This dissertation has been
microfilmed exactly as received

69-8131

NICHOLAS, Jr., Hugh Blackard, 1941-
THE STRUCTURE OF ZINC GLYCYLGLYCYL-GLYCINE; A PATTERSON SOLUTION METHOD; AND THE ATTEMPTED SOLUTIONS OF ACETYL-PROLINE.

The University of Oklahoma, Ph.D., 1969
Chemistry, physical

University Microfilms, Inc., Ann Arbor, Michigan

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/2510

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE STRUCTURE OF ZINC GLYCYLGLYCYLGLYCINE; A PATTERSON SOLUTION METHOD; AND THE ATTEMPTED SOLUTIONS OF ACETYLPROLINE

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

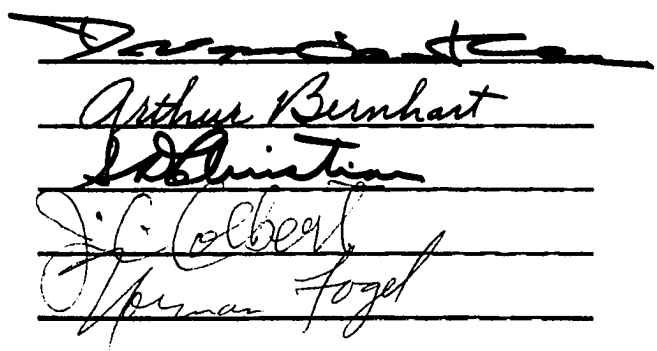
HUGH BLACKARD NICHOLAS, JR.

Norman, Oklahoma

1969

THE STRUCTURE OF ZINC GLYCYLGLYCYLGLYCINE; A PATTERSON SOLUTION METHOD; AND THE ATTEMPTED SOLUTIONS OF ACETYLPROLINE

APPROVED BY



DISSERTATION COMMITTEE

ACKNOWLEDGEMENT

The author wishes to express his appreciation to Dr. D. van der Helm, who suggested and directed the research topics presented in this dissertation. The author also wishes to express his appreciation to Dr. van der Helm for initiating the discussions concerning the erroneous standard deviations which result from the least square method as it is usually applied to polar space groups. These discussions led to the formulation of the vector refinement concept which is presented in part II of this dissertation.

The author wishes to acknowledge the support of the National Institute of Health. This support came from N.I.H. grant GM-10514 obtained by Dr. van der Helm and from an N.I.H. Fellowship, F1 GM-28,189.

The author wishes to extend his thanks to Dr. Tom Karns for taking and interpreting the N.M.R. spectra used in the identification of N-Acetyl-l-Proline $\cdot H_2O$, as well as for several helpful discussions on a wide variety of problems.

The author also wishes to thank Miss Brenda Bailey for her patience in typing the dissertation.

Finally the author wishes to thank his parents, Hugh and Nevola Nicholas for their patience and encouragement. Also the author wishes to thank his wife, Ann, and children, Brad, Carol Lea, and Dianna for their patience and encouragement.

TABLE OF CONTENTS

	Page
LIST OF TABLES	v
LIST OF ILLUSTRATIONS	vii
PART I. THE CRYSTAL AND MOLECULAR STRUCTURE OF DIAQUO ZINC GLYCYLGLYCYLGLYCINATO HEMISULFATE DIHYDRATE	
Chapter	
I. INTRODUCTION	1
II. EXPERIMENTAL	4
III. NOMENCLATURE	7
IV. SOLUTION OF THE STRUCTURE	10
V. REFINEMENT OF THE TRIAL STRUCTURE	27
VI. DISCUSSION OF THE STRUCTURE	41
PART II. A PROPOSAL FOR VECTOR REFINEMENT AND SOLUTION OF CRYSTAL STRUCTURES	99
PART III. ATTEMPTED SOLUTIONS OF THE CRYSTAL STRUCTURE OF N-ACETYL-L-PROLINE MONOHYDRATE	115
PART IV. SUMMARY AND CONCLUSIONS	133
REFERENCES	135
APPENDIX A: MATHEMATICAL ANALYSIS OF CONFORMATIONAL ANGLES	139
APPENDIX B: THE SYMBOLIC ADDITION PROCEDURE	144
APPENDIX C: LIST OF COMPUTER PROGRAMS USED DURING THESE INVESTIGATIONS	151

LIST OF TABLES

Table — Page

PART I

1. Zinc-Sulfur Vectors for the Trial Solutions..... 24

2a. Electron Density of the Atoms..... 32

2b. Electron Density of the Atoms from the Difference Fourier..... 33

3a. Atomic Parameters of $Zn(ggg).1/2SO_4 \cdot 4H_2O$..... 34

3b. Anisotropic Temperature Valves..... 36

3c. Principal Axes and Direction Cosines of Anisotropic Ellipsoids.. 37

3d. List of Observed and Calculated Structure Factors Using Final Parameters of All Atoms..... 39

4a. Comparison of Zinc Surroundings with Values for Ideal Trigonal Bipyramid and Tetragonal Pyramid..... 44

4b. Distances of the Ligands from the Zinc..... 45

4c. Known Zinc Complexes with Penta-Coordination..... 46

4d. Least Squares Planes of the Zinc Surrounding..... 47

5. Ligand Fields and Coordination Numbers in Cu(II) Complexes..... 51

6a. Hydrogen Bond Distances and Angles..... 60

6b. Other Relevant Distances and Angles in the Hydrogen Bonds..... 61

6c. van der Waal's Contacts Less than 3.50 A°..... 62

6d. Minimum Contact Distances..... 62

6e. Geometry of the Possible Bonding Contacts of the Water Molecules..... 63

7. Distances and Angles in Sulfate Ion..... 69

8a. Bond Lengths and Angles in Tripeptide Molecule..... 71

LIST OF TABLES

(Continued)

