, an analytic procedure was developed.

It is assumed that when the atoms are well resolved in the Fourier representation of the electron density, the atoms have a Gaussian shape, at least near the center of the atom. Then the electron density along any line passing near an atomic center is

$$\rho = \rho_0 e^{-k(x-x_0)^2}$$

where  $\rho_0$ ,  $x_0$ , and  $k$  are parameters permitting the equation to be fitted to particular sets of data and  $\rho$  is the electron density at the point  $x$  along the line. Let the density at three adjacent points along the line be  $\rho_1$  at  $x_1$ ,  $\rho_2$  at  $x_2 = x_1 + 1$ , and  $\rho_3$  at  $x_3 = x_1 + 2$ , where the coordinate is measured in tabular units (usually sixtieths of a cell edge). Fitting the equation to these three points and solving for the coordinate  $x_0$  of the maximum electron density yields

$$x_0 = x_1 + 1/2 + \left(1 + \frac{\log \rho_2 - \log \rho_3}{\log \rho_2 - \log \rho_1}\right)^{-1}$$

This equation can be easily modified if it is convenient to choose points separated by more than one tabular unit.

The equation is found to represent the centers of carbon, oxygen and nitrogen atoms surprisingly well. If the three points of highest density along the line are chosen to be fitted, and if the intervals are such that none of them has a density less than about half that at the peak center, then the calculated atomic center is in complete agreement with the result obtained from inspection of large-scale graphs. Furthermore, in the analytic procedure there are fewer opportunities for errors in judgment and it is significantly faster.

In practice, the coordinates of atomic centers were obtained from the electron density values along nine parallel lines through each atom along each axis. Each set of nine centers establishes a nearly plane surface which is a function of three coordinates in general. The three surfaces thus obtained from the sets of lines along the three axes all pass through the atomic center and their intersection determines the desired coordinates. In the case of acetylglucine, the  $y$  coordinates should be independent of the  $x$  and  $z$  coordinates so the nine values for the  $y$  coordinate of the center along the lines parallel to the  $y$ -axis for each atom were simply averaged, with weights proportional to the maximum density along the corresponding line. The  $x$  and  $z$

coordinates are measured along non-perpendicular axes and so should be linearly related. Figure 4 illustrates the steps in the calculation of the  $x$ ,  $z$  coordinates of the center of atom  $C_{IV}$  in the third Fourier synthesis, as an example. The level  $y = -1/60$  is shown. The tabular values

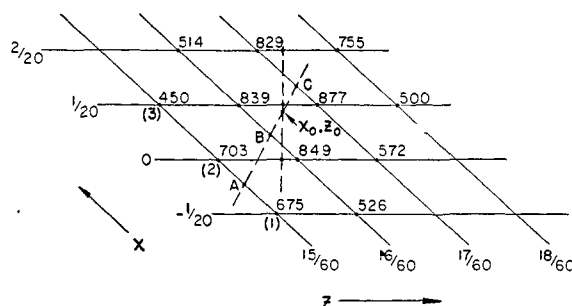


Fig. 4.—Steps in the calculation of the center of atom  $C_{IV}$  in the third three-dimensional Fourier synthesis, illustrated for  $y = -1/60$ .

proportional to the electron density are known at each intersection of the grid; only the values used in the calculation of the atomic center are shown. The three values at (1), (2), and (3) are used to determine the  $x$  coordinate of the point of highest density along the grid line  $z = 15/60$  on which they lie; this is the point A. Values for the positions of the points corresponding to A on the adjacent layers  $y = -2/60$ ,  $-1/60$ , and  $0/60$  are obtained in the same way. Since these three values for the positions of A are approximately independent of  $y$ , they may be averaged, with weights about proportional to the maximum density in each layer, to give a single value for the position of A to be used as described below. Similarly, the coordinates of the centers B and C are determined from the electron density values along the lines  $z = 16/60$  and  $17/60$ , respectively, in the layers  $y = -2/60$ ,  $-1/60$  and  $0/60$ . The least-squares-best straight line through A, B, and C, dotted in Fig. 4, passes through the center of the atom. In the same manner the centers along the lines  $x = 0$ ,  $1/20$ , and  $2/20$  determine another straight line through the atomic center; this line also is dotted. The intersection of these two straight lines gives the coordinates  $x_0$ ,  $z_0$  of the atomic center.

**Agreement of Calculated and Observed Structure Factors.**—In order both to secure good agreement between the magnitudes of calculated and observed structure factors (respectively,  $|F|_{\text{calcd.}}$  and  $|F|_{\text{obs.}}$ ) and to be able to make accurate use of the "back-shift" correction for non-convergence,<sup>12</sup> it was necessary to determine the temperature factor to be applied in the computation of  $|F|_{\text{calcd.}}$  and the scale factor which places  $|F|_{\text{obs.}}$  on the absolute scale. No definite evidence of absorption and few evidences of extinction were found, so no corrections were made for these effects.

(12) A. D. Booth, *Proc. Roy. Soc. (London)*, **A188**, 77 (1946).

Scale and temperature factors may be obtained simultaneously by the method of Wilson.<sup>13</sup> This method compares the sum of the relative  $F^2$  values in a small range of  $\sin \theta$  with the sum of the squares of the atom form factors in this range. In consequence of the layer structure of the acetylglycine crystal, the  $(0kl)$  zone data are representative of the entire  $(hkl)$  data; hence this method can be applied successfully to the  $(0kl)$  data alone. The preliminary scale factor (for  $F$  values) so obtained was extended to the  $(hkl)$  data as a first approximation; it was subsequently found to be only about 6% too large for the  $(hkl)$  data.

The first three-dimensional Fourier synthesis of the electron density exhibited ellipsoidal atoms elongated in a direction nearly perpendicular to the plane of the molecules. This indicates that the correct temperature factor is anisotropic. The direction of maximum elongation (nearly the same for all atoms) and the average atom eccentricity required the anisotropic part of the temperature factor to be  $e^{-(0.16h + 0.044l)^2}$  under the assumption that the preliminary isotropic part for the  $(0kl)$  zone represents the minimum temperature factor. The physical basis for this assumption is that the thermal motion of the atoms is least in the plane of the molecules. Once the anisotropic part was fixed, it was included in the Wilson procedure to calculate the isotropic part with greater accuracy; again only  $(0kl)$  data were employed.

