

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_α -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 -Threonine¹

BY DAVID P. SHOEMAKER,² 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,³ glycine,⁴ and DL-alanine⁵ 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.^{6,7}

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.⁷ 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*- α -amino- β -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 rat⁸ and more recently to be a necessary dietary constituent for the human.⁹ It is a significant constituent of many common proteins,¹⁰ such as egg albumin, lactalbumin, bovine and human serum albumin, human γ -globulin, β -lactoglobulin, gelatin, casein, insulin, silk fibroin,¹¹ and hemoglobin.¹²

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).

and the relative configurations of the asymmetric centers as deduced by Rose and co-workers,¹³ and has also given detailed information concerning bond lengths, bond angles, hydrogen bonding, and van der Waals packing. Because of our high confidence in the experimental data, we have striven toward the fullest possible utilization of their precision. In this connection we have developed, improved, or used for the first time a number of new techniques: the use of the entire three-dimensional Patterson function, a more rapidly convergent non-centrosymmetric Fourier method, an analytical method for locating Fourier maxima, a "three-dimensional" least squares procedure for the simultaneous refinement of all positional parameters (except those of hydrogen atoms), punched card methods for calculating structure factors and for least squares refinement, and methods for estimating the precision of the parameter determination. Most of these will be described only briefly here, and some will be described in detail elsewhere.

Experimental

The Unit Cell and Space Group.—We are indebted to Professor Carl Niemann for samples of DL- and L-threonine, prepared respectively by Dr. P. L. Nichols and Dr. H. B. Milne.

Crystals of both materials were prepared by slow evaporation of aqueous solutions. The crystals prepared from DL-threonine were needles; those prepared from L-threonine were laths.

By optical examination and Laue photography, crystals obtained from both preparations were found to be of orthorhombic symmetry and to be identical from the standpoint of X-ray diffraction. It was apparent that crystallization from solutions of DL-threonine was accompanied by spontaneous resolution into optical isomers. No distinction was made between the DL and L preparations in later stages of the work.

The only well defined forms observed in the optical examination were {011}, {100}, and {210}. There was no morphological evidence of hemihedral symmetry. The only observed cleavage was parallel to (100). The density of the crystals at room temperature, as measured by flotation, was found to be 1.464 g. cm.⁻³.

Preliminary values of the lattice constants were obtained by measurement of layer-line spacings on heavily exposed rotation photographs, and these values were used to index several oscillation photographs. Measurement of twenty-nine reflections on these photographs, most of them at large angles, gave the following values and limits of error for the lattice constants

$$\begin{aligned} a_0 &= 13.611 \pm 0.020 \text{ \AA.} \\ b_0 &= 7.738 \pm 0.008 \text{ \AA.} \\ c_0 &= 5.142 \pm 0.005 \text{ \AA.} \end{aligned}$$

(λ for CuK α = 1.5418 \AA). The axial ratios, a_0 :

$b_0:c_0 = 1.759:1:0.665$, agree well with the optical values (1.748:1:0.669). There are four asymmetric molecules in the unit cell (4.01 calculated from the density). The orthorhombic cell is therefore primitive, and the space group is hemihedral.

Over 1800 reflections were observed and indexed on one asymmetric and three symmetric Laue photographs. Of these, over 800 were first order, and none required a unit cell larger than that given above.

Three equatorial Weissenberg photographs taken around the three respective crystallographic axes showed the absence of all ($h00$), ($0k0$), and ($00l$) reflections having odd values of h , k , and l respectively within the respective ranges of observation 0 to 16, 0 to 9, and 0 to 6 in these indices. No systematic extinctions of other kinds were found. These results lead to the space group D_{2h}^4 -P2₁2₁2₁, and thereby confirm our conclusion that when solutions of DL-threonine crystallize, the optical isomers are spontaneously resolved into separate crystalline individuals, or possibly into separate domains in twinned crystals.

Albrecht and co-workers¹⁴ have reported that the space group is C_{2v}^1 -Pmm2, with two molecules of D-threonine and two of L-threonine in the unit cell, although the values they reported for the lattice constants (13.64, 7.75 and 5.16 \AA .) are in substantial agreement with our own. These workers prepared their crystals from the racemic mixture and apparently determined the space group from morphology and Laue data alone. Hence, they were unable to carry out a satisfactory examination of pinacoid reflections for systematic extinctions. The results of the present investigation leave no doubt that the space group assigned by these workers is incorrect, and indeed there seems to be no valid evidence for the existence of crystals that can properly be termed DL-threonine. It should perhaps be pointed out that space groups having mirror planes provide highly unfavorable packing for molecules not having the corresponding planes of symmetry, and that in the case of threonine the space group Pmm2 can also be ruled out on steric grounds from a knowledge of the lattice constants and certain minimum molecular dimensions.

Collection of Intensity Data.—Two cleaved fragments of a large crystal of L-threonine were mounted on glass fibers with shellac, one with the a axis and the other with the b axis parallel to the Weissenberg rotation axis, and reduced to rough cylinders 0.4 mm. or less in diameter with a camel's hair brush moistened with water. The crystal selected for rotation around the c axis was a needle, 0.25 and 0.18 mm. in its largest and smallest cross-sectional dimensions, obtained from the DL-threonine preparation. Weissenberg photographs of all three crystals were taken for all layer lines with equi-inclination angles less than

(13) R. H. McCoy, C. E. Meyer and W. C. Rose, *J. Biol. Chem.*, **112**, 283 (1935); C. E. Meyer and W. C. Rose, *ibid.*, **115**, 721 (1936).

(14) G. Albrecht, G. W. Schnakenberg, M. S. Dunn and J. D. McCullough, *J. Phys. Chem.*, **47**, 24 (1943).

30° ($h = 0, 1, 2, \dots, 8$; $k = 0, 1, 2, \dots, 5$; $l = 0, 1, 2, 3$). Each exposure was taken with a set of four or five sheets of Eastman No-Screen film in the camera. The films in each set were developed uniformly and the relative intensities of the reflections were estimated visually, using the multiple-film technique of Robertson and co-workers.¹⁵ It was found convenient to estimate the intensities on a logarithmic scale, with the film factor as the base of logarithms. For the film factor, or ratio between intensities on successive films, a value of 3.7, kindly determined for us by Dr. Herman E. Seeman of the Eastman Kodak Co., was used.¹⁶ The intensities were corrected for Lorentz and polarization factors to obtain relative values of F_{hkl}^2 , and the values obtained from each set of films were multiplied by an appropriate scale factor to bring them to the same arbitrary intensity scale. The strongest reflections were corrected for secondary extinction by use of powder photographs of very finely divided material. The corrections were not large.

Of the 636 non-equivalent planes for which $(2 \sin \vartheta)/\lambda < 1.20 \text{ \AA}^{-1}$, reflections from 16 were extinguished by the space group, 65 gave reflections too weak to be observed, the intensities of 74 were estimated once, those of 303 twice (*i. e.*, on two separate photographs taken around different axes), and those of 178 three times. Where more than one estimate of F_{hkl}^2 was made the arithmetic mean was taken for each plane. From the residuals among the two or three measurements on 481 planes standard errors for the individual measurements were calculated.¹⁷ These estimates of the standard errors will be referred to as *external estimates* in later sections, primarily to distinguish them from *internal estimates* based on comparisons of observed and calculated structure factors. For the 240 strongest planes (ranging from 7 to 53 in the structure factor) the relative standard error of a single intensity measurement is 6.9% (corresponding to a probable error of 4.7%) on the assumption that the film factor is not in error and that it is approximately constant over this intensity range. For the remaining planes the relative standard error increases approximately in inverse proportion to the structure factor.

On the basis of considerations too lengthy to be given here, the root mean square uncertainty in the F_{hkl}^2 due to uncertainty in the film factor has been estimated to be about five per cent.

Values obtained for F_{hkl}^2 on a relative scale were placed on an absolute scale and the temperature

factor was determined by a method proposed by E. W. Hughes. This method is based on the fact that over a sufficiently large and appropriately chosen collection of planes there exists a relation

$$\sum_{hkl} F_{hkl}^2 = \sum_{hkl} t_{hkl} \sum_j f_{j,hkl}^2 \quad (1)$$

where t_{hkl} is the temperature factor for the plane (hkl), $f_{j,hkl}$ is the atomic form factor for atom j and plane (hkl), and the second sum on the right-hand side is taken over all atoms in the unit cell. In practice the method is applied by requiring the equation to be satisfied in each of a small number of regions in the reciprocal lattice obtained by dividing the space with concentric spheres. When carefully used the method appears to be good to one per cent. in the scale factor. It is very similar to the method described by Wilson.¹⁸

Finally, square roots were taken in order to obtain the moduli of the structure factors, termed the *observed structure factors*. These are given in Table III, where they are compared with the moduli of the final *calculated structure factors*. The estimated experimental uncertainty in the structure factors, including that due to uncompensated errors in the film factor, may be given in the form of relative standard deviations as 3.5% for the medium and strong reflections and somewhat more for the weak ones and the few very strongest ones, the minimum absolute standard deviations for any class of reflections being about 0.25. For the complete collection of data an absolute standard deviation of 0.4 may be taken as representative.

Determination of the Trial Structure

The success with which Patterson projections¹⁹ and Harker sections²⁰ were used in the determination of the structures of glycine⁴ and alanine⁵ suggested that these methods might be useful in the present work, although additional complications were anticipated because of the increased complexity of the threonine molecule. Therefore a Harker section $P(x, y, 1/2)$ was prepared. The degree of complexity was indeed such that it was not possible to distinguish Harker peaks (*i. e.*, peaks due to interactions between symmetry-equivalent atoms) from the more numerous non-Harker peaks, and after careful study the attempt to interpret this section was abandoned without useful clues to the structure having been obtained from it.

At this point the use of a Patterson function in three dimensions was considered. Although the Patterson method has been available since 1935, it has not been used in three dimensions until recently because the computations have been regarded as much too involved. However, a great increase in computational efficiency had been gained through the introduction of methods involving the use of International Business Machines

(15) J. J. de Lange, J. M. Robertson and I. Woodward, *Proc. Roy. Soc. (London)*, **A171**, 398 (1939); J. M. Robertson, *J. Sci. Inst.*, **20**, 175 (1943).

(16) No correction was made in the film factor for high layer lines to take account of the oblique incidence of the X-rays on the photographic film. Because the precision of visual intensity estimation turned out to be considerably greater than was anticipated, it has become apparent that such a correction would have been appropriate.

(17) E. T. Whittaker and G. Robinson, "The Calculus of Observations." Blackie and Son, London, 1944, Ch. VIII.

(18) A. J. C. Wilson, *Nature*, **150**, 151 (1942).

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

(20) D. Harker, *J. Chem. Phys.*, **4**, 381 (1936).

and punched cards,²¹ and more recently a new, shorter set of punched cards corresponding to the Beevers-Lipson strips²² had been found to afford a further considerable reduction in the computational time for most Fourier work. Moreover, other workers, by these or similar methods, had succeeded in carrying out electron density syntheses in three dimensions. It was therefore decided to calculate the three-dimensional Patterson function for threonine. Although three-dimensional Patterson functions have been calculated recently by other workers, the present investigation appears to be the first instance of the successful use of a *complete* three-dimensional Patterson function in the determination of a crystal structure.

The punched card computations were accomplished in about sixty man hours. The function was calculated from 0 to $\frac{1}{2}$ in each of the three coordinates (u, v, w); the function elsewhere in space is related to that in the given octant by the space group D_{2h}^1 -Pmmm of the Patterson function. The intervals between calculated points in the three axial directions are 0.23 Å., 0.26 Å., and 0.26 Å., or $\frac{1}{60}$, $\frac{1}{30}$ and $\frac{1}{20}$ of the respective axial lengths a_0 , b_0 and c_0 . The number of points at which the function was calculated is 5456. In the calculation of the Patterson function a modification function^{23,24} was applied to the F_{hkl}^2 data in order to improve the resolving power of the function, and the peak at the origin was removed by subtracting its Fourier transform from the coefficients in the summation. The details of calculating and interpreting the three-dimensional Patterson function will be published elsewhere,²⁴ and only an outline of the procedure will be given here.

The number of resolved maxima in the function as a whole turned out to be only about half the total number of interactions, due to accidental overlapping and superposition of peaks, but this proved to be no such hindrance to interpretation as it had in the case of the Harker section.

Perhaps the most obvious mode of interpretation of a three-dimensional Patterson function for a molecular crystal is that of first deducing the spatial orientation and configuration of one of the molecules by interpretation of Patterson peaks corresponding to intramolecular vectors, particularly vectors between chemically bonded atoms, and of then determining the position of the molecule in the crystal by reference to the Harker sections. Though this procedure might be a successful one in some cases, it was unsuccessful in the present instance, because the peaks in the neighborhood of the origin were, with one exception, diffuse and unresolved. A strong, well resolved

peak was found on the section $v = 0$ at a distance of 2.25 Å. from the origin. Since in glycine⁴ and alanine⁵ the carboxyl oxygen atoms were found to be 2.20 Å. apart, and since no other pair of atoms in the crystal, excluding hydrogen atoms from consideration, can be expected to have this separation, it could be concluded with confidence that this peak is due to the two oxygen atoms, O_I and O_{II}, of the carboxyl group.

Before any success was obtained in attempts to restrict the configuration of the molecule further, attempts were made to utilize other features of the function, notably the Harker sections. It was soon found possible to determine unambiguously the positions of the two carboxyl oxygen atoms relative to the screw axes in the unit cell. For a structure of this complexity it is difficult to correlate unambiguously on the three Harker sections the Harker peaks due to individual atoms, in spite of restrictions imposed by the space group on the relative coordinates of these peaks. When, however, the relative positions of two atoms can be obtained with good precision from a Patterson peak, whether or not these two atoms can be identified immediately as to kind, the six Harker peaks corresponding to the two atoms may be more easily found. In the present case the positions of the six Harker peaks due to O_I and O_{II} were easily and unambiguously found, in spite of the rather poor conditions of resolution existing in the Harker sections. It was then possible to assign parameters to atoms O_I and O_{II}. Moreover, three additional non-Harker peaks were found which are located with good precision at the positions required, by the parameter assignments and symmetry, for the remaining three interactions involving carboxyl oxygen atoms.