Table		Page
8b.	Average Values in Peptides and Metal Peptide Complexes.....	72
8c.	Least Squares Planes of the Peptide Chain.....	73
9a.	Least Squares Planes for the Chelate Ring.....	83
9b.	Distances and Angles in the Chelate Ring.....	84
10.	Conformational Angles of Three Triglycine Metal Complexes.....	86
11.	Carbonyl Bond Lengths of Structures Included in Marsh and Donohue's Averages.....	98
PART II		
1.	Polar Crystallographic Point Groups.....	100

LIST OF ILLUSTRATIONS

Figure Page

PART I.

1. The Zinc Ion and Its Ligands..... 43

2. The ac Projection of Zn(ggg) $1/2SO_4 \cdot 4H_2O$..... 70

3. Bond Distances in the Tripeptide Molecule..... 74

4. Bond Angles in the Tripeptide Molecule..... 75

5. A Partial ab Projection of Zn(ggg) $1/2SO_4 \cdot 4H_2O$..... 82

6. Allowed Conformational Angles for Glycine Residues..... 87

7. Conformational Angles from Known Structures..... 88

8. Reference Peptide Configuration..... 89

PART III

1. N-Acetyl-l-Proline Monohydrate Data Crystal..... 116

2a. Proton Magnetic Resonance Spectrum of N-Acetyl-l-Proline Monohydrate..... 119

2b. Proton Magnetic Resonance Spectrum of N-Acetyl-l-Proline After Drying..... 119

Appendix B

1. Geometrical Interpretation of the Inequality..... 142

I. THE CRYSTAL AND MOLECULAR STRUCTURE OF DIAQUO ZINC(II) GLYCYLGLYCYLGLYCINATO HEMISULFATE DIHYDRATE

CHAPTER I

INTRODUCTION

During the past 20 or 30 years remarkable progress has been made in the isolation of large biologically important molecules, especially proteins. These techniques have progressed to the point where relatively large single crystals of active protein can now be obtained. This has allowed the biochemist to characterize these with reasonably certainty that he has the molecular species that is active in the living organism. In turn these characterizations have revealed that the metals of the fourth period are singularly important in biological systems. The loss of a metal ion or even its replacement by another metal ion will render an enzyme of 30,000 molecular weight impotent.

The fourth period includes Potassium, Calcium, Iron, Copper, and Zinc among others. All of these are necessary to human life. For instance only about 100 mg. of Copper are found in an adult human but its importance may be seen from the fact that cytochrome oxidase is among the many enzymes it activates. Copper enzymes are efficient O_2 carriers. Copper deficiency also leads to a deficiency of heme proteins.

Zinc is involved in biological oxidation-reduction with many of the flavoproteins although it apparently is not directly involved in the electron transfer. Zinc is also essential in many hydrolytic

enzymes, for example alkaline phosphotase, carbonic anhydrase and carboxy-peptidase. Zinc is also unique among the biologically active metals; it is the only post-transition metal ion with a well established catalytic role.

Further during the course of the characterization of the bio-metallic systems it became obvious that the metal ions behaved differently from the usual metal-inorgnaic ion complexes. With these facts in mind the decision was made to isolate and determine the structure of complexes of 4th period metals in protein like environment. It is hoped that these experiments will help elucidate the differences between the fourth period ions and thereby explain their specifity. So that the results would be relevant to biological systems, the complexes have been isolated almost exclusively from solutions of approximately neutral pH.

The role of zinc ion in biochemical phenomena was of special interest. Virtually all zinc enzymes, except those involving flavinoid co-enzymes, are hydrolytic. Further the mechanisms of many of them appear similar.[1] Metal replacement studies show a similar pattern in carbonic anhydrase, alkaline phosphotase and carboxy-peptidase.[2,3,4] The zinc and cobalt complexes are active while the copper compounds are completely inactive. Spectral studies show a similar bonding geometry in carboxy-peptidase and carbonic anhydrase. Also carbonic anhydrase has been shown to have esterase activity.[2] In view of Freeman's comparative structural[5] studies, which indicated a marked similarity between copper and zinc geometries, the inactivity of the copper enzyme complexes was disturbing. Therefore when crystals of diaquo zinc glycyl glycyl

glycinato hemi-sulfate dihydrate were obtained it was decided that this structure would offer an excellent opportunity to investigate the differences of copper and zinc ions in biological surroundings, since the structure of copper(II) glycyl glycyl glycinato chloride sesquihydrate had been reported.[46]

An obvious question concerning the validity of the results of conformational studies in the crystalline state arises, mainly to what extent does the process of crystallization affect the geometry of the metal surroundings? While it is impossible to prove that there is no effect, it may be safely said that the effect is very small or perhaps even negligible. This conclusion is reached after examining the metal surroundings with different ligands and the different metals with the same ligand. These investigations show a large variation in the crystal system and the number of solvent molecules included in order to fill space. This indicates that the chemistry of the system determines the metal geometry rather than considerations of efficient space filling.