Structure factors were computed for a series of structures related as described below, in the course of Fourier refinement. The over-all agreement of  $|F|_{\text{calcd.}}$  with  $|F|_{\text{obs.}}$  is measured, in each case, by the average discrepancy, values of which are listed in Table I. The calculation of the structure factors was performed with IBM punched card equipment.<sup>14</sup>

Structure factors trial 1 were calculated with the parameters measured from the first three-dimensional Fourier synthesis and were modified by the combination of isotropic and anisotropic temperature factors described above. Structure factors trial 2 are based on parameters from the second three-dimensional synthesis and the same temperature factors as in trial 1. A significant improvement is indicated by the reduced average discrepancy in Table I. Comparison of individual observed and calculated structure factors suggested that a slight modification of the anisotropic temperature factor would further improve the agreement. By trial, the new factor  $e^{-(0.20h + 0.045l)^2}$  was selected. Structure factors with this new anisotropic temperature factor but otherwise identical with those from trial 2 are designated trial 3 in Table I. The change in scale factor which is required by this change was expected to be so small that it was considered unnecessary at this stage.

(13) A. J. C. Wilson, *Nature*, **150**, 152 (1942).

(14) J. Donohue and V. Schomaker, *Acta Cryst.*, **2**, 344 (1949).

The contribution of the seven hydrogen atoms per molecule was then calculated for planes with  $\sin \theta < 0.55$ . The approximate positions of the hydrogen atoms were fixed by the general configuration of the molecule, except that the orientation of the methyl group was uncertain. The methyl hydrogen atoms were placed at tetrahedral positions with the C-H distances 1.09 Å. and so that one hydrogen atom lay in the plane of the molecule and *trans* to the carbonyl oxygen atom. Addition of the hydrogen contribution to the trial 3 structure factors gave those designated trial 4 in Table I. Trial 4 appeared to be no better than trial 3, and from the average discrepancies it may even be worse.

TABLE I  
AVERAGE DISCREPANCIES  $\Delta$  BETWEEN CALCULATED AND OBSERVED STRUCTURE FACTORS

$$\Delta = (\sum |F|_{\text{obs.}} - |F|_{\text{calcd.}}) / \sum |F|_{\text{obs.}}$$

$F_{\text{calcd.}}$ trial number	Description	$\Delta$ Values (%)	
		First scale factor	Final scale factor
1	Parameters from first synthesis	19.0	
2	Parameters from second synthesis	16.3	
3	2 with new temperature factor	15.0	
4	3 with all hydrogen contributions	15.1	
5	3 with non-methyl hydrogen contributions	14.5	15.4
6	Parameters from third synthesis with non-convergence correction	14.2	
7	6 with non-methyl hydrogen contributions	13.9	13.4

Since there was little or no structural basis for the orientation of the methyl hydrogen atoms, their contribution was subtracted from trial 4 to obtain trial 5. This appeared to be better than the same structure for the carbon, nitrogen, and oxygen atoms either with all hydrogen atoms or with no hydrogen atoms. Therefore, the assumed methyl hydrogen atom positions are definitely wrong.

The third three-dimensional Fourier synthesis, which included all but six small values of the approximately  $800|F|_{\text{obs.}}$  values, gave new parameters for the atoms other than hydrogen, none of which differed by as much as 0.02 Å. from the previous set. Small maxima corresponding to three of the hydrogen atoms of the molecule were observed in this synthesis. Since these parameters are very nearly the best that can be obtained from the usual Fourier refinement procedure, it appeared profitable to apply the correction<sup>12</sup> for missing terms in the Fourier summation by the calculation of another three-dimensional Fourier synthesis from the best calculated structure factors. The parameter corrections thus obtained were applied to the parameters from the third Fourier synthesis; the maximum correction to any parameter was 0.013 Å. and the average correction was 0.004 Å. The corrected parameters were used to calculate structure factors

trial 6. Structure factors trial 6 appeared to be slightly better than those from the most nearly comparable set, trial 3, as was indicated also by the average discrepancies in Table I.

The contribution of the non-methyl hydrogen atoms, the same as in trial 5, was added to trial 6 to obtain trial 7. The latter is slightly better than the most nearly corresponding set, trial 5, and also slightly better than the set without hydrogen contributions, trial 6. The decision that trial 7 was in best agreement with the observations required that a final scale factor appropriate to it be applied to the observations to get the best agreement possible with parameters of trial 7. Comparison of  $\Sigma|F|_{\text{calc'd.}}$  with  $\Sigma|F|_{\text{obs.}}$  showed that an additional scale factor of 0.94 must be applied to  $|F|_{\text{obs.}}$ . The average discrepancies for the two best structures, calculated after application of the final scale factor, are shown in the last column of Table I.

So far as the carbon, nitrogen, and oxygen atoms are concerned, the parameters used for the computation of the structure factors trial 7 are considered to be the best that can be obtained from the data. The final parameters for carbon, nitrogen, and oxygen atoms are listed in Table II as fractions  $x$ ,  $y$ ,  $z$  of the cell edges. They are compared with the final  $y$  and  $z$  parameters from zone data alone and with the parameters of the

original trial structure. This table also contains the electron density at atomic centers, measured from the third three-dimensional synthesis.

The observed structure factors, upon which this determination was based, are presented in Table III, together with the calculated structure factors trial 7.

### Discussion of the Structure

**The Acetylglucine Molecule.**—The dimensions of a single molecule are presented in Fig. 5 and Table IV. The probable error of a bond length, on the basis of experience with organic structure determinations, is estimated to be about 0.010 Å. It is thus unlikely that any bond length is in error by more than 0.030 Å. The corresponding probable error in a bond angle is about 1.0°.

The atoms of a single molecule, exclusive of hydrogen atoms, are within 0.1 Å. of a common plane. The two halves of the molecule are more accurately planar: the atoms C<sub>I</sub>, C<sub>II</sub>, O<sub>I</sub>, O<sub>II</sub> (in the group -CH<sub>2</sub>-COOH) are less than 0.001 Å. from a common plane, and the atoms N, C<sub>III</sub>, C<sub>IV</sub>, O<sub>III</sub> (in the group CH<sub>3</sub>-CO-NH-) are less than 0.005 Å. from another common plane. The dihedral angle between these planes through the two halves of the molecule is 5.2°.

The general configuration of the acetylglucine molecule is like that of glycylglycine.<sup>4</sup> The doubly-bonded oxygen atoms are on opposite sides of the molecule.

The identification of the atoms is seen to be unambiguous from the Fourier section shown in Fig. 3, where the equi-electron-density contours are drawn at levels arbitrary but equally spaced and the same throughout the section, and from the electron densities in Table II. It is obvious that the three oxygen atoms and the nitrogen atom have the highest electron densities and the four carbon atoms have the lowest. The atoms C<sub>IV</sub> and O<sub>II</sub>, however, have maximum densities lower than others of their kind (Table II); C<sub>IV</sub>, especially, is very low. Since low values for these atoms persist in the Fourier series made using  $F_{\text{calc'd.}}$  values, although the differences in this case are not so large, only part of the differences may be ascribed to non-convergence errors. Since these atoms are distinguished from others in the crystal in that they are bonded firmly to other atoms on one side only, they would be expected to be subject to relatively larger thermal motion which would readily account for the observed differences in peak densities.

