THE CRYSTAL STRUCTURES OF TWO COMPOUNDS

I. A BASE DERIVED FROM VEATCHINE

II. α-TRIGLYCINE

Thesis by
Robert Nathan

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California
1956

To Anita,

for many reasons

ACKNOWLEDGEMENT

I wish to express my sincere gratitude to Dr. Edward Hughes for his careful, pertinent yet friendly observations and criticisms during this period of work.

I also wish to thank Dr. Harry Yakel and Dr. John Rollett for their generous technical aid in both the X-ray and digital computing laboratories.

Abstract

Organic chemists at the University of New Brunswick, while investigating the alkaloid veatchine derived from Garrya veatchii, had isolated a degradation product $C_{16}H_{15}$ N which they believed to be a substituted phenanthridine. A sample was submitted to us for investigation by X-ray diffraction. Using the phenanthrene-like nucleus as a "heavy atom" it was possible to determine the substituents. Difference syntheses indicated that the skeleton is not a phenanthridine but an aza-phenanthrene. The final structure obtained is 1-methyl-7-ethyl-3-aza-phenanthrene. At one time the chemists were planning to synthesize each of nineteen isomers considered by them to be possible, but all of them in fact incorrect; the X-ray analysis directed them to the correct structure.

A desire to investigate the bond lengths and angles of a peptide group located between two other peptide groups prompted the investigations of alpha-triglycine. In the h0l projection all atoms are resolved. The molecule takes on the extended chain configuration, and is twisted only to an extent necessary to make fullest use of all possible hydrogen bonds. The present stage of refinement shows no particular conflict with those dimensions for polypeptide chains presented by Pauling and Corey.

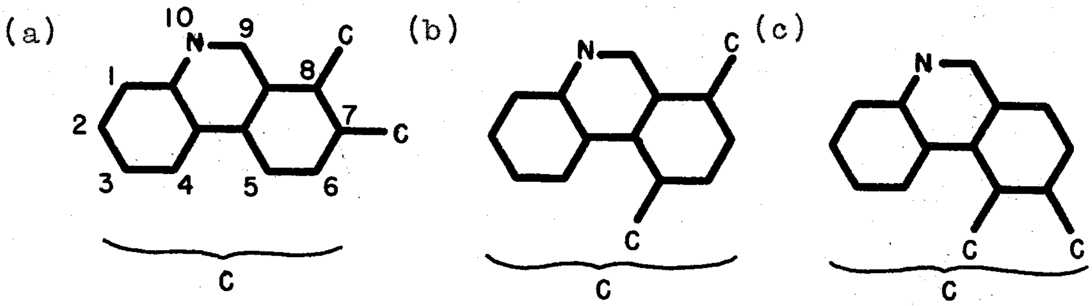
Another approach to the phase problem through the intensity continuum is also described.

I. A BASE DERIVED FROM VEATCHINE

In seeking a quinine substitute Oneto (1946) isolated two crystalline alkaloids, veatchine and garryine, from the bark of Garrya veatchii. They both have been given the empirical formula $C_{22}H_{33}O_2N$. Structural analysis of these compounds was undertaken by organic chemists at the University of New Brunswick under the direction of Professor Wiesner (1952,3,4). These chemists noted a similarity between the chemical reactions of these compounds and those of the aconite alkaloids, atisine and isoatisine, as reported by Jacobs (1943). It was suggested by Wiesner (1952) that the solution of the structure of the garrya alkaloids would give important clues about the structure of the aconite alkaloids.

A selenium dehydrogenation of veatchine at 340°C gave rise to two compounds, 1-methyl-7-ethyl-phenanthrene and an unidentified base of empirical formula $C_{16}H_{15}N$. Comparison of the ultra-violet absorption spectrum of this base with that of phenanthridine led the Canadians to believe that the unidentified base was a substituted phenanthridine. Oxidative degradation of this compound afforded 1,2,3,4-benzene tetracarboxylic acid, which seemed to fix the substitution to one of the three following types:

Figure 1



The problem remaining was to locate the three remaining carbon atoms as substituents either as three methyl groups or as one methyl and one ethyl group.

There were still 19 possible isomers to consider as possible structures. Professor Wiesner and his group were seriously considering methods for synthesizing all of them when contact was made with Professor Pauling with regard to the possibility of supplementing the analysis by means of X-ray crystallography.

A 120 mgm. sample of the crystalline base was mailed to me in January 1953 and the investigation was begun.

The crystals were noted to be very soft plates (m.p. 114±1°C). Their interference figures as seen in a polarizing microscope mounted on an optical goniometer placed them most probably in the monoclinic system. Oscillation pictures about the b axis (parallel to the face of the plate and parallel to a crystal edge) in the 5 cm. camera and $h0\ell$ and $h1\ell$ Weissenberg pictures were taken. From the Weissenberg pictures one very long axis was noted and was taken as the c axis. Many choices were possible for the a axis and the choice was taken which made β almost 90°. Using a right angle holder to remount the same crystal so as to rotate about a gave an $0k\ell$ Weissenberg photograph which showed 2/m symmetry.

The consistent extinctions observed are $0k0$ when k is odd and $h0\ell$ when h is odd. These give unambiguously the space group $P2_1/a$.

All X-ray photographs were made with CuK_α rays with λ taken as 1.542Å. In the region of high $\sin\theta$ where K_{α_1} and K_{α_2} begin to separate in an optical micrometer, only the mean apparent $\sin\theta$ was taken using the weighted λ of 1.542Å. By using the Weissenberg photographs to index the oscillation picture of the 5 cm. camera $h0\ell$ reflections, it was possible to measure b and to obtain a, c and β by least squares. The unit cell dimensions are $a=9.54\pm0.01$Å, $b=5.54\pm0.02$Å, $c=23.57\pm0.01$Å and $\beta=90°20'\pm10'$.

When the crystals were first sent to this laboratory, there was still some uncertainty about the empirical formula, i.e. it was not certain whether we were given crystals of $C_{16}H_{15}N$ (MW=221.2) or $C_{17}H_{17}N$ (MW=235.2). From the unit cell dimensions and the density, which was found by flotation to be $\rho=1.196$, an experimental MW of 224.2 was determined, assuming four molecules per unit cell. This calculated molecular weight confirmed the empirical formula of $C_{16}H_{15}N$ which was confirmed in a subsequent communication by Professor Wiesner from a better elemental analysis.

It is desirable in the estimation of the intensities of diffraction maxima to have the shape of the crystal uniform with regard to its axis of rotation in the camera, or else to use a very small crystal. After three months of searching for a crystal of the desired perfection and after fruitless attempts to grind a crystal into cylindrical shape, a very small crystal was finally selected.

The use of a very small crystal in the Weissenberg camera required an extremely long exposure time. The first set of useful intensity photographs required a 90 hour exposure. The result was somewhat unsatisfactory since there appeared to be an excessive number of absences, but enough reflections were visible (146) to start initial calculations.

All intensities were made by visual comparison of the spot in question with a calibrated intensity strip. This strip was prepared from an $h0\ell$ reflection with increasing intensity based approximately on a logarithmic scale. After correction for Lorentz and polarization factors the reflections were placed on an approximate absolute scale using the f curves of James and Brindley (1935) according to the method developed by A.J.C. Wilson (1942). A temperature factor of $B=4.0$ was thus disclosed, which accounted for the large number of absences.

From the shape of the unit cell it was considered that the phenanthridine skeleton was oriented nearly parallel to the c-axis. When the skeleton is projected onto the c-axis, it has an apparent center of symmetry which is not being used by the space group of the crystal. But Dr. Hughes noticed that this apparent symmetry caused effective weakening of the odd order reflections along the c-axis (fig. 2). A set of signs for the 00ℓ planes was determined by a calculation of a set of structure factors for the skeleton located with its center at $z=\frac{1}{4}$. These signs were used to make a one-dimensional Fourier along c. The resultant Fourier showed a large spurious peak near the origin which was separated by a sharp dip from the peaks representing the skeleton. A substantial part of the projected symmetry could still be roughly maintained if the skeleton were shifted along the c-axis by an amount equal to half the distance between the centers of two of the skeleton's benzene rings. A set of structure factors for the skeleton was calculated from this new position and another one-dimensional Fourier was made using the newly calculated signs (fig. 3). This time a peak again appeared in addition to the skeleton, but unlike the first trial it was close enough to the skeleton so that it might be the result of a substituent carbon in the 7 position. According to the analysis by the organic chemists, there should also be a substitution in the 8 position as seen in figure 1a. Further refinement of the one-dimensional Fourier did not support this conclusion; therefore, the skeleton used in making the trial structure for the first

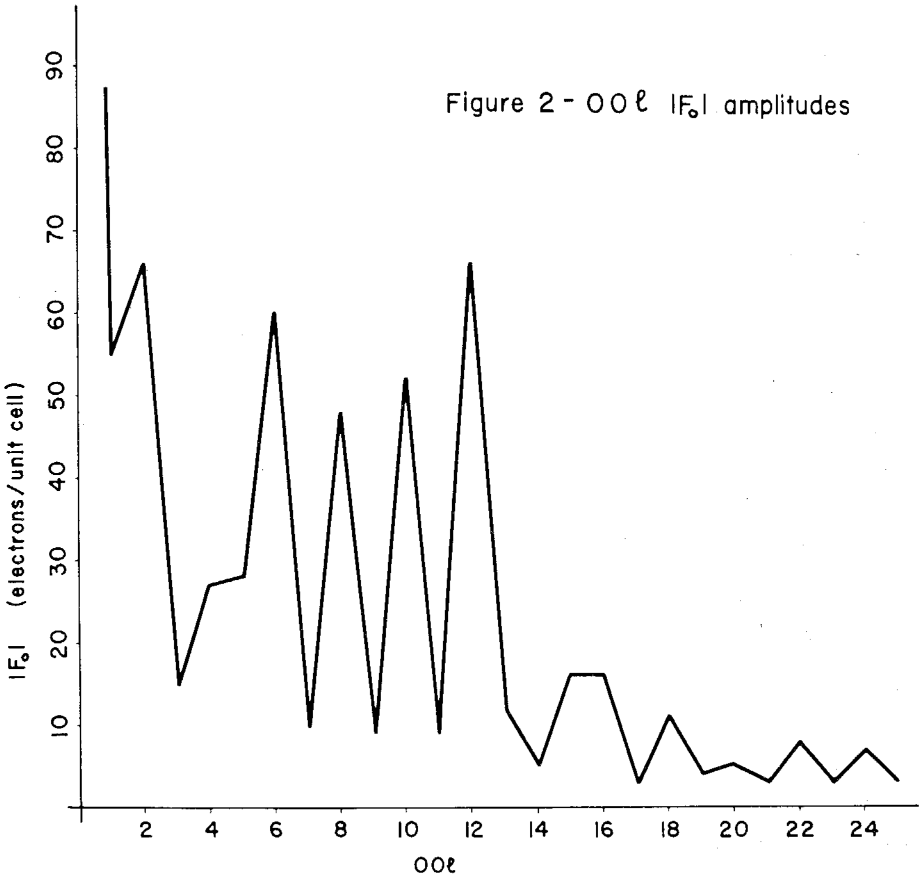


Figure 2 - 00ℓ |F$_o$| amplitudes

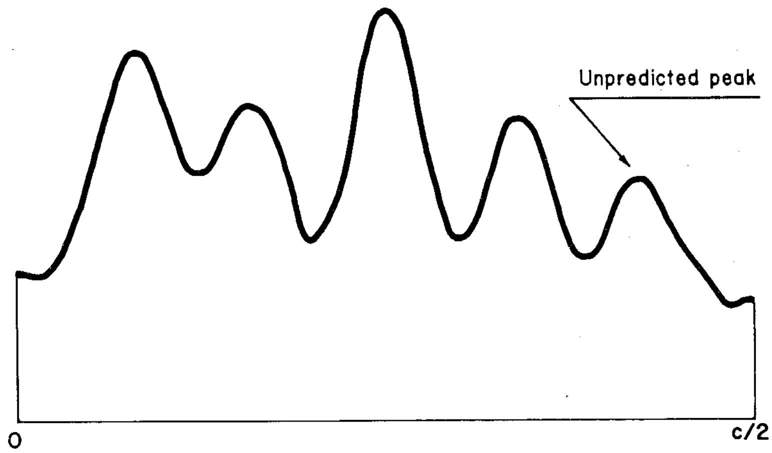
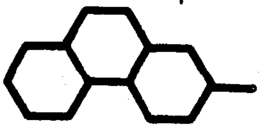
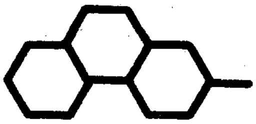
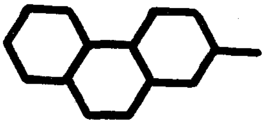


Figure 3 - Electron density projected onto c-axis

h0ℓ Fourier was  . It was now necessary to determine the tilt of the skeleton and also to decide whether the central ring was in the upper or lower position, i.e.,

 or  .

Dr. Rollett pointed out that the skeleton rings would disclose their tilt by the appearance of three regions of higher intensity in reciprocal space, corresponding approximately to the Fourier transform of a benzene ring. Two of these regions were located while the third appeared to be beyond the range of CuKα radiation. By coupling this information with the peak shapes of the one-dimensional Fourier it was possible to postulate a reasonable trial structure which was used with Bragg-Lipson charts to determine 56 signs in the h0ℓ zone. The resultant Fourier projection is seen in figure 4.

The projection at this stage does not show resolved peaks, but does give general agreement with regard to the position of the skeleton. There is certainly no way of determining from the map whether the nitrogen is in the 9 or 10 position. Two other features of the map deserve attention. A methyl group has appeared at position 1 where none was put into the trial structure. Also, a disordered ethyl group has appeared in the 7 position where only a methyl carbon had been put into the trial

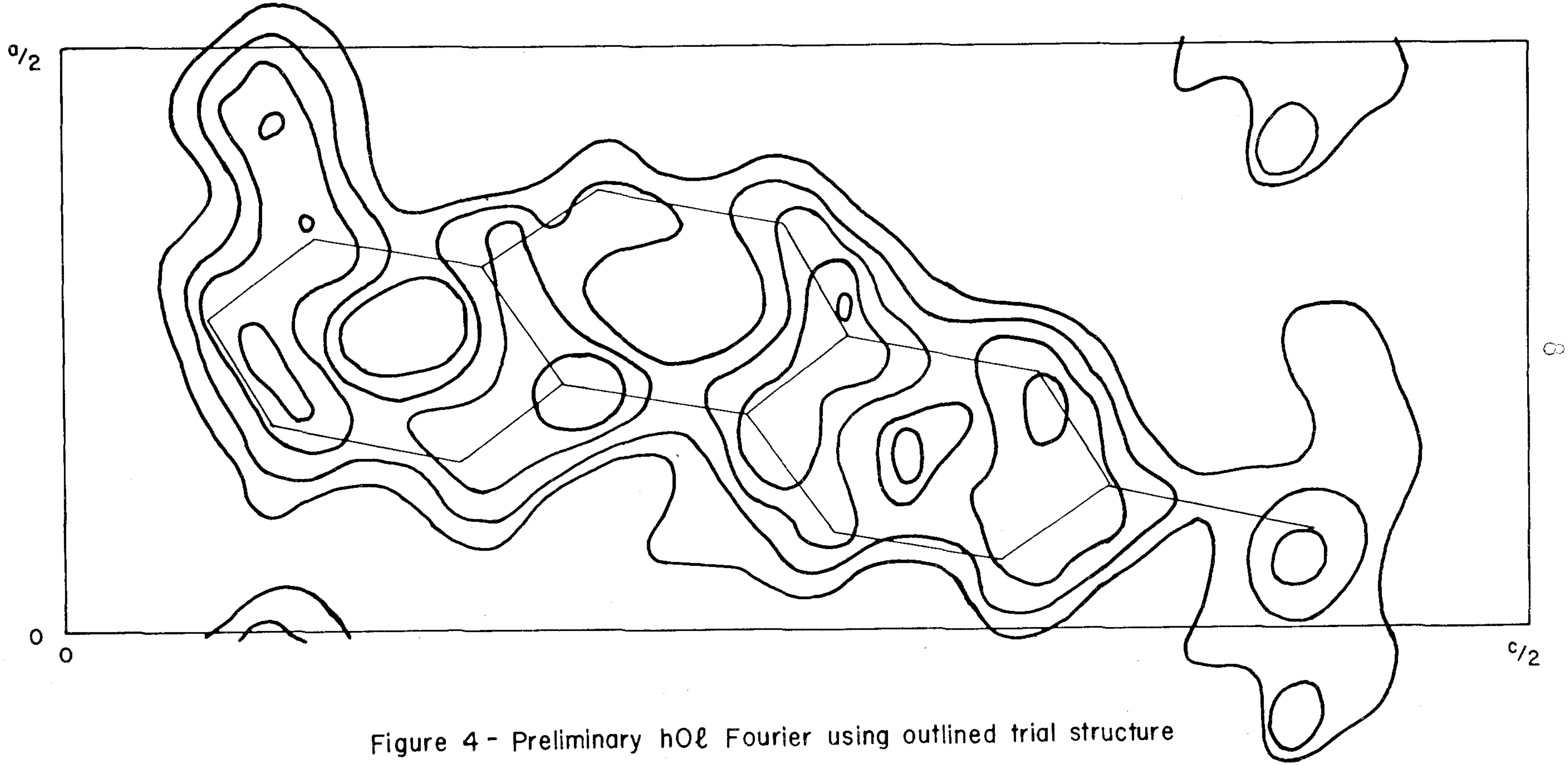


Figure 4 - Preliminary h0ℓ Fourier using outlined trial structure

structure. The heights of the two peaks representing the beta-carbon of the ethyl group were about half the heights of the peaks of the rest of the molecule. The position of this atom appeared to be disordered.