CHAPTER II

EXPERIMENTAL

During attempts to grow crystals of the mixed chelate, zinc glycyl glycyl glycinate l-serinate, long boat-shaped plates crystallized from several of the solutions. The solutions were prepared by dissolving equi-molar quantities of zinc sulfate, glycyl glycyl glycine and l-serine in distilled water and precipitating the sulfate by adding an equi-molar amount of barium hydroxide. Because identical crystals grew from solutions containing not only the desired compound alone but also from solutions which contained either an equivalent of hydrochloric acid or an equivalent of sodium hydroxide per equivalent of zinc, the presence of an anion other than the peptide and amino acid was not expected. Also it was thought that all of the sulfate ion had been removed by precipitation with barium hydroxide during the preparation of the chelate solution. Therefore when the crystals were found to be of a centrosymmetric space group it was thought that crystals of the compound zinc glycyl glycyl glycinate hydroxide hexahydrate had been grown.

However, structure analysis by X-ray diffraction revealed that the crystals were of the compound diaquo zinc glycyl glycyl glycinate hemi-sulfate dihydrate, $Zn(C_6H_{10}N_3O_4).\frac{1}{2}SO_4.4H_2O$. Crystals of this compound may be readily grown by layering a five-hundredths

to one-tenth molar aqueous solution of the compound with ethanol. This solution is prepared by dissolving equi-molar quantities of zinc sulfate and glycyl glycyl glycine. The sulfate is partially precipitated by the addition of a one-half molar quantity of barium hydroxide. The long, boat-shaped plates appear within twenty-four hours. The crystals are elongated along the b-axis and the plate face is perpendicular to the a-axis. Approximate chemical analysis showed a ratio of 2.3 atomic weights of zinc to one ionic weight of sulfate. The zinc was determined by titration with ethylene diamine tetraacetate[6] in basic media and the sulfate ion was determined gravimetrically by precipitation as barium sulfate.[7] The density was measured by flotation and found to be 1.763 grams per cubic centimeter, the density claculated assuming eight molecules in the unit cell is 1.762 grams per cubic centimeter. The agreement between the observed and calculated densities indicates that the ratio of zinc to sulfate is actually 2.0 to 1 as expected rather than 2.3 to 1 shown by approximate analysis.

The reflections 0kl when k = 2n+1, h0l when l = 2n+1, hk0 when h+k = 2n+1, h00 when h = 2n+1, 0k0 when k = 2n+1, and 00l when l = 2n+1 were found to be systematically absent. This uniquely determines the space group to be Pbcn. The cell parameters were fitted by least squares to thirty-four observed reflections in four octants of reciprocal space and are:

$$a = 25.86 \pm 0.03 \text{ A}$$

$$b = 8.011 \pm 0.006 \text{ A}$$

$$c = 13.59 \pm 0.01 \text{ A}$$

The intensities were measured with nickel filtered copper Kα radiation on a General Electric XRD-5 diffractometer, using the theta-two theta scan technique. 2095 independent reflections were measured out to a two theta of one hundred twenty degrees; of these 317 were too weak to be observed. Near the end of the refinement ninety-two reflections were retaken because of poor agreement between their observed and calculated values. Of these, only fifty were different enough from their original values to justify replacement by their new values. Lorentz-polarization, absorption, and anomalous dispersion corrections were applied to the data. The absorption corrections were redone when the sulfate ion was found to be present. The final linear absorption coefficient including the contribution from the sulfate ion is 37.1 cm^{-1}.

CHAPTER III

NOMENCLATURE

In order to describe the crystal and molecular structure of diaquo zinc glycyl glycyl glycinato hemi-sulfate dihydrate clearly it is necessary to introduce a standard nomenclature for the description of peptides.[8]

The term glycyl residue shall refer to the sequence of atoms: $-NH-C^{\alpha}H_2-C'O-$. The sequence of atoms, $-C^{\alpha}-C'O-NH-(C^{\alpha})$, shall be referred to as a peptide unit. The residues and units will be numbered starting at the amine terminal end of the peptide chain. The tetrahedral carbon atoms will be referred to as C^{α}_i where i is the number of the residue in which the carbon atom is found. The trigonal carbon atom will be referred to as C'_i, the nitrogen as N_i and the oxygen as O_i. The oxygen atoms in the carboxyl terminal residue will be referred to as O^1_3 and O^2_3 where O^1_3 will be the atom bonded to the C'_3 with the most double bond character, that is the oxygen with the shortest of the two C'_3-O_3 bonds. Thus $N_1-C^{\alpha}_1-(C'_1O_1)-N_2-C^{\alpha}_2-(C'_2O_2)-N_3-C^{\alpha}_3-(C'_3O^1_3O^2_3)^-$ shows the numbering of the skeletal atoms of the glycyl glycyl glycinate ion.

While the molecular structure can be described in terms of a single molecule, the description of the crystal structure requires that the interactions between the several molecules within the unit cell be discussed. Therefore it is also necessary to introduce a standard notation for referring to different molecules. This will be done with prescripts, both subscripts and superscripts.

Atoms which have neither a pre-subscript nor a pre-superscript shall be understood to have the co-ordinates given in the list of atomic positions and belong to a continuous molecule. Superscripts shall be used to designate symmetry operations within the unit cell. Subscripts shall be used to indicate unit translations in the directions of the cell axes.