The peptide C-N bond is found to have a length of 1.32 Å. This value is nearly that found in diketopiperazine,<sup>1</sup> 1.33 Å., and somewhat longer than that reported for glycylglycine,<sup>7</sup> 1.29 Å. In view of the relatively large probable error in the latter, this agreement is quite satisfactory and provides evidence that the peptide bond length does not vary greatly. The result of this determination leaves little doubt as to the length of the

TABLE II

ACETYLGLYCINE ATOMIC PARAMETERS				
Atom	Maximum density (electrons per cu. Å.)	Fractions of cell edges <sup>a</sup>		
		<i>x</i>	<i>y</i>	<i>z</i>
C <sub>I</sub>	8.50	0.0113	0.7167	0.5165
			.716	.516
			.706	.509
C <sub>II</sub>	8.25	0.0031	0.7293	0.4120
			.731	.413
			.712	.403
C <sub>III</sub>	8.54	-0.0167	0.6087	0.2739
			.609	.276
			.588	.272
C <sub>IV</sub>	6.97	-0.0493	0.4900	0.2242
			.490	.222
			.468	.230
N	10.39	-0.0172	0.6165	0.3640
			.617	.365
			.601	.365
O <sub>I</sub>	11.34	0.0099	0.8179	0.5570
			.820	.558
			.798	.553
O <sub>II</sub>	10.36	0.0192	0.6262	0.5577
			.626	.557
			.610	.546
O <sub>III</sub>	11.79	0.0053	0.6966	0.2304
			.695	.230
			.675	.223

<sup>a</sup> Numbers in bold-face are the final parameters; these are followed in order by the final parameters from the (0*kl*) projections and by the parameters of the original trial structure.

TABLE III  
OBSERVED AND CALCULATED STRUCTURE FACTORS

In each column, the left hand column is the  $l$  index, the middle column the observed structure factor, and the right hand column, the calculated structure factor, trial 7. All structure factors have been multiplied by ten.

00 $l$	6 <21 +1	8 <30 +14	10 $l$	15 $l$	3 <40 +29
0 +2360	7 37 -36	9 <25 -18	0 1215 +1416	0 <10 +2	4 <37 +66
2 13 +47	8 23 +36	10 <19 +1	2 <11 -1	1 152 -162	5 56 +74
4 286 +235	9 30 -29		4 124 +120	2 89 -105	6 <26 -20
6 257 -264	10 67 +77	09 $l$	6 102 -135	3 131 +120	
8 32 -1	11 40 -39	1 211 +231	8 20 -2	4 95 -108	1.11 $l$
10 52 -63	12 32 +36	2 74 -65	10 12 -40	5 39 -40	0 <20 +6
12 12 -15		3 46 +38		6 29 -48	1 123 -138
	05 $l$	4 24 +23		7 60 +89	2 <20 -14
01 $l$	1 281 -265	5 <33 -37	11 $l$	8 27 +30	3 <36 -17
1 232 +251	2 186 -161	6 31 -38	0 34 +31	9 9 -10	4 <32 +52
2 13 +28	3 215 +168	7 <30 -8	1 100 +98		5 <26 +3
3 <16 -18	4 203 -194	8 26 +35	2 65 +87	16 $l$	
4 230 +234	5 68 -65	9 <21 -11	3 <24 0	0 58 +86	1.12 $l$
5 82 -91	6 88 -102		4 117 +132	1 36 -58	0 83 -87
6 276 +278	7 148 +170	0.10 $l$	5 39 -53	2 18 -20	1 54 -54
7 56 +63	8 42 +43	0 185 +207	6 115 +138	3 76 -85	2 <18 -15
8 58 +59	9 17 -17	1 <19 +7	7 <31 +36	4 46 +50	3 <30 +18
9 <31 -26	10 47 +48	2 95 -100	8 37 +32	5 91 -106	4 <25 -9
10 <29 +13	11 104 +125	3 35 +32	9 <23 -14	6 <41 -19	1.13 $l$
11 <25 -5		4 88 +103	10 <14 +6	7 <37 +6	0 <16 +2
12 <18 -19	06 $l$	5 108 +124		8 <30 -25	1 <10 +12
	0 89 +102	6 <30 -39	12 $l$		2 14 -18
02 $l$	1 32 -58	7 70 +72	0 376 -360	17 $l$	3 <20 +39
1 471 -493	2 47 -35	8 44 +39	1 323 -293	0 26 +20	
2 384 -316	3 139 -123	9 <16 +12	2 158 -141	1 42 -19	1.14 $l$
3 140 +142	4 82 +96		3 80 +91	2 103 +102	0 23 -20
4 84 -81	5 188 -200	0.11 $l$	4 55 -55	3 45 -24	1 33 +50
5 <11 +24	6 33 -53	1 155 -164	5 <14 +3	4 <42 -2	
6 70 +69	7 <32 -9	2 20 -18	6 34 +48	5 <42 -29	20 $l$
7 114 +108	8 34 -36	3 35 -38	7 46 +55	6 <39 +5	0 526 +604
8 18 +18	9 73 +69	4 61 +65	8 14 +3	7 <34 -8	2 <15 -7
9 41 -42	10 27 -35	5 <29 +5	9 21 -25	8 <26 -20	4 42 +46
10 <14 -4	11 46 +53	6 <26 +39			6 32 -54
11 12 +6		7 46 +62		18 $l$	
12 24 -27	07 $l$	8 <17 +14		0 81 +88	21 $l$
	1 <16 +25		13 $l$	1 107 +119	0 15 +29
03 $l$	2 182 +170	0.12 $l$	0 20 -18	2 <21 -10	1 <27 +20
1 159 -113	3 54 -44	0 92 -98	1 120 -102	3 <42 -47	2 <29 +35
2 66 -61	4 <19 -12	1 60 -58	2 55 -49	4 <42 -48	3 <31 +9
3 72 +31	5 33 -32	2 25 -27	3 43 +44	5 <40 +11	4 44 +57
4 233 +216	6 <32 +16	3 23 +16	4 102 +118	6 <36 -3	5 <29 -27
5 <15 -13	7 <33 -26	4 <18 -10	5 <20 -16	7 44 -49	6 39 +54
6 26 -24	8 31 -40	5 28 -39	6 35 -27	8 <18 +4	7 <16 +17
7 54 -42	9 46 -50	6 <22 -7	7 35 -12		
8 <21 -8	10 40 +32	7 <16 -17	8 18 -12	19 $l$	22 $l$
9 45 -44	11 39 -41		9 23 -24	0 <20 +7	0 175 -171
10 106 -109		0.13 $l$		1 158 +163	1 147 -135
11 16 +6	08 $l$	1 <16 +15	14 $l$	2 51 -53	2 69 -68
12 25 +25	0 108 +98	2 16 -15	0 119 -122	3 <42 +31	3 36 +56
	1 152 +146	3 38 +42	1 156 +171	4 <40 +18	4 22 -22
04 $l$	2 <18 -3	4 17 +15	2 28 -17	5 <37 -30	5 <14 -1
0 83 -121	3 67 -72	5 <19 -14	3 27 +18	6 <32 -25	6 <12 +26
1 293 +239	4 72 -71		4 58 +61	7 <24 -3	7 19 +26
2 71 -68	5 <32 +24	0.14 $l$	5 81 +97		
3 <9 +6	6 <33 -8	0 25 -21	6 <24 +9	1.10 $l$	23 $l$
4 121 +112	7 88 -83	1 53 +48	7 22 -21	0 163 +179	0 26 -16
5 194 +193		2 23 +20	8 20 +8	1 <15 +11	1 54 -67
		3 67 -72	9 24 -21	2 63 -76	2 29 -8
		4 15 +23			