An $h0\ell$ Patterson map was made and is shown in figure 5. In this drawing F_{000}^2 has been omitted and contours are shown only for positive regions. The most distinct feature of the Patterson is the set of peaks near the origin representing the intramolecular vectors corresponding to the distance between the centers of the rings of the skeleton. These vectors agreed with the above trial structure.

Four successive Fourier maps and structure factor calculations using all 146 reflections failed to fix the beta-carbon of the ethyl group in either the upper or lower position and also failed to fix the nitrogen at 9 or 10. At this time Dr. Hughes checked with Dr. Wiesner to see whether he really had gotten 1,2,3,4-benzene tetracarboxylic acid upon oxidation. Dr. Wiesner answered that he had confirmed this oxidation. The organic chemists reexamined their spectroscopic data in the light of our structure and concluded that the compound was not a phenanthridine, but that the nitrogen must now be at either 2, 3 or 6.

Since it takes at least twenty hours of laborious calculations to make a two-dimensional Fourier using Beavers-Lipson strips, and at least a week of even more

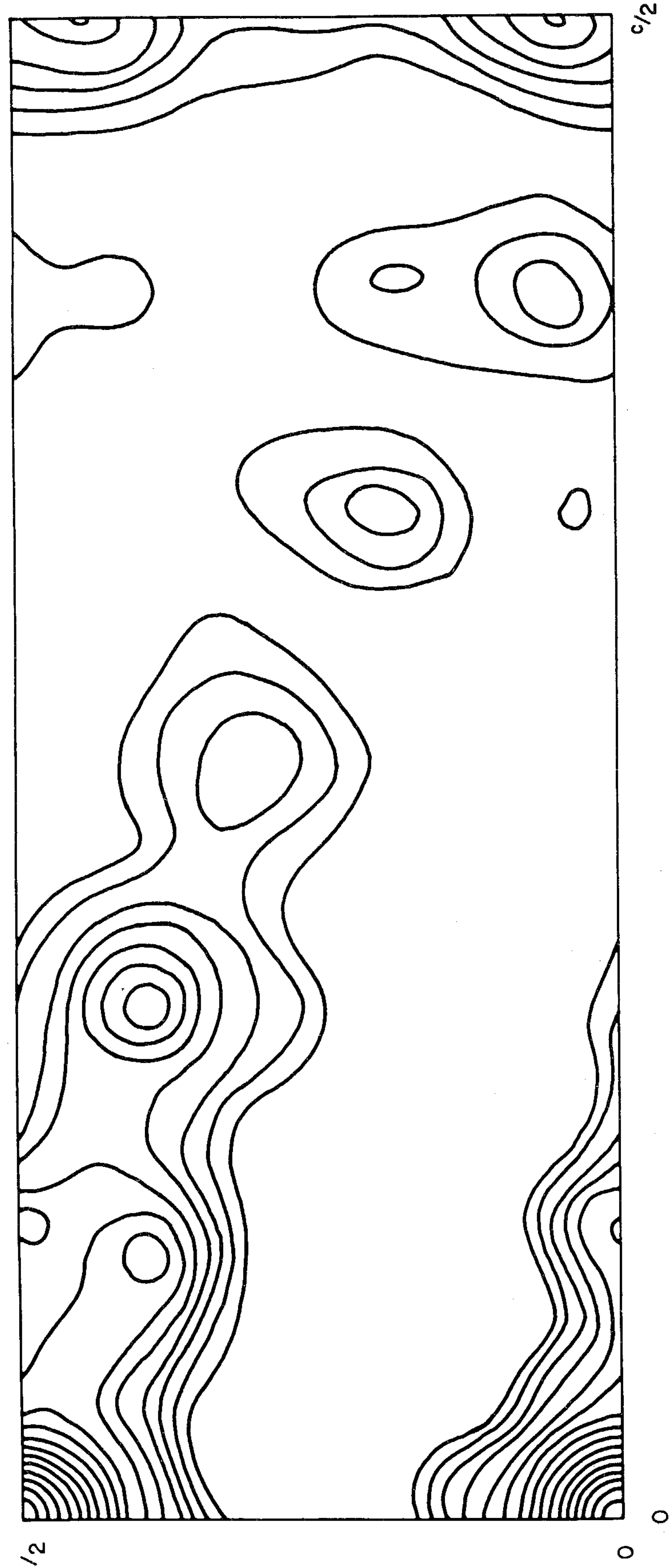


Figure 5 - h0ℓ Patterson

monotonous calculations to determine a set of structure factors for 18 atoms and 230 planes, it appeared very desirable to introduce IBM techniques at this point.

Once a set of trial coordinates was chosen and the data placed on IBM cards, a computing control panel was wired which made it possible to calculate a set of structure factors for some chosen position of the nitrogen in half an hour. Since $\cos 2\pi\,(hx+lz)$ for each position was already on the cards, it was necessary to shift only one or two wires on the control panel in order to switch the f values of the nitrogen atom with those of a carbon atom. Since half an hour was such a large reduction in the time requirement as compared to the hand calculation, it now was feasible to calculate sets of structure factors for the nitrogen in various positions in the azaphenanthrene skeleton. All possible positions were tried, giving the following R factors (where the subscript indicates the position of the nitrogen):

$R_- = 29.2\%$ (all carbon) $R_4 = 33.5\%$ $R_8 = 40.4\%$
$R_2 = 36.8\%$ $R_5 = 36.8\%$ $R_9 = 38.1\%$
$R_3 = 38.1\%$ $R_6 = 29.2\%$ $R_{10} = 32.9\%$

where R equals $\frac{100\Sigma|F_o-F_c|}{\Sigma|F_o|}$ for observed planes only.

The R factor for the all carbon skeleton is low presumably because that structure was used in the refinement. Examination of these sets of structure factors suggested that the nitrogen might be in the 6 position. A difference synthesis was made using F_o-F_c (for N at 6) which showed a large hole in the 6 position and a large peak equidistant from the 2 and 3 positions, thereby indicating that the nitrogen was certainly not at 6 and probably was at 2 or 3. The difference map is shown in figure 6 with contours shown at 0.5 electrons $Å^{-2}$.

At this stage the data appeared to be inadequate for further refinement because of the large number of unobservably small reflections. Since the earlier search for the desired shape of crystal was fruitless, it was decided to use one of the larger crystal plates. The use of a plate has the disadvantage that the spots on different portions of the film have different shapes which make it difficult to estimate their intensities. I offset this disadvantage somewhat by taking off my glasses and observing the film at such a distance that the spots were blurred to the point where their shapes were unrecognizable. They were then compared to an intensity strip to give reasonably good intensity estimations.

The new photographs required a 69 hour exposure and gave about a ten-fold increase in intensity over the previous pictures. Of the total of 253 possible reflections 180 were now observed as compared with the 146 reflections seen on the previous set of pictures.

Use of the new intensity data in comparison with two sets of structure factor calculations (one with nitrogen at 3 and one with nitrogen at 2) favored the nitrogen at 3. A difference synthesis was made with N at 3. The size of the peaks and troughs indicating shifts of other carbon

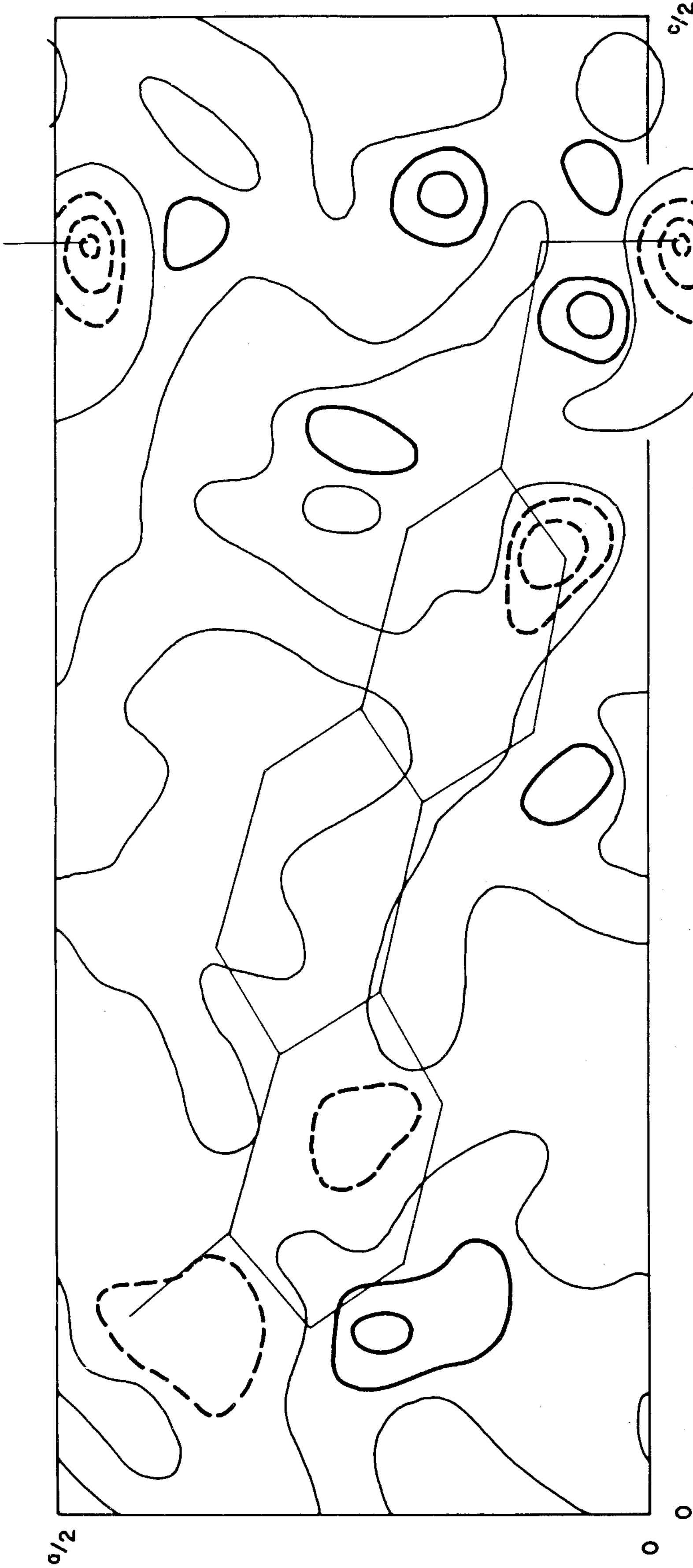


Figure 6 - hOℓ $F_o - F_c$ difference synthesis using N in 6 position

atoms were of the same order of magnitude as the trough associated with the nitrogen atom, indicating a shift rather than a switch of nitrogen with the carbon in the 2 position.

Least squares refinement techniques as set forth by Hughes (1941) were added to the refinement procedures. When the cross terms were used, it required about a day of hand calculations to solve the sets of normal equations. Similar calculations required about an hour when only the diagonal terms were used. It was possible to make two cycles of calculations using only the diagonal terms in the same time that would otherwise have been required for one refinement cycle using the cross terms. Therefore, it was felt that the refinement proceeded faster by using the shift in parameters obtained from the use of the diagonal terms only.

With the combined use of least squares and difference syntheses three refinements brought the R factor for the projection down to 17.8%. The Fourier is shown in figure 7 with contours drawn at one electron $\mathring{A}^{-2}$ with the lowest contour representing 1 e/$\mathring{A}^2$. The beta-carbon now appeared to be approximately one-quarter up and three-quarters down, which merely denotes the average position among all the molecules and does not necessarily imply rotation.

A set of intensity photographs was taken in the Weissenberg camera for the $0k\ell$ zone. Of the possible 155 planes only 93 were observed because of the high temperature factor and the position of the molecules in

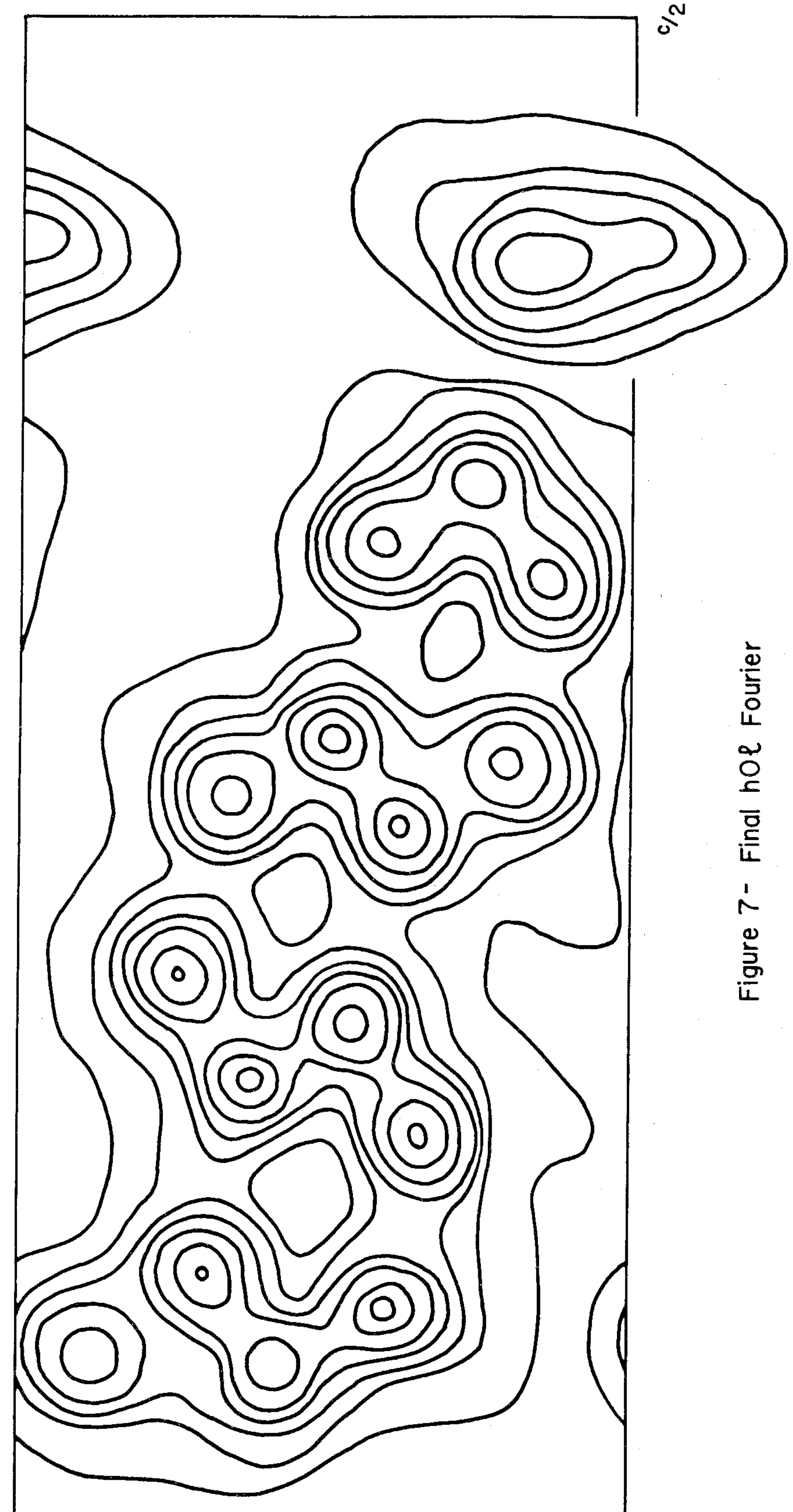


Figure 7- Final h0ℓ Fourier

the unit cell. There were large gaps in the sequence of planes along a particular reciprocal lattice row, thereby making the Weissenberg very difficult to index. A photograph of the $0k\ell$ zone taken by the precession camera (with Mo K radiation) proved to be of tremendous aid in indexing the planes. A Patterson map is shown in figure 8 with negative contours omitted. It indicated that there would be sufficiently favorable atomic resolution to warrant refinement of the projection. One subsequent electron density Fourier brought the R factor for the $0k\ell$ projection down to 22.3%. The projection is shown in figure 9 where the contours are drawn at intervals of two electrons Å^{-2} with the lowest contour drawn at 1 $e/\text{Å}^2$.