The eight equivalent positions given in the section on the solution of the structure result from seven symmetry operations. Superscripts one through seven are used to indicate the symmetry operations one through seven respectively. The equations for these symmetry operations are:

$^{1}X=\frac{1}{2}-X$	$^{1}Y=\frac{1}{2}-Y$	$^{1}Z=\frac{1}{2}+Z$
$^{2}X=\frac{1}{2}+X$	$^{2}Y=\frac{1}{2}-Y$	$^{2}Z=1-Z$
$^{3}X=1-X$	$^{3}Y=Y$	$^{3}Z=\frac{1}{2}-Z$
$^{4}X=1-X$	$^{4}Y=1-Y$	$^{4}Z=1-Z$
$^{5}X=\frac{1}{2}+X$	$^{5}Y=\frac{1}{2}+Y$	$^{5}Z=\frac{1}{2}-Z$
$^{6}X=\frac{1}{2}-X$	$^{6}Y=\frac{1}{2}+Y$	$^{6}Z=Z$
$^{7}X=X$	$^{7}Y=1-Y$	$^{7}Z=\frac{1}{2}+Z$,

The subscripts will have the values ±a, ±b, or ±c, indicating positive or negative translations along the indicated axis. Also more than one direction of translation may be indicated. The use of the subscripts may be expressed by the following equations:

$_{a}X=X+1$	$_{a}Y=Y$	$_{a}Z=Z$
$_{-a}X=X-1$	$_{-a}Y=Y$	$_{-a}Z=Z$
$_{b}X=X$	$_{b}Y=Y+1$	$_{b}Z=Z$

and so on with c operation on the z co-ordinate.

The order of operation should always be the symmetry operation followed by the translation. Thus an atom $_{-c}^{2}W$ with co-ordinates of W at X, Y, Z will have co-ordinates found by applying symmetry operation 2 to yield ½+X, ½-Y, 1-Z followed by a -c translation, Z-1 to yield final co-ordinates of ½+X, ½-Y, -Z.

Additionally the oxygen atoms in the water molecules shall be referred to as O_1^W, i = 1,2,3,4. O_1^W and O_2^W shall be the water molecules bound to the zinc ion. O_i^S shall refer to the oxygen atoms of the sulfate ion. Also, for convenience the compound diaquo zinc glycyl glycyl glycinato hemi-sulfate dihydrate shall be referred to as zinc triglycine.

CHAPTER IV

SOLUTION OF THE STRUCTURE

In the Patterson function for space group Pbcn one would normally expect to find seven peaks representing zinc-zinc vectors, the Patterson inversion peak and six Harker peaks.[9] It is then possible to find the co-ordinates of the zinc atom by using the co-ordinates of these seven peaks to fit the seven sets of equations generated by the symmetry operations of the space group. These seven symmetry operations are expressed in terms of an independent set of atoms with co-ordinates X, Y, Z and the seven symmetry equivalent positions. For space group Pbcn the eight symmetry equivalent positions and the seven resulting equations for the symmetry peaks are as follows:

Equivalent Positions

X	Y	Z
-X	-Y	-Z
-X	Y	$\frac{1}{2}$-Z
$\frac{1}{2}$-X	$\frac{1}{2}$-Y	$\frac{1}{2}$+Z
$\frac{1}{2}$-X	$\frac{1}{2}$+Y	$\frac{1}{2}$+Z
X	-Y	$\frac{1}{2}$+Z
$\frac{1}{2}$+X	$\frac{1}{2}$-Y	-Z
$\frac{1}{2}$+X	$\frac{1}{2}$+Y	$\frac{1}{2}$-Z

Symmetry Peaks

a)	u = 2x,	v = 2y,	w = 2z
b)	u = -2x,	v = 0,	w = ½-2z
c)	u = ½-2x,	v = ½-2y,	w = ½
d)	u = ½-2x,	v = ½,	w = 0
e)	u = 0,	v = -2y,	w = ½
f)	u = ½,	v = ½-2y,	w = -2z
g)	u = ½,	v = ½,	w = ½-2z

However in this case the peaks corresponding to equations a and b both had v co-ordinates of zero and the peaks corresponding to equations f and g had v co-ordinates of one-half. This caused the sets of equations a and b, and f and g to become degenerate; that is, it is not possible to distinguish the peak which corresponds to equation a from the peak which corresponds to equation b. This results in the x and y co-ordinates of the zinc atom being determined but not the z co-ordinate. More exactly, this situation results in two equally probable z co-ordinates, z_1 and z_2, related to each other by the equation:

$$z_1 = \tfrac{1}{4}-z_2$$

This is the equation of a mirror plane parallel to the xy plane and with z = 1/8.

An attempt was made to remove this ambiguity by calculating the relevant sections of the Patterson function on a grid of eight one-hundredths Angstroms in the direction of the v co-ordinate. However, even at this resolution the v co-ordinates of neither of the peaks could be distinguished from zero. Thus fitting the seven peaks

marked as Zn-Zn vectors in Table A, the two possible zinc positions are found to be: $x = 0.125$, $y = 0.0$, and $z_1 = 0.062$ or $z_2 = 0.188$.

It was decided to attempt to solve the structure by interpreting the Patterson function in terms of both possible zinc positions. It was hoped that the ambiguity in the zinc positions would be resolved when one position resulted in a satisfactory structure and the other led to results which were nonsensical, either chemically or from packing considerations.

The first glycyl residue was readily found since it is chelated to the zinc atom through the amine terminal nitrogen, N_1 and the carbonyl oxygen, O_1. However, these two atoms, N_1 and O_1, could not be located accurately because of overlap in the Patterson peaks from which their co-ordinates were derived. This overlap was caused by the fact that these atoms were approximately mirrored across the $y = 0$ plane, the plane in which the zinc atom is situated.

At this point a difficulty was encountered which caused the abandonment of this method of solving the crystal structure. This difficulty was caused by the Patterson function containing both the crystal structure and its mirror image. This is true of all Patterson maps. However combined with the zinc atom having a zero y co-ordinate results in a pseudo-mirror plane in the crystal structure derived from the zinc peptide vectors of the Patterson function. That is any zinc to light atom vector which determines an atom at x, y, z also determines an atom at x, -y, z. However this type of mirror plane can be resolved.