TABLE III (Continued)

3 24 +33	2·11 <i>l</i>	10 131 -136	9 32 -31	9 <41 -22	$\bar{1}\cdot 14\bar{l}$
4 39 +56	0 <18 +7	12 29 -32	10 110 +118	10 <38 -4	1 47 +46
5 <19 -15	1 68 -92	14 62 +61	11 66 -69	11 <33 -21	2 14 +16
6 19 -15	2 <23 -9		12 78 +81	12 <25 -15	3 72 -77
7 <9 -4		$\bar{1}1\bar{l}$	13 26 -27		4 24 +29
	2·12 <i>l</i>	1 306 +279	14 20 -21	$\bar{1}9\bar{l}$	5 <8 -1
24 <i>l</i>	0 54 -62	2 35 -8		1 217 +240	
0 62 -68	1 32 -43	3 37 -50	$\bar{1}5\bar{l}$	2 72 -61	$\bar{2}0\bar{l}$
1 62 +82		4 252 +271	1 225 -225	3 40 +38	2 92 +90
2 46 +18	30 <i>l</i>	5 127 -97	2 146 -136	4 20 +22	4 303 +299
3 25 +10	0 220 +258	6 454 +473	3 194 +159	5 <15 -29	6 424 -395
4 <23 +22	2 <14 -6	7 88 +86	4 252 -246	6 39 -44	8 86 +62
5 27 +35	4 <8 +14	8 122 +98	5 94 -65	7 <42 -18	10 206 -186
6 <16 +9		9 <28 -38	6 150 -141	8 <41 +52	12 68 -58
	31 <i>l</i>	10 30 +28	7 261 +245	9 <39 -11	14 80 +80
	0 <31 +18	11 <31 -9	8 55 +44	10 <35 -49	16 <9 -4
25 <i>l</i>	1 <30 -4	12 <30 -42	9 23 -21	11 <28 -18	
0 <15 +2	2 <27 +12	13 <26 -15	10 84 +75		$\bar{2}1\bar{l}$
1 41 -76	3 <22 +6	14 <19 -19	11 197 +218	$\bar{1}\cdot 10\bar{l}$	1 156 +163
2 41 -60	4 <15 +24		12 31 +37	1 <19 -3	2 13 +21
3 70 +74		$\bar{1}2\bar{l}$	13 42 -49	2 93 -98	3 71 -49
4 <39 -48	32 <i>l</i>	1 466 -406		3 24 +22	4 170 +185
5 <14 -21	0 81 -85	2 443 -323	$\bar{1}6\bar{l}$	4 110 +122	5 38 -51
6 <11 -20	1 60 -56	3 118 +134	1 34 -50	5 145 +162	6 434 +470
	2 27 -29	4 43 -54	2 54 -36	6 44 -55	7 81 +76
26 <i>l</i>	3 <11 +26	5 52 +39	3 139 -116	7 90 +101	8 170 +142
0 <20 +36	4 <7 -11	6 117 +89	4 150 +144	8 51 +60	9 25 -28
1 <36 -23		7 201 +181	5 270 -272	9 <35 +29	10 57 +43
2 <37 -8	33 <i>l</i>	8 38 +42	6 70 -83	10 <29 -28	11 37 -19
3 <36 -45	0 24 -12	9 72 -60	7 <38 -38	11 <20 -18	12 78 -81
4 <33 +15	1 19 -36	10 16 -12	8 57 -47		13 <31 -32
5 <27 -45	2 <18 -2	11 20 +0	9 94 +94	$\bar{1}\cdot 11\bar{l}$	14 <28 -28
6 <18 -4	3 <14 +17	12 35 -36	10 66 -75	1 138 -157	15 35 +46
		13 44 +37	11 86 +84	2 <23 -17	16 40 -41
	34 <i>l</i>	14 10 -0	12 <34 -22	3 53 -58	
27 <i>l</i>	0 23 -24		13 <25 +15	4 67 +63	$\bar{2}2\bar{l}$
0 28 +26	1 <22 +29	$\bar{1}3\bar{l}$		5 <15 +6	1 225 -203
1 42 -35	2 30 +23	1 90 -71	$\bar{1}7\bar{l}$	6 59 +55	2 176 -157
2 48 +49	3 16 +6	2 74 -74	1 24 +41	7 75 +73	3 38 +46
3 <34 -12		3 18 -16	2 199 +190	8 <34 +22	4 78 -77
4 <30 +1	35 <i>l</i>	4 330 +288	3 63 -58	9 <29 +7	5 34 +37
5 <24 -19	0 <17 +5	5 29 -12	4 27 -22	10 <22 -31	6 79 +56
	1 15 -24	6 41 -40	5 50 -27		7 214 +202
28 <i>l</i>	2 13 -28	7 100 -86	6 24 +28	$\bar{1}\cdot 12\bar{l}$	8 55 +54
0 42 +58	3 29 +38	8 <18 +10	7 <40 -54	1 50 -47	9 66 -60
1 52 +67		9 60 -63	8 66 -71	2 31 -36	10 41 -28
2 <35 -11	40 <i>l</i>	10 184 -184	9 59 -74	3 17 +13	11 15 -13
3 <32 -25	0 79 +101	11 <21 +3	10 47 +55	4 18 -11	12 39 -35
4 <27 -28		12 50 +43	11 64 -71	5 37 -41	13 70 +60
5 <18 +3	41 <i>l</i>	13 <17 -21	12 60 -54	6 <18 -16	14 14 -11
	0 <18 +9	14 22 -26	13 <18 -18	7 <32 -30	15 22 -29
				8 <28 -20	16 <8 +2
	42 <i>l</i>	$\bar{1}4\bar{l}$		9 42 -52	
29 <i>l</i>	0 30 -38	1 238 +214	$\bar{1}8\bar{l}$		$\bar{2}3\bar{l}$
0 <21 +8		2 120 -110	1 140 +142	$\bar{1}\cdot 13\bar{l}$	1 45 -33
1 76 +83	43 <i>l</i>	3 26 -8	2 <21 +5	1 <15 +16	2 85 -70
2 36 -33	0 <10 -8	4 150 +125	3 78 -75	2 <19 -7	3 28 -28
3 <28 +19		5 309 +279	4 80 -77	3 34 +37	4 199 +196
4 <22 +10	44 <i>l</i>	6 46 -32	5 36 +34	4 16 +15	5 34 +26
	0 9 -11	7 37 -41	6 <21 -15	5 11 -11	6 22 -28
		8 68 +80	7 122 -120	6 43 -48	7 140 -119
	45 <i>l</i>		8 <42 +28	7 <24 -24	8 16 +25
	0 39 +103				
2·10 <i>l</i>	$\bar{1}0\bar{l}$				
0 87 +112	2 39 +103				
1 <32 +5	4 330 +336				
2 32 -44	6 435 -423				
3 <23 +22	8 38 +20				