On the section $v = 0$ two other strong peaks were observed which are separated by a vector having the same length and direction as one of the intra-molecular O_I-O_{II} vectors. The inference was strong that these peaks represent interactions of a third atom with O_I and O_{II}. That this is the case was proved beyond reasonable doubt by the success with which peaks were found at positions required by all interactions involving this new atom and those already found, including the Harker interactions. The position of this atom was fixed, with only its kind left in question. A fourth atom was found in approximately the same way, the position of which made it apparent that it can be nothing but C_{II} (the α -carbon atom), and that it is directly bonded to the third atom. The third atom must then be either C_{III} or N, and the former assumption was made provisionally in order to provide acceptable hydrogen bonding and to satisfy steric requirements. This assignment proved later to be correct. There remained to be fixed only the relative positions of hydroxyl and methyl groups. In Fig. 1 certain features of the three-dimensional Patterson function are shown diagrammatically, and all interactions involving

(21) P. A. Shaffer, Jr., V. Schomaker and L. Pauling, *J. Chem. Phys.*, **14**, 648 (1946).

(22) V. Schomaker, unpublished work.

(23) J. Waser, "Doctoral Dissertation," California Institute of Technology, 1944, p. 70.

(24) V. Schomaker and D. P. Shoemaker, to be published in *Acta Cryst.*

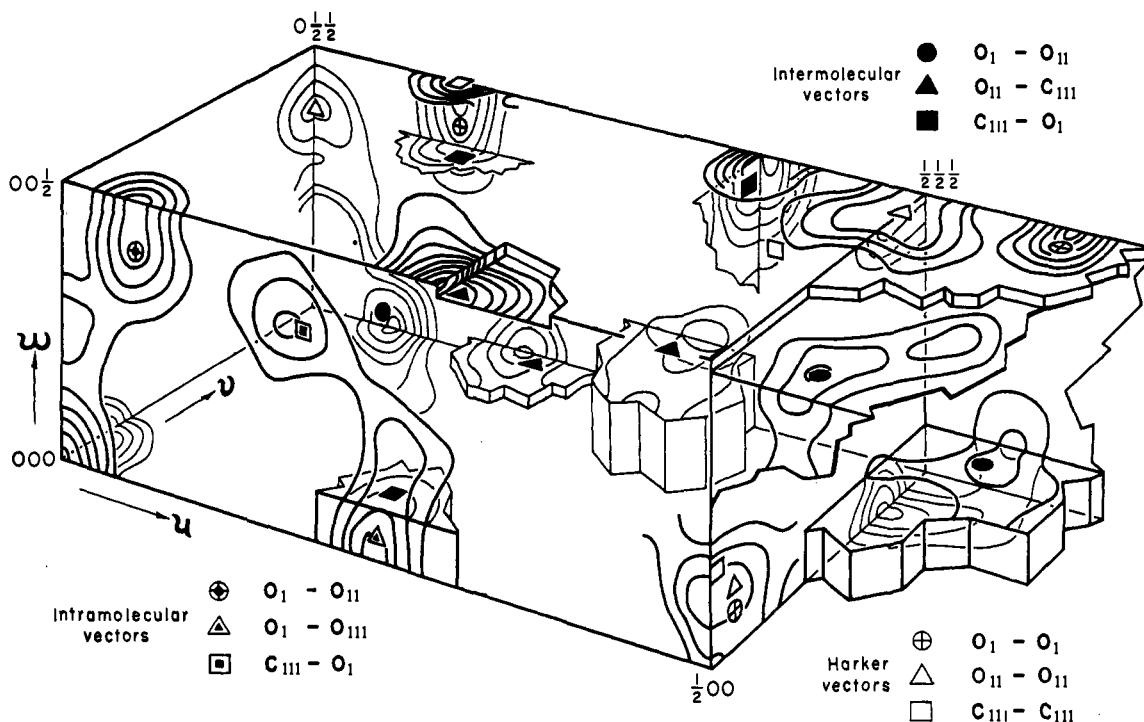


Fig. 1.—A representation of the three-dimensional Patterson function. Since the Patterson function has symmetry $Pmmm$, only one octant of the Patterson unit cell is shown. Only the peaks which correspond to interactions among the atoms O_I , O_{II} and C_{III} are shown; other peaks were omitted for clarity.

atoms of kinds O_I , O_{II} and C_{III} are marked. The corresponding interatomic vectors may be compared with those involving the atoms shown in the drawings of the structure, Figs. 6a, 6b, and 7.

The few remaining steric possibilities were tested by a trial and error procedure. This procedure involved calculating structure factors for a succession of trial structures and making parameter adjustments by the use of structure factor maps²⁵ drawn to the scale of the unit cell.

A trial structure was soon found which gave satisfactory agreement between observed and calculated (hkl) structure factors, but no parameter adjustments appeared capable of improving the bad agreement in the ($h0l$) and ($0kl$) structure factors, unless the methyl group and the hydrogen atom on the β -carbon atom were interchanged, thereby reversing the relative stereochemical configuration around the α and β asymmetric carbon atoms. We then discovered that we had made an accidental misinterpretation of the stereochemical formula of threonine; the interchange which the structure factors had compelled us to make brought the molecule into conformity with the chemically assigned stereochemical formula of threonine. The trial structure with the correct relative configurations around the α and β positions was taken as the starting point for the refinement procedure described below, which converged to give the final structure without essential change

in any structural features save for small shifts in atomic parameters.

Refinement of Parameters

The steps and procedures used in refining the parameters may be followed conveniently by reference to the scheme given in Fig. 2. To indicate the quality of the fit between observed and calculated structure factors, *percentage discrepancies* calculated with the expression

$$P. D. = 100 \frac{\sum_{hkl} ||F_{hkl}|_{obs} - |F_{hkl}|_{calc}||}{\sum_{hkl} |F_{hkl}|_{obs}} \quad (2)$$

are given in Fig. 2 for some of the sets of structure factors to illustrate the progress of the refinement. Unobserved planes have been omitted in the calculation of the percentage discrepancy, with the justification that in the later stages of refinement the effect of these planes has been compensated by means of the *back-shift rule*, which will be described later on. In Table I are given some of the sets of parameters obtained in the course of the refinement process.

Two-Dimensional Fourier Syntheses.—Starting with the first correct trial structure (Parameters Set 1), obtained by the procedure described in the preceding section, successive parameter shifts were made with the use of structure factor maps²⁵ until the agreement between ob-

(25) W. L. Bragg and H. Lipson, *Z. Krist.*, **95**, 323 (1936).

served and calculated prismatic ($hk0$, $h0l$, $0kl$) structure factors was good enough to enable signs to be assigned for the purpose of preparing Fourier projections. Since poor resolution was expected in a projection on (100), this projection was not prepared. The Fourier projection on (001), which was made with 72 of the 104 observed ($hk0$) reflections, is shown in Fig. 3a. The projection on (010), shown in Fig. 3b, was made with 56 of the 69 observed ($h0l$) reflections. Unfortunately the structure is such that only a few of the atoms in each projection are well enough resolved to allow accurate estimation of their parameters. Parameter Set 2 includes the parameters capable of accurate estimation from the projections, and are otherwise the best values obtained with the structure factor maps.

Three-Dimensional Fourier Syntheses.—Since the structure factors of the three prism zones calculated with these parameters were in fairly good agreement with those observed, the structure factors of the entire set of 618 reflections were calculated by a punched card method.²⁶ The atomic form factors of James and Brindley²⁷ were used in all structure factor calculations. These values were multiplied by the isotropic temperature factor $\exp(-B \sin^2 \theta / \lambda^2)$, where B was found to have the value 2.35 \AA^2 by the procedure which was used in the determination of the scale factor. No significant tendency toward anisotropy in the temperature factor was observed. The phases obtained were then used to calculate a "part-cell" projection on (010) of the scattering matter in the cell between $y = 0$ and $y = 1/2$. This projection is shown in Fig. 3c. All eight of the heavy atoms in the molecule are well resolved. Since the y -parameter of the nitrogen atom is very near 0, this atom appears twice at about half weight in each of two places. The x and z parameters obtained from this projection are in good agreement with those used to calculate the structure factors. It was therefore decided to proceed with three-dimensional Fourier analysis without first recalculating the structure factors.

(a) **Techniques of Three-Dimensional Fourier Refinement.**—Let it be assumed that a nearly correct trial structure, with parameters ξ_i^0 or (x_j^0, y_j^0, z_j^0) (*i. e.*, $\xi_1^0, \xi_2^0, \dots, \xi_{3N}^0 = x_1^0, y_1^0, z_1^0, x_2^0, \dots, z_N^0$) exists at some stage of the work. We shall here trace the various steps belonging to one complete Fourier refinement cycle, leading successively to ξ_i' , the coordinates of the maximum of a Fourier peak; ξ_i'' , these coordinates corrected for errors due to finite summation (by a *back-shift rule*); and, after a further correction for the retarding effects of phase-angle errors (by a *double-shift* or *n-shift rule*), the parameters ξ_i''' , which may be accepted as final or used as the new starting point for a repetition of the entire process.

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

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

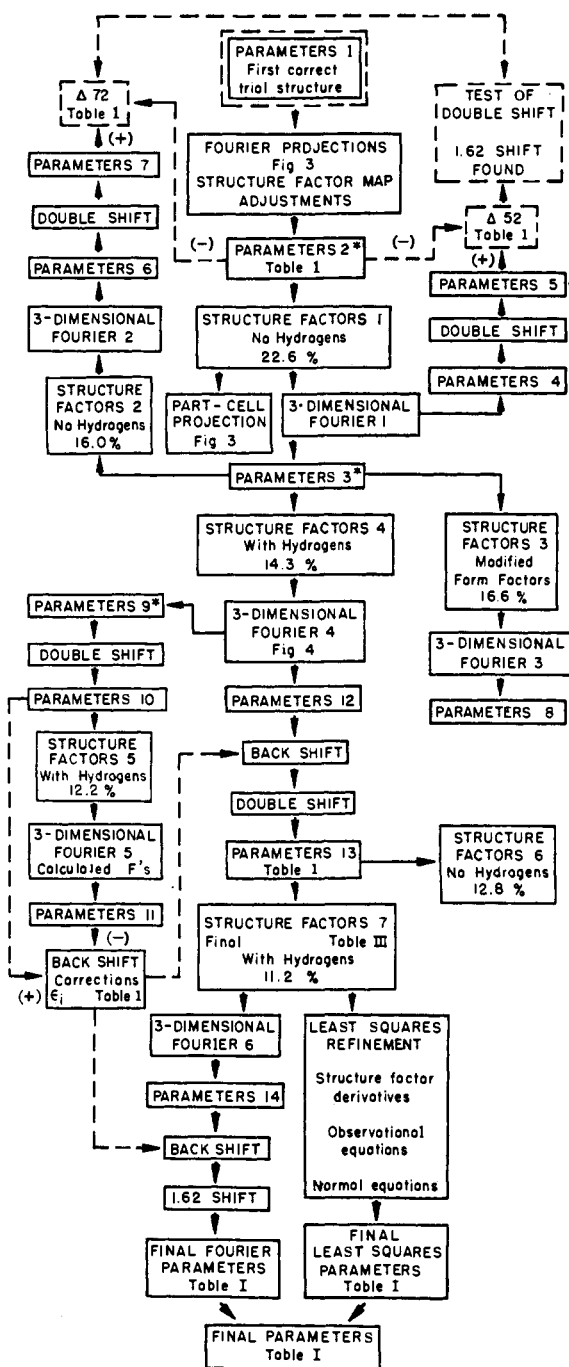


Fig. 2.—Diagram showing sequence of operations involved in the refinement of parameters. Where indicated by an asterisk (*) a graphical method was used in determining the coordinates of Fourier maxima; where the asterisk is absent, an analytical method was used.

In each case, structure factors and phase angles were computed from the ξ_i^0 , and the phase angles together with the observed structure factor moduli were used to calculate a three-dimensional electron density function $\rho(x, y, z)$. The calculations for

TABLE I
 ATOMIC PARAMETERS OF MOLECULE *M*: ATOMS C, N, O

Atom	Set no. 2 $\xi_i^{(2)}$	Δa_i $\times 10^4$	Δb_i $\times 10^4$	ϵ_i $\times 10^4$	Set no. 13 $\xi_i^{(13)}$	Final Fourier F_i^F	Final least squares ξ_i^{LS}	Final set ξ_i
C ₁	<i>x</i>	0.4920	22	22	-4	0.4958	0.4953	0.4956
	<i>y</i>	.1840	10	-16	3	.1846	.1837	.1836
	<i>z</i>	.3000	-34	-26	-8	.2970	.2985	.2984
C ₁₁	<i>x</i>	.3920	50	64	0	.3980	.3988	.3990
	<i>y</i>	.1170	-114	-84	-13	.1063	.1074	.1074
	<i>z</i>	.3950	-4	30	-20	.3906	.3906	.3908
C ₁₁₁	<i>x</i>	.3100	94	90	-2	.3186	.3175	.3178
	<i>y</i>	.2500	-130	-126	21	.2428	.2427	.2445
	<i>z</i>	.4200	92	84	26	.4346	.4320	.4299
C _{1v}	<i>x</i>	.2800	178	146	16	.2904	.2901	.2909
	<i>y</i>	.3300	38	28	1	.3340	.3336	.3357
	<i>z</i>	.1600	256	244	-23	.1814	.1829	.1821
N	<i>x</i>	.4120	3	12	2	.4128	.4133	.4129
	<i>y</i>	.0120	-2	10	2	.0130	.0139	.0139
	<i>z</i>	.6330	152	148	-9	.6444	.6410	.6420
O ₁	<i>x</i>	.5600	-38	-40	-8	.5574	.5574	.5579
	<i>y</i>	.2150	58	44	-2	.2170	.2165	.2171
	<i>z</i>	.4600	78	74	8	.4676	.4667	.4669
O ₁₁	<i>x</i>	.4980	42	52	-4	.5034	.5026	.5026
	<i>y</i>	.2000	130	116	6	.2124	.2111	.2115
	<i>z</i>	.0500	106	110	-14	.0588	.0598	.0581
O ₁₁₁	<i>x</i>	.2360	30	20	-2	.2364	.2367	.2365
	<i>y</i>	.1550	-58	-52	-19	.1460	.1464	.1473
	<i>z</i>	.5100	212	214	-19	.5276	.5272	.5275

this function were carried out with the same set of punched cards as was used in the calculation of the Patterson function. The function was calculated from 0 to 1 along *a*, and from 0 to 1/2 along *b* and *c*. The intervals between calculated points were the same as in the Patterson function; more than ten thousand points were calculated. Approximately 80 man hours was required for the punched card operations in the calculation of each of the six three-dimensional electron density functions that were used in the refinement procedure. This does not include the considerable time involved in calculating the phases, preparing the coefficients, plotting the results, and determining the positions of the maxima of the peaks.