This is done by taking advantage of Friedel's law which states that $I(h,k,l) = I(-h,-k,-l)$ where $I(h,k,l)$ is the intensity of the

reflection associated with the Miller indices or reciprocal lattice indices h,k,l. This law is extended by the symmetry operations of the Laue group of the crystal. As a consequence of this, Friedel's Law is extended for space group Pbcn to: $I(h,k,l) = I(-h,-k,-l) = I(-h,k,l) = I(h,-k,l) = I(h,k,-l) = I(-h,-k,l) = I(-h,k,-l) = I(h,-k,-l)$. This law may also be expressed in terms of the structure factors, F, by replacing I(h,k,l) in the above expression by $|F|$ (h,k,l) where $|F|$ is the absolute value of the structure factor. By examining the general structure equation:

$$F = A + iB$$

where

$$A = \sum_n f_n \cos 2\pi (hx_n + ky_n + lz_n)$$

$$B = \sum_n f_n \sin 2\pi (hx_n + ky_n + lz_n)$$

it is apparent that one may write Friedel's Laws in terms of the atomic positions in real space as well as the h,k,l indices in reciprocal space. In terms of the atomic positions Friedel's Law is $|F|(x,y,z) = |F|(-x,-y,-z)$ where (x,y,z) are the co-ordinates of the atoms in the structure and $|F|$ (x,y,z) is the contribution made to $|F|$ (h,k,l) by the atom at (x,y,x). This law may also be extended by use of the symmetry operations of the Laue group of the crystal so that: $|F|(x,y,z) = |F|(-x,-y,-z) = |F|(-x,y,z) = |F|(x,-y,z) = |F|(x,y,-z) = |F|(-x,-y,z) = |F|(-x,y,-z) = |F|(x,-y,-z)$ is valid for space group Pbcn.

Thus the mirror may be broken by use of the equality $|F|(x,y,z) = |F|(x,-y,z)$. At this point the only atoms used in the structure factor calculations have y co-ordinates of zero so that $y = -y$. Therefore one is free to choose the y co-ordinate of any other atom as either y or -y,

subject only to the constraint that all further atoms be of the same mirror image across the false mirror plane at y = 0.

Either position may be arbitrarily chosen for the first atom which is placed out of the mirror plane. This however fixes the mirror image of the structure and all subsequent atomic positions must be of this same mirror image. Therefore the mirror image was fixed by the choice of image for N_1, the first atom found in a location not in the pseudo-mirror plane.

After the mirror image was chosen in this manner the remaining atoms of the first glycyl residue were readily found. Another problem arose at this point. N_2 and N_3 were found in positions which were their own mirror images across the pseudo mirror plane in the crystal structure. This appears, at first to cause no problem. However, the problem arises in placing the next atoms in the sequence along the chain, the C_2 and C_3 atoms which are bonded to N_2 and N_3 respectively. This is because the two positions x, y, z and x, -y, z for the carbon atoms are equidistant from the nitrogen atom to which an atom at either of those positions would be bonded, also the bond angles for both positions were reasonable for the accuracy with which the atoms were placed. Further, at this stage there were not enough atoms placed to make use of packing considerations. Therefore it was not possible to choose the mirror image of the second glycyl residue, located between N_2 and N_3, such that it was necessarily of the same mirror image as the first glycyl residue. Because both N_2 and N_3 were their own mirror image a similar ambiguity existed between glycyl residues one and three as well. As a consequence a similar ambiguity also existed between the

second and third glycyl residues. Thus it was not possible to place any two glycyl residues such that they were sure to be of the same mirror image.

Therefore all attempts to solve the major portion of the structure from the Patterson function were abandoned, even though several possible peptide chains and one water molecule could be readily located for either zinc position.

After abandonment of the Patterson function the structure analysis was continued using standard heavy atom methods and successive Fourier syntheses. Because of the overlap in the Patterson of the peaks corresponding to vectors between the zinc atom and its surrounding ligands, we decided to start by using only the two possible zinc positions. Therefore two sets of structure factors were calculated, one for each possible zinc positions.

These structure factor calculations did not successfully determine the signs of enough structure factors to warrant the calculation of an electron density map. The reason for this may be seen by examining the structure factor equations for space group Pbcn, these are:

a) for $h+k = 2n$ and $l = 2n$

$$A = 8\cos 2\pi hx \cos 2\pi ky \cos 2\pi lz$$

b) for $h+k = 2n$ and $l = 2n+1$

$$A = 8\cos 2\pi hx \sin 2\pi ky \sin 2\pi lz$$

c) for $h+k = 2n+1$ and $l = 2n$

$$A = 8\sin 2\pi hx \cos 2\pi ky \sin 2\pi lz$$

d) for $h+k = 2n+1$ and $l = 2n+1$

$$A = 8\sin 2\pi hx \sin 2\pi ky \cos 2\pi lz$$

Also B is equal to zero in all cases and F, the structure factor is equal to A+iB. Since y=0 for both zinc positions equations b and d are zero for any h, k, and 1. This leaves the signs of the half of the data with 1 an odd number undetermined. Further because x is approximately 1/8, equation c will be zero when h=4n and equation a will be zero when h=2n but not 4n. These two conditions cause another eighth of the data to be calculated as zero when the zinc position alone was used. These phenomena associated with the zinc positions, along with the two lighter atoms which have y co-ordinates of nearly zero, explain the relatively high percentage of weak and unobservable reflections.