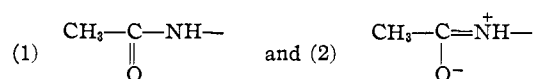
TABLE III (Continued)

$\bar{4}2l$	5 55 +67	$\bar{4}7l$	8 <10 +11	15 28 -33	14 <14 +11
1 29 -37	6 21 -27	4 <18 -11	9 <10 +10	16 26 +27	15 12 +20
2 33 -36	7 21 -15	5 <13 +16	10 21 -18	17 17 +14	16 46 +46
3 <14 +1	8 39 +62	6 31 +22	11 18 +27	18 25 -18	17 93 -89
4 26 -16	9 <19 -5	7 36 -40		$\bar{5}0l$	$\bar{5}3l$
5 <15 +4	10 65 +71	8 70 -66			$\bar{5}6l$
6 <15 +3	11 71 -78	9 30 -36	4 < 7 +31	5 <12 +14	6 <10 -12
7 65 +76	12 180 +180	10 68 +61	6 20 -36	6 15 -6	7 < 9 -20
8 21 +35	13 44 -36	11 115 -105	8 30 +29	7 29 -33	8 <14 -2
9 18 -5	14 48 -44	12 78 -62	10 20 -47	8 18 +11	9 26 +21
10 44 -46	15 48 -40		12 40 -35	9 31 -18	10 53 -65
11 29 -27	16 39 -36	$\bar{4}8l$	14 41 +33	10 88 -80	$\bar{5}7l$
12 21 -10	17 59 -46	4 18 -17	16 <12 +12	11 28 -22	7 20 -25
13 83 +77		5 <12 +13	18 76 +58	12 33 +33	8 40 -41
14 40 -37	$\bar{4}5l$	6 <17 -17		13 <19 -1	9 12 -16
15 45 -41	1 <10 -17	7 60 -60	$\bar{5}1l$	14 <18 -17	10 40 +39
16 27 +22	2 <13 -6	8 34 +34	4 <14 +7	15 <17 -8	$\bar{5}8l$
17 18 +15	3 32 +12	9 19 -16	5 <20 +12	16 <16 -10	
18 15 -8	4 35 -48	10 <19 -10	6 37 +55	17 <14 +8	9 11 -11
	5 14 -16	11 27 -19	7 <13 -4	18 79 -65	$\bar{6}0l$
$\bar{4}3l$	6 42 -46	12 <18 -11	8 50 +48		$\bar{5}4l$
1 <15 -5	7 83 +89		9 35 +22	5 15 +28	10 14 -17
2 17 -10	8 <17 -2	$\bar{4}9l$	10 28 +27	6 <14 -13	12 30 -19
3 <20 -20	9 <17 -3	4 <12 -4	11 <29 -23	7 <11 -7	14 < 9 +14
4 43 +48	10 68 +56	5 10 +13	12 <29 -42	8 <17 +30	16 < 8 +11
5 15 +25	11 238 +244	6 <15 -6	13 <29 -28	9 <16 -2	$\bar{6}1l$
6 20 -13	12 43 +43	7 15 -19	14 <28 -19	10 30 +30	11 9 -15
7 69 -66	13 35 -30	8 30 +30	15 39 +44	11 38 -47	12 <14 -22
8 25 +24	14 18 +16	9 30 +22	16 71 -59	12 105 +126	13 <18 -13
9 46 -35	15 29 +26	10 31 -34	17 <22 -25	13 20 -20	14 <18 -10
10 151 -151	16 68 +53	11 34 -27	18 46 +32	14 35 -35	15 <17 +24
11 42 -28	17 119 -103	12 28 +32		15 36 -34	16 43 -42
12 60 +54			$\bar{5}2l$	16 26 -22	$\bar{6}2l$
13 20 -7		$\bar{4} \cdot 10 \cdot l$	4 < 6 -9	17 54 -42	11 16 -13
14 35 -33	$\bar{4}6l$	5 39 +55	5 <10 +2		12 16 -2
14 <19 -6	4 19 +38	6 <12 -22	6 <11 -35	$\bar{5}5l$	13 41 +36
16 <17 -9	5 50 -65	7 68 +59	7 25 +33	6 10 -19	14 19 -26
17 <15 +8	6 19 -28	8 39 +43	8 <13 +15	7 43 +36	15 < 8 -21
18 88 -67	7 18 -39	9 44 +48	9 <14 +5	8 <14 -3	16 18 +24
	8 <23 -8	10 23 -25	10 30 -36	9 <14 +2	$\bar{6}3l$
$\bar{4}4l$	9 54 +45	11 32 -32	11 25 -21	10 40 +31	
1 <16 +20	10 92 -104		12 17 -5	11 118 +140	12 17 +16
2 <19 -12	11 87 +79	$\bar{4} \cdot 11 \cdot l$	13 68 +61	12 20 +22	13 <12 +5
3 <22 +4	12 21 -5	7 13 +20	14 32 -35	13 <14 -6	14 <10 -6
4 25 +23					

peptide C-N bond. The bond angles at both ends of the peptide bond are close to  $120^\circ$  but show slight differences in the expected directions: the angle opposite the carbonyl double bond is  $118^\circ$  whereas the angles of which it forms one side are each  $121^\circ$ .