At first a graphical method was used in locating the peak maxima. An analytical method was then proposed²⁸ and was later modified by the present authors.²⁹ In the modified method the 27 calculated points in a 3 × 3 × 3 rectangular parallelepiped as close as possible to the peak maximum are fitted by least squares to a three-dimensional ten parameter Gaussian function

$$\rho = \exp \left(p - \frac{r}{2} x^2 - \frac{s}{2} y^2 - \frac{t}{2} z^2 + ux + vy + wz + lyz + mxz + nxy \right) \quad (3)$$

representing approximately the shape of the peak.

(28) G. B. Carpenter and J. Donohue, *THIS JOURNAL*, **72**, 2315 (1950).

(29) V. Schomaker and D. Shoemaker, to be published in *Acta Cryst.*

Tests have shown that the Gaussian function fits the peak shapes very well in the region of the maximum. It is convenient to take the central point of the 27 as origin and to take as units of *x*, *y* and *z* the intervals between calculated points in the corresponding directions (so that these variables are restricted to the integral values -1, 0 and 1). By taking the logarithm of each side of equation (3) and substituting for the 27 calculated points in turn, 27 observational equations or "equations of condition," linear in the ten parameters, are obtained. With the assumption that these equations have equal weight it is very easy to reduce them³⁰ to ten normal equations which, being almost diagonal, are easily solved to give the ten parameters. The coordinates ξ_i' of the maximum are then obtained by solving three simultaneous linear equations, obtained by equating to zero the *x*, *y* and *z* derivatives of $\log \rho$; these equations contain all of the ten parameters except *p* as coefficients. In the application of this method the lattice need not be orthogonal and the peaks may deviate considerably from spherical shape.

Having obtained the coordinates ξ_i' of the maxima, what we have called the *back-shift rule* was used in two cases to correct them for the effects due to finite summation. From a set of nearly correct trial parameters $\xi_i^{(10)}$ (*i. e.*, Parameters 10, see Fig. 2) a set (no. 5) of structure factors and phases was calculated (with the inclusion of hy-

(30) E. T. Whittaker and G. Robinson, *op. cit.*, Ch. IX.

drogen atoms) and a special Fourier synthesis (no. 5) was carried out with these phases and with these *calculated* structure factors instead of the observed structure factors. The coordinates of the maxima of this special synthesis, $\xi_i^{(11)}$, would be identical with the starting parameters if the effects of finite summation were negligible. The desired finite summation corrections are therefore the differences between these two sets of parameters, with signs changed

$$\xi_i'' = \xi_i' + \epsilon_i \quad (4)$$

where

$$\epsilon_i = \xi_i^{(10)} - \xi_i^{(11)} \quad (5)$$

The values of the ϵ_i are given in Table I. This procedure is similar to that described by Booth.³¹

Corrections were then applied to the ξ_i'' in order to compensate for the slowness of convergence of the Fourier refinement process. If the residual parameter errors in a *centrosymmetric* structure are so small that the calculated signs of the structure factors within the sphere of reflection are correct, a single Fourier synthesis will bring the parameters to their final (*i. e.*, converged) values. However, if the residual parameter errors in a *non-centrosymmetric* structure are not zero, errors must exist in the calculated phases, with the result that a single Fourier synthesis only reduces parameter errors instead of eliminating them altogether. It has been conjectured from experience that the parameter errors in an asymmetric structure are reduced to one-half their former values by a single Fourier synthesis. This conjecture was verified theoretically, by two of the present authors²⁹ and independently by Dr. E. W. Hughes, for a *completely asymmetric* structure containing a fair number of more or less randomly distributed atoms, none of which by itself dominates the structure factors, with small and random parameter errors.^{31a} This forms the basis of the *double-shift rule*, according to which the parameter shifts indicated by a single Fourier step (after the back-

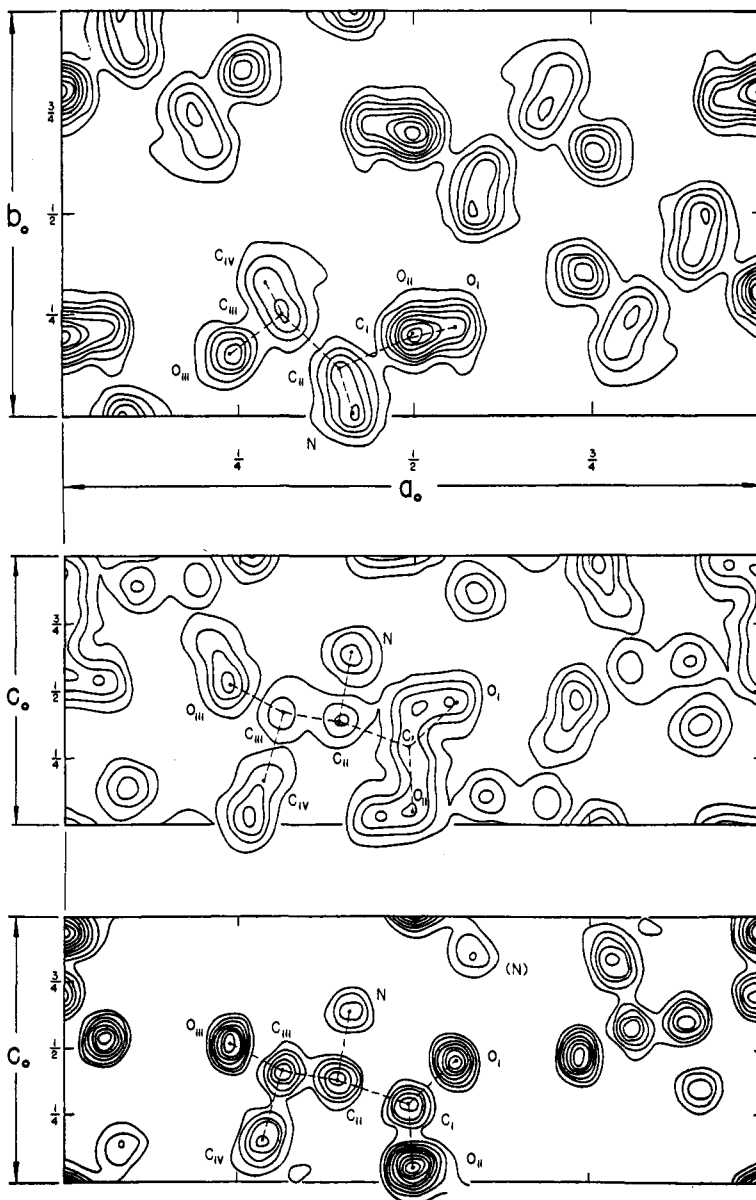


Fig. 3.—(a) Above: Fourier projection on (001); (b) center: Fourier projection on (010); (c) below: part-cell projection on (010) of all scattering matter from $y = 0$ to $y = 1/2$.

shift rule has been applied) should be doubled to obtain the best values obtainable for the parameters from the one Fourier step

$$\xi_i''' = \xi_i'' + 2(\xi_i'' - \xi_i'') \quad (6)$$

These parameters are subject to random errors (in threonine, estimated theoretically to be about one-quarter of the double shift) introduced by the approximations in the rule; hence the double-shift rule is no substitute for repeated Fourier syntheses but rather only an aid to more rapid convergence.

In order to test the double-shift rule two successive Fourier syntheses (Numbers 1 and 2, see

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

(31a) NOTE ADDED IN PROOF.—Lately D. W. J. Cruikshank (*Acta Cryst.*, **3**, 10 (1950)) has independently found the same result by theoretical considerations, but has not taken account of the presence of symmetry elements, the effects of which are described briefly in the present communication and will be discussed in more detail later.²⁹

Fig. 2) were carried out. The double-shift rule was applied to both to give Parameter Sets 5 and 7. The starting parameters, Set 2, were subtracted from the respective parameters in Sets 5 and 7, giving quantities which are given in Table I under the headings Δ_{52} and Δ_{72} . If the double-shift rule and the computational procedures were exact, the numbers given under these headings for a given parameter should be equal. (The omission of the back-shift correction affects the two parameter sets equally and hence has no effect on the present discussion.) In fact the mean difference between the Δ_{52} and the Δ_{72} is only 0.009 Å., while the mean shift Δ_{72} is 0.061 Å. A weak correlation between the signs of these differences and those of the shifts Δ_{72} suggested that the double-shift rule shifted the parameters a little too far. On the assumption of an n -shift instead of a double-shift, where n is the same in all syntheses and for all parameters, *e. g.*

$$\xi_i^{(5)} = \xi_i^{(2)} + n(\xi_i^{(4)} - \xi_i^{(2)}) \quad (7)$$

a rough calculation gave a value 1.8 for n . This was not considered significantly different from 2, so the double-shift rule was used in the early stages of the work.

A later and more careful analysis of these data by least squares gave 1.62 for n , and a similar calculation based on later parameters gave 1.63. It appeared that the differences between these results and 2.00 are significant and probably ascribable to the effect of symmetry since a large proportion of the structure factors (the prism zones) have signs instead of phases and would be expected to bring the convergence rate closer to that corresponding to a centrosymmetric structure. Indeed, by considering the Fourier synthesis to be the sum of two parts of which one, calculated with only the general planes, is subject to the double-shift rule while the other, calculated with only the prism zones, is not, a theoretical calculation of n was made which gave an average value of 1.66, in good agreement with the experimental values.

(b) **Outline of the Fourier Refinement Process.**—The first two Fourier syntheses were carried out without regard for the hydrogen atoms in the calculation of phases. An effect of this omission was that even after application of the double-shift rule to the trial parameters the agreement between observed structure factors for the three prism zones and ones calculated without regard for the hydrogen atoms was not nearly as good as might have been expected from the quality of the data. A preliminary attempt was made to account for the hydrogen atoms by modifying the form factors of the heavy atoms by the method of Hughes.³² No significant improvement in the structure factor agreement was obtained, however. A third Fourier synthesis was carried out with phases determined by use of these modified

form factors. Both with and without the double-shift correction the resulting parameters gave no significant improvement in structure factor agreement.

There is theoretical reason,²⁹ quite analogous to that which leads to the double-shift rule, to expect that small peaks, of approximately half the amplitude that would correspond to a true hydrogen peak, should exist in the three-dimensional electron density functions calculated with phases containing no contributions from the hydrogen atoms. Examination of the second three-dimensional Fourier synthesis at the positions at which hydrogen atoms were expected to lie revealed small peaks in seven of the nine plausible positions, but there were also present a few additional minor peaks at positions which could not reasonably correspond to the positions of any atoms. The amplitudes at the maxima of all of these minor peaks were of the order of 0.7 electron per Å³, whereas the heights of the carbon, nitrogen, and oxygen peaks were respectively about 8.9, 10.4 and 11.8 electrons per Å.³

Parameters for the nine hydrogen atoms were then assigned in such a way as to obtain a fair compromise between the positions of the minor maxima and the positions expected from bond lengths and bond angles in molecules of known structure. These parameters are listed in Table II. The contributions of the hydrogen atoms to the structure factors in the range $0 < \sin \vartheta < 0.55$ were then calculated, and these contributions were added to the structure factors that had been calculated from the results of the first three-dimensional synthesis. The contributions of the hydrogen atoms outside of this range are negligible. This procedure not only removed most of the major discrepancies in the structure factor agreement but also improved the general agreement significantly. The new phase angles were then used to calculate the fourth Fourier synthesis.

TABLE II
ATOMIC PARAMETERS OF MOLECULE *M*: HYDROGEN ATOMS^a

	<i>x</i>	<i>y</i>	<i>z</i>
H (C ₁₁)	0.387	0.007	0.260
H (C ₁₁₁)	.332	.353	.555
H' (C _{1V})	.230	.427	.235
H'' (C _{1V})	.275	.253	.040
H''' (C _{1V})	.363	.380	.090
H' (N)	.432	-.100	.585
H'' (N)	.488	.050	.760
H''' (N)	.340	-.057	.710
H (O ₁₁)	.155	.226	.535

^a These parameters correspond to positions predicted for the hydrogen atoms from previous structural work on other compounds and from the positions of minor peaks on Fourier Synthesis 4.

The positions of the maxima of carbon, nitrogen and oxygen atoms in the fourth synthesis were essentially identical to those in the second synthe-

(32) E. W. Hughes, *THIS JOURNAL*, **63**, 1737 (1941).

sis, which was calculated without consideration of the hydrogen atoms. The average atomic parameter shift for the heavy atoms was 0.008 Å., and the maximum was 0.027 Å. The positions and heights of the minor peaks on the other hand were found to have changed considerably. The heights of the spurious peaks dropped below 0.7 electron per Å.³, and peaks of height averaging 1.0 electrons per Å.³ were found at all of the positions predicted for hydrogen atoms. That the hydrogen peaks should be increased in height by a factor less than two is a phenomenon which presumably is closely related to the occurrence of a value for n less than two in the n -shift. If parameters are taken directly from the positions of these maxima the lengths of the bonds involving hydrogen atoms differ by an average of 0.11 Å. from the expected values (1.09 Å. for C-H, 1.01 Å. for N-H, and 0.97 Å. for OH), and the bond angles differ by an average of 9° from the expected tetrahedral value (109.5°). Naturally the accuracy with which parameters can be obtained from the hydrogen atoms in this way is not high. Since the positions of the hydrogen peaks conform satisfactorily to the previously assigned positions, and since the uncertainties in these positions are much larger than the remaining errors in the heavy atom positions, the originally chosen hydrogen parameters were used without change in all of the subsequent work.

In Figs. 4a and 4b the fourth Fourier synthesis is summarized. Each peak is represented by contour lines of electron density on a plane which passes through the center of the peak and is parallel to the (010) plane, and all contour lines are projected on the (010) plane. In Fig. 4a the heavy atoms are represented, and in Fig. 4b the hydrogen peaks are given relative to the heavy atom framework.