At this point two of the lighter atoms which lay very nearly at y=0 were placed using the Patterson function. These atoms were later identified as N_2 and O^2_3. These two atoms were then included in the structure factor calculations along with the zinc atom. Because all three atoms have y co-ordinates of zero, half of the structure factors again calculate as zero. However a sufficient number of the signs of the structure factors were well enough determined to warrant calculating Fourier electron density maps for both trial zinc positions. These maps were calculated using less than half of the total data and contained a false mirror plane at y=0. The false mirror plane is a consequence of the zero co-ordinates of the three atoms used in the structure factor calculations to determine the signs of the observed data. It is caused by half of the data calculating as identically zero because of the form of the structure factor equations, which was previously discussed. However even with these

disadvantages the Fourier maps resolved the overlap and resulting ambiguity in placement of the atoms of the first glycyl residue which had been present in the Patterson map.

With this information it was now possible to begin attempts to resolve the false mirror plane. This is done by arbitrarily choosing the mirror image of one atom as explained previously. At this point then the first glycyl residue was placed by arbitrarily choosing one mirror immage. The peptide chain past the first peptide nitrogen, N_2, could not be placed at this time because the ambiguity which prevented the solution from the Patterson function still existed.

At this time another set of structure factors was calculated for both zinc positions and the seven lighter atoms associated with each of them. The atoms, other than the zinc, which had been located for both trial structures were: N_1, $C^{\alpha}{}_1$, C'_1, O_1, N_2, N_3 and $O^1{}_3$. N_3 was mistakenly assumed to be O_2 and entered as an oxygen in the structure factor calculation. $O^1{}_3$ was assumed to be an oxygen; however, it was not identified as part of the peptide at this time.

Only four of these light atoms, N_1, $C^{\alpha}{}_1$, C'_1, and O_1 were not their own mirror image across the y=0 false mirror plane. Therefore only these four contributed to resolving this false mirror plane.

Fourier electron density maps were calculated using the observed structure factor magnitudes and the signs of the calculated structure factors. Only those structure factors which had observed magnitudes of 9.9 electrons or greater and which had calculated magnitudes of three-fourths or more of their observed magnitudes, were included in the calculation of the Fourier maps.

It was hoped that these maps would have the two mirror images resolved for at least the peptide chain. However this was not the case. Even so three more atoms were resolved from their mirror images for each trial structure and the positions of the atoms in the first glycyl residue were improved. For trial solution one, with the zinc z co-ordinate of 0.062, the three newly resolved atoms included the carbonyl oxygen of the second glycyl residue, O_2, which was mistakenly identified as a nitrogen. This was a consequence of having misidentified N_3 as O_2 in the previous Fourier map. An oxygen later found to be O^2_3 was also resolved as was an atom on the two-fold symmetry axis. This atom was thought to be an oxygen from a water molecule. It was later found to be the sulfur atom in the sulfate ion. For trial solution two with the zinc z co-ordinate of 0.188 the three new atoms were the remainder of the second glycyl residue, C^{α}_2, C'_2, and O_2. In this case the previous misidentification of N_3 as O_2 was recognized at this point and corrected.

At this point in the investigation each trial solution contained eleven atoms, the zinc plus ten light atoms. For trial solution one, with zinc at z=0.062, they were: N_1, C^{α}_1, C'_1, O_1, N_2 and O^1_3 which were correctly identified. O^1_3 and O^2_3 were included as oxygens but had not been assigned to the peptide at this point. Also included but misidentified at this point were O_2 and N_3 which were identified as N_3 and O_2 respectively and an unidentified atom which was thought to be an oxygen from a water molecule. For the second trial solution all of the atoms placed at this point were also correctly identified at this time. These atoms were the zinc with a z co-ordinate of 0.188, N_1, C^{α}_1,

C'_1, O_1, N_2, C^{α}_2, C'_2, O_2, N_3 and O^1_3. O^1_3 was not yet assigned to the peptide chain. Once again structure factors were calculated for both trial solutions and Fourier electron density maps were then calculated using the results of the structure factor calculations in the manner previously described.

The Fourier map for the structure including zinc position one showed the complete peptide chain clearly resolved from its false mirror image. Also two prominent peaks appeared at a distance of 1.5 Angstroms from the supposed water molecule on the two fold axis. It seemed likely that this was caused by the presence of disordered water molecules in the crystal structure. Therefore this atom was removed to allow this region to resolve itself without the inclusion of an outside bias. Also three new peaks were tentatively assigned as water molecules. However, their mirror images were not resolved. The oxygens O^1_3 and O^2_3, were now correctly assigned to the peptide chain. The misidentification of O_2 and N_3 was also corrected as a result of this Fourier map.

The Fourier electron density map for the trial structure containing zinc position two failed to resolve the third glycyl residue of the peptide or any of the other remaining atoms from their false mirror image. In fact a complete peptide chain could not be found for either mirror image in this map. If the two images are called mirror image A and mirror image B then this may be explained as follows. Glycyl residues one and two plus N_3 were present in the Fourier map because they were used for the preceding structure factor calculations. Then C^{α}_3 was present in image A but not for image B. C'_3 was present

in image B but not in image A. O^1_3 was present for both images since it was its own mirror image but O^2_3 was not found for either mirror image A or B. This was a net regression as the three atoms C^{α}_3, C'_3, and O^2_3 were present for both mirror images but unresolved in the previous Fourier.

On this somewhat arbitrary basis zinc position one was taken as the correct zinc position and attempts to complete that trial solution were continued. All attempts to complete the trial structure including zinc position two were discontinued. Later more conclusive evidence was found to indicate that the second zinc position was indeed wrong.