The  $C_{II}-N$  distance of 1.45 Å. is not significantly different from the sum of the single bond radii<sup>15</sup> 1.47 Å. The  $N-C_{II}-C_I$  angle is the tetrahedral angle to within the probable error.

The large difference in the two carbon-nitrogen bond lengths and also the slightly long carbonyl bond, 1.24 Å. (normal 1.215 Å.) are the results of resonance between the structures



If the contributions of structures (1) and (2) are chosen as 60 and 40%, respectively, to get the best agreement, the predicted lengths are carbon-oxygen = 1.25 (observed 1.24 Å.) and carbon-nitrogen 1.33 (observed 1.32 Å.). These lengths were calculated by use of the standard single and double bond radii<sup>15</sup> in Pauling's equation<sup>16</sup> relating bond character and length. The good agreement between observed and calculated distances may be fortuitous because the effects of hydrogen bond formation by the atoms involved and of the

(15) L. Pauling, "The Nature of the Chemical Bond," Cornell University Press Ithaca, N. Y., 1940, p. 184.

(16) *Ibid.*, p. 175.

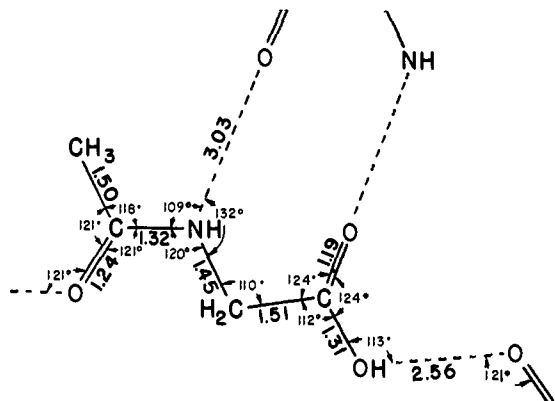


Fig. 5.—Dimensions of the acetylglycine molecule.

TABLE IV

BOND LENGTHS AND ANGLES IN ACETYLGLYCINE

Bond	Length, Å.	Angle	Value (°)
C <sub>I</sub> -C <sub>II</sub>	1.506	O <sub>I</sub> -C <sub>I</sub> -C <sub>II</sub>	111.5
C <sub>I</sub> -N	1.448	O <sub>I</sub> -C <sub>I</sub> -O <sub>II</sub>	124.2
N-C <sub>III</sub>	1.323	O <sub>II</sub> -C <sub>I</sub> -C <sub>II</sub>	124.3
C <sub>III</sub> -C <sub>IV</sub>	1.503	C <sub>I</sub> -C <sub>II</sub> -N	110.2
C <sub>I</sub> -O <sub>I</sub>	1.312	C <sub>II</sub> -N-C <sub>III</sub>	119.6
C <sub>I</sub> -O <sub>II</sub>	1.192	N-C <sub>III</sub> -C <sub>IV</sub>	117.7
C <sub>III</sub> -O <sub>III</sub>	1.243	N-C <sub>III</sub> -O <sub>III</sub>	121.3
O <sub>I</sub> ...O <sub>III</sub>	2.558	O <sub>III</sub> -C <sub>III</sub> -C <sub>IV</sub>	121.0
N...O <sub>II</sub>	3.030	C <sub>II</sub> -N...O <sub>II</sub>	131.8
C <sub>II</sub> ...C <sub>IV</sub>	3.707	C <sub>III</sub> -N...O <sub>II</sub>	108.5
		C <sub>I</sub> -O <sub>I</sub> ...O <sub>III</sub>	113.3
		C <sub>III</sub> -O <sub>III</sub> ...O <sub>I</sub>	121.4

contribution of hyperconjugated structures was neglected.

The C-C bonds have lengths 1.50 and 1.51 Å., just significantly shorter than the normal single bond distance 1.54 Å. The shortening in both cases can be explained by assuming hyperconjugation of the C-H bonds with the neighboring C-O double bond.

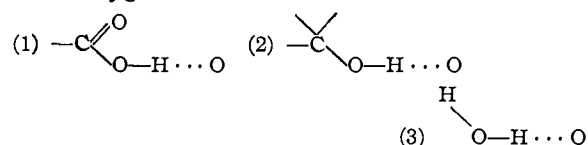
Unlike glycylglycine,<sup>7</sup> acetylglycine is, of course, not an internal salt (zwitterion), so the carboxyl C-O distances are quite different, respectively 1.19 and 1.31 Å. The latter C-O bond is much shorter than the sum of the single bond radii, 1.43 Å., indicating that carboxyl ion resonance is present to a considerable extent. The shorter C-O distance, 1.19 Å., may be significantly shorter than the sum of the double bond radii, 1.215 Å., and certainly is significantly shorter than the somewhat greater C-O distance which would be expected from the carboxyl ion resonance structures.

**The Structure of the Crystal.**—The more important distances between atoms in adjacent molecules are given in Fig. 5 and Table IV. In the latter, atoms in the reference molecule are designated unprimed whereas those in other molecules are designated with one, two, or three primes respectively, if they are related to the corresponding atoms in the reference molecule by the first, second, or third transformation of

parameters given in the description of the space group.

The strongest attractive forces between adjacent molecules in the crystal of acetylglycine are the hydrogen bonds formed between the carboxyl oxygen O<sub>I</sub> of one molecule and the carbonyl oxygen O<sub>III</sub> of the next molecule. The O-O distance is only 2.56 Å. whereas twice the van der Waals radius<sup>17</sup> of oxygen is about 2.8 Å.; the shortening is indicative of a strong hydrogen bond. These bonds link acetylglycine molecules head-to-tail to form infinite chains along the *c*-axis, as can be seen from Fig. 6, which illustrates the packing of molecules in the crystal.