At this time a letter was received from Professor Wiesner stating that the compound had been synthesized with the nitrogen in the 3 position and had proved to be identical with the dehydrogenation product.

The main purpose of this investigation, i.e., the determination of the structure of the veatchine fragment, had been accomplished. Once this stage of progress had been reached, it appeared to be just as easy to continue refining the parameters somewhat further and thereby to contribute some knowledge regarding the bond distances. But because the molecule is disordered in the orientation of the ethyl group, it did not seem worth the effort to obtain full three-dimensional data because of the probable effect on the apparent bond lengths caused by the disorder.

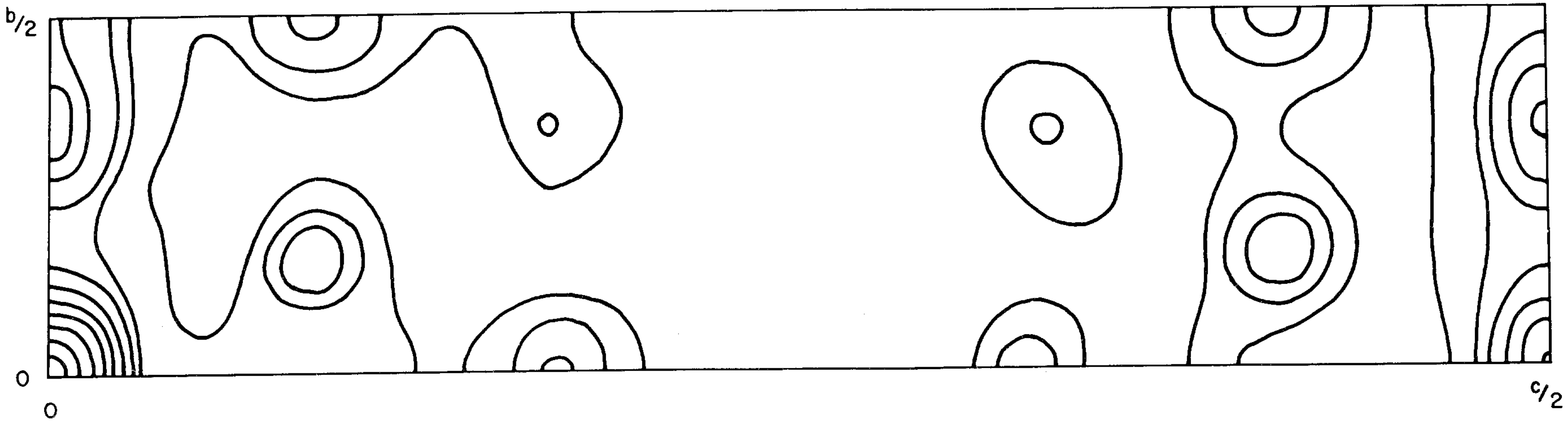


Figure 8 - 0kℓ Patterson

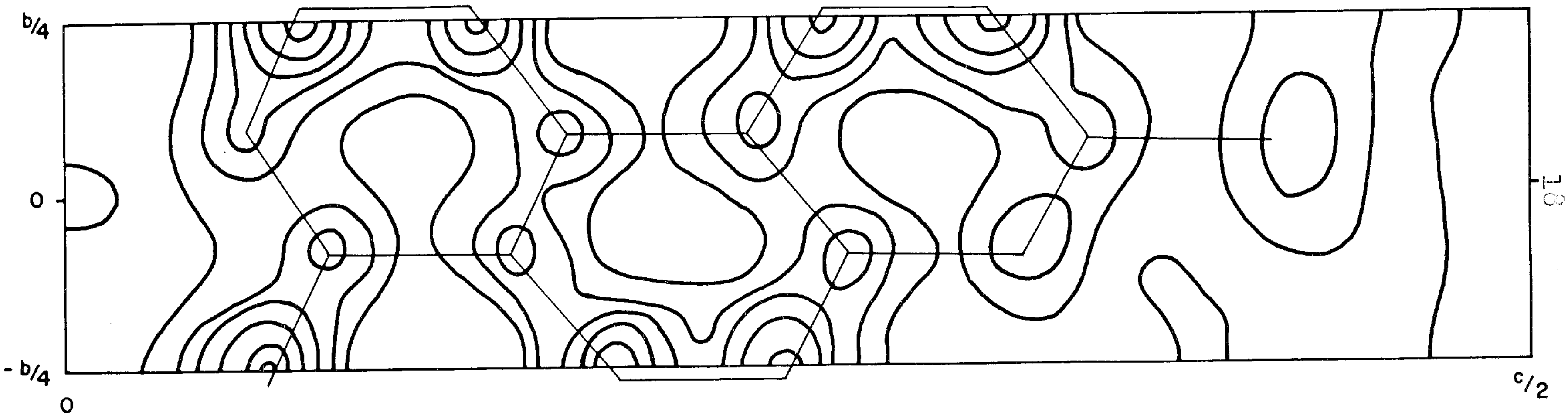


Figure 9 - 0kℓ Fourier

Several changes were made **which** improved the agreement between calculated and observed structure factors. Conversion was made from James and Brindley (1935) to McWeeny (1951) f curves. Hydrogen atoms were added to the skeleton rings with a C-H bond length of 1.1 Å. A corresponding change was made in the temperature and scale factors so that B was now equal to 4.7.

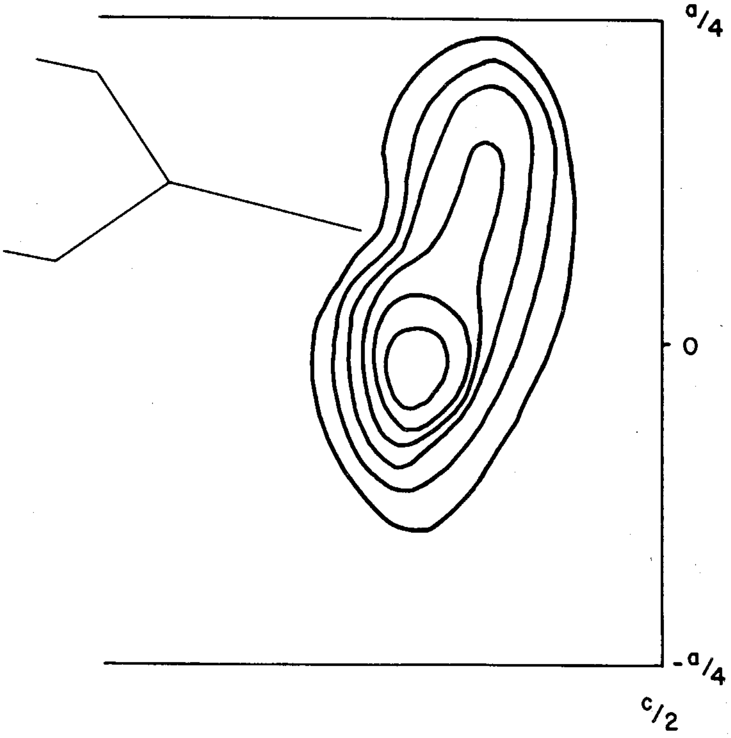
In the $0k\ell$ projection there are many poorly resolved atoms very close to the glide plane which appears as a mirror in this projection. Application of least squares in the usual manner to these atoms does not result in rapid refinement, but it was noticed that the difference synthesis is very sensitive in this region to any shift from true position. Also, in the case of a disordered structure such as is the case with this compound, the use of least squares fails to elucidate the nature of the disorder while such discrepancies are very evident on the difference map. Therefore, the use of the difference synthesis is preferred.

In order to investigate the nature of the disorder in the ethyl group, sets of structure factors for the two zones were calculated from which the contribution from the beta-carbon was omitted. For these sets of structure factors the alpha-carbon was placed 1.54 Å from the skeleton carbon. (See discussion of bond lengths.) The difference syntheses for these zones were then made. They are shown in figures 10 and 11 where the contours are drawn at intervals of one electron Å^{-2} and the lowest contour is drawn at one electron Å^{-2}. A very clear representation of the disorder is now visible which was only implied by the normal Fourier. The toroidal shape representing the beta-carbon in the $0k\ell$ projection seems to imply a uniform distribution on the circle generated by the motion of the beta-carbon by rotation about the C_7-C_α bond. But the shape of the electron density representing the beta-carbon in the $h0\ell$ projection is not symmetrically distributed with respect to the C_7-C_α bond, which implies either a shift in the alpha-carbon or that the beta-carbon is more frequently in the "down" position.

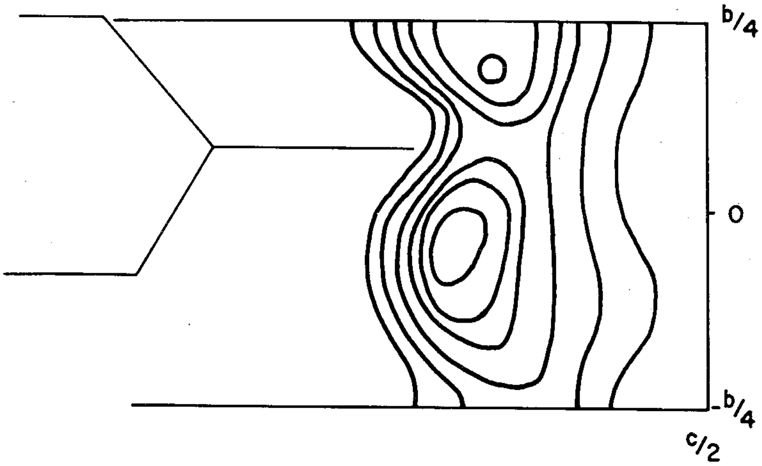
Several trial structures were chosen using various combinations of atomic coordinates in regions of higher electron density. The weighting of a combination of these positions was used so that the total density added up to one carbon atom. It was still difficult to obtain satisfactory results since there was no real way of estimating the positions of the five ethyl hydrogen atoms.

In the initial calculation of bond distances the C-C single-bond distance for the skeletal carbon and alpha-carbon of the ethyl group turned out to be 1.68 Å which was certainly too long by comparison with the present accepted value of 1.54 Å. The extra length as derived from the $h0\ell$

hOℓ Projection



Okℓ Projection



Figures 10 & 11 - $F_o - F_c$ difference syntheses

showing β-carbon atom

projection is easily explained by the lack of resolution of the peak representing the combined electron density of both carbon atoms in the ethyl group (figure 7). In subsequent trial structures this C-C distance was put in as 1.54 Å, but because of extreme libration effects a large hole was always evident at this position in the difference synthesis. It seemed at this point that no further refinement from the given data could be accomplished without the introduction of asymmetric temperature factors.

Since the x and z parameters were more accurate than the y parameters **because** of the lack of resolution caused by the mirror plane symmetry in the 0kℓ projection, a plane passed through the skeleton rings by least squares was used in conjunction with the resolvable y coordinates to determine the y coordinates of the unresolved atoms. An R factor of 17.9% for the 0kℓ zone and 17.1% for the h0ℓ zone was finally reached.

An estimation of error was determined for the x and z coordinates as indicated in Whittaker and Robinson (1940) in their description of the least squares method. The mean square deviation was taken as $\sigma = \sqrt{\frac{\Sigma\omega(F_o - F_c)^2}{D}} = 0.016$ Å for both x and z. ω is the weighting function and D is the diagonal term of the normal equation, thereby ignoring the effects of the cross terms.

σ_y would normally be much larger for those atoms near the mirror planes. But, since their y coordinates were determined partly by the restriction of making the ring planar, it was assumed that $\sigma_y \approx 0.02$ Å.

For the bond lengths we then have approximately that σ is equal to 0.028 Å, and for the bond angles σ is equal to 1.5° from the method of Cruickshank and A.P.Robertson (1953).

The molecule is shown in figure 12. Calculated bond lengths (in Ångstroms) and angles are shown. A comparison of these bond lengths is made with those calculated theoretically for phenanthrene by Oakley and Kimball (1949):

bond	veatchine	phenanthrene
C9 -C10	1.32 Å	1.36 Å
C12-C13	1.48	1.47
C9 -C14 C10-C11	1.46 1.43	1.46
C1 -C11 C8 -C14	1.425 1.41	1.43
C1 -C2 C7 -C8	1.365 1.36	1.37
C2 -N3 C6 -C7	1.37 1.41	1.40
N3 -C4 C5 -C6	1.34 1.39	1.37
C4 -C12 C5 -C13	1.43 1.41	1.42
C11-C12 C13-C14	1.40 1.395	1.39

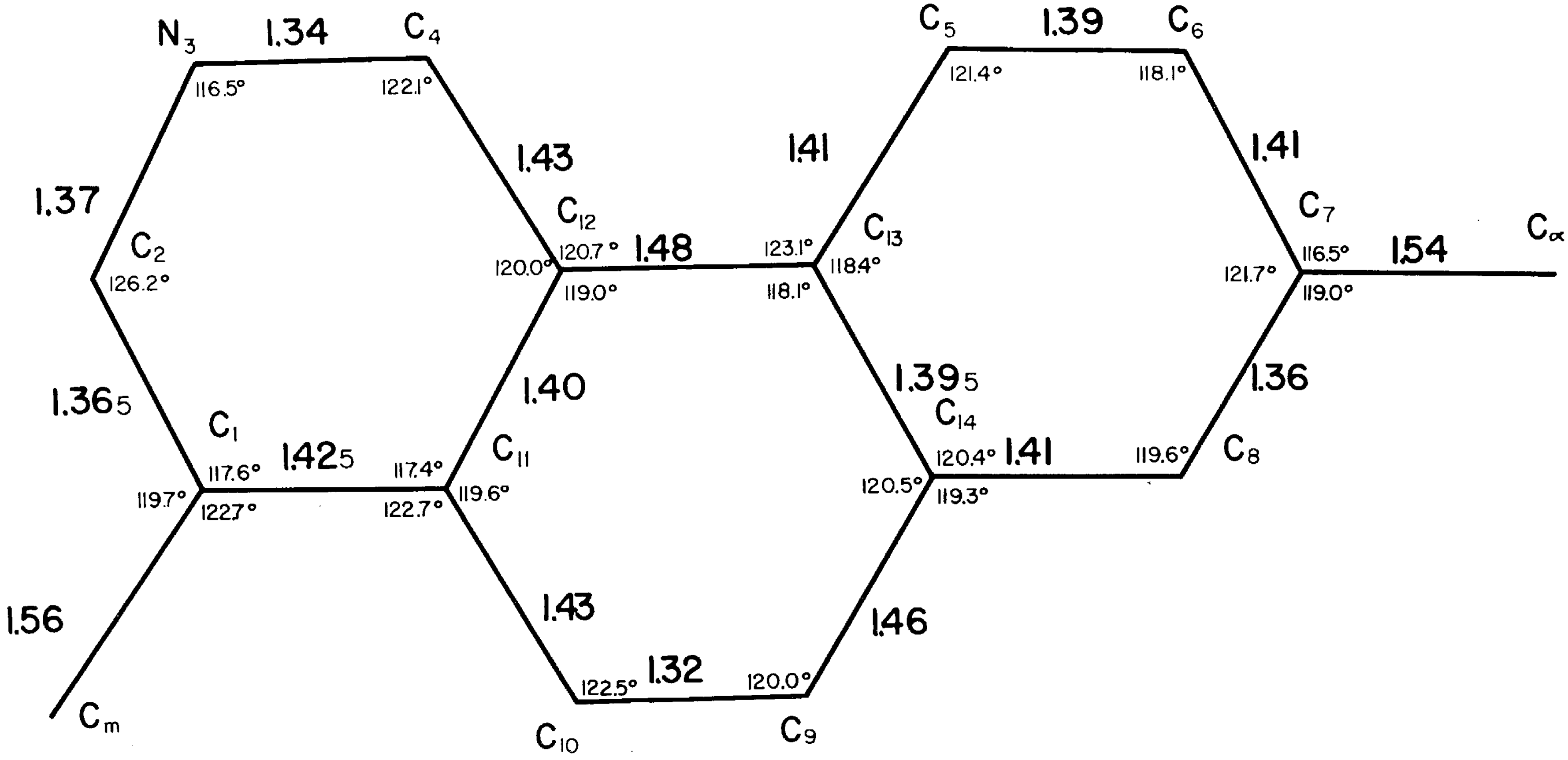


Figure 12 - Dimensions of the molecule (not showing β-carbon atom)

The deviations from the plane of the molecule of all atoms except those of the ethyl group are less than 0.03 Å and well within the limit of experimental error. The alpha-carbon of the ethyl group is 0.05 Å out of the molecular plane, which might be real and caused by intermolecular steric hindrance. But, since the cross terms of the normal equations relating the alpha- and beta-carbons are significantly large with respect to the diagonal term, it must be admitted that 0.05 Å could also be within experimental error.