Structure factors were again calculated and a Fourier electron density map computed for the trial solution containing zinc position one. The atom on the two-fold axis, which had been deleted from the structure factor calculation, returned in the Fourier map. Its electron density was high enough in this map to indicate that it was an atom considerably heavier than oxygen. Also it had an extremely regular tetrahedral surrounding at a distance of 1.5 Angstroms. The peaks for these surrounding atoms were slightly higher than any of the four identifiable water molecules. These facts indicated that the previous supposition of disorder was incorrect. A careful consideration of the circumstances of preparation indicated that sulfate ion was the most probable identity for this grouping of atoms. Qualitative test on a solution of redissolved crystals disclosed the presence of sulfate ion. In addition to the sulfate ion four molecules of water were found. Two of these were ligands of the zinc atom.

Examination of a three dimensional model of the compound in its crystal environment indicated that all of the available space was filled by these atoms. Subsequent difference Fourier maps indicated that there were no more atoms, other than hydrogen atoms, in the crystal structure. Structure factor calculations over all data, using the atomic positions revealed by the previous Fourier map yielded a residual error, R of 0.229. The residual R is defined as

$$R = \frac{\Sigma||kF_o| - |F_c||}{\Sigma|kF_o|}$$

where F_o is the observed structure factor and F_c is the structure factor calculated from the assumed atomic parameters. This is compared to a value of 0.828, the most probable value for random arrangement of incorrectly placed atoms.[10]

It is necessary to examine at this point the final choice for the correct zinc position. This is necessary because the rejection was made on an arbitrary basis. That is the phenomena on which the choice was made could conceivably have been caused by picking glycyl residues one and two of opposite mirror images. It is unlikely that this is the case, since the free choice of image for glycyl residue one led to the resolution of the two mirror images for glycyl image two and thus dictated which of the two possible images was chosen. However it is useful to consider what other evidence is available on which to base the choice between the two possible zinc positions.

The most powerful of this additional evidence results from the presence of the sulfur atom on a special position, the two-fold symmetry axis, in one of the trial solutions. From the Fourier maps

calculated using the results of the structure factor calculations involving the zinc positions and the seven associated light atoms, it was obvious that the two trial structures were related to each other by the same equation by which the two possible zinc positions were related. That is: $z_1 = \frac{1}{4} - z_2$ for any atom, not only the zinc atoms. Thus it is obvious that only the sign of the w component, the component in the z direction, of the interatomic vectors has been changed. Patterson[11] has shown that the intensities, I(h,k,l) are dependent only on the atomic composition and the interatomic vectors of a crystal structure. They are related by the equation:

$$|F^2|\ (hkl) = kI_{hkl} = \sum_{j=1}^{n} f^2\ j + \sum_{j=1}^{n}\sum_{i=1}^{n} f_i f_j \cos h(x_i - x_j) + k(y_i - y_j) + l(z_i - z_j)$$

Since the interpretation of Friedel's Law discussed earlier holds for the interatomic vectors as well as atomic positions the two structures are then homometric.[12,13] That is they have the same magnitudes in the components of their interatomic vectors and therefore the same set of intensities. However, these arguments hold only for structures whose homometric pairs have their corresponding atoms in the same symmetry.

The consequence of an atom being situated on a special position is that there are fewer of this species in the unit cell than would be required if the atom were located on a general position. In this case there are only four sulfur atoms required by the symmetry operations of the unit cell, rather than the eight required of atoms in general positions. However the location of the sulfur atom in trial solution two was not on the two-fold axis. This is predicted by the equation

relating the two trial solutions and seen in the trial Fourier electron density maps. This fact then leads to a significant difference in the interatomic vector set for the two trial structures, with the consequence that they are in fact not homometric.

For trial solution one, with zinc position one and the sulfur on the two-fold axis, one expects four vectors from sulfur to any atom in the unit cell, since there are four sulfur atoms per unit cell. For trial solution two, with zinc positon two there are eight sulfur atoms and one expects eight vectors from sulfur to any atom. Therefore it should be possible to use the Patterson function, which is simply a mapping of the magnitudes of the components of the interatomic vectors, to distinguish between the two posibilities.

Generally, the higher the peak height of the peaks used to make the distinction the more reliable the results. The highest peaks suitable for this purpose are the zinc-sulfur vectors. However because of the y co-ordinate of the zinc atom being zero, the vectors which should be missing if trial solution one is correct instead of trial solution two are of the type u,-v,w while the vectors present for both solutions are of the type u,v,w. Because the Patterson function reveals only the absolute value of u,v, and w the peaks from the two possible structures coincide. The atomic positions and derived Patterson peaks for the zinc-sulfur vectors in the two possible structures are listed in Table 1. Thus, because of the coincidence of the vectors in Table 1 it is obvious that the distinction between the two structures cannot be made using the zinc-sulfur vectors.

TABLE 1

Zinc-Sulfur Vectors for the Trial Solutions

Correct Structure

	|u|	|v|	|w|
Zinc Position	.128	.000	.062
Sulfur Positions			
a	1/2	-y	1/4
b	1/2	y	3/4
c	0	1/2-y	1/4
d	0	1/2+y	3/4
Patterson vectors			
1	.372	y	.188
2	.372	y	.312
3	.128	1/2-y	.188
4	.128	1/2-y	.312

Alternate Structure

	|u|	|v|	|w|
Zinc Position	.128	.000	.188
Sulfur Positions			
a	1/2	-y	0
b	1/2	y	0
c	1/2	-y	1/2
d	1/2	y	1/2
e	0	1/2-y	0
f	0	1/2+y	0
g	0	1/2-y	1/2
h	0	1/2+y	1/2
Patterson vectors			
1	.372	y	.188
2	.372	y	.188
3	.372	y	.312
4	.372	y	.312
5	.128	1/2-y	.188
6	.128	1/2-y	.188
7	.128	1/2-y	.312
8	.128	1/2-y	.312