The double-shift rule was applied to the coordinates (Parameters 9) of the maxima of the carbon, nitrogen, and oxygen peaks of this synthesis (as obtained by a graphical method) to obtain Parameters 10, which were used as the starting point in the calculation of the back-shift corrections ϵ_i . By use of the analytical method better coordinates, Parameters 12, were obtained for the maxima of the fourth synthesis, and the back-shift and double-shift rules were applied to give Parameters 13 (see Table I).

(c) **Comparison of Observed and Calculated Structure Factors.**—Two final sets of structure factors were calculated from Parameter Set 13.

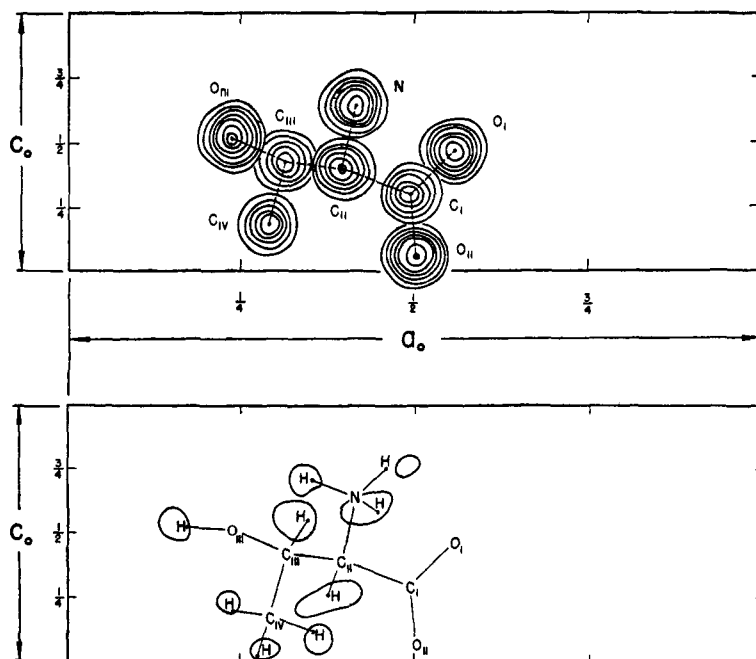


Fig. 4.—Representation of three-dimensional Fourier synthesis 4. The contours given are those on planes parallel to (010) and passing near the maxima of the peaks: (a) above, atoms C, N and O; (b) below, hydrogen atoms.

In one of these, Set No. 6, the hydrogen atoms were not taken into account. In the other, Set No. 7, terms corresponding to the hydrogen atoms at their expected positions (see Table II) were included in the calculation of the structure factors for all reflections for which $\sin \vartheta < 0.55$, and were omitted (on account of their smallness) outside this range. The structure factors of Set No. 7 are compared with the observed structure factors in Table III.

Examination of the table shows very good general agreement and very few instances of pronounced disagreement. As a quantitative measure of the quality of the agreement we have chosen, in accord with the usual practice, to use the percentage discrepancy as defined by Equation (2), although this is not necessarily the best measure that could be devised. The over-all figure of 11.2% ranks favorably with values obtained in other investigations and may be considered as entirely satisfactory by current standards, although it must be realized that the interpretation of differences in the percentage discrepancies for different structures is very complex and is not simply a matter of comparative observational accuracies or of the relative accuracies of the final parameters. This figure is about twice as large as might have been expected solely on the basis of our previous estimates of observational precision. This is true even of the data for low scattering angles, where the calculated structure factors are relatively insensitive to errors in the parameters;

TABLE III
OBSERVED AND CALCULATED (FINAL, SET No. 7) STRUCTURE FACTORS ($|F_{hkl}|$)

The three columns comprising each vertical section are, from left to right, the h index and the observed and calculated structure factors, respectively. All structure factors have been multiplied by ten to eliminate decimal points.

	$h00$	10	22	1	5	91	117	16	41	43	3	50	49	2	17	17	7	118	125
0	2560	11	<25	22	6	27	41		$h21$		4	129	133	3	14	9	8	88	84
2	222 228	12	20	14	7	62	41	0	56	47	5	56	57	4	165	149	9	113	113
4	431 426	13	67	73	8	<22	12	1	492	464	6	58	41	5	200	177	10	83	76
6	124 94	14	22	9	9	<25	18	2	357	324	7	123	132	6	88	62	11	<22	12
8	60 78	15	66	46	10	22	26	3	201	210	8	166	182	7	58	48	12	20	24
10	158 178		$h40$			$h80$		4	316	272	9	86	106	8	27	15	13	20	17
12	166 152	0	123	134	0	48	38	5	164	139	10	28	20	9	25	35	14	41	27
14	37 49	1	210	222	1	67	97	6	123	114	11	52	59	10	104	108		$h42$	
16	69 55	2	242	272	2	60	60	7	178	168	12	42	36	11	92	59	0	65	50
		3	306	302	3	<17	4	8	136	128	13	28	21	12	45	44	1	85	96
	$h10$	4	66	58	4	76	74	9	71	71		$h61$		14	50	48	2	106	138
1	278 297	5	135	118	5	58	67	10	59	46	0	45	44	15	47	44	3	206	206
2	176 130	6	130	116	6	70	70	11	100	93	1	63	81		$h12$		4	72	67
3	278 257	7	184	178	7	22	34	12	109	114	2	40	50	3	46	33	0	23	40
4	403 350	8	136	143	8	51	42	13	101	82	4	133	148	1	141	121	7	113	104
5	132 110	9	33	35		$h90$		14	<20	15	5	131	146	2	115	124	8	92	91
6	124 125	10	22	24	1	46	42		$h31$		6	88	84	3	225	210	9	71	67
7	261 213	11	71	69	2	33	32	0	25	18	7	58	42	4	147	131	10	37	39
8	42 57	12	45	57	3	37	6	1	49	44	8	52	39	5	114	98	11	<22	9
9	41 34	13	50	43	4	<17	4	2	96	75	9	52	47	6	202	204	12	25	18
10	47 44	14	25	23		$h01$		3	67	62	10	<27	23	7	101	93	13	32	31
11	20 26		$h50$			$h01$		4	148	152	11	<17	46	8	46	43		$h52$	
12	181 169	1	30	49	1	321	297	5	73	64	12	<22	25	9	71	77	0	100	79
13	109 111	2	37	14	2	140	114	6	95	79		$h71$		10	74	80	1	52	40
14	<20 1	3	41	40	3	35	44	7	62	61	0	51	63	11	20	10	2	72	77
15	20 15	4	63	80	4	20	11	8	219	226	1	36	15	12	115	107	3	114	121
16	39 32	5	71	76	5	353	321	9	107	116	2	45	44	13	27	25	4	130	142
	$h20$	6	<17	18	6	280	243	10	111	107	3	66	71	14	17	15	5	81	91
0	527 593	7	<17	14	7	95	81	11	<25	28	4	27	29	15	32	40	6	92	73
1	392 379	8	36	21	8	<14	16	12	62	61	5	<20	21		$h22$		7	33	42
2	117 130	9	<20	18	9	150	142	13	80	77	6	25	14	0	361	337	8	66	65
3	360 362	10	20	27	10	112	89	14	63	59	7	74	77	1	65	68	9	62	54
4	503 467	11	17	19	11	49	34	15	52	33	8	45	46	2	166	167	10	75	74
5	170 123	12	17	6	12	226	213		$h41$		9	46	57	3	100	80	11	14	24
6	319 284	13	42	43	13	<25	11	0	91	86	10	30	30	4	66	65	12	53	49
7	165 147		$h60$		14	44	49		$h81$		5	71	53		$h22$		12	53	49
8	<14 0	15	40	44	15	40	44	0	83	68	6	25	18	5	71	53		$h82$	
9	<20 3	16	39	30	16	39	30	1	83	93	6	25	18	7	32	14	0	83	92
10	22 11	0	172	199		$h11$		2	196	189	8	89	105	8	89	105	1	130	143
11	173 165	2	28	46	3	28	46	3	201	204	9	95	79	9	95	79	2	80	87
12	<22 18	3	110	131	0	473	438	4	181	179	2	52	56	10	77	69	3	30	37
13	<22 5	4	88	104	1	301	280	5	113	93	3	52	44	4	<20	6	4	65	58
14	28 48	5	117	129	2	258	243	6	109	94	5	71	81	11	35	21	5	35	42
15	46 29	6	67	67	3	83	75	7	79	84	6	32	25	12	52	50	6	73	62
16	79 109	7	27	3	4	113	114	8	81	75	7	<22	11	13	92	90	7	60	42
		8	73	63	5	250	211	9	88	81	8	32	18	14	30	27	8	66	67
	$h30$	9	20	20	6	271	257	10	33	25		$h91$		15	39	37	9	81	70
1	49 63	10	<27	29	7	90	75	11	47	34		$h32$			$h32$		10	64	62
2	154 144	11	<25	16	8	178	170	12	79	78	0	59	59	0	208	175	11	30	35
3	308 294	12	<17	14	9	70	64	13	35	28	1	24	30	1	181	180		$h72$	
4	323 318		$h70$		10	148	154	14	69	57	2	30	39	2	57	54		$h72$	
5	17 13		$h70$		11	121	100		$h51$		3	32	33	3	192	189	0	41	37
6	58 23	1	61	71	12	22	14		$h51$			$h02$		4	186	173	1	78	83
7	105 99	2	36	27	13	27	31	0	22	24	0	386	362	5	53	33	2	42	40
8	35 20	3	35	58	14	86	84	1	20	27	1	162	163	6	85	85	3	76	74
9	160 148	4	22	20	15	17	16	2	105	109		$h02$			$h02$			$h02$	

TABLE III (Continued)

4	40	51	11	52	46	6	52	38	0	78	92	11	36	36	2	<24	26			<i>h</i> 35
5	83	87	12	65	67	7	69	78		<i>h</i> 04		12	<17	31	3	22	17	0	47	44
6	71	70	13	63	60	8	40	40	0	12	32			<i>h</i> 34	4	32	34	1	51	60
7	63	58	14	14	12	9	56	56	1	48	53	0	142	155	5	17	20	2	73	71
8	32	40				10	44	40	2	17	23	1	57	65	6	<20	18	3	60	60
9	27	30		<i>h</i> 23		11	22	28	3	<17	4	2	27	23			<i>h</i> 74	4	78	79
	<i>h</i> 82		0	103	71	12	<22	15	4	147	153	3	85	89	0	<10	16	5	61	67
0	56	60	1	168	161			<i>h</i> 53	5	89	92	4	131	122	1	44	49	6	20	27
1	65	76	2	98	103	0	56	65	6	48	30	5	42	39	2	40	64	7	<20	11
2	42	36	3	119	123	1	88	77	7	25	29	6	61	51			<i>h</i> 05	8	22	19
3	37	43	4	76	76	2	112	88	8	25	20	7	69	76	1	139	136			<i>h</i> 45
4	49	45	5	73	70	3	87	89	9	89	93	8	96	90	2	101	94	0	52	55
5	27	28	6	57	41	3	87	89	9	89	93	8	96	90	3	47	43	1	93	94
6	35	50	7	54	52	4	99	102	10	92	88	9	101	101	4	<22	7	2	48	44
			8	85	76	5	56	59	11	65	65	10	73	78	5	32	33	3	20	23
	<i>h</i> 03		9	76	82	6	66	58	12	25	8	11	27	24	6	40	41	4	32	30
1	198	184	10	66	54	7	51	41			<i>h</i> 14			<i>h</i> 44	7	20	13	5	20	8
2	173	148	11	64	70	8	45	41	0	33	41	0	17	14	8	66	65	6	37	38
3	<14	13	12	48	56	9	60	67	1	39	25	1	69	84	9	<20	8			<i>h</i> 55
4	217	212	13	17	32	10	14	11	2	86	92	2	62	65						
5	79	74				11	54	58	3	220	229	3	40	48			<i>h</i> 15	0	<10	10
6	37	14		<i>h</i> 33				<i>h</i> 63	4	107	111	4	48	38	0	25	32	1	84	92
7	<25	4	0	33	40	0	20	52	5	59	39	5	28	23	1	28	37	2	31	24
8	70	58	1	33	22	1	95	94	6	60	63	6	36	43	2	84	79	3	35	47
9	27	7	2	74	61	2	23	18	7	174	187	7	<27	8	3	32	27	4	<17	21
10	<25	19	3	129	112	3	20	19	8	70	60	8	28	26	4	77	78			
11	<25	28	4	104	104	4	59	73	9	122	103	9	22	1	5	93	99			<i>h</i> 06
12	45	40	5	194	194	5	63	69	10	<17	9	10	<20	9	6	22	20	0	<11	20
13	<22	11	6	66	60	6	40	35	11	62	62				7	33	30	1	33	34
14	55	51	7	71	56	7	<27	4	12	22	21			<i>h</i> 54	8	61	65	2	<20	23
			8	119	124	8	<20	5			<i>h</i> 24	0	53	46	9	25	27	3	14	10
	<i>h</i> 13		9	121	131	9	<25	11	0	123	140	1	30	25			<i>h</i> 25	4	17	20
0	117	160	10	25	17			<i>h</i> 73	1	14	24	2	51	41						<i>h</i> 16
1	155	136	11	<22	26				2	82	72	3	22	21	0	25	22	0	<10	7
2	183	197	12	22	21	0	25	11	3	96	104	4	78	77	1	57	65	1	14	24
3	36	42	13	25	31	1	74	70	4	<22	9	5	35	42	2	51	56	2	30	32
4	52	53				2	50	54	5	92	88	6	68	74	3	88	92	3	<14	10
5	141	140		<i>h</i> 43		3	42	39	6	79	79	7	22	27	4	25	22			
6	137	146	0	28	51	4	25	16	7	59	59	8	85	82	5	88	84			<i>h</i> 26
7	56	63	1	32	42	5	22	23	8	22	16			<i>h</i> 64	6	56	58	0	28	24
8	123	128	2	71	71	6	39	42	9	59	56	0	<10	11	7	45	37	1	20	24
9	115	118	3	22	37	7	71	87	9	59	56	1	47	43	8	25	27	2	<17	18
10	111	103	4	77	88			<i>h</i> 83	10	<20	10				9	<17	27			
			5	143	123															

for $\sin \vartheta < 0.31$ the percentage discrepancy is 7.7. It seems likely that over the entire range of scattering angles a large contribution is made to the discrepancies by errors in the atomic form factors. The data are probably sufficiently accurate to justify the introduction of additional parameters for the specification of atomic shape and the determination of these parameters by, say, the method of least squares. This has not been done for threonine because of the press of time.