The C-N-C bond angle of 116.5° is at first sight unexpected, since there is a shortening of the C-N bonds relative to the C-C bonds in the rest of the ring. One would expect an angle greater than 120°. But, in melamine (Hughes 1941), the C-N-C angles are also 116-117°. In cyanuric triazide (Knaggs 1935) the C-N-C angle is even smaller being 113°, as is also the case in s-triazine (Wheatley 1955).

The intermolecular distances are shown in figures 13 and 14. The packing of the phenanthrene type molecule is rather common as is evidenced by the compilation of unit cell dimensions for various hydrocarbons related to sterols made by Bernal and Crowfoot (1935). The space group $P2_1/c$ with $a \approx 9$ Å and $b \approx 6$ Å with a variation in the length of the c axis is very common.

The more important feature visible from the packing is the steric hindrance which does not allow the two beta-

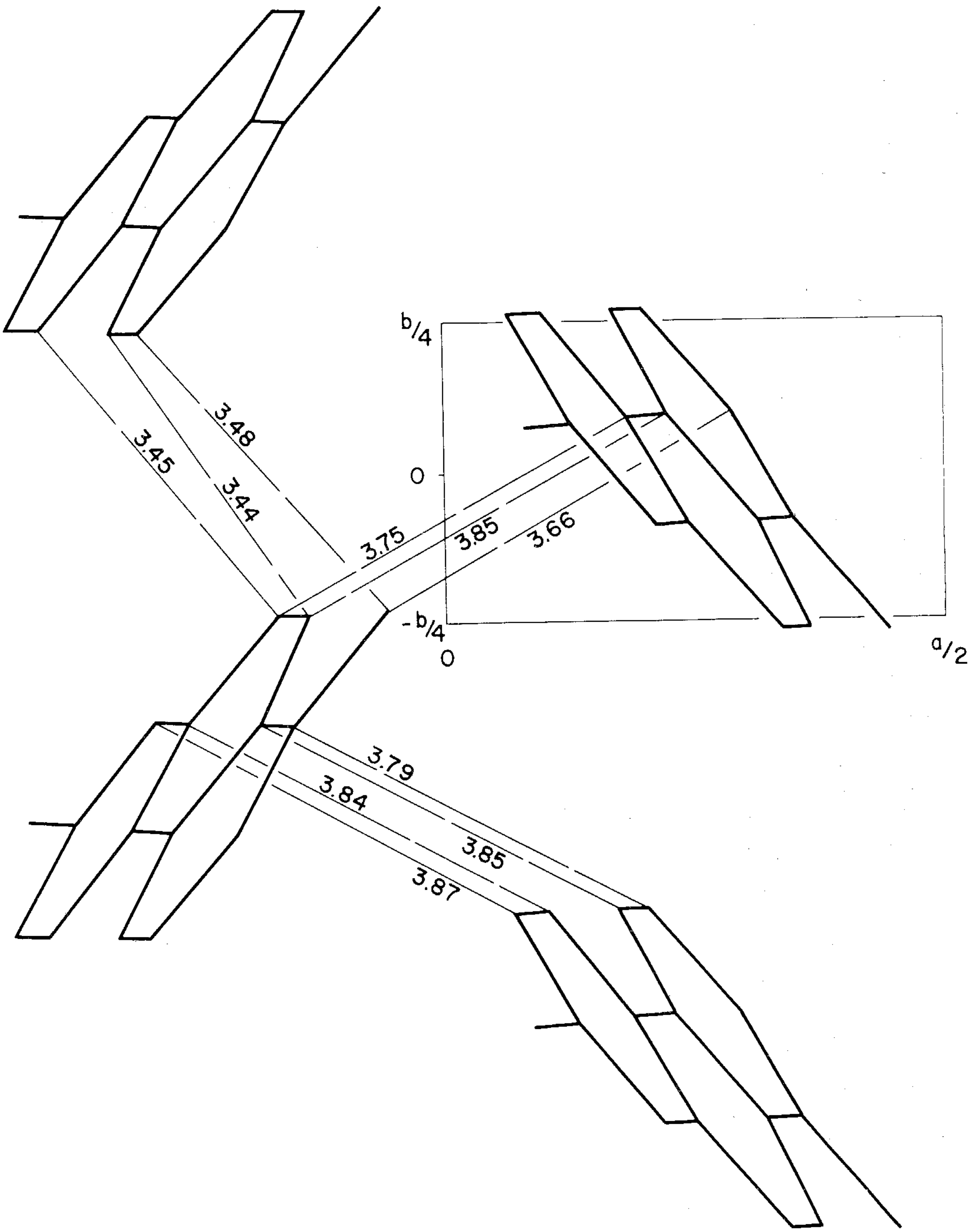


Figure 13 - Intermolecular dimensions showing 1/2 cell hk0 projection

(not showing β-carbon atom)

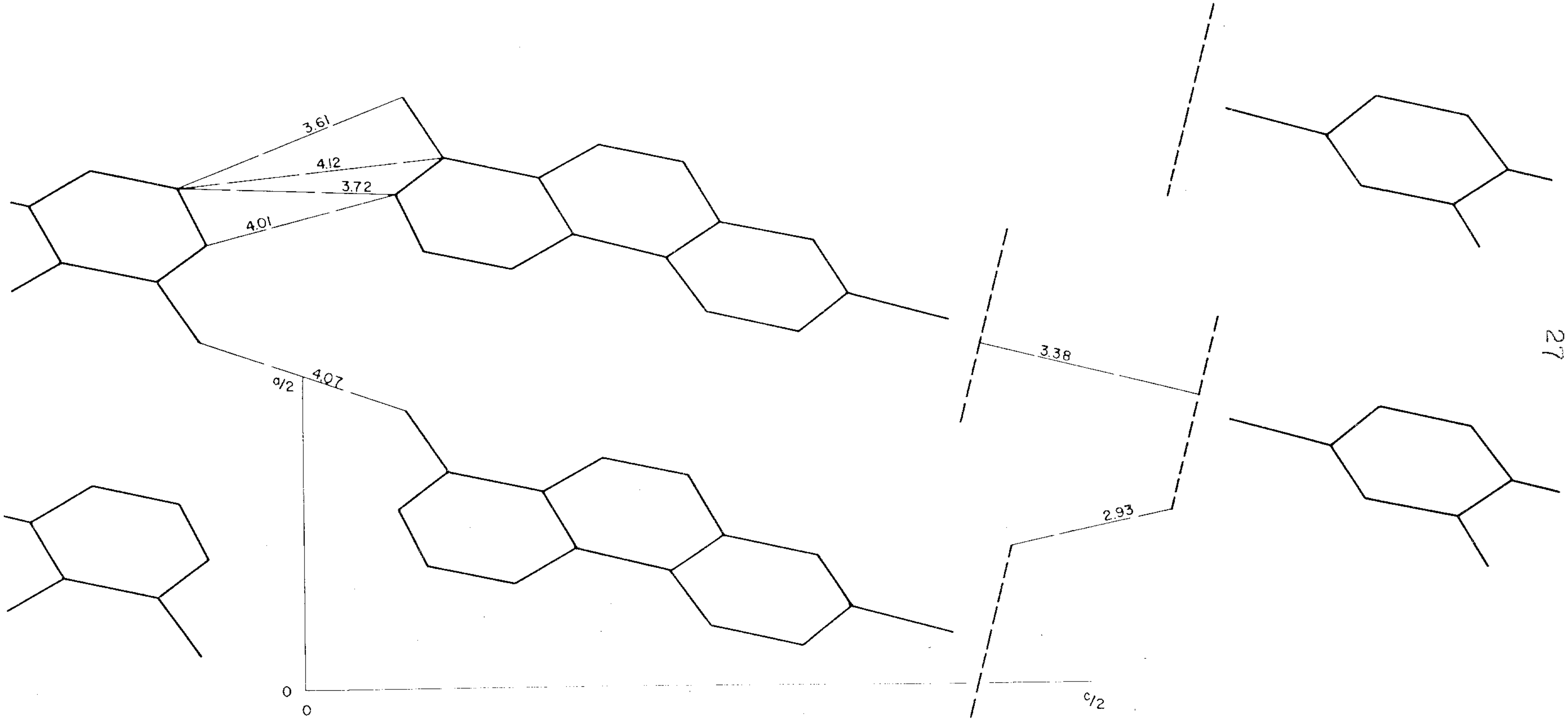


Figure 14 - Intermolecular dimensions showing $h0\ell$ projection

carbon atoms related by the screw axis to be simultaneously in the "up" configuration. The closest approach of the beta-carbon atoms would be 2.93 Å if free rotation occurred, whereas the allowable minimum is about 3.95 Å (Robertson 1953, pg. 226).

I would like to make a final comment with regard to the verification of the compound by the organic chemists. When it was mentioned to Dr. Wiesner that difficulty was being encountered with regard to the ethyl group, he suggested it might be because the sample he sent us may have been contaminated by an impurity, say to the extent of 5 or 10%, which would be a compound having an isopropyl group instead of the ethyl group as the result of the selenium degradation. To see if such was really the case, he had some of the synthetic compound sent to us so that its X-ray diagrams could be compared with those of the natural compound. Optical inspection showed that the two crystals are different. Although both crystals are flat plates, the synthetic crystal has a pseudo-tetragonal shape while the natural crystal has a pseudo-hexagonal outline. The melting points are also different, with the synthetic compound melting 4°C higher than the natural one. Yet, these two compounds are claimed to be identical as checked by means of mixed melting points. Moreover, an oscillation photograph and $h0\ell$ and $h1\ell$ Weissenberg photographs show that the synthetic compound has a different crystal structure.

Although the space group is the same, $P2_1/a$, the axial lengths are quite different:

	Natural	Synthetic
a	9.54 Å	8.94 Å
b	5.54	11.78
c	23.57	12.02
β	90°20'	102°

A screw axis now passes near the center of the molecule making it impossible to obtain resolution by the use of a b-axis projection. The general appearance of the packing as seen in figure 13 is very likely to be the same (with every other molecule turned end for end) since the 00ℓ reflections of the synthetic compound bear a very strong resemblance to the even orders of the 00ℓ reflections of the natural compound. It is, therefore, indicated that the molecule is still oriented lengthwise along the c-axis.

Because no resolution could be expected without going into a three-dimensional analysis, it was not considered worth the trouble to do any further work with this new crystal modification.

Final atomic coordinates:

Atom	x	y	z
C 1	0.3455	0.915	0.0922
C 2	.2860	.092	.0599
N 3	.1960	.268	.0784
C 4	.1670	.268	.1339
C 5	.0960	.265	.2570
C 6	.0615	.258	.3142
C 7	.1215	.074	.3480
C 8	.2080	.904	.3255
C 9	.3350	.728	.2436
C 10	.3651	.730	.1890
C 11	.3117	.910	.1511
C 12	.2220	.089	.1716
C 13	.1820	.087	.2322
C 14	.2408	.911	.2672
C m	.4448	.727	.0640
C α	.0781	.070	.4110
C β (.2)	.0630	.167	.4300
C β (.2)	.0000	.133	.4283
C β (.3)	.1567	.017	.4417
C β (.3)	.4417	.938	.4250

h0ℓ Structure Factors (M denotes maximum value)

h	ℓ	$\lvert F_o \rvert$	F_c	h	ℓ	$\lvert F_o \rvert$	F_c
	1	54	66	4		20	18 -
	2	66	94 -	4	1	6	1 -
	3	15	2 -	4	2	24	32
	4	27	29 -	4	3	24	24
	5	28	30 -	4	4	12	15
	6	60	55 -	4	5	9	11 -
	7	10	15 -	4	6	12	9
	8	48	41 -	4	7	16	14
	9	9	11 -	4	8	29	35 -
	10	52	40 -	4	9	26	22 -
	11	9	10 -	4	10	8	8 -
	12	66	68	4	11	8	8 -
	13	12	9	4	12	3	3 -
	14	4	3	4	13	3	5 -
	15	16	16	4	14	3 M	3
	16	16	14	4	15	5	5
	17	3 M	2	4	16	6	6
	18	12	10 -	4	17	3	2 -
	19	4	1 -	4	18	22	24
	20	5	4 -	4	19	20	26
	21	3 M	1 -	4	20	15	18 -
	22	9	6	4	21	16	15 -
	23	3 M	1 -	4	22	6	6 -
	24	6	7 -	4	23	3 M	4 -
	25	3 M	2 -				
				6		8	4
2		80	87 -	6	1	2 M	2
2	1	106	105 -	6	2	4	1 -
2	2	115	112	6	3	5	6 -
2	3	37	35	6	4	5	2 -
2	4	12	12 -	6	5	3	2
2	5	40	40	6	6	2 M	5
2	6	25	19	6	7		12 -
2	7	20	23	6	8	22	16
2	8	2	5 -	6	9	36	35
2	9	6	2 -	6	10	3 M	4 -
2	10	5	9 -	6	11	12	11 -
2	11	2 M	2 -	6	12	7	6 -
2	12	12	8 -	6	13	4	5 -
2	13	15	18 -	6	14	5	4 -
2	14	9	8	6	15	3 M	3 -
2	15	7	9 -	6	16	4	4 -
2	16	2 M	2 -	6	17	4	5 -
2	17	8	5	6	18	7	2 -
2	18	12	9 -	6	19	10	14 -
2	19	4	5	6	20	7	7
2	20	4	1	6	21	16	24
2	21	6	6	6	22	2 M	3
2	22	8		6	23	3	2 -
2	23	3 M		6	24	2 M	3
				6	25	2 M	1
				6	26	2 M	3 -

hOℓ Structure Factors (continued)

h	ℓ	$\lvert F_o \rvert$	F_c
8		13	15 -
8	1	7	8
8	2	4 M	5 -
8	3	3	4 -
8	4	3 M	5
8	5	7	8
8	6	5	9
8	7	3 M	5 -
8	8	3 M	4
8	9	8	6 -
8	10	3 M	4 -
8	11	8	9
8	12	4	1
8	13	3	
8	14	3 M	3 -
8	15	3 M	
8	16	3	2 -
8	17	2 M	
8	18	2 M	
8	19	2 M	4 -
8	20	2 M	3
8	21	2	2 -
8	22	2 M	1 -
10		3	
10	1	3 M	
10	2	3 M	1
10	3	3 M	3
10	4	3 M	3 -
10	5	6	5
10	6	5	3
10	7	2 M	3 -
10	8	2 M	1
10	9	2 M	5 -
10	10	2 M	1
10	11	2 M	2
10	12	2 M	1 -
10	13	2 M	1 -
10	14	3	1 -
10	15	2	1 -
10	16	2 M	2 -
12		2 M	
12	1	2	2 -
12	2	2 M	4 -
12	3	2 M	3 -
12	4	2 M	
12	5	4	4 -
12	6	1 M	8 -