It is interesting to note that the lengths of O-H...O bonds in different crystals vary, apparently depending chiefly on the situation of the donor oxygen atom



Examples of type (1), hydrogen bonds from a carboxyl group, are found in acetylglycine and three dicarboxylic acid dihydrates<sup>18</sup>; these oxygen-oxygen distances vary from 2.52 to 2.56 Å. Examples of type (2), hydrogen bonds from a hydroxyl group, are found in threonine,<sup>19</sup> pentahydroxythritol,<sup>20</sup>  $\alpha$ -resorcinol,<sup>21</sup>  $\beta$ -resorcinol,<sup>22</sup> and hydroquinone-sulfur dioxide<sup>23</sup>; the oxygen-oxygen distances vary from 2.66 to 2.75 Å. Examples of type (3), hydrogen bonds from a water molecule, are found in the dicarboxylic acid dihydrates<sup>18</sup>; the oxygen-oxygen distances vary from 2.82 to 2.90 Å. Although this classification overlooks the situation of the acceptor oxygen atom, there is good reason to believe that the observed differences are significant. In examples of type (1), the contribution of the second carboxyl resonance structure would enable the hydrogen atom to form a stronger bond with the acceptor oxygen atom, thus leading to a shortening of the O-H...O distance. In one limiting case, a symmetrical hydrogen bond with the hydrogen atom midway between the two oxygen atoms and forming a half-bond with each of them, the oxygen-oxygen distance is expected<sup>24</sup> to be 2.28 Å. The other extreme, where only one covalent bond and a van der Waals contact would be considered, leads to an expected oxygen-oxygen distance of

(17) Ref. 15, p. 189.

(18) J. D. Dunitz and J. M. Robertson, *J. Chem. Soc.*, 142, 148, 1145 (1947).(19) D. Shoemaker, J. Donohue, V. Schomaker, and R. B. Corey, *THIS JOURNAL*, **72**, 2328 (1950).(20) F. J. Llewellyn, E. G. Cox, and T. H. Goodwin, *J. Chem. Soc.*, 883 (1937).(21) J. M. Robertson, *Proc. Roy. Soc. (London)*, **A157**, 79 (1936).(22) J. M. Robertson and A. R. Ubbelohde, *ibid.*, **A167**, 122 (1938).(23) D. E. Palin and H. M. Powell, *J. Chem. Soc.*, 208 (1947).(24) L. Pauling, *THIS JOURNAL*, **69**, 542 (1947).

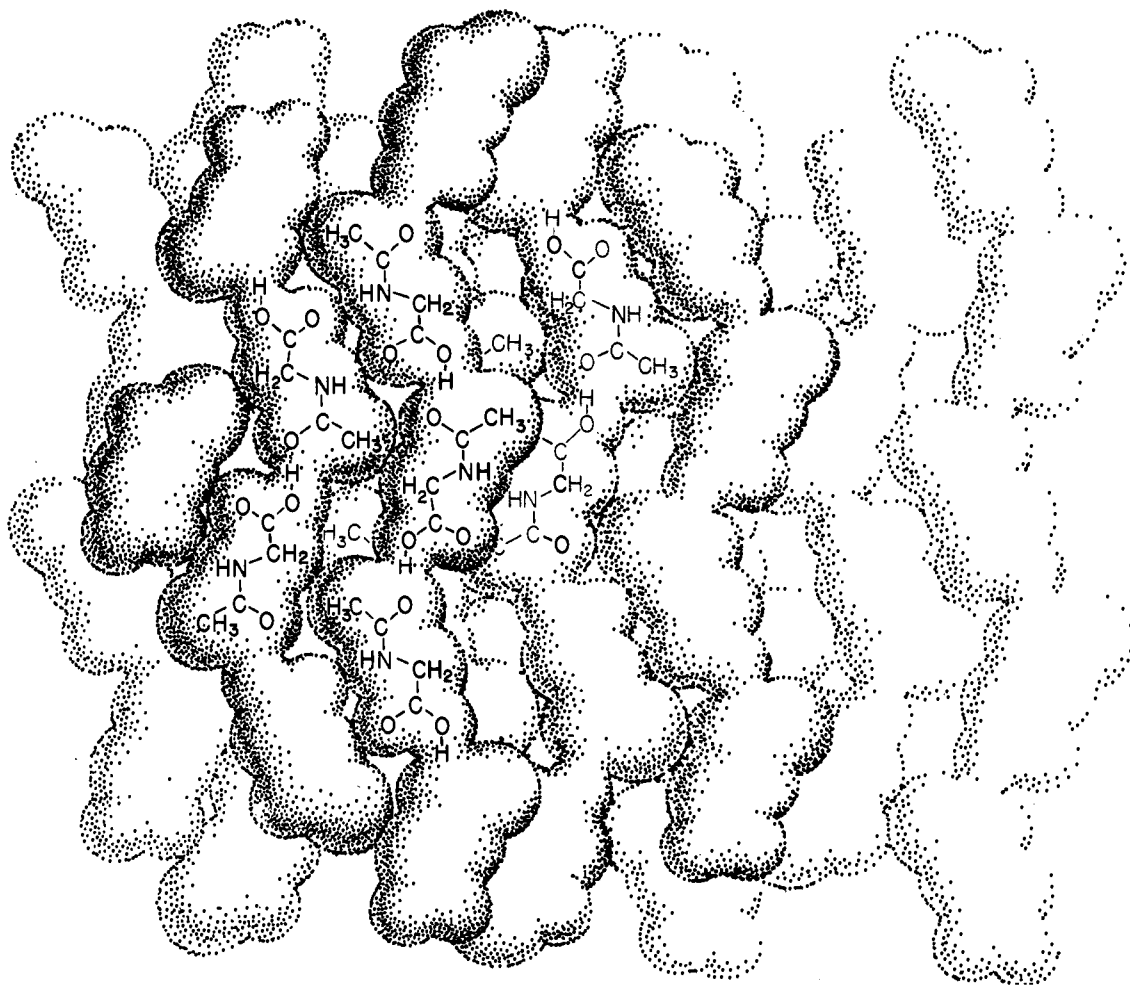


Fig. 6.—Arrangement of acetylglycine molecules in the crystal.

3.56 Å.<sup>15,17</sup> No examples of these two special cases are known at present, but the variations in distances referred to above correspond to reasonable intermediate conditions.

Less strong than the O—H···O bonds in acetylglycine, but still important in determining the structure of the crystal, are the hydrogen bonds formed between the nitrogen atom N of one molecule and the carboxyl oxygen O<sub>II</sub> of an adjacent molecule. The N—H···O distance is 3.03 Å., rather longer than the N—H···O distances of about 2.8 Å. found in other amino acids<sup>4,5,6,7,19</sup>; therefore, this is a relatively weak link. Nevertheless, these bonds tie the molecular chains together laterally to form the infinite layers parallel to (100) which are characteristic of the acetylglycine crystal. Parts of three layers in the crystal are shown in Fig. 6.

These layers appear to be bound to one another only by van der Waals forces to form the three-dimensional crystal.