The prismatic zone classes of reflections, rather surprisingly, all have larger percentage discrepancies (14.4 for $(0kl)$, 11.2 for $(h0l)$, and 12.2 for $(hk0)$) than the remaining general (hkl) reflections (10.2).^{32a}

(32a) NOTE ADDED IN PROOF.—One might easily expect the reverse effect, since the general (hkl) data contain a higher proportion

of weak reflections at high scattering angles than do the prism data. However, at least a partial explanation is now possible: the predominating difference seems to lie in the fact that Fourier projections made with zonal data have centers of symmetry in the case of threonine. By use of the two different structure factor amplitude distributions found by A. J. C. Wilson (*Acta Cryst.*, 2, 318 (1949)) for centrosymmetric and non-centrosymmetric F 's, it can be shown that, other things being equal, the denominator in the P.D. expression for a collection of centrosymmetric F 's (having signs) is smaller than that for non-centrosymmetric F 's (having phases) by the factor $2\sqrt{2/\pi}$. The same would be approximately true for the numerator if experimental errors in the F 's were solely responsible for the observed residuals, though the factor in this case may be greater than $2\sqrt{2/\pi}$ because of the fact that the distribution function for centrosymmetric F 's predicts a higher fraction of weak reflections which have larger percentage errors. (The effect of omission of unobserved planes from the P.D. must also be considered.) But if, instead, errors in the atomic form factors (or indeed positional parameters if these have not been adjusted so as to minimize the P.D.) were solely responsible for the observed residuals, it can easily be shown from the theory of propagation of errors that the numerator in the P.D. for

Within the region 0–0.55 in $\sin \vartheta$ the over-all percentage discrepancies with and without hydrogen atoms are respectively 9.6 and 13.0. The difference between these two figures is large enough to show that the omission of hydrogen atoms results in errors in the calculated structure factors that are comparable in order of magnitude with the experimental errors in the observed structure factors. It shows also that the positions assigned to the hydrogen atoms are, in the main, correct.

(d) **Final Fourier Refinement.**—With the use of phases from the last set of structure factors (no. 7) a sixth and final Fourier synthesis was carried out and the peak maxima for the carbon, nitrogen, and oxygen atoms were found by the analytical method. It was from a study of the correlation between the shifts obtained from this synthesis and those obtained from Synthesis 4 that the value 1.63 was obtained, by a least squares analysis, for n in the n -shift, in agreement with the previous value 1.62. By applying the back-shift and the 1.62 shift, the Final Fourier Set of parameters, ξ_i^F was obtained (see Table I)

$$\xi_i^F = \xi_i^{(13)} + 1.62(\xi_i^{(14)} + \epsilon_i - \xi_i^{(13)}) \quad (8)$$

Least Squares Refinement.—Our experience with the double-shift, 1.62 shift, and back-shift rules left us uncertain as to the magnitudes of the uncertainties inherent in these techniques. It therefore seemed desirable to carry out a parallel refinement of the parameters by a method which is subject to errors largely different in character from those peculiar to the Fourier method, in the hope that the magnitudes of the discrepancies between the results obtained from this and the Fourier method could serve as a basis for conclusions regarding the reliability of the determination as a whole. The method chosen was the method of least squares.

The procedure used in applying the method of least squares³⁰ to the refinement of parameters followed that described by Hughes.³² The observational equations (equations of condition) chosen were of the form

the centrosymmetric case should be larger by a factor of $\sqrt{2}$ than for the non-centrosymmetric case, while the factor $2\sqrt{2}/\pi$ still applies to the denominator; the P.D. for the centrosymmetric case is therefore larger than that for the other case by a factor of $\pi/2$ or 1.57. Where errors in both the F 's and the f 's are responsible for the residuals the ratio of the P.D.'s should be intermediate between $\pi/2$ and slightly more than one. A good example seems to be given by *m*-tolidine hydrochloride (F. Fowweather and A. Hargreaves, *Acta Cryst.*, 3, 81 (1950)), where the ratio of the P.D.'s for a centrosymmetric and a non-centrosymmetric projection is 1.3. The situation for threonine is complicated by the circumstance mentioned at the beginning of this note, and yet the ratio of the P.D.'s averages about 1.35, indicating perhaps that the most important contributions to the residuals are the errors in the f 's. It must be emphasized, however, that although the understanding of these effects may be an important aid in interpreting different P.D.'s for a given structure there are certainly other factors which must be taken into account before significant comparisons of P.D.'s for different structures can be made.

$$\sum_{i=1}^{24} \left(\sqrt{w_{hkl}} \frac{\partial |F_{hkl}|_{\text{calcd}}^2}{\partial \xi_i} \right) \Delta \xi_i = \sqrt{w_{hkl}} (|F_{hkl}|_{\text{obs}}^2 - |F_{hkl}|_{\text{calcd}}^2) = \sqrt{w_{hkl}} \Delta F_{hkl}^2 \quad (9)$$

where the $\Delta \xi_i$ are the corrections to be solved for and added to the values of ξ_i used to calculate the quantities $|F_{hkl}|_{\text{calcd}}^2$. No terms were included for the variation of the hydrogen atom parameters, but the contributions of the hydrogen atoms are contained in the $|F_{hkl}|_{\text{calcd}}^2$ and in their derivatives. The derivation of Equation (9) is based on the assumption that the $\Delta \xi_i$ are small enough to be treated as differentials; the left-hand side of (9) is in the form of the standard expression for the differential of a function of several variables. The quantities $|F_{hkl}|^2$ were used rather than $|F_{hkl}|$ in order to simplify the computational procedure and to provide a convenient basis for the assignment of weights. The weighting factors, $\sqrt{w_{hkl}}$, were taken equal to $1/\sigma_{hkl}$,¹⁷ where the σ_{hkl} are our external estimates of the standard errors in the $|F_{hkl}|_{\text{obs}}^2$ as computed from residuals among independent intensity measurements. This weighting procedure is more appropriate than that described by Hughes.³²

The derivatives of the structure factors with respect to the x_j ($j = 1, 2, \dots, 8$) are, for example, of the form

$$\frac{\partial |F_{hkl}|_{\text{calcd}}^2}{\partial x_j} = -16\pi f_{i,hkl} h(A_{hkl} \sin 2\pi h x_i \cos 2\pi k y_j \cos 2\pi l z_j + B_{hkl} \cos 2\pi h x_i \sin 2\pi k y_j \sin 2\pi l z_j) \quad (10)$$

if h , k and l are either all even or all odd. Different combinations of sines and cosines appear in the corresponding expressions for other planes. The calculation of these derivatives and of those with respect to the y_j and z_j was carried out by a punched card method.

The parameter set which was used in the calculation of the derivatives was Set. No. 13, the same set that was used in the calculation of the final set of structure factors (Set No. 7) and as the starting point of the last Fourier synthesis (no. 6). This choice considerably facilitated the calculation of derivatives largely because it was possible to make use of the punched cards that had been used to calculate the final set of structure factors. The derivatives were computed with an IBM 602 multiplying punch in about forty hours of machine operating time. These values were then transferred to a second set of cards such that on each card was punched, in addition to identifying code punches, the twenty-four derivatives associated with a single (hkl) plane.

The normal equation for $\Delta \xi_i$ is, for example

$$\sum_{i=1}^{24} D_{ij} \Delta \xi_i = \sum_{hkl} \left(\sqrt{w_{hkl}} \frac{\partial |F_{hkl}|_{\text{calcd}}^2}{\partial \xi_i} \right) \left(\sqrt{w_{hkl}} \Delta F_{hkl}^2 \right) \quad (11)$$

where

$$D_{ij} = \sum_{hkl} \left(\sqrt{w_{hkl}} \frac{\partial |F_{hkl}|_{\text{calcd}}^2}{\partial \xi_i} \right) \left(\sqrt{w_{hkl}} \frac{\partial |F_{hkl}|_{\text{calcd}}^2}{\partial \xi_j} \right) \quad (12)$$

From the form of Equation (11) it is apparent that this equation is obtained by multiplying each ob-

servational Equation (9) through by the coefficient of $\Delta\xi_i$ in it, and then summing the equations thereby obtained.³⁰ The coefficients D_{ij} were calculated with the 602 punch, about fifty hours of machine time being necessary to calculate the 300 coefficients of the $\Delta\xi_i$ in the 24 normal equations. Since the normal equations proved to be very nearly diagonal, they could be solved quite quickly by an iteration method. The roots computed in this way differed from the roots computed with neglect of the off-diagonal coefficients by an average of only 0.002 Å., the maximum difference being 0.006 Å. Possibly in future work with the three-dimensional least squares method the computation of off-diagonal coefficients might be omitted, as E. W. Hughes has suggested on theoretical grounds, with a consequent material reduction in computational effort.

The final parameters resulting from the least squares procedure are presented in Table I. These new parameters differ from those resulting from the final Fourier refinement by an average of 0.005 Å., the maximum difference being 0.016 Å. and the root-mean-square difference being 0.0069 Å. These differences are of the same order of magnitude as the shifts themselves. At this point the parameters obtained by both methods seem to be about as close to ultimate convergence as they will ever be, and further work would probably only reduce somewhat the effect of certain computational errors, which at this stage are presumably no larger than the errors inherent in the Fourier and least square methods and are significantly smaller than the irreducible errors of the determination as a whole.

Final Parameters and Precision of their Determination.—To obtain the final set of parameters, given in Table I, the arithmetic mean of the results of the final Fourier refinement and of the least squares refinement was taken, on the assumption that the two methods, at this stage of refinement, are equally reliable.

The ultimate attainable accuracy of the parameters as determined by either the Fourier method or the least squares method of interpreting the experimental data depends upon two factors: (1) the accuracy of the experimental data, and (2) the correctness of the method. Among the errors in the interpretational methods are computational errors, including especially errors due to rounding off the trigonometric functions on the punched cards, and also errors in the assumed atomic form factors, in the temperature factor, in the determinations of the positions of Fourier maxima, in the n -shift and back-shift rules, in the assumption of linearity in the least squares treatment, and perhaps others. Most of these are peculiar to one or the other of the two interpretive methods or have effects which are not correlated between the two methods. If this were true of all errors of interpretation an estimate of their collective effect could be derived from an analysis of the differ-

ences between the results given by the two methods. It might then be quite proper, in assessing the dependence of the parameter accuracy on the accuracy of the experimental data, to make use of the previously mentioned external estimates of error. However, the effects of errors in the atomic form and temperature factors in the two methods are presumably related in such a way that they are not adequately represented in the differences between the two sets of results. Since the errors in the form factors constitute an important part of the residuals between observed and calculated structure factors, and since it is difficult to see how otherwise to take proper account of these errors, it seems appropriate and reasonably conservative to regard these residuals as *internal estimates* of errors in the data, thereby in effect lumping the form factor errors together with the actual observational errors, and to use these internal estimates in place of the external estimates in the equations given below.

The effect of these internally estimated experimental errors is taken into account in the Fourier method as follows.²⁹ It is assumed that the Fourier procedure, carried out with the erroneous observed structure factors, has converged. If phases computed from the final parameters ξ_i^F thereby obtained are used in a new Fourier synthesis together with the *correct* structure factors, the shifts obtained will be $1/n$ times the differences between the correct parameters and the erroneous ones, by virtue of the n -shift rule. The magnitudes of the shifts themselves can be calculated from the differences between the erroneous and correct structure factors by a straight-forward application of the theory of propagation of errors. By this means we find that

$$\sigma_i^F = \frac{2\pi n}{V \left| \frac{\partial^2 \rho}{\partial \xi_i^2} \right|_0} \sqrt{\sum_{hkl} h_{i,hkl}^2 (|F_{hkl}|_{\text{obs}} - |F_{hkl}|_{\text{calcd}})^2} \quad (13)$$

where σ_i^F is the standard error in a parameter ξ_i^F , n is the parameter of the shift rule, V is the volume of the unit cell, $h_{i,hkl}$ is the magnitude of the reciprocal lattice vector \mathbf{h}_{hkl} in the direction corresponding to increasing ξ_i , the second derivative in the denominator is evaluated at the maximum of the peak for the atom having ξ_i as one of its three coördinates, and the summation is taken over all planes in the sphere of reflection except those too weak to estimate. The derivation does not take into account the effects of the symmetry of the crystal, except as it allows n to be given an appropriate value less than 2, but these will not change the order of magnitude of the results. The expression is closely related to those derived by Hughes and Lipscomb³³ and Cox and Cruickshank,³⁴ but differs from them by

(33) E. W. Hughes and W. N. Lipscomb, *THIS JOURNAL*, **68**, 1970 (1946).

(34) E. G. Cox and D. W. J. Cruickshank, *Acta Cryst.*, **1**, 92 (1948); D. W. J. Cruickshank, *ibid.*, **2**, 65 (1949); **2**, 154 (1949).

inclusion of the factor n (which is equal to unity for a centrosymmetric crystal).^{34a}

The expression for the standard error of a parameter ξ_i^{LS} in the least squares method is³⁰

$$\sigma_i^{LS} = \sqrt{\frac{\sum_{hkl} w_{hkl} (|F_{hkl}|_{\text{obs}}^2 - |F_{hkl}|_{\text{calcd}}^2)^2}{m - s}} D_{ii}^{-1} \quad (14)$$

where m is the number of observational equations, s is the number of parameters, and D_{ii}^{-1} is the i^{th} diagonal element of the matrix inverse to that of the D_{ij} . Since the off-diagonal matrix elements D_{ij} , and hence also the off-diagonal D_{ij}^{-1} , are small, we write as an approximation that

$$D_{ii}^{-1} \cong \frac{1}{D_{ii}} = \frac{1}{\sum_{hkl} w_{hkl} \left(\frac{\partial |F_{hkl}|_{\text{calcd}}^2}{\partial \xi_i} \right)^2} \quad (15)$$

In Table IV are presented the standard errors as calculated with the above expressions for both the Fourier and the least squares determinations. They are, as might be expected, comparable in magnitude. No significance should be attached to the fact that the figures obtained for the Fourier method are usually somewhat smaller than those obtained for the least squares method; indeed, it should be expected that the errors in the Fourier method are larger than those in the least squares method since the weighting of the data is more appropriate in the latter than in the former.