h	ℓ	$\lvert F_o \rvert$	F_c
2	1 -	20	22 -
2	2 -	26	35 -
2	3 -	22	24
2	4 -	10	13
2	5 -	3	2
2	6 -	35	28 -
2	7 -	7	5
2	8 -	24	15 -
2	9 -	32	27
2	10 -	52	56
2	11 -	27	26 -
2	12 -	18	21 -
2	13 -	14	9 -
2	14 -	5	5 -
2	15 -	4	2
2	16 -	10	8
2	17 -	7	7
2	18 -	5	2 -
2	19 -	8	9
2	20 -	8	9
2	21 -	12	13 -
2	22 -	17	19 -
2	23 -	3 M	1 -

h0ℓ Structure Factors (continued)

h	ℓ	|F$_o$|	F$_c$	h	ℓ	|F$_o$|	F$_c$
4	1-	7		8	1-	20	21-
4	2-	5	7-	8	2-	4	3-
4	3-	2 M	1	8	3-	7	5
4	4-		1-	8	4-	4	3
4	5-	10	9-	8	5-	4	2
4	6-	4	1-	8	6-	5	7
4	7-	6	4-	8	7-	4	3
4	8-	3	1	8	8-	6	7
4	9-	11	13-	8	9-	4	4
4	10-	3	3	8	10-	11	7
4	11-	2 M		8	11-	19	22
4	12-	2 M		8	12-	11	11-
4	13-	16	18	8	13-	24	20-
4	14-	3 M	1	8	14-	8	4-
4	15-	3 M		8	15-	3 M	1
4	16-	3 M	3-	8	16-	3 M	4-
4	17-	3 M	2-	8	17-	2 M	1-
4	18-	4	3	8	18-	4	3
4	19-	8	8-	8	19-	2 M	1
4	20-	4	4-	8	20-	2 M	2
4	21-	6	4	8	21-	2 M	2
4	22-	3 M	4	8	22-	2 M	
4	23-	3 M	1				
				10	1-	3 M	
6	1-	31	32	10	2-	3 M	1-
6	2-	5	4	10	3-	3 M	1-
6	3-	33	30-	10	4-	3 M	1
6	4-	20	11-	10	5-	3 M	3-
6	5-	9	7-	10	6-	2	1
6	6-	6	5-	10	7-	2 M	
6	7-	4	5-	10	8-	3	
6	8-	3 M	3	10	9-	2 M	3
6	9-	4	3-	10	10-	7	7-
6	10-	3 M	1-	10	11-	4	6-
6	11-	4	6	10	12-	4	6
6	12-	3	1-	10	13-	4	5
6	13-	27	20	10	14-	2 M	3
6	14-	16	13	10	15-	2 M	2
6	15-	12	13-	10	16-	2 M	2-
6	16-	5	2				
6	17-	4		12	1-	2 M	1-
6	18-	5	6-	12	2-	2 M	1
6	19-		10-	12	3-	2 M	1-
6	20-	5	3-	12	4-	2	5
6	21-	3	2-	12	5-	2 M	5
6	22-	2 M	1-	12	6-	1 M	2-
6	23-	2 M					
6	24-	2 M	2-				
6	25-	4	5				
6	26-	2 M	3				

$0k\ell$ Structure Factors

k	ℓ	$\lvert F_o \rvert$	F_c
1	1	18	21 -
1	2	14	17
1	3	22	30 -
1	4	13	17 -
1	5	8	6
1	6	21	21
1	7	8	10
1	8	3 M	2 -
1	9	3	
1	10	18	10 -
1	11	6	8
1	12	13	15
1	13	5	5
1	14	4	2 -
1	15	7	6 -
1	16	2 M	
1	17	6	7 -
1	18	5	4 -
1	19	3 M	
1	20	4	5 -
1	21	7	8
1	22	12	15
1	23	3 M	2
1	24	4	6 -
1	25	2 M	2 -
2		36	33 -
2	1	29	36 -
2	2	16	20
2	3	3	6
2	4	15	16 -
2	5	19	24
2	6	15	16
2	7	13	15
2	8	7	4
2	9	4	7
2	10	2	1
2	11	2 M	2
2	12	2 M	
2	13	9	10 -
2	14	5	5
2	15	13	13 -
2	17	4	4
2	18	13	16
2	19	8	6
2	20	2 M	2
2	21	2 M	2
2	22	4	3
2	23	4 M	
3	1	6	7
3	2	2 M	6
3	3	8	1
3	4	24	32
3	5	12	12 -
3	6	45	47 -
3	7	14	11 -
3	8	5	4
3	9	3 M	3
3	10	3 M	4
3	11	5	6
3	12	3 M	
3	13	5	6
3	14	4	8
3	15	8	6
3	16	27	27
3	17	4 M	5
3	18	23	25 -
3	19	6	5 -
3	20	4 M	
3	21	6	8 -
3	22	5	6 -
3	23	2 M	1 -
4		8	10
4	1	5	7
4	2	2 M	2 -
4	3	2 M	1 -
4	4	2 M	5
4	5	5	5 -
4	6	8	7 -
4	7	3	3 -
4	8	3 M	
4	9	3 M	
4	10	3 M	1
4	11	3 M	
4	12	3 M	1 -
4	13	3 M	3
4	14	3 M	1 -
4	15	5	6
4	16	10	13
4	17	3 M	1 -
4	18	8	10 -
4	19	4	5 -
4	20	2 M	1
4	21	3	2 -
4	22	2 M	1 -

okℓ Structure Factors (continued)

k	ℓ	$|F_o|$	F_c
5	1	4 M	5 -
5	2	4 M	4
5	3	3 M	2 -
5	4	4	5 -
5	5	3 M	4
5	6	4	3
5	7	3 M	1
5	8	3 M	
5	9	3 M	1 -
5	10	3 M	3 -
5	11	3 M	1
5	12	3 M	4
5	13	3 M	1 -
5	14	3 M	1
5	15	3 M	2 -
5	16	3 M	2 -
5	17	3 M	1 -
5	18	3 M	2 -
5	19	2 M	2
6	1	5	6 -
6	2	4 M	8
6	3	4 M	
6	4	3 M	2
6	5	3 M	4
6	6	3 M	3
6	7	3 M	3
6	8	3	2
6	9	4	2
6	10	2 M	4

α -Triglycine

The important problem of protein structure has been under intensive investigation at this Institute for twenty years, and there is no necessity to justify the work on the structure of triglycine other than to say it is one more step toward understanding the larger problem.

Until now, single crystal X-ray investigations which have been completed include studies of either single amino acid molecules or dipeptide molecules which contain only one peptide link. A significant contribution toward predicting the bond distances and angles in a polypeptide would come from the solution of the structure of a tripeptide where the dimensions of a peptide group linked between two other amino acid residues could be determined.

Dr. Harry L. Yakel, Jr. had started this investigation in 1950 when crystals of triglycine ($NH_3CH_2CONHCH_2CONHCH_2COO$) were received from Professor Edwin J. Cohn. Complete three-dimensional data were gathered, using unfiltered Cu Kα radiation (λ=1.542 Å for Cu Kα). The space group was found by Dr. Yakel (1952, pg. 16) to be $P2_1/c$ with a=8.55 Å, b=4.36 Å, c=23.01 Å, β=105°36' and Z=4. The density was determined by flotation as 1.56±0.03 gm/cm^3, while the calculated density was 1.52 gm/cm^3.

Since the b axis is so short, it was expected that the main attack on the structure could be made through this projection alone. Dr. Yakel made a Patterson projection for this zone and from its peaks concluded that the triglycine chain was stretched parallel to the c-axis, and along a line roughly at $x=\frac{1}{4}$. He had proposed several trial structures in accordance with these criteria, but none gave calculated intensities in agreement with the observed, and the investigation was dropped.

I started to investigate the structure by estimating the intensities of the $h0\ell$ zone from later photographs of greater intensity which Dr. Yakel had taken, but which he had never used. A determination of the scale and temperature factors using the Wilson method (1942) gave B=2.6.

My $h0\ell$ Patterson is shown in figure 15 where the negative contours are omitted. It was essentially identical to that which Dr. Yakel obtained from intensities estimated from earlier pictures. Another Patterson was made, but this time the origin peak was removed and a sharpening function was used which reduced the contribution of the lower order reflections. The sharpening or modification function used was of the form

$$M(\sin\theta) = 64\pi^2\left(\frac{\sin\theta}{\lambda}\right)^4 \exp\frac{-8\pi a \sin^2\theta}{\lambda^2},$$

where θ is the Bragg angle, as suggested by Waser and Schomaker (1953, pg. 685). The sharpened Patterson is shown in figure 16 where the negative contours are omitted.

Two new features were revealed on this map. The most outstanding feature is the hexagonal array of peaks which are highest around the origin and which extend to the

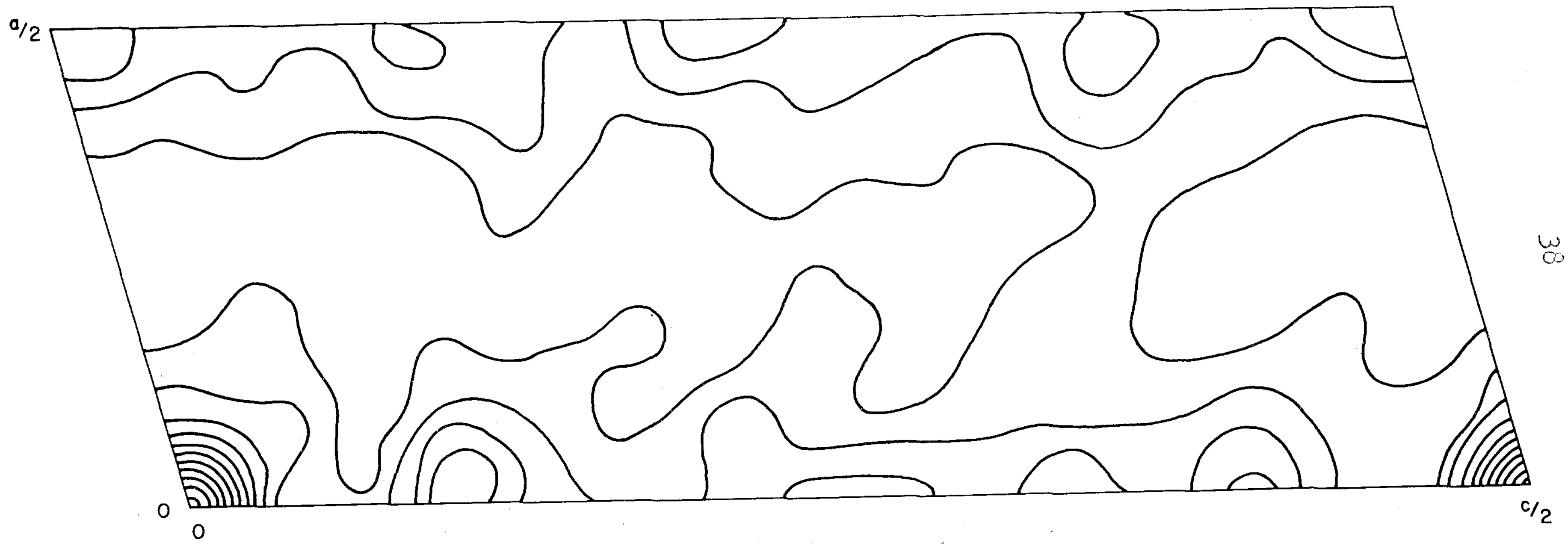


Figure 15- Normal h0ℓ Patterson

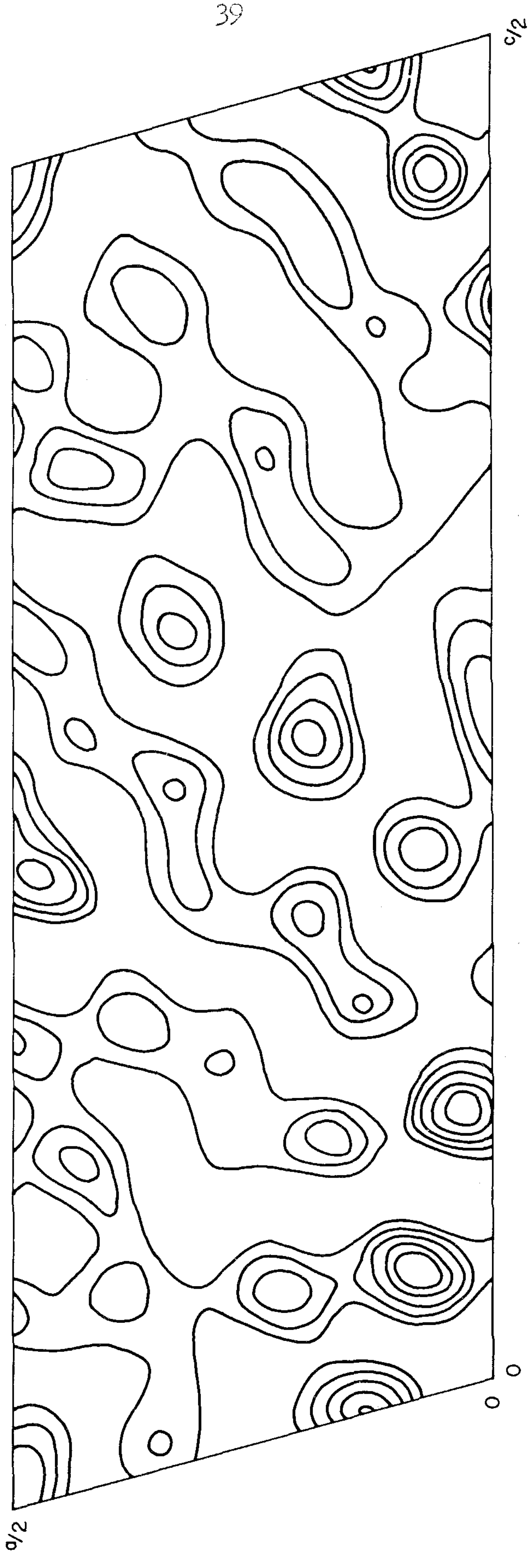
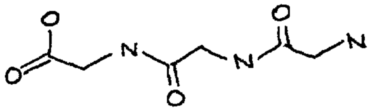
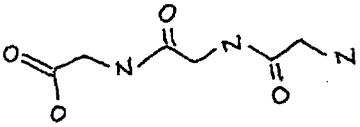


Figure 16- Sharpened h0ℓ Patterson with origin peak removed

right and slightly upward on the figure. Because b is short, no intermolecular vectors fall near the origin. This array was presumed to be caused by a zig-zag chain which did not lie exactly parallel to the c-axis, but which was tilted somewhat away from the axis as stated above. The other useful feature is the difference in the magnitudes of the heights of those peaks which immediately surround the origin. It was necessary to determine in which of two positions the molecule was oriented:

I 

or II 

.

By using each model in turn the vector contributions to these central Patterson peaks were calculated and the first model illustrated above gave better agreement with the observed pattern. The actual position of the molecule within the unit cell had to be determined by the vector interactions which appear near $x=\frac{1}{2}$ in the $h0\ell$ Patterson, since those peaks near $x=0$ arise in the main from intramolecular vectors.

A trial structure was used which utilized all possible intermolecular hydrogen bonds and appeared to satisfy the Patterson peaks. It differed in general by a shift along the c-axis of about nearly half the length of a carbon-carbon bond from the last model used by Dr. Yakel.

The first set of calculated structure factors gave an R factor of 63% which was not very encouraging. A Fourier projection was made using 51 planes (of a total of 172 observed planes). The resolution was surprisingly good and the map suggested a second trial structure with minor shifts from the first. Another set of structure factors was calculated. The R factor from these calculations was 60%. Two successive cycles of the same sort yielded 56% and 52% respectively. The Fourier for the last of these trials, which includes 151 terms, is illustrated in figure 17. The contours are drawn at intervals of one electron $Å^{-2}$ with the lowest contour at one electron $Å^{-2}$. Complete atomic resolution had been obtained, but the carboxyl O-C-O angle had become very large, and there were some indications of false peaks.