Other factors helping to determine the arrangement of the molecules in the crystal structure can

be seen by examination of Fig. 6, in which the atoms (and atom groups in the cases of CH<sub>3</sub>, CH<sub>2</sub>, and NH) are drawn with van der Waals radii.<sup>17</sup> The feature which prevents the formation of shorter N—H···O bonds between adjacent chains appears from the figure to be the methyl-methylene contacts between chains. The actual distances between the carbon atoms C<sub>II</sub> and C<sub>IV</sub> is 3.71 Å., which is appreciably less than the sum of the van der Waals radii<sup>17</sup> for CH<sub>2</sub> and CH<sub>3</sub>, 4.0 Å. A further weakening influence is the unfavorable acceptance angle of the oxygen atom: for maximum strength the hydrogen bond would be expected to be formed along the direction in which the oxygen would form a covalent bond, rather than along a direction nearly collinear with the C—O bond, as found here. This situation cannot be improved without violating more stringent requirements: the hydrogen atom must remain on the normal N—H bond direction, and the position of the oxygen atom cannot be adjusted by relative displacement of adjacent chains parallel to their axes because of the same methyl-methylene contact de-

scribed above. By comparison, the orientation at atom  $O_{III}$  of the strong hydrogen bond between  $O_I$  and  $O_{III}$  is very favorable.

It is apparent also from Fig. 6 that an important factor in determining the mode of stacking of the layers in the crystal is the bulk of the methyl group. These relatively large groups fit neatly into gaps in the adjacent layer above or below, so that a very efficient packing is achieved.

This picture of the forces which determine the crystal structure provides a convincing explanation of the observed cleavages. The crystals show no tendency to cleave in directions that involve rupture of the strong  $O-H \cdots O$  bonds. The good cleavage into layers parallel to the molecular layers requires overcoming of van der Waals attraction only. The less marked cleavage of these layers into fibers along the  $c$ -axis requires also the breaking of the weak  $N-H \cdots O$  bonds between the molecular chains.

**Acknowledgments.**—We are indebted to Dr. Robert B. Corey for the original suggestion of this problem and for continued interest and helpful discussion during the course of this work. Most of the calculations were performed by

Misses Lillian Casler and Jeanne Mainwaring and Mrs. Jean Dewey Smith.

### Summary

The crystal structure of N-acetylglycine has been determined by single crystal methods. The unit cell is monoclinic with  $a_0 = 4.86$ ,  $b_0 = 11.54$ ,  $c_0 = 14.63$  Å., and  $\beta = 138^\circ 12'$ . The space group is  $P2_1/c$  and there are four molecules per unit cell.

Atomic positions were determined by three-dimensional Fourier refinement, with use of structure factors obtained from visually estimated intensities on Weissenberg photographs taken with  $Cu K\alpha$  radiation. A correction for non-convergence has been applied. A convenient analytical method for interpolating the atomic centers among the arbitrary points at which the electron density is calculated is described.

The crystal is composed of nearly plane layers of acetylglycine molecules joined together by a two-dimensional network of hydrogen bonds. The molecules are essentially planar. The peptide C-N bond length is 1.32, the  $C_\alpha$ -N bond length is 1.45 Å.

PASADENA 4, CALIFORNIA RECEIVED SEPTEMBER 6, 1949

[CONTRIBUTION FROM THE GATES AND CRELLIN LABORATORIES OF CHEMISTRY, CALIFORNIA INSTITUTE OF TECHNOLOGY, No. 1331]

## The Crystal Structure of $L_s$ -Threonine<sup>1</sup>

BY DAVID P. SHOEMAKER,<sup>2</sup> JERRY DONOHUE, VERNER SCHOMAKER AND ROBERT B. COREY

### Introduction

In the years immediately preceding the war determinations were made in these Laboratories of the crystal structures of diketopiperazine,<sup>3</sup> glycine,<sup>4</sup> and DL-alanine<sup>5</sup> by X-ray diffraction methods. The bond distances and bond angles found in these molecules have been used in discussions of the probable configuration of polypeptide chains in proteins.<sup>6,7</sup>

The structures of these crystals also provided considerable information concerning the part played by hydrogen bonds and other interatomic forces in determining the steric relationships between the amino acid molecules. When extrapolated into the field of protein structure, this information suggests a rather definite picture of the probable interactions between portions of the

polypeptide chains in solid proteins.<sup>7</sup> These earlier studies of the structures of amino acids together with the present investigation form a part of a program of research on the structure of protein molecules which is in progress at this Institute.

$L_s$ -Threonine (*threo*- $\alpha$ -amino- $\beta$ -hydroxy-*n*-butyric acid) is an especially important and interesting amino acid. It has been shown to be necessary for the normal growth of the rat<sup>8</sup> and more recently to be a necessary dietary constituent for the human.<sup>9</sup> It is a significant constituent of many common proteins,<sup>10</sup> such as egg albumin, lactalbumin, bovine and human serum albumin, human  $\gamma$ -globulin,  $\beta$ -lactoglobulin, gelatin, casein, insulin, silk fibroin,<sup>11</sup> and hemoglobin.<sup>12</sup>

In the three-year investigation described below, we have determined by X-ray diffraction methods the crystal structure of  $L$ -threonine. This work has confirmed the molecular structure

(1) Aided by a grant from the National Foundation for Infantile Paralysis, Inc.

(2) National Research Council Predoctoral Fellow, 1946-1947. Part of the work described here was included in a dissertation submitted by D. P. Shoemaker to the California Institute of Technology in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

(3) R. B. Corey, *THIS JOURNAL*, **60**, 1598 (1938).

(4) G. Albrecht and R. B. Corey, *ibid.*, **61**, 1087 (1939).

(5) H. A. Levy and R. B. Corey, *ibid.*, **63**, 2095 (1941).

(6) R. B. Corey, *Chem. Rev.*, **26**, 227 (1940).

(7) R. B. Corey, *Advances in Protein Chemistry*, **4**, 385 (1948).

(8) W. C. Rose, *Physiol. Rev.*, **18**, 109 (1938).

(9) W. C. Rose, W. J. Haines and J. E. Johnson, *J. Biol. Chem.*, **146**, 683 (1942); W. C. Rose, W. J. Haines, J. E. Johnson and D. T. Warner, *ibid.*, **148**, 457 (1943).

(10) E. Brand, *Ann. N. Y. Acad. Sci.*, **47**, 187 (1946).

(11) B. H. Nicolet and L. A. Shinn, *J. Biol. Chem.*, **140**, 685 (1941).

(12) J. Roche and R. Michel, *Compt. rend. soc. biol.*, **141**, 303 (1947).