Let us now consider errors due to defects inherent in the interpretive methods for which we shall take, as a measure, the differences between the results obtained by the two different methods. If it is assumed that the inherent errors of the two methods are approximately the same on the average and are independent, a value of 0.00345 Å. may be assigned to the standard error in the final parameters due to these defects. On combining this figure with the internal estimates of standard errors due to errors in the experimental data and form factors, we arrive at what may be regarded as sufficiently conservative final values for the standard errors σ_i in the parameters. These are listed in Table IV. Their mean, 0.0075 Å., may be taken as representative of the determination as a whole. The corresponding standard error in an interatomic distance is 0.0106 Å.; the probable error in an interatomic distance is therefore 0.0071 Å. If we take three times this probable error as the limit of error in the determination of an interatomic distance, we arrive at a value of 0.021 Å. for this quantity. The corresponding limit of error for bond angles is about 1.0° if the angle is close to tetrahedral and the bond lengths are close to 1.5 Å.

(34a) NOTE ADDED IN PROOF.—Cruikshank (*Acta Cryst.*, 3, 72 (1950)) has recently revised his previous treatments and has obtained 2 for the ratio of estimated parameter errors in non-centrosymmetric and symmetric crystals. However, it must be emphasized, the value 2 is correct only under the rather restricted conditions of applicability of the double shift rule. In practical cases a ratio of n , the parameter of the n -shift, would be more nearly correct.

TABLE IV
STANDARD ERRORS OF PARAMETERS

Atom	$\sigma_i^{F^a}$ Å., $\times 10^3$	$\sigma_i^{LS^a}$ Å., $\times 10^3$	σ_i^b Å., $\times 10^3$	
C ₁	x	6.4	7.9	8.6
	y	6.8	6.6	7.6
	z	5.1	7.1	7.9
C ₁₁	x	7.5	7.8	8.5
	y	7.3	7.1	8.1
	z	5.8	7.9	8.6
C ₁₁₁	x	6.3	7.9	8.6
	y	6.7	6.6	7.5
	z	6.2	7.2	8.0
C _{1v}	x	7.9	7.4	8.6
	y	9.1	6.9	9.7
	z	7.1	7.4	8.2
N	x	5.3	6.1	7.0
	y	6.3	6.0	7.2
	z	5.2	6.9	7.7
O ₁	x	5.2	5.4	6.4
	y	6.7	4.7	7.5
	z	4.8	5.2	6.3
O ₁₁	x	5.3	5.6	6.6
	y	5.7	4.6	6.6
	z	4.1	4.8	5.9
O ₁₁₁	x	5.0	5.2	6.3
	y	5.9	4.5	6.8
	z	4.8	5.2	6.2
Average	6.1	6.3	7.5	

^a Standard errors as calculated from internal estimate of experimental errors by means of equations (13) and (14). ^b Calculated from the larger of the values in the first two columns, and containing an additional contribution to the error due to defects in the methods on the assumption of a corresponding standard deviation of 0.00345 Å. for the final parameters (calculated from differences in the parameters obtained by the two methods).

Discussion of the Structure

The Threonine Molecule.—The shape of the threonine molecule, as it exists in the crystal, is given in Fig. 5. The bond lengths and the bond angles are given both in Fig. 5 and in Table V.

TABLE V
INTRAMOLECULAR INTERATOMIC DISTANCES AND BOND ANGLES

	Distance		Angle
C ₁ —C ₁₁	1.517 Å.		
C ₁₁ —C ₁₁₁	1.542		
C ₁₁₁ —C _{1v}	1.505	O ₁ —C ₁ —O ₁₁	126.9°
C ₁₁ —N	1.490	O ₁ —C ₁ —C ₁₁	117.0°
C ₁ —O ₁	1.236	O ₁₁ —C ₁ —C ₁₁	116.1°
C ₁ —O ₁₁	1.253	C ₁ —C ₁₁ —C ₁₁₁	113.4°
C ₁₁₁ —O ₁₁₁	1.424	C ₁ —C ₁₁ —N	110.4°
		C ₁₁ —C ₁₁₁ —C _{1v}	112.5°
		C ₁₁ —C ₁₁₁ —O ₁₁₁	104.1°
N ... O ₁	2.672	N—C ₁₁ —C ₁₁₁	108.0°
N ... O ₁₁₁	2.678	O ₁₁₁ —C ₁₁₁ —C _{1v}	110.5°
C ₁ ... C _{1v}	3.084		
O ₁₁ ... C _{1v}	3.106		

The shape of the molecule, relative to atoms C_{II} and C_{III}, corresponds to the staggered (D_{3d}) configuration shown by ethane. Of the three positions available for a group attached to C_{III} presumably the most crowded would be that one which is close to both C_I and N, and it might be expected that in the most favorable configuration of the molecule this position would be occupied by the hydrogen atom, while the other two positions on C_{III} would be occupied by the hydroxyl group and the methyl group. This expectation is realized in the molecule as it actually exists in the crystal.

The relative stereochemical configuration around the two asymmetric carbon atoms is identical to that assigned by Meyer and Rose¹⁸ from the results of their chemical work. The conclusions of Meyer and Rose, taken by themselves, have been generally regarded as being of well established validity. However, there seems to be reason for assuming that additional confirmatory evidence, such as that provided by the present crystallographic investigation, is not altogether superfluous, since the possibility of a Walden inversion in the reaction with nitrous acid, partly on the basis of which the *threo*-configuration was assigned, cannot be altogether ruled out. The results of the present investigation, by proving that (–)threonine has the *threo*-configuration, definitely rules out the Walden inversion in this particular case. It appears that the present unambiguous establishment of the relative configuration of the two centers has, in other ways as well, important bearing on the whole problem of the establishment of stereochemical relationships between the amino acids on one hand and lactic acid and the sugars on the other.³⁵

The α C–N distance was found to be 1.490 Å., in agreement with the sum of the covalent radii.³⁶ This result has helped materially in clearing up a doubtful feature of previously published dimensions of amino acid molecules: in alanine,⁵ the α C–N distance was reported to be 1.43 Å., and in the less precisely determined glycine structure⁴ the distance was reported as 1.39 Å. Trial and error refinement methods were necessarily used in these earlier investigations, and it was not feasible to establish objective limits of error for the results. The experience obtained in the present investigation has made it possible for one of the present authors³⁷ to redetermine the atomic parameters for DL-alanine from the data originally published.⁵ The revised dimensions thus obtained for the alanine molecule are substantially the same as those previously reported except for the α C–N distance, which is found to be 1.50 Å., in entire agreement with that of threonine. The value 1.51 Å. has been reported for this distance in β -glycylglycine.³⁸

(35) See A. Neuberger, *Advances in Protein Chem.*, **4**, 297 (1948).

(36) L. Pauling, "The Nature of the Chemical Bond," The Cornell University Press, Ithaca, N. Y., 1940, p. 164.

(37) J. Donohue, *THIS JOURNAL*, **72**, 949 (1950).

(38) E. W. Hughes, *ibid.*, **71**, 2618 (1949).

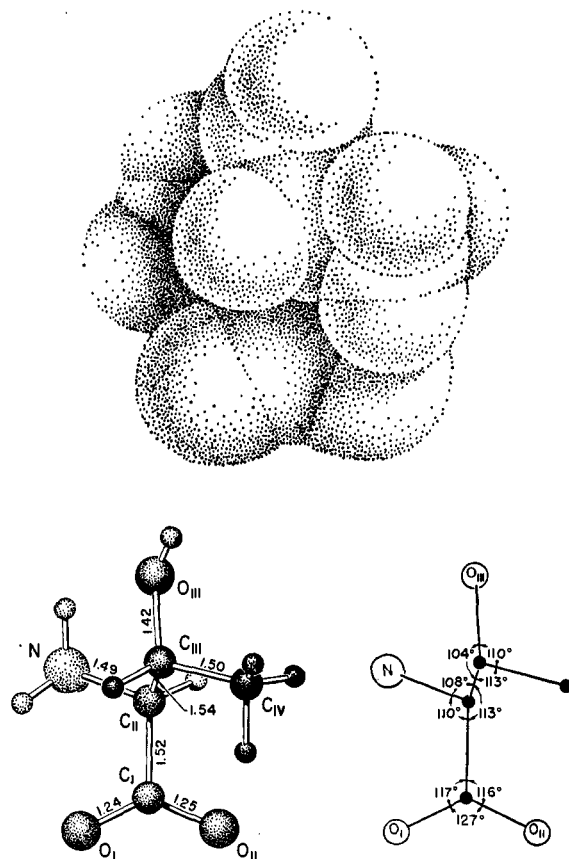


Fig. 5.—The threonine molecule, showing interatomic distances and bond angles.

The C_{III}–C_{IV} bond seems to be abnormally short with a distance of 1.505 Å.; no reason for this has been found. Two of the bond angles show very considerable deviations from the expected tetrahedral angle of 109.5°. These are C_{II}–C_{III}–O_{III} with 104.1°, and C_I–C_{II}–C_{III} with 113.4°. Although no explanation has been found for these values, there can be no doubt of the significance of their differences from the tetrahedral angle, as these differences are several times the limits of error in these two bond angles.

The Molecular Environment.—In Figs. 6a, 6b, and 7 are given, respectively, views of the structure looking along the *c* axis in the direction of decreasing *x*, along the *b* axis in the direction of increasing *y*, and along the *a* axis in the direction of decreasing *x*. The origin and cell edges are the same in every drawing, and the system of axes is right-handed. The positions of screw axes conform to the usage given in the International Tables.²⁷

In all three drawings the particular molecule the environment of which is to be discussed in detail is designated *M*, with atomic coordinates (*x*, *y*, *z*) (see Table I). The three other molecules situated mainly in the same unit cell, and related to the molecule *M* by three screw axes

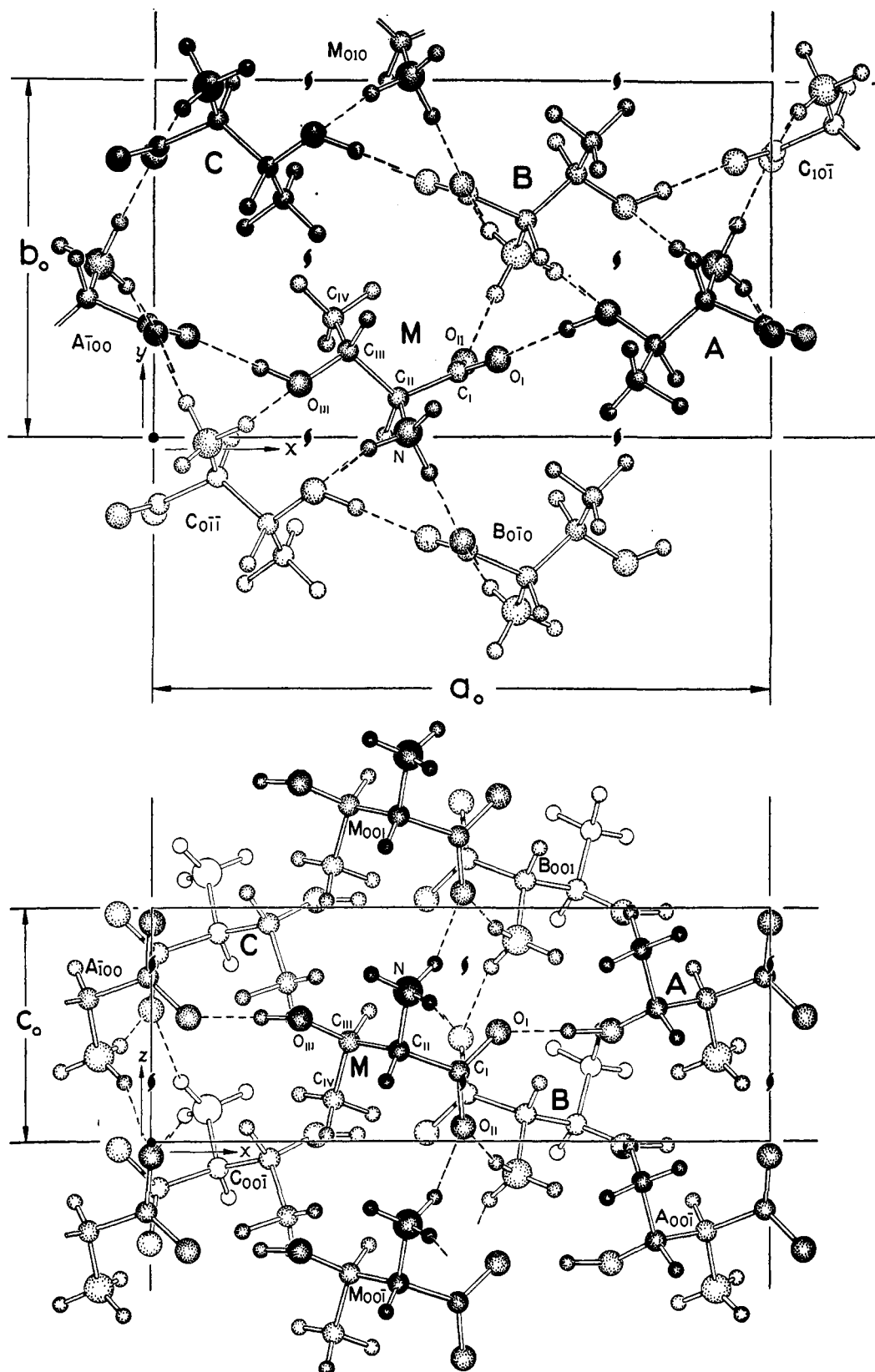


Fig. 6.—Views of the structure: (a) above, looking along the c -axis in the negative direction; (b) below, looking along the b -axis in the positive direction. Hydrogen bonds are shown as broken lines.