It was felt strongly that there must be a large degree of truth in the structure, but there was no obvious means of bringing about further improvement from examination of the Fourier because the peaks very nearly coincided with the previous trial model and did not indicate any shifts of parameters.

An entirely different approach was called for and several methods were applied simultaneously. These methods included the continued use of cycling between Fourier maps and structure factor calculations, but with the restriction imposed that only low order planes were included in the Fourier maps and an F_o-F_c difference synthesis was made simultaneously with each Fourier. An entirely new method was also introduced at this time. This method involved a mathematical approximation to an intensity continuum.

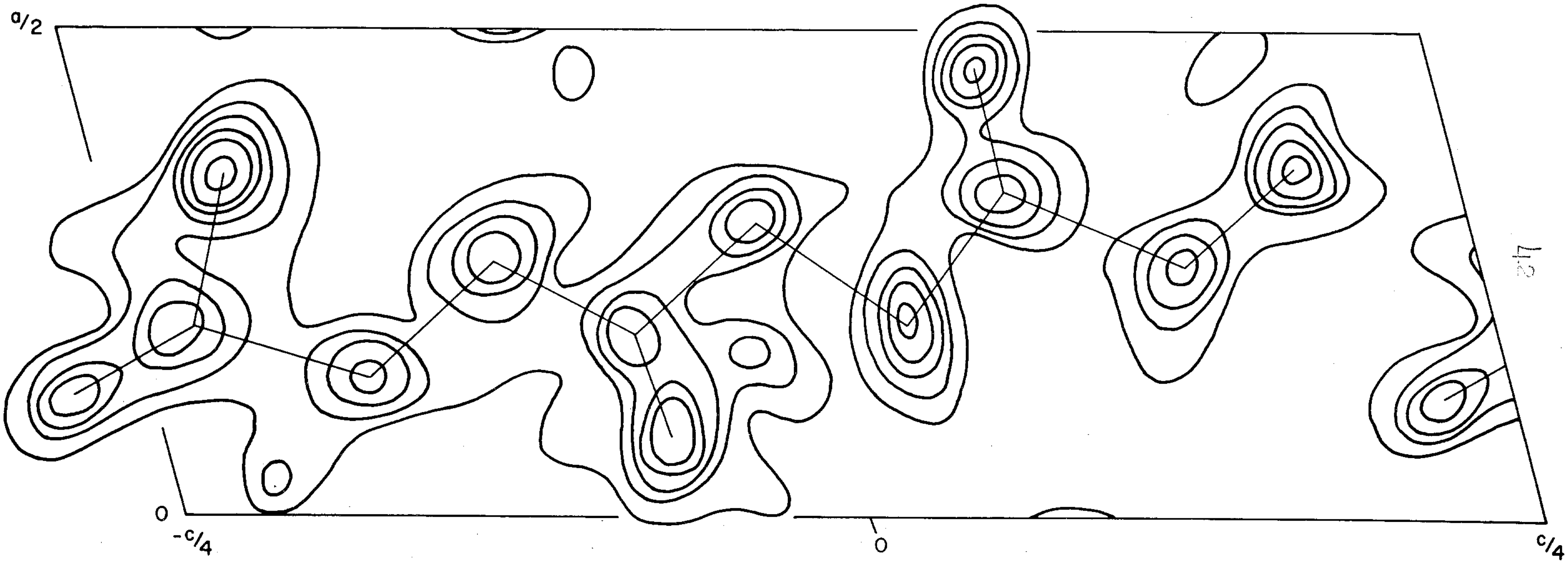


Figure 17 - h0ℓ Fourier showing resolved atoms but incorrect structure (R=52%)

The particular intensity continuum which was sought, represents the square of the Fourier transform for a single unit cell. A more complete description of the method and an example as applied to the h00 intensities for triglycine is given in the Appendix.

After approximately seven trial Fouriers a set of signs chosen for the h00 reflections implied that there was a greater electron density near $x=\frac{1}{3}$ than had appeared previously. Also, a large shift was required to bring a greater contribution of electron density to the relatively strong $1,0,\overline{2}$ reflection. The combination of these two observations implied that the carboxyl group as a whole needed to be shifted upward. A set of structure factors were calculated and the corresponding Fourier was made. The R factor was still high, around 60%, but the low order agreement had improved greatly.

Dr. Rollett then suggested that I use the ElectroData Electronic Computer to refine this trial structure by rapid cycling between structure factor calculations and least squares refinement. Each iteration required only about half an hour on the ElectroData Computer compared to one to two days on the IBM equipment. Using the latest trial structure a set of structure factors was calculated out to about $\sin\theta = 0.5$. The R factor for this trial was 38%. Successive cycles reduced the R factor to 27, 24, and 23% respectively. In the two next cycles all the data were included, with corresponding R factors of 38% and 36%. At no time during these cycles was any correction made to the scale and temperature factors.

The resulting parameters were used to calculate a set of structure factors on IBM equipment and a Fourier was made (with an R factor of 33.5%). The Fourier is shown in figure 18 with contours drawn at intervals of one electron Å^{-2} and where the lowest contour represents one electron Å^{-2}. Complete resolution of all atoms is evident. From the bond lengths as tabulated for peptide groups by Corey and Pauling (1953), an estimation was made of the relative y coordinates and approximate x and z parameters were estimated for the hydrogen atoms. Using the same coordinates as before for the heavier atoms, and adding the contribution of the hydrogen atoms, a temperature and scale factor calculation was made which resulted in reducing the R factor to 28%.

Again the ElectroData Computer was used, while the hydrogen atoms were kept fixed during each refinement cycle. Three iterations reduced the R factor to 26, 23.5 and 23.2% respectively. For the next two cycles the calculated changes in isotropic temperature factors as well as the positional parameters for each atom were included in the refinement. The R factor was thereby reduced to 20.0 and 19.0% respectively.

A projection of the molecule appears in figure 19. The theoretical bond lengths of Corey and Pauling (1953)

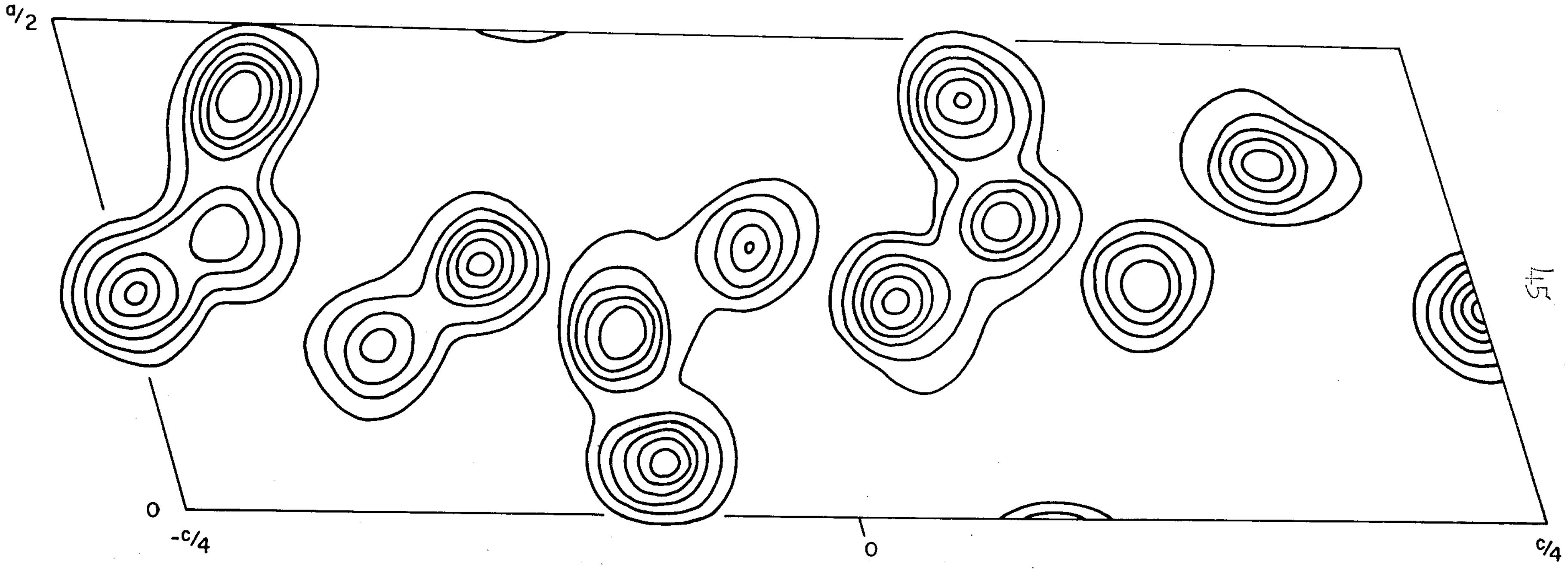


Figure 18 - h0ℓ Fourier using more nearly correct trial structure (R = 33 %)

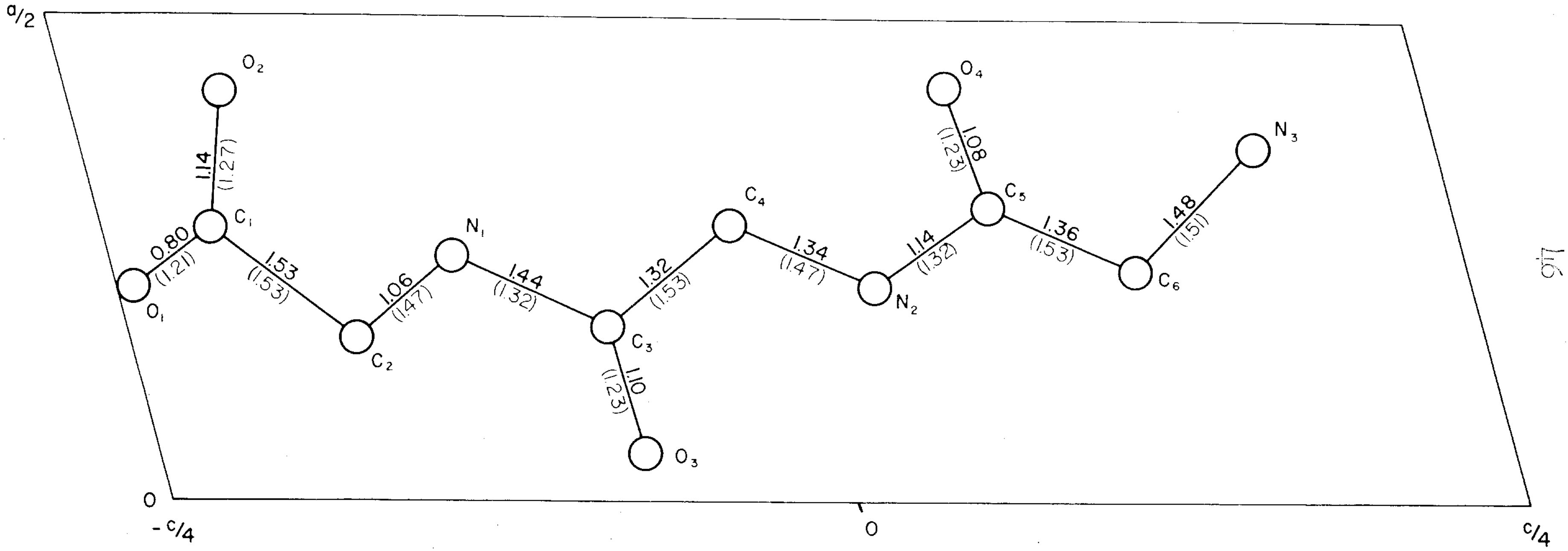


Figure 19- h0ℓ projected dimensions from trial structure (R=19%)

are shown in parentheses in Ångstroms and the calculated projected bond distances are shown without parentheses. After four cycles of refinement, one bond N1-C3 has persisted in remaining significantly greater in length than the predicted value. It is expected that three-dimensional refinement will greatly improve this situation.

Lack of further refinement was attributed to asymmetry in the temperature factors of each atom and the rough approximation of the hydrogen parameters. Since the computer was not programmed to handle asymmetric temperature factors, the refinement was stopped at this point.

Before proceeding to complete three-dimensional refinement, it seemed desirable to obtain approximate values of the y coordinates by solving the $0k\ell$ projection at least to the extent of bringing the $R_{0k\ell}$ factor down to about 30%.

The $0k\ell$ intensities (about 85 observed planes) were estimated and corrected for Lorentz and polarization factors. Their scale factor was adjusted so that the 00ℓ reflections as estimated from the $h0\ell$ and $0k\ell$ zones matched each other. A Patterson projection was made (fig. 20) which showed peaks which because of their distances from the origin seemed to correspond to carbonyl interactions. Relative y parameters were obtained as mentioned above when obtaining hydrogen parameters in the $h0\ell$ projection. It was difficult to determine the relative y coordinates of those interatomic vectors where the y contribution was small, and a large amount of error was introduced into the

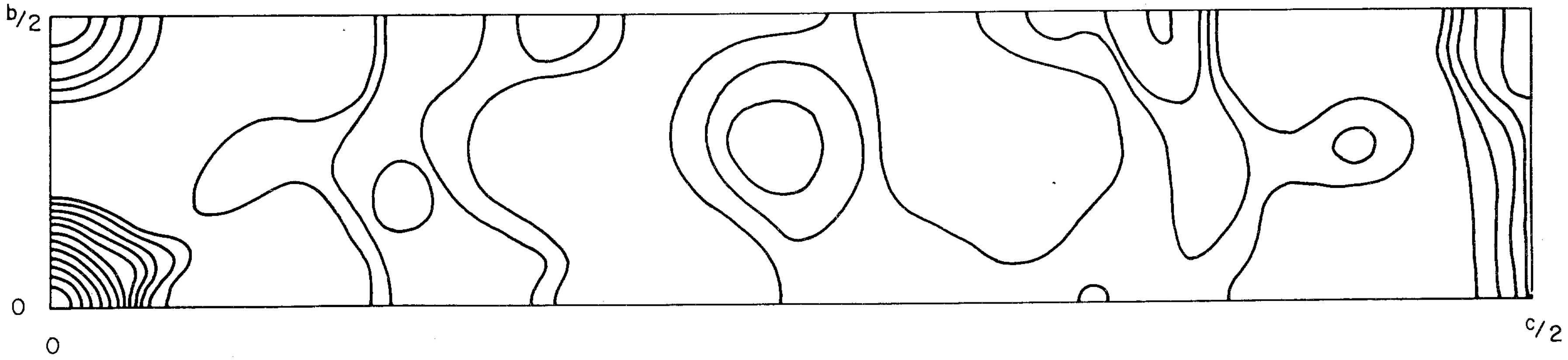


Figure 20 - 0kℓ Patterson

model in this way. Once the trial model was estimated, it was only necessary to find its translation relative to the y axis since its z coordinates were fixed. Bragg-Lipson charts quickly established a reasonable choice of y parameters. Refinement proceeded by calculating several sets of structure factors and trial Fouriers. For each trial Fourier an F_o-F_c difference synthesis was also calculated simultaneously. Because of the expected lack of resolution in this projection, it was not felt that least squares refinement (using only the diagonal terms) would accomplish anything. Therefore, the ElectroData Computer was not used. After fourteen cycles (and five weeks work) of structure factor calculations and Fourier and difference syntheses the R factor had been slowly reduced to 32%. During the refinement the amide groups were each kept planar, but nevertheless, because of non-resolution, it was very difficult to refine the model. Also, it appeared that an extreme asymmetry in the temperature factor was present. It could be approximated by multiplying the isotropic temperature factor, which was B=2.3, by an additional factor exp $-0.05k^2$, which is equivalent to about B=5 for the y direction. The final Fourier and difference synthesis which includes the asymmetric temperature correction (R=32%) appear in figures 21 and 22. In figure 21 the contours are marked at intervals of two electrons $\AA^{-2}$ with the lowest contour at four electrons $\AA^{-2}$. In figure 22 the contours are drawn at intervals of one electron $\AA^{-2}$.

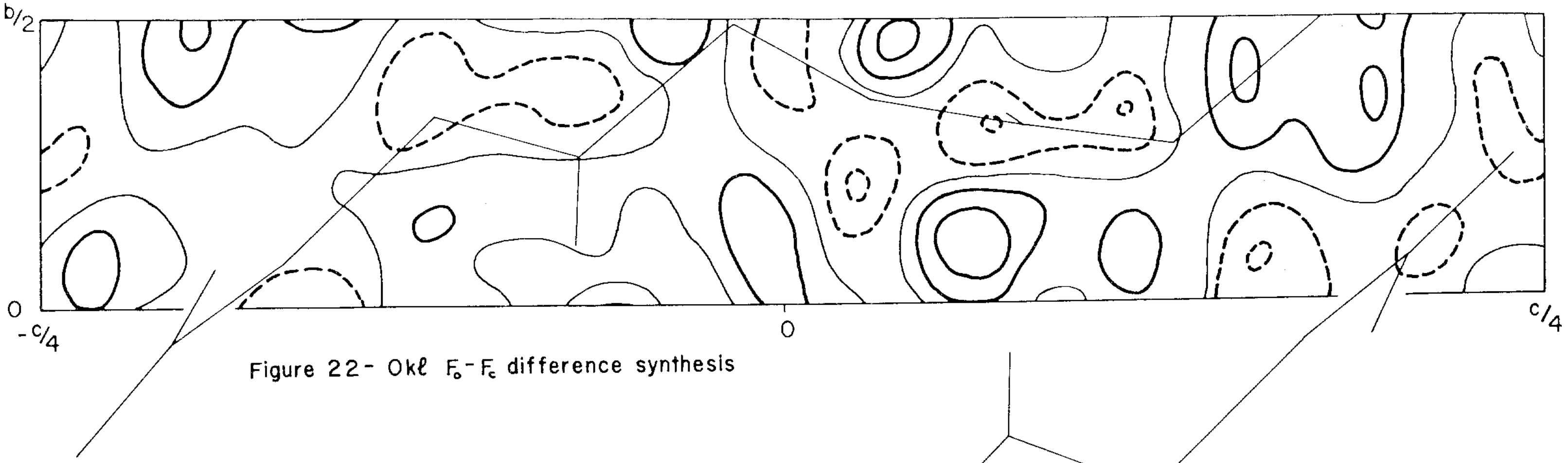


Figure 22- Okℓ $F_o - F_c$ difference synthesis

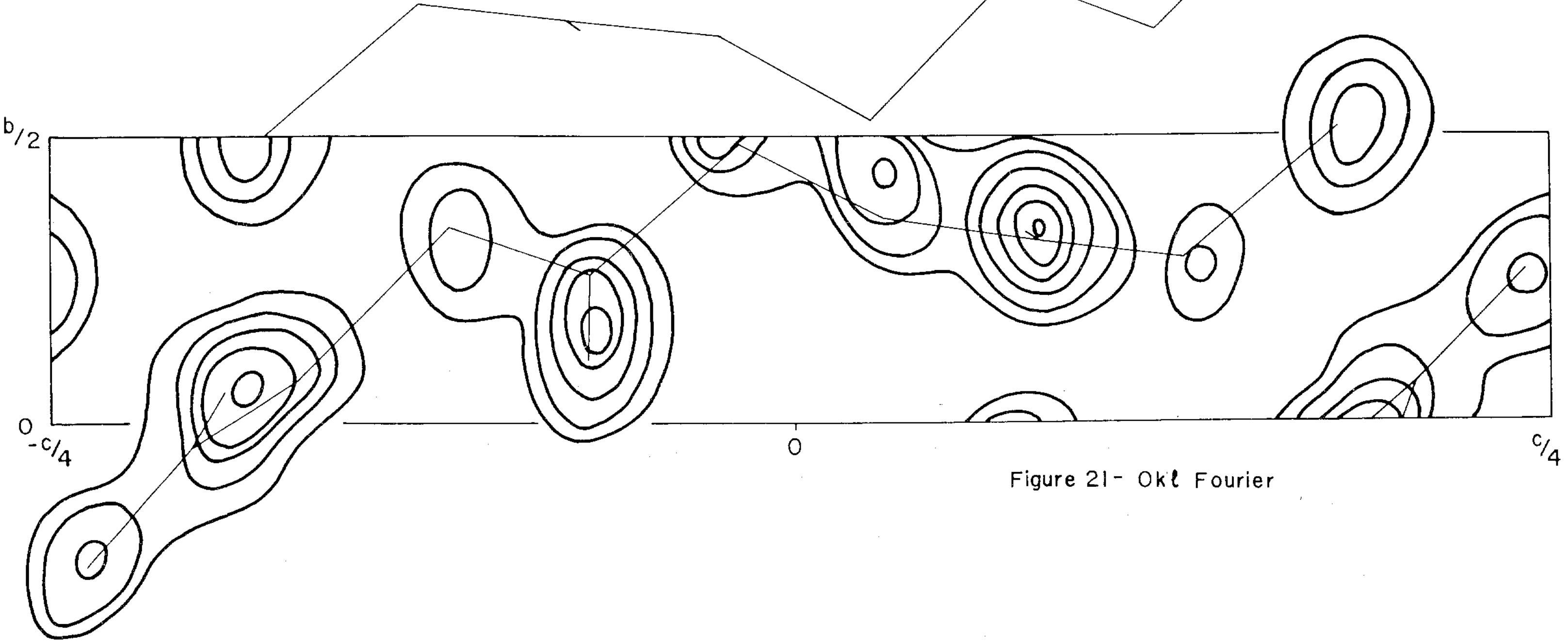


Figure 21- Okℓ Fourier

It is not expected that the y parameters can be trusted to much better than 0.2 Å, but it is considered that sufficient refinement has taken place to enter into three dimensional work as the next stage of refinement.

There is a network of hydrogen bonds approximately perpendicular to the c-axis which link together the amide groups. This network gives rise to a section of an anti-parallel pleated sheet as described by Pauling and Corey (1953). The sheet passes obliquely through the unit cell in such a way that one sheet passes through the centers of symmetry located at (000), ($\frac{1}{2}$,$\frac{1}{2}$,0), (1,1,0), etc., with other sheets parallel to it but translated by multiples of b. By the action of the c glide symmetry plane a similar set of sheets arises, one of which passes parallel to (1,0,$\frac{1}{2}$), ($\frac{1}{2}$,$\frac{1}{2}$,$\frac{1}{2}$), (0,1,$\frac{1}{2}$), etc. The sheets are theoretically of infinite length along the direction of hydrogen bonding and have the width of a triglycine molecule. The present stage of refinement seems to indicate that the terminal NH_3 group establishes three hydrogen bonds with adjacent carboxyl groups. The intramolecular dihedral angles of the amide groups are by no means 180°, but no accurate determinations of their exact angles are yet available.

The next step proposed in this investigation is to estimate the intensities of the reflections on the non-zero layer Weissenberg photographs and to use the ElectroData Computer for successive cycling of structure factor and least squares calculations. It is expected that the computer will be reprogrammed to handle asymmetric temperature factors before the final parameters are calculated.

The present stage of refinement has yielded the following parameters:

	x	y	z
O1	0.2199	0.767	0.7585
C1	.2825	.908	.7930
O2	.4196	.067	.8096
C2	.1679	.100	.8342
N1	.2502	.292	.8818
C3	.1823	.250	.9290
O3	.0481	.108	.9296
C4	.2739	.500	.9848
N2	.2289	.388	.0329
O4	.4349	.350	.0786
C5	.3088	.325	.0823
C6	.2418	.255	.1292
N3	.3750	.500	.1844

h0ℓ Structure Factors (M denotes maximum value)

h	ℓ	$\lvert F_o\rvert$	F_c
	2	15	8
	4	15	18 -
	6	36	24 -
	8	62	56 -
	10	16	16
	12	21	26
	14	5 M	
	16	35	31
	18	31	29 -
	20	24	27 -
	22	17	14 -
	24	7	8
	26	10	7
	28	8	4
1		29	34 -
1	2	47	48
1	4	33	35 -
1	6	8	4 -
1	8	4 M	
1	10	36	35 -
1	12	17	9 -
1	14	14	12
1	16	9	7 -
1	18	25	24
1	20	28	24
1	22	6	2
1	24	5	7 -
1	26	3	3
2		116	110 -
2	2	46	58
2	4	14	13
2	6	7	2 -
2	8	58	76 -
2	10	36	41 -
2	12	5	4 -
2	14	32	33
2	16	6 M	7
2	18	23	20
2	20	6 M	
2	22	6 M	
2	24	11	14 -
2	26	3 M	
3		8	2 -
3	2	7	4
3	4	6	7
3	6	7	5
3	8	10	9
3	10	22	19
3	12	13	9 -
3	14	6 M	9 -
3	16	7 M	
3	18	13	10 -
3	20	12	6 -
3	22	5 M	
4		38	42
4	2	22	25
4	4	36	42
4	6	20	16 -
4	8	37	33
4	10	6 M	
4	12	19	16 -
4	14	7 M	
4	16	6 M	11
4	18	9	8 -
4	20	5 M	
4	22	7	12 -
5		32	38
5	2	5 M	9 -
5	4	13	15 -
5	6	12	9
5	8	23	24 -
5	10	12	9 -
5	12	15	17
5	14	13	17 -
5	16	11	11
5	18	5 M	
5	20	4 M	
6		33	34 -
6	2	8	6
6	4	24	23
6	6	9	8 -
6	8	25	26 -
6	10	19	25 -
6	12	6 M	7 -
6	14	12	11
6	16	5 M	
6	18	10	15

h0ℓ Structure Factors (continued)

h	ℓ	$\lvert F_o \rvert$	F_c
7		6 M	7
7	2	11	12
7	4	7 M	
7	6	11	9
7	8	6 M	
7	10	6 M	7
7	12	5 M	
7	14	4 M	6
8		6	7 -
8	2	9	7
8	4	5 M	
8	6	22	26
8	8	19	21
8	10	5 M	
8	12	8	8 -
9		6 M	8 -
9	2	5 M	
9	4	11	11
9	6	9	12 -
9	8	3 M	3
10		8	3 -
10	2	8	12 -

h	ℓ	$\lvert F_o \rvert$	F_c
1	2 -	17	15
1	4 -	30	36
1	6 -		3 -
1	8 -	41	47
1	10 -	45	60 -
1	12 -	17	8
1	14 -	5 M	7
1	16 -	6 M	
1	18 -	9	5
1	20 -	37	31 -
1	22 -	26	27 -
1	24 -	12	10
1	26 -	10	11 -
1	28 -	12	10
2	2 -	56	54 -
2	4 -	44	51 -
2	6 -	19	15 -
2	8 -	39	54 -
2	10 -	67	77
2	12 -	46	54
2	14 -	11	4 -
2	16 -	6 M	
2	18 -	22	18 -
2	20 -	6 M	7
2	22 -	24	21
2	24 -	5	8 -
2	26 -	11	10
2	28 -	4 M	
3	2 -	7	9
3	4 -	7	3 -
3	6 -	9	5
3	8 -	44	50 -
3	10 -	52	48
3	12 -	13	12 -
3	14 -	34	29
3	16 -	6 M	
3	18 -	6 M	
3	20 -	21	16
3	22 -	7	11
3	24 -	12	6 -
3	26 -	7	5
3	28 -	4 M	

$h0\ell$ Structure Factors (concluded)

h	ℓ	$\lvert F_o \rvert$	F_c
4	2-	19	20
4	4-	30	29-
4	6-	33	34-
4	8-	44	49
4	10-	33	28-
4	12-	14	13-
4	14-	6 M	6
4	16-	22	23-
4	18-	6 M	
4	20-	7	14-
4	22-	6	6-
4	24-	10	10
4	26-	5 M	
4	28-	4 M	
5	2-	22	22-
5	4-	17	24
5	6-	18	16-
5	8-	10	1
5	10-	20	16-
5	12-	19	1
5	14-	5 M	
5	16-	19	20
5	18-	22	21-
5	20-	11	13-
5	22-	16	18-
5	24-	6	6-
5	26-	11	11
6	2-	6	8-
6	4-	39	36-
6	6-	13	11
6	8-	15	13
6	10-	18	16
6	12-	48	50
6	14-	5	7-
6	16-	20	16-
6	18-	7 M	
6	20-	12	5
6	22-	11	3
6	24-	11	11
6	26-	4 M	

h	ℓ	$\lvert F_o \rvert$	F_c
7	2-	6	8-
7	4-	8	9-
7	6-	5 M	
7	8-	30	24-
7	10-	34	37
7	12-	21	13
7	14-	22	21
7	16-	6	4-
7	18-	10	10-
7	20-	6 M	
7	22-	5 M	
7	24-	4 M	7
7	26-	8	8
8	2-	25	23-
8	4-	6	3
8	6-	6 M	
8	8-	18	15
8	10-	7 M	
8	12-	30	32-
8	14-	18	14-
8	16-	13	7-
8	18-	6 M	
8	20-	8	5-
8	22-	13	13
8	24-	3 M	
9	2-	6 M	6
9	4-	14	11-
9	6-	6 M	
9	8-	20	18
9	10-	7	9-
9	12-	6 M	
9	14-	21	18-
9	16-	5 M	
9	18-	11	6-
9	20-	10	3-
10	2-	4	4
10	4-	21	18
10	6-	5 M	
10	8-	5 M	
10	10-	5 M	5
10	12-	4 M	
10	14-	9	7

0kℓ Structure Factors

k	ℓ	$\lvert F_o \rvert$	F_c	k	ℓ	$\lvert F_o \rvert$	F_c
1	1	24	29 -	2	10	18	18
1	2	8	19 -	2	11	20	19
1	3	29	47 -	2	12	9	3 -
1	4	11	9	2	13	11	9 -
1	5	3 M	12	2	14	8	6 -
1	6	7	11	2	15	15	8 -
1	7	28	24 -	2	16	8	5 -
1	8	11	11 -	2	17	10	9 -
1	9	7	12 -	2	18	5 M	1
1	10	16	23	3	1	5	5
1	11	4 M	4 -	3	2	11	4 -
1	12	8	6 -	3	3	5	
1	13	11	3	3	4	15	9
1	14	14	19	3	5	11	13 -
1	15	17	13	3	6	5 M	1
1	16	8	8	3	7	5	5 -
1	17	19	4	3	8	6	1 -
1	18	9	7 -	3	9	5 M	2
1	19	10	13	3	10	5 M	6
1	20	26	22 -	3	11	5 M	1 -
1	21	7	1 -	3	12	5 M	3
1	22	8	5 -	3	13	10	
1	23	12	5 -	3	14	11	1 -
1	24	5	3	4		6	1
1	25	4 M	3	4	1	6	2 -
1	26	7	11	4	2	5 M	3 -
2		3	10 -	4	3	10	1
2	1	24	21 -	4	4	5 M	1
2	2	10	4 -	4	5	7	2
2	3	30	38 -	4	6	6	1
2	4	17	17	4	7	4 M	2 -
2	5	16	18 -	4	8	5 M	5
2	6	49	47	4	9	10	2 -
2	7	24	20 -	4	10	5 M	
2	8	26	28 -	4	11	9	1 -
2	9	11	20				

Appendix

Another Approach to the Phase Problem through the Intensity Continuum

Throughout the following discussion it is assumed that there is a center of symmetry located at the origin, thereby reducing the phase problem to a problem of sign determination.

Since the electron densities $\rho(r)$ of crystals are periodic, their Fourier transforms F(h) are discrete functions of h. Two kinds of information are lost when taking X-ray pictures of a crystal, either of which would suffice to elucidate its structure: The signs of F(h) are lost since only intensity data are available, and because of the periodicity, the continuous function $F_c(h)$ corresponding to the contents of one unit cell of the crystal is also unavailable.

Let us suppose that the square of this continuous function $F_c(h)$ is known (which here shall be called the intensity continuum $I_c(h) \equiv F_c^2(h)$). When a complete node appears in this intensity continuum $I_c(h)$ between two regions of finite intensity, only then may the two corresponding regions in the Fourier transform continuum $F_c(h)$ have opposite signs. A complete node is defined as a region of zero intensity which completely separates the above-mentioned regions of finite intensity.

Methods have been described for obtaining experimental data concerning the intensity continuum by W.L.Bragg and M.F.Perutz (1952), W.H.Zachariasen (1945), and Yü (1949). D. Sayre (1952) has pointed out that to obtain the intensity continuum it is only necessary to obtain these intensities at half-integral and integral values of h. All of these methods require information in addition to that supplied by the value of the intensity at integral values of h.

The present method uses the intensity continuum of some trial structure and derives from it a folding function. This function is then folded with the observed intensities to give an approximation to the correct intensity continuum. From this approximate intensity continuum, signs of successive Fourier amplitudes are chosen by noting when the intensity continuum approaches zero.