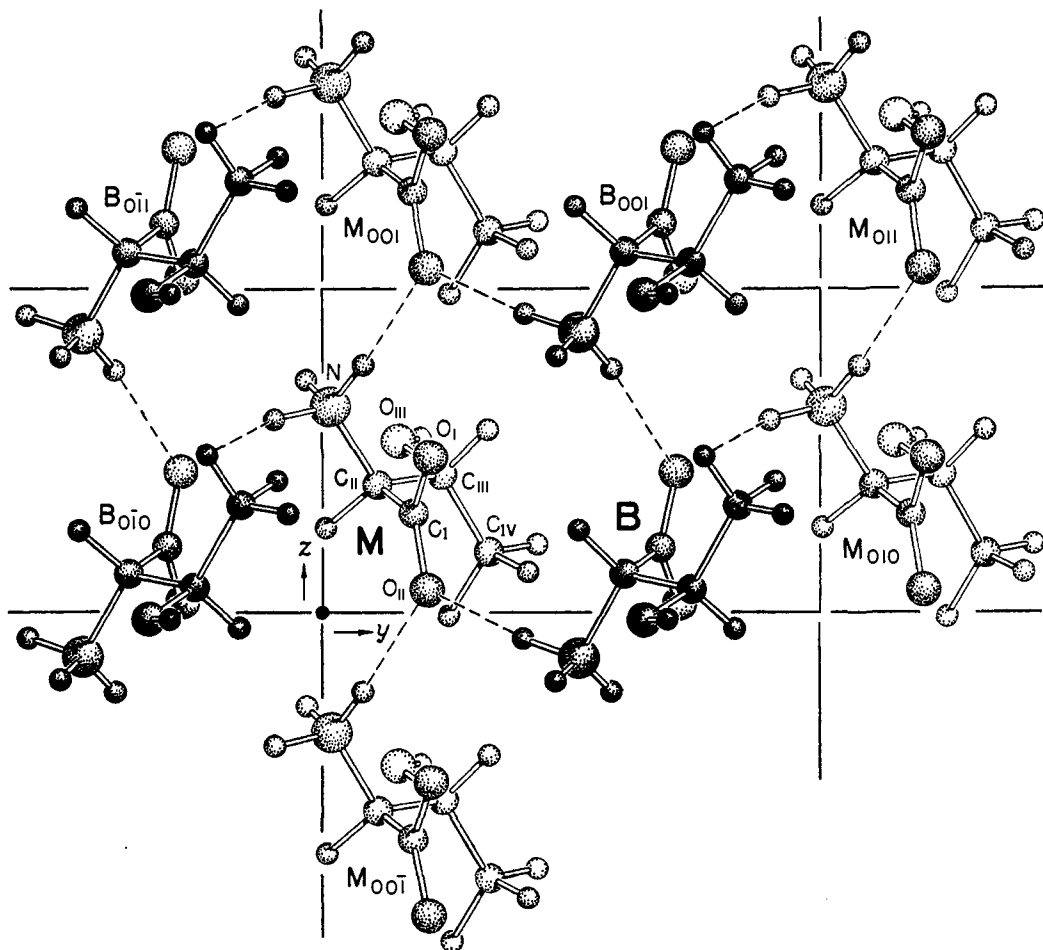


Fig. 7.—View of the structure looking along the *a*-axis in the negative direction. Only molecules of types *M* and *B* are shown.

parallel respectively to *a*, *b*, and *c*, are designated *A*, *B*, and *C*; the atomic coordinates are $(\frac{1}{2} + x, \frac{1}{2} - y, 1 - z)$, $(1 - x, \frac{1}{2} + y, \frac{1}{2} - z)$, and $(\frac{1}{2} - x, 1 - y, \frac{1}{2} + z)$. Other molecules, in adjacent unit cells, are designated by the same letters but with subscripts indicative of the lattice translation vectors relating these molecules with the unsubscripted molecules *M*, *A*, *B*, and *C*; M_{mnp} , for example, occupies the position which would be occupied by molecule *M* if the latter were translated through the vector $ma + nb + pc$.

The intermolecular contacts, including hydrogen bonds, involving molecule *M* are summarized in Table VI. We shall adopt the convention of disregarding distances involving hydrogen atoms in this table and, for the most part, in the ensuing discussion, for it was not possible to make a precise determination of the parameters of the hydrogen atoms in the present investigation.

In crystals of related compounds, such as glycine and alanine, it was found that an important part in intermolecular cohesion is played by hydrogen bonds between the amino groups and the carboxyl oxygen atoms. The same would cer-

tainly be expected to be true for threonine. Indeed the threonine structure appears to be tied together in two dimensions by hydrogen bonds from amino groups to carboxyl oxygen atoms and by salt linkages (see Fig. 7), but in the third dimension mainly by hydrogen bonds from hydroxyl groups to carboxyl oxygen atoms (Fig. 6).

We shall begin our discussion of the molecular environment with a description of the environment of the amino groups, with particular attention to the hydrogen bonds. Cohesion in the *c* direction in the structure apparently derives largely from hydrogen bonds which unite the molecules into infinite vertical chains, *e. g.*, $M, M_{001}, M_{002} \dots$ (see Fig. 6b), and cohesion in the *b* direction appears to derive largely from hydrogen bonds which unite the molecules into chains such as $M, B, M_{010}, B_{010}, M_{020}, \dots$ (see Fig. 6a). These conclusions follow from the following facts. The nitrogen atom of molecule *M* and the carboxyl oxygen atom O_{II} of molecule M_{001} (or, alternatively, atom *N* of $M_{00\bar{1}}$ and atom O_{II} of *M*) are 2.90 Å. apart, and atom *N* of *M* and atom O_{II} of $B_{0\bar{1}0}$ are 2.80 Å. apart. These distances

TABLE VI

INTERMOLECULAR INTERATOMIC DISTANCES AND ANGLES

From atom (X) on molecule M	To atom (Y) on molecule (1)	Equiv. contact to molecule (2)	Distance X...Y Å.	Angle CX...Y	Angle X...YC	
A. Hydrogen bond contacts						
N	O ₁₁	M ₀₀₁	M _{00i}	2.90	116°	128°
O ₁₁₁	O ₁	A ₁₀₀	A	2.66	120°	136°
N	O ₁₁	B ₀₁₀	B	2.80	98°	122°
N	O ₁₁₁	C ₀₁₀	C _{01i}	3.10	132°	120°
B. Other distances ^b						
N	C ₁₁	M ₀₀₁	M _{00i}	3.92		
C ₁₁₁	C _{1V}	M ₀₀₁	M _{00i}	3.94		
N	C _{1V}	M ₀₀₁	M _{00i}	3.73		
O ₁	O ₁₁	M ₀₀₁	M _{00i}	3.14		
O ₁	C _{1V}	A	A _{i00}	3.67		
C _{1V}	O ₁	B	B ₀₁₀	3.69		
C ₁	C ₁₁	B	B ₀₁₀	3.71		
O ₁	C ₁₁	B	B ₀₁₀	3.59		
O ₁	O ₁₁	B	B ₀₁₀	3.92		
O ₁₁	O ₁	B	B ₀₁₀	4.00		
N	O ₁	B ₀₁₁	B ₀₀₁	3.08	161°	134°
C _{1V}	C _{1V}	C	C _{00i}	3.79		
C ₁₁₁	C _{1V}	C	C _{00i}	3.81		
O ₁₁₁	C ₁₁	C ₀₁₀	C _{01i}	3.28		
O ₁₁₁	C ₁₁₁	C ₀₁₀	C _{01i}	3.74		
O ₁₁₁	C _{1V}	C ₀₁₀	C _{01i}	3.83		
N	C _{1V}	C ₀₁₀	C _{01i}	3.87		

^a In every case where the contact is from X on M to Y on molecule (1), the equivalent contact is from Y on M to X on molecule (2). ^b All distances of 4 Å. or less, excepting those listed in part A and those involving hydrogen atoms, are given here.

are comparable with N...O distances in N—H...O contacts ascribed to hydrogen bonds in other crystals (*e. g.*, 2.98 and 3.03 Å. in urea,³⁹ 2.85 Å. in diketopiperazine,³ 2.76 and 2.88 Å. in glycine,⁴ and 2.84 and 2.88 Å. in alanine⁵). The assumption of hydrogen bonds of course requires that hydrogen atoms lie nearly between the nitrogen and oxygen atoms, and that the H...O distances be considerably shorter than the expected van der Waals contact distance (2.6 Å.)⁴⁰ between hydrogen and oxygen atoms. It was found that the conditions could be fulfilled in both cases by an assignment of hydrogen atoms to positions 1.0 Å. from the nitrogen atom in conformity with tetrahedral coordination; the C—N...O angles in the two supposed hydrogen bonds are respectively 116° and 98°, neither of which is so far from the tetrahedral angle 109.5° as to preclude the existence of hydrogen bonds. This consideration played an important part in deciding the probable positions of the hydrogen atoms in the crystal. Strong confirmation of these assignments was obtained, as already described, from the existence of small peaks at or near the expected positions in the Fourier syntheses, both before and after inclusion of terms

(39) R. W. G. Wyckoff and R. B. Corey, *Z. Krist.*, **89**, 462 (1934).

(40) L. Pauling, *op. cit.*, p. 189.

due to hydrogen in the computation of structure factors, and from the improved agreement obtained in the structure factors by the inclusion of the hydrogen atoms.

The third hydrogen atom on the amino group of M lies not very far from a line connecting atom N with atom O₁₁₁ of the hydroxyl group of molecule C₀₁₁ (see Fig. 6a). The N...O distance is 3.10 Å., or somewhat longer than the N—H...O hydrogen bonds already described, and the angle C—N...O is 132°, which differs considerably from the tetrahedral angle. The calculated H...O₁₁₁ distance is approximately 2.1 Å. If a hydrogen bond exists here, it is a much weaker one than those already discussed.

Another close contact involving the amino group on M is with atom O₁ of the carboxyl group of molecule B₀₁₁ (see similar contact of O₁ of M with N of B₀₀₁, Fig. 6b). The N...O distance is 3.08 Å., which exceeds the sum of the nitrogen and oxygen van der Waals radii, 2.9 Å. None of the three hydrogen atoms on the amino group is in the neighborhood of the straight line between the oxygen and nitrogen atoms. The shortest H...O distance is approximately 2.7 Å., which is larger than 2.6 Å., the sum of the hydrogen and oxygen van der Waals radii. For a hydrogen bond the H...O distance should not greatly exceed 1.8–2.0 Å. There is certainly no hydrogen bond here; the closeness of the nitrogen and oxygen atoms is merely a consequence of normal van der Waals packing, aided by the electrostatic force of attraction between the positively charged amino group and the negatively charged oxygen atom.

The only remaining hydrogen bond to be discussed is that from atom O₁₁₁ of the hydroxyl group of molecule M to atom O₁ of the carboxyl group of molecule A₁₀₀ (see Figs. 6a and 6b). The O₁₁₁...O₁ distance is 2.66 Å., which is somewhat less than twice the 1.4 Å. oxygen van der Waals radius. The C₁₁₁—O₁₁₁...O₁ angle is 120°, which is close enough to the tetrahedral angle to be favorable for a hydrogen bond. The distance may be compared with the O—H...O distance of 2.76 Å. in ice,⁴¹ 2.69 Å. in pentaerythritol,⁴² 2.67 Å. in gaseous formic acid dimer,⁴³ and about 2.6 Å. in oxalic acid.⁴⁴ Presumably this hydrogen bond plays the predominant role in providing cohesion in the a-direction in the structure.

Only the normal van der Waals contacts need to be mentioned to complete the description of the environment of the amino and hydroxyl groups. The atom closest to the nitrogen atom of molecule M, excepting hydrogen atoms and other atoms discussed above, is atom C_{1V} of the methyl group of molecule M₀₀₁. The N...C distance

(41) W. H. Barnes, *Proc. Roy. Soc. (London)*, **A125**, 670 (1929).

(42) F. J. Llewellyn, E. G. Cox and T. H. Goodwin, *J. Chem. Soc.*, 883 (1937).

(43) J. Karle and L. O. Brockway, *This Journal*, **66**, 574 (1944).

(44) J. D. Dunitz and J. M. Robertson, *J. Chem. Soc.*, 142 (1947).

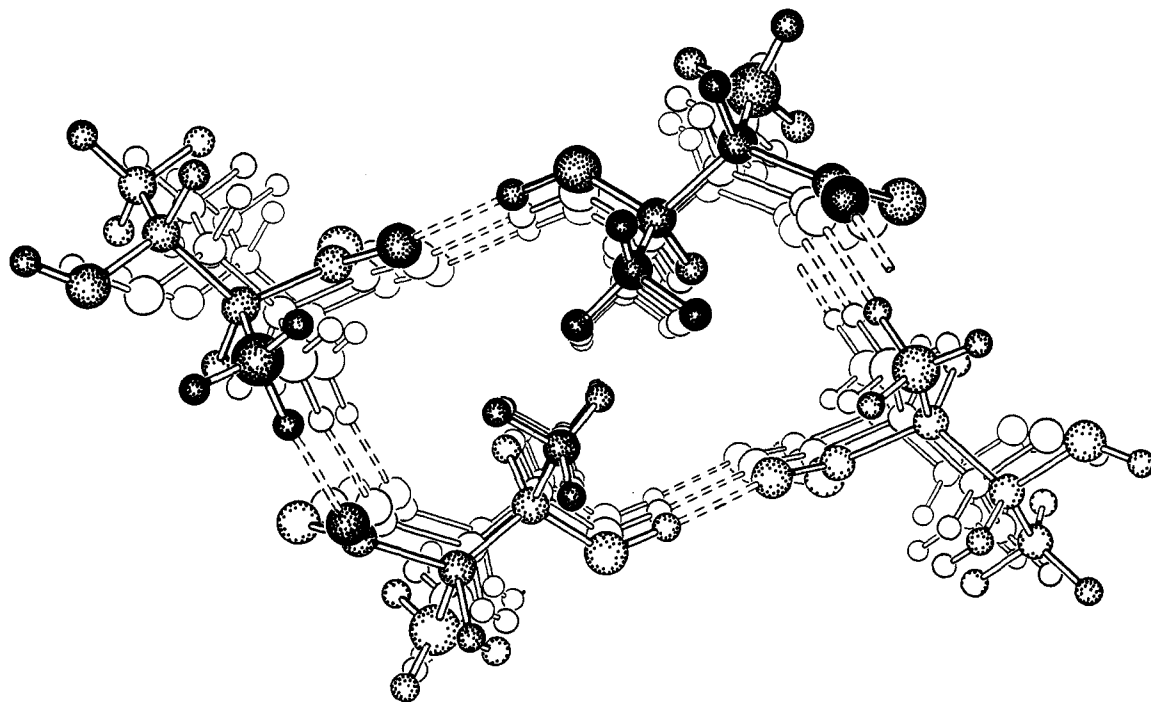


Fig. 8.—View of the structure looking along a screw axis parallel to the *c*-axis, showing part of an infinite spiral of molecules connected by hydrogen bonds.

is 3.73 Å., which is somewhat shorter than the sum of the expected van der Waals radii of the methyl group, 2.0 Å., and the amino group, 1.9 Å. Other neighbors are atom C_{IV} of the methyl group of molecule C₀₁₀, 3.87 Å. away, and atom C_{II} of M₀₀₁, 3.92 Å. away. Near atom O_{III} on *M* is atom C_{II} of molecule C₀₁₀, at the rather short distance of 3.28 Å. The H ··· O distance here is about 2.4 Å., which is less than the expected distance of 2.6 Å. Atom C_{III} and C_{IV} of the same molecule are at distances of 3.74 and 3.83 Å., respectively, from atom O_{III} on *M*.