It is assumed that the electron density within the unit cell is nowhere negative and that the shapes of the atoms require that the Fourier transform be convergent. If necessary, more rapid convergence can be introduced by the addition of an artificial temperature factor.

The intention of this discussion is to describe a method of obtaining successive approximations to the correct folding function mentioned above. In practice, it is not the folding function itself which is used. Instead, it is more convenient to work with its Fourier transform. This transform is the ratio of the Patterson map which is obtained by using only the interatomic vectors of one unit cell to the complete Patterson obtained by making a Fourier transform of the observed intensities. The approximation to this folding function transform is that obtained by using the ratio of the vector map of the unit cell of the trial structure to the normal calculated Patterson of this same trial structure. When this ratio, which is a finite and non-periodic function, is multiplied by the normal observed Patterson, an approximation to the vector map corresponding to one unit cell is obtained. The folding function transform is, therefore, a device which extracts from the full Patterson only those vectors belonging to one unit cell. The Fourier transform of this finite vector map then gives rise to an approximate intensity continuum, which now has the property that the intensities at all Bragg points of this continuum are identical to those of the observed intensities, even though the calculated trial structure intensities may have had either zero or very different magnitudes for the same points.

Although the method has its greatest power in three dimensions, since a complete node in the higher dimension intensity continuum can be more positively identified, it will be simpler to discuss examples of the method by referring to one-dimensional cases.

Two extreme cases may first be mentioned. In one case, if all the electron density in the unit cell were located in the region $-\frac{1}{4}<x<\frac{1}{4}$, then the vector map corresponding to this structure would contain no vectors outside of the region $-\frac{1}{2}<x<\frac{1}{2}$, and no vectors representing interactions between neighboring unit cells would fall inside this region. Therefore, a folding function transform of the form $\begin{cases} T(x)=1, \text{ where } -\frac{1}{2}<x<\frac{1}{2} \\ T(x)=0, \text{ everywhere else} \end{cases}$ could be used to multiply the normal Patterson function. The Fourier transform of the resulting product would then be the desired intensity continuum.

Another extreme case arises when the electron density within the unit cell is almost uniformly distributed throughout the entire unit cell, i.e. $U_{ooo} \gg U_{hkl}$. The vector map corresponding to this structure would be the fold of this nearly constant density function with itself giving rise to a triangular wedge with the maximum at the origin and decreasing linearly to zero at $x=\pm 1$. The normal periodic Patterson obtained for this structure would instead appear to consist of an approximate constant vector density. The ratio of these two vector maps would give a triangular folding function transform approximated by

$$\begin{cases} T(x)= 1-|x|, \text{ for } -1<x<1 \\ T(x)= 0 \quad, \text{ everywhere else} \end{cases} .$$

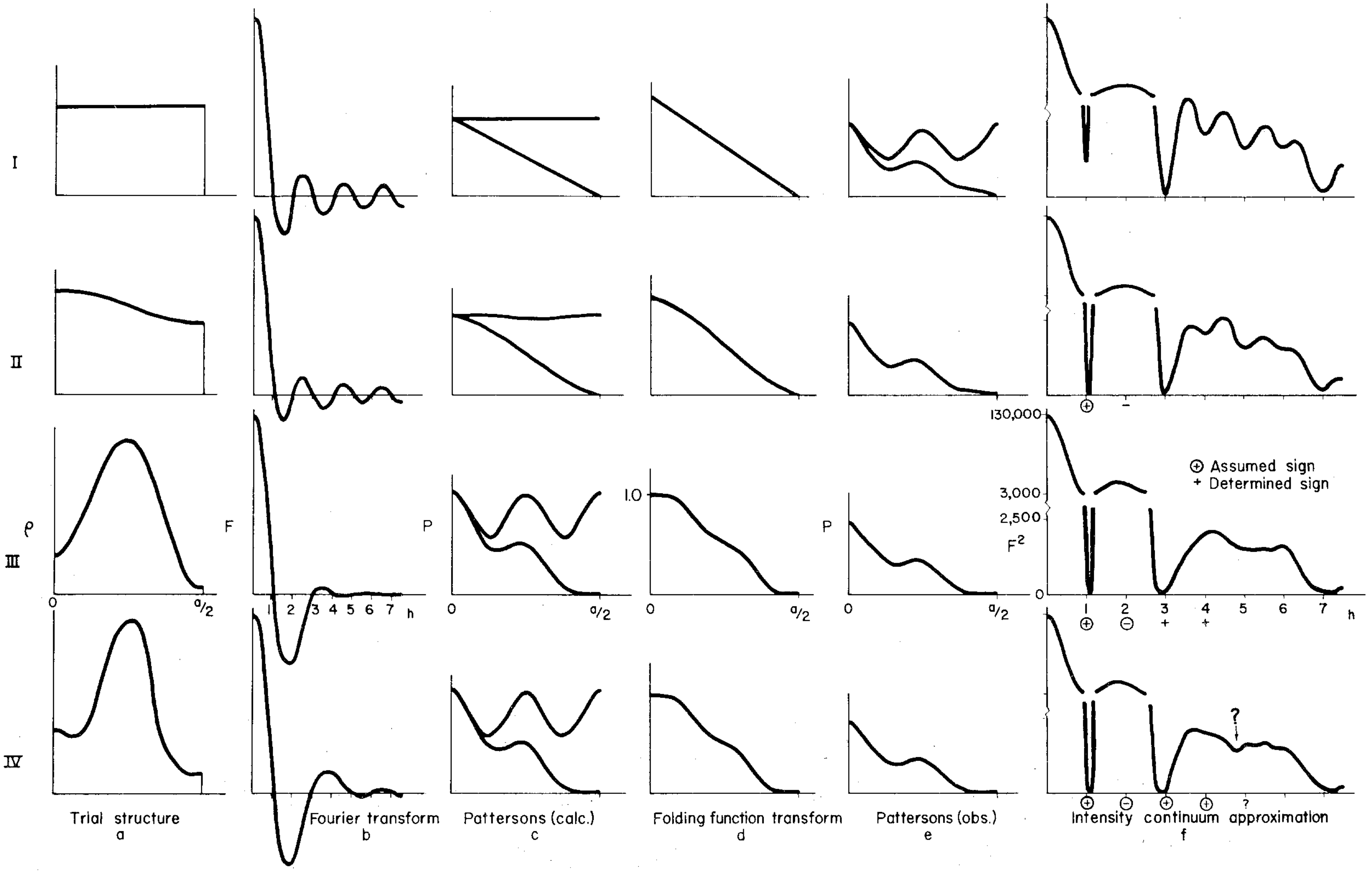
The folding function itself would be everywhere positive;

$$M(h)= \left(\frac{\sin \pi h}{\pi h}\right)^2$$

(see Waser and Schomaker 1953). The result of the folding operation will give rise to an intensity continuum which can not in general be zero except at a Bragg point, and then only if the corresponding Bragg intensity is itself zero. It is implied, therefore, that the sign can change only when the observed intensity is zero.

A third case arises when chemical evidence or such devices as the inequalities or statistical methods for sign determination have given some information which can be used in a trial structure. It was, in fact, used for alpha-triglycine for the $h0\ell$ projection trial structure of R=52%. The approximate intensity continuum obtained in this way was very complex and although many hints with regard to choice of sign were present, my experience with the method at that time very likely did not enable me to utilize fully and correctly the information which was present. There were some conclusions drawn and used, but since the other refinement techniques mentioned in the discussion of that structure were used simultaneously, it has not yet been shown conclusively that this method was the deciding factor in solving the structure.

The fourth and most general approach, when no other information is available, is perhaps the most powerful, and shall be illustrated in an actual case by taking the h00 intensities for alpha-triglycine. (All figure references in this discussion are to figure 23. The scales for the figures shown in row III hold for all the other rows.) Figure (a) of each row of curves illustrates the electron density of the trial structure from $x=0$ to $x=\frac{1}{2}$. The second figure (b) illustrates its continuous Fourier transform. The third figure (c) is the Patterson map for the trial structure where the upper curve is for the periodic structure and the lower curve is for one unit cell.



Trial structure
a

Fourier transform
b

Pattersons (calc.)
c

Folding function transform
d

Pattersons (obs.)
e

Intensity continuum approximation
f

Figure 23 - Illustration of intensity continuum method for h00 row of α-triglycine

The fourth digure (d) is the ratio of these two functions, and is, therefore, the approximate folding function transform. The fifth figure (e) shows the Patterson function in the upper curve as calculated from the observed intensities of the h00 row of alpha-triglycine, while the lower figure is this Patterson multiplied by the folding function which was just derived. The final figure of the row shows the intensity continuum which is the transform of the approximate vector map for the unit cell interactions.

For the sake of illustration the first trial structure used is that of a uniform distribution of electron density and is seen as row I of the figure.

One choice of phase for the one-dimensional case is arbitrary. In order that the intensity continuum be as slowly changing a function as possible, it is desirable that the lower orders of its Fourier transform be large with respect to its higher orders. One can accommodate this desire by choosing the sign of the first order as plus. This choice of sign has the effect of placing the greater electron density nearer the origin and is similarly reflected in its vector map. (Or, as shall be seen later, if the lower orders are very weak, then one can choose as having a positive sign the odd order with the largest intensity.) The first trial structure will then consist of the Fourier transform of the two known Fourier components, i.e. F_{000} and that F whose sign has just been chosen, which here is F_{100}. The trial structure for this choice appears in IIa. The same operations are performed on it as for the first row, and its resulting intensity continuum approximation appears in figure IIf. Visual examination of this curve indicates the presence of a node between the first and second orders implying a change of sign with F_{200} now having a negative sign. The next approximation (row III) uses for a trial structure the amplitudes F_{000}, F_{100}, and $-F_{200}$. Progress along the reciprocal row is indicated in the remainder of the figure. A point of ambiguity is finally reached between the fourth and fifth orders, and the method apparently falters.

Closer examination of the method reveals that the sign of the fifth order is being determined by the relatively weaker first order (since the third order is extremely weak). One would then suppose that perhaps the initial choice of sign could have been made by choosing the fifth order reflection as having a positive sign. Other arguments can also be presented by considering the magnitudes of only the even orders. But these cases are meant to be only illustrative, and too much emphasis should not be put on the one-dimensional case.

I will say that by using the above method, I was able to derive a set of signs for all the significant intensities of the h00 row which agreed with those ultimately calculated for the $h0\ell$ projection of alpha-triglycine. Since I had "a posteriori" information, I will not vouch for myself as being completely unprejudiced in my visual judgement of the presence or absence of an intensity continuum node.

The method requires more work per iteration of each trial structure as compared to the conventional Fourier refinement technique, but it may be possible to progress with a poor trial structure by this method where the conventional approach may not refine at all. Also, all the information which is contained in the Patterson function is automatically utilized in this method, whereas the inequalities formulae pick out isolated points from reciprocal space without using the information contained in the other reflections. It is, therefore, suggested that when all the other methods fail or only give partial information, the intensity continuum method should be tried.

No claim can be made that what is proposed is the solution to the phase problem. It is simply another approach.

Bibliography

Bernal, J.D., and Crowfoot, D. -- J. Chem. Soc. 1935 93

Bragg, W.L., and Perutz, M.F. -- Proc. Roy. Soc. A 213 425 (1952)

Corey, R.B., and Pauling, L. -- Proc. Roy. Soc. B 141 10 (1953)

Cruickshank, D.W.J., and Robertson, A.P. -- Acta Cryst. 6 698 (1953)

Hughes, E.W. -- J.A.C.S. 63 1737 (1941)

Jacobs, W.A. -- J. Biol. Chem. 147 567 (1943)

James, R.W., and Brindley, G.W. -- Int. Tab. 571 (1935)

Knaggs, J.E. -- Proc. Roy. Soc 150 576-602 (1935)

McWeeny, R. -- Acta Cryst. 4 513 (1951)

Oakley, M.B., Kimball, G.E. -- J. Chem. Phys. 17 706 (1949)

Oneto, J.F. -- J. Am. Pharm. Assoc. 35 204 (1946)

Pauling, L., and Corey, R.B. -- Proc. Roy. Soc. B 141 21 (1953)

Robertson, J.M. -- Organic Crystals and Molecules (Cornell University Press 1953)

Sayre, D. -- Acta Cryst. 5 843 (1952)

Waser, J. and Schomaker, V. -- Rev. Mod. Phys. 25 671 (1953)

Wheatley, P.J. -- Acta Cryst. 8 224 (1955)

Whittaker, E.T., Robinson, G. -- The Calculus of Oservations (Blackie and Son Ltd. 1940)

Wiesner, K., Armstrong, R., Bartlett, M.F., Edwards, J.A. -- J.A.C.S. 76, 6068 (1954)

Wiesner, K., Taylor, W.I., Figdor, S.K., Bartlett, M.F., Armstrong, J.R., and Edwards, J.A. -- Ber. 86 800 (1953)

Wiesner, K., Figdor, S.K., Bartlett, M.F. and Henderson, D.R. -- Can. Jour. Chem. 30 608 (1952)

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

Yakel, H.L. -- Ph.D. thesis, California Institute of Technology (1952)

Yü, S.H. -- A paper which was to have been presented at the meeting of the ASXRED in Ithica, N.Y., June 1949

Zachariasen, W.H. -- Theory of X-ray Diffraction in Crystals (Wiley and Sons 1945)

Propositions

1a. Contrary to the conclusion of Cowley*, it is my belief that the present-day electron microscope is capable of being used to derive a high-resolution image corresponding to an experimental solution of the crystallographic phase problem.

b. For those crystals which cannot be subjected to the high vacuum of the electron microscope, it is suggested that attempts be made to extend the above method to the X-ray microscope of Kirkpatrick and Baez**.

* J.M. Cowley -- Proc. Phys. Soc B 46 1096 (1953)

** P. Kirkpatrick and A.V. Baez -- J. Opt. Soc. Am. 38 766 (1948)

2. Depending on the validity of Jacobson's work a practical mind reading device is proposed.

* E. Jacobson -- Am. J. Psych. 44 677 (1932)
Am. J. Psysiol. 97 200 (1931)

3. A rheological model in the form of a sliding block can be used to represent structural breakdown.

4. Where there are either heavy atoms or disordered atoms in a structure, a method of weighting is proposed which should bring about a smaller root mean deviation for the parameters determined for the lighter atoms.

5. A method is proposed for speeding up the present IBM M-card system for performing Fourier transforms. This method automatically decreases the chances for making mistakes.

6. It is suggested that F_o-F_c difference maps are an efficient means of determining the identities of hetero atoms in ring systems of molecules whose structures are being determined by X-ray diffraction techniques.

7. As a means of testing the hypothesis of von Muralt* concerning the structure of myelinated nerve fibers, it is proposed that Ag^+ be injected by a microelectrode into the nerve axon in the region between two Ranvier nodes.

* von Muralt, A. -- Die Signalübermittlung im Nerven (1946) Verlag Birkhaüsen

8. A practical smog control for buildings is proposed. Such a method would be capable of removing the ozone which is postulated by Haagen-Smit* to be the source of eye irritation.

*Haagen-Smit, A.J. -- Eng. and Sci. Mo. 18 11 (1954)

9. The ASTM method for the determination of oxygen dissolved in water below the region of 100 ppb of O_2 is open to a question of accuracy.

10. Arguments will be presented as to why I feel work should be started immediately upon the crystal structure of alpha-chymotrypsinogen.

11. It is proposed that a text of the nature of Robertson's "Organic Crystals and Molecules" (1953 Cornell University Press) be adopted in conjunction with a summary of calculating techniques in a course of crystallography for organic analysts where practical technique is emphasized with minimum physical theory.

12. Stokes has invented a device used in psychotherapy called the "Auroratone", wherein the colors of thin crystals viewed between crossed "Nicols" are varied by rotation in time with the rhythm of background music. It is proposed that liquid crystals be used in order that a greater variety of color be obtained by applying appropriate impulses of pressure on the liquid and also in order to remove the distractions caused by the appearance of sharp edges associated with solid crystals.

13. From personal experience and observation it is suggested that some form of group psychotherapy be adopted, perhaps in the form of a course in psychology, by the Institute for the graduate students. It is hoped that such action would help the students become more aware of their own drives and motivations in such a way that they would become better, more efficient scientists.