Much of the environment of the carboxyl group has already been described in connection with the hydrogen bonds. Atom O_I forms one hydrogen bond, with the hydroxyl group of molecule *A*, and atom O_{II} forms two hydrogen bonds, one with the amino group of molecule M₀₀₁ and the other with the amino group of molecule *B*. The only other close contact aside from those with amino groups is between atom O_I of *M* and atom O_{II} of M₀₀₁, with a distance of 3.14 Å. The next nearest neighbor of atom O_I of *M* is atom C_{II} of *B*, at a distance of 3.59 Å. Other distances involving the carboxyl group are listed in Table VI.

The methyl groups lie fairly close to one of the vertical screw axes (see Figs. 6a and 8). In consequence, the methyl groups of molecules *M* and *C* or C₀₀₁ are near neighbors, with a C ··· C distance of 3.79 Å., somewhat less than twice the customary van der Waals radius of 2.0 Å. for the methyl group. This distance is larger,

however, than it is in alanine, in which the methyl groups, which are also close to a screw axis, are 3.64 Å. apart.

Though the strong cohesive forces which give the crystal a relatively high density, hardness, and resistance to melting are largely provided by hydrogen bonds, forces of electrostatic attraction must also play a very important part. As the negatively charged carboxyl groups lie very close to planes parallel to (100) and 6.8 Å. apart, and as the positively charged amino groups are only 1.3 Å. from the centers of the nearly rectangular figures defined on these planes by the carboxyl groups (see Fig. 7), it is apparent that the electrostatic forces are confined closely to these planes and contribute to cohesion in the *c*- and *b*-directions but not appreciably to cohesion in the *a*-direction. These forces, as well as the hydrogen bonds between amino and carboxyl groups, connect the molecules into infinite sheets parallel to the (100) plane. Such a sheet is, for example, formed by the molecules M_{0*np*} and B_{0*n'**p'*}. These are the molecules shown in Fig. 7, from which molecules *A* and *C* have been omitted for clarity. Since the cohesive forces within such a sheet are presumably stronger than those between molecules in different sheets, the structure may well be regarded as consisting of parallel sheets of this kind fitted together snugly and tied firmly by hydrogen bonds from hydroxyl to carboxyl groups.

The conclusion that the cohesive forces within the sheets are stronger than the forces between

them appears to provide a satisfactory explanation of the observed cleavage parallel to the (100) plane.

The distribution of hydrogen bonds and electrostatic forces which we found in crystals of L-threonine conforms to the same general principles which were suggested by previous structural work on amino acids and related compounds and therefore provides additional confirmation of previously formed conceptions⁷ as to the way in which these bonds and forces influence and maintain the configurations assumed by protein molecules under various conditions. For example, recent experiments on the binding of organic ions by proteins⁴⁵ have indicated that in many native proteins there is a preferential binding of hydroxyl groups to carboxyl groups rather than to quaternary nitrogen groups. Some confirmation is lent to this indication by the crystal structure of L-threonine, in which the only strong hydrogen bond involving the hydroxyl group of a given molecule is formed with a carboxyl oxygen atom. Further evidence concerning this possible generalization should be provided by the crystal structures of other hydroxyamino acids now under investigation in these Laboratories.

Acknowledgments.—Most of the computational work was done by Miss Lillian Casler and Mrs. Jean Smith, and the figures were prepared by Mrs. Maryellin Reinecke. The authors are grateful to Professor Linus Pauling for his interest, encouragement, and helpful criticism. They wish to express their thanks to Professor Werner Nowacki of Bern, Switzerland, who assisted in the work with the structure factor maps and participated in the computation of the two-dimensional Fourier syntheses and of a second Patterson function. Grateful acknowledgment is also expressed for the financial support of the National Foundation for Infantile Paralysis, Inc.

Summary

Crystals of threonine (*threo*- α -amino- β -hydroxy-*n*-butyric acid), whether prepared from the L isomer or the racemic mixture, were found to give the same X-ray diffraction patterns and to have an orthorhombic unit ($a_0 = 13.61 \text{ \AA}$., $b_0 = 7.74 \text{ \AA}$., $c_0 = 5.14 \text{ \AA}$.) with the space group D_2^4 - $P2_12_1$. Evidently crystallization from the racemic mixture was accompanied by resolution of the isomers.

By means of Weissenberg photography with copper radiation, intensity data were obtained for all planes represented by reciprocal lattice points lying in a sphere of radius 1.2 \AA^{-1} . Generally more than one independent measurement was made for each plane, thereby providing im-

proved over-all precision and enabling an estimate to be made of the observational uncertainties. A standard error of about 7% was found for individual measurements of intensity for strong and medium reflections with neglect of errors in the film factors; inclusion of these errors gives about 9%. The approximate structure was found with the aid of a complete three-dimensional Patterson function. The refinement of parameters was carried out principally by means of three-dimensional Fourier syntheses, of which a total of six were computed. A novel feature of one of the electron density functions, even though the phases used for it were calculated without regard for hydrogen atoms, is the appearance of small peaks at positions expected for most of the hydrogen atoms. When hydrogen atoms at the expected positions were taken account of in the calculation of structure factors and phases the subsequently calculated electron density function showed the presence of small peaks corresponding to all the hydrogen atoms, and the agreement between observed and calculated structure factors was very much improved. A three-dimensional least-squares refinement was also carried out. In the course of the Fourier work the rate of convergence of the Fourier procedure was studied; it was found theoretically that for an asymmetric structure the average parameter shift obtained in a single Fourier step is about one-half that needed to obtain the final structure, and for threonine it was found both theoretically and experimentally that the corresponding fraction is about $1/1.6$. A careful study was made of the factors entering into the precision of the final results, and it was found that the standard error of a parameter determination is 0.008 \AA . The limit of error in the determination of an interatomic distance (defined as three times the probable error in such a distance) was found to be 0.02 \AA .

The molecule was found to have the *threo* structure, in conformity with previous chemical evidence. The bonded interatomic distances found within the molecule are carboxyl carbon to α -carbon, 1.517 \AA .; α -carbon to β -carbon, 1.542 \AA .; β -carbon to methyl carbon, 1.505 \AA .; carbon to nitrogen, 1.490 \AA .; carbon to carboxyl oxygen, 1.236 and 1.253 \AA .; and carbon to hydroxyl oxygen, 1.424 \AA . The molecules are tied together by a three-dimensional network of hydrogen bonds. The amino group forms two hydrogen bonds of lengths 2.90 and 2.80 \AA ., and the hydroxyl group forms a hydrogen bond of length 2.66 \AA . All of these hydrogen bonds are to carboxyl oxygen atoms; one of the carboxyl oxygen atoms is hydrogen bonded to two amino groups and the other to a hydroxyl group. The methyl groups, as in DL-alanine, are close to a two-fold screw axis, but are farther apart (3.79 \AA .) than they are in alanine (3.64 \AA .). Electrostatic binding forces are essentially confined

(45) I. M. Klotz and J. M. Urquhart, *This Journal*, **71**, 1597 (1949).

to layers of molecules parallel to the (100) plane, and these layers are bonded to each other by hydroxyl to carboxyl hydrogen bonds. The

observed cleavage parallel to (100) seems to be satisfactorily explained.

PASADENA 4, CALIFORNIA RECEIVED SEPTEMBER 12, 1949

[CONTRIBUTION NO. 50 FROM THE INSTITUTE FOR ATOMIC RESEARCH AND FROM THE DEPARTMENT OF CHEMISTRY, IOWA STATE COLLEGE]

The Separation of Rare Earths by Ion Exchange.¹ IV. Further Investigations Concerning Variables Involved in the Separation of Samarium, Neodymium and Praseodymium

BY F. H. SPEDDING, E. I. FULMER, T. A. BUTLER AND J. E. POWELL

I. Introduction

Variables concerned with the separation of certain rare earth mixtures by elution from synthetic ion exchange resins with 5% citric acid solutions have been studied extensively.²⁻¹⁰ Lower citric acid concentrations have received very little attention, although 0.5% citrate solutions have been used successfully on a pilot plant scale.¹¹ The present communication presents systematic studies on the effect of such variables as temperature, pH of eluant, size of resin particles and flow rate using 0.5% citric acid as the eluting agent.

II. Materials, Apparatus and General Procedure

(1) **Materials.**—The pure samarium, neodymium and praseodymium oxides were prepared as previously described¹¹ for the pilot plant scale operations. The mixtures employed consisted of pure oxides in equimolar ratios. The Amberlite IR-100 resin was prepared by passing commercial grade resin over standard screens. Unless otherwise stated -30 + 40 mesh size particles were used in the columns.

(2) **Apparatus.**—The columns, constructed of 22 mm. i. d. Pyrex glass tubing, were closed near the bottom with coarse fritted glass discs. Each column was filled with distilled water and then tapped as resin was added; the bed was given several regeneration cycles with periodic tapping and then adjusted to the desired height by removing the excess resin. Unless otherwise stated, all resin beds were 120 cm. long.

(3) **General Procedure.**—Prior to each new experiment each column was regenerated with the following solutions in the order given: 4 liters of 5% citrate solution at a pH

value of 5.0, 4 liters of 5% sodium chloride solution and 4 liters of 5% hydrochloric acid. After removing the excess acid with a distilled water rinse, a sample, consisting of 0.005 mole of R₂O₃ (about 1.70-1.80 g. depending on the rare earths involved), dissolved in 2.6 ml. of concentrated hydrochloric acid and a liter of distilled water, was adsorbed on the top of the resin bed.

(4) **The Eluting Solution.**—The eluant, designated as 0.5% citrate solution, contained 5 g. of citric acid monohydrate per liter of distilled water initially and was adjusted to the required pH with concentrated ammonium hydroxide. In order to prevent the growth of mold, 1 g. of phenol was added per liter of solution.¹¹

(5) **Recovery and Analysis.**—The rare earths were recovered from the eluate as oxalates and ignited to the oxides for weighing. The fractions were analyzed with a Beckman Quartz Spectrophotometer. The solutions for analysis were prepared by dissolving 50 mg. of the oxide in 5 ml. of 5% hydrochloric acid and diluting to 10 ml. with the same acid. The Nd was measured at 740 m μ , the Sm at 401 m μ and the Pr at 444 m μ .^{3,10} The values for the molar extinction coefficients for Nd, Sm and Pr at the above wave lengths were redetermined using the purest materials on hand; the data in Table I compare the new values with those used previously.¹¹

TABLE I

DATA EMPLOYED IN THE SPECTROPHOTOMETRIC ANALYSIS OF THE RARE EARTHS

Element	Abs. band, m μ	Band width, Å.	Extinction coefficient, l. X g. moles ⁻¹ X cm. ⁻¹	
			Old value	New value
Nd	740	10	6.27	6.53
Sm	401	5	3.09	3.09
Pr	444	5	9.80 ^a	10.07

^a The value of 7.30 previously given was due to a typographical error. The extinction coefficient value should have been 9.80 and the band width 5 Å. instead of 10 Å.

III. Experimental

(1) **The Effect of pH on the Elution of Pure Samarium and Neodymium in the Range 3.80 to 4.20.**—Columns were loaded with 1.744 g. each of pure Sm₂O₃. The solution of the oxides in hydrochloric acid prior to adsorption on the resin bed, as described in the general procedure, is assumed in all subsequent discussions. The samples were eluted at flow rates of 0.5 and 2.0 cm./min. at pH values of 4.20, 4.10, 4.00, 3.90 and 3.80. Another experiment, identical in all respects to that described above, except for the substitution of pure Nd for Sm, was performed.

The data for these experiments are plotted in Figs. 1 and 2. The pH value of the eluant is recorded beside the curve with which it is associated. From the elution curves, it was observed that decreasing the pH of the eluant increased the volume required for the break-through to occur. At pH values of 4.00 and above this effect is

(1) This work was supported, in part, by a grant from the Atomic Energy Commission.

(2) F. H. Spedding, A. F. Voigt, E. M. Gladrow and N. R. Sleight, *THIS JOURNAL*, **69**, 2777 (1947).

(3) F. H. Spedding, A. F. Voigt, E. M. Gladrow, N. R. Sleight, J. E. Powell, J. M. Wright, T. A. Butler and P. Figard, *ibid.*, **69**, 2786 (1947).

(4) D. H. Harris and E. R. Tompkins, *ibid.*, **69**, 2792 (1947).

(5) B. H. Kettle and G. E. Boyd, *ibid.*, **69**, 2800 (1947).

(6) E. R. Tompkins, J. X. Khym and W. E. Cohn, *ibid.*, **69**, 2769 (1947).

(7) E. R. Tompkins and S. W. Mayer, *ibid.*, **69**, 2859 (1947).

(8) S. W. Mayer and E. R. Tompkins, *ibid.*, **69**, 2866 (1947).

(9) E. R. Tompkins, D. H. Harris and J. X. Khym, *ibid.*, **71**, 2504 (1949).

(10) G. E. Boyd, J. Schubert and A. W. Adamson, *ibid.*, **69**, 2818 (1947).

(11) F. H. Spedding, E. I. Fulmer, T. A. Butler, E. M. Gladrow, M. Gobush, P. E. Porter, J. E. Powell and J. M. Wright, *ibid.*, **69**, 2812 (1947).