The next highest peaks are those caused by the sulfur-sulfur vector. These seven peaks indeed prove suitable. The co-ordinates of the three peaks common to both structures and their heights are as follows:

u	v	w	height
½	½	0	412
½	0.018	½	293
0	0.482	½	331

Because the sulfur atom is on a two-fold axis for trial solution one these three are the only sulfur-sulfur vectors expected. The other four which are normally expected coincide with these three and the origin peak. The four additional peaks which would be expected if trial solution two were correct are as follows:

u	v	w	height
0	0	½	3500
0	0.482	0	-21
½	0.018	0	-35
½	½	½	60

The peak heights are on a scale such that 130 is the expected height of a single sulfur-sulfur vector. The very high value of the peak at 0,0,½ is caused by four multiple interactions of all atoms which lay approximately in the y=0 plane. This of course includes the zinc atoms. The other three peaks for trial solution two are below the predicted peak height for a sulfur-sulfur vector. Since these vectors are not present in the vector map trial structure two may be eliminated as a possible solution.

The evidence from the Patterson map alone is conclusive in eliminating the second zinc possibility. When combined with the experimentally found two to one ratio of zinc to sulfur, which is predicted by the structure containing zinc position one and the behavior of the Fourier electron density maps, during elucidation of the structure, the evidence in favor of zinc position one is overwhelming. Thus it can be safely stated that there is no possibility that the wrong structure has been chosen.

CHAPTER V

REFINEMENT OF TRIAL STRUCTURE

After all of the atoms were found by Fourier methods their positional and thermal parameters were refined by the block diagonal least squares method. To save computer time all atoms except the zinc were assigned isotropic temperature factors initially. At this stage of refinement the contribution from any reflection to the least squares sums was included only if $||F_o| - |F_c| \; / \; |F_o| < 0.50$. The accepted contributions to the least squares sums were weighted by the function

$$w = |F_o|^2/625 \text{ if } |F_o|^2 < 625$$

and

$$w = 625/|F_o|^2 \text{ if } |F_o|^2 > 625.$$

The $|F_o|$ used was on the observed rather than the absolute scale. This weighting scheme was used throughout the refinement.

This refinement was continued until the shifts in the parameters were less than one-half of the estimated standard deviation. At this point all atoms were given anisotropic temperature factors and the refinement continued. The weighting scheme and rejection criteria employed for the least squares sums were the same as were used previously. This refinement was continued until the shifts for all positional and temperature parameters were less than one-third

of the estimated standard deviation.

When the refinement of the structure had reached this level it was appropriate to make corrections for anomalous dispersions. The corrections were made on the calculated structure factors, F_c, rather than on the observed structure factors F_o. This method was employed because it was included in the structure factor-least square program which was being used.[14] This method is equivalent to the method of Patterson[15] applied to either $|F_o-|$ or $|F_o+|$ rather than the mean value, $(|F_o+| + |F_o-|)/2$. While making the corrections in this manner is perhaps esthetically less pleasing than the method of Patterson because of this method's noncorrespondence to the physical model, it has the advantage of allowing the correction to be made during each cycle of refinement. Patterson indicated the desirability of making anomalous dispersion corrections during successive cycles of refinement. However the labor involved in making the correction had in practice prevented the acceptance of this procedure. The refinement was continued, making anomalous dispersion corrections on each cycle, until the shifts were once again less than one-third of the estimated standard deviations.

When the shifts had diminished to this level a difference Fourier was calculated. This Fourier, with co-efficients F_o-F_c, was calculated using only those data with a sine squared of theta of less than 0.45. Using this Fourier difference map 14 of 18 hydrogen atoms were found. These hydrogen atoms were then included in the parameters list for the structure factor calculations but were not refined by least squares. They were assigned isotropic temperature factors of 0.5

plus the last isotropic temperature factor value calculated for the atom to which they were bonded.

At this point in the refinement the rejection criteria for the least squares sums was changed. Let F_m be the minimum value of an observed F_o which is considered to be distinguishable from the background intensity. Then reflections were included in the least squares sums only if the following pairs of conditions were obeyed.[54]

For $F_o < F_m$

include if $||F_o| - |F_c|| / |F_o| > 2$;

for $F_m < F_o < 2F_m$

include if $2.1 \times |F_c| > |F_o|$;

for $2\ F_m < F_o < 3F_m$

include if $1.8 \times |F_c| > |F_o|$;

for $3\ F_m < F_o < 4F_m$

include if $2.0 \times |F_c| > |F_o|$;

for $4F_m < F_o$

include if $2.5 \times |F_c| > |F_o|$.

F_m was assigned a value of 5.0 on the absolute scale. After two cycles of least square the shifts were once again less than one-third of the estimated standard deviations. The hydrogen atoms were removed from the parameter list and a structure factor was calculated. A difference Fourier was calculated as before. In this difference map all hydrogen atoms were located.

These 18 atoms were assigned temperature factors in the same manner as was previously used. They were included in the structure factor calculations but not the least squares scheme as before. The

least squares was continued as before using the new rejection scheme.

At this stage it was decided to make the anomalous dispersion corrections using Patterson's[5] method rather than the method used previously. The corrections were applied to the $|F_{o+}|^2$ rather than the mean value of $(|F_+|^2 + |F_-|^2)/2$. The corrections were applied on each successive cycle of structure factor-least squares rather than only once as is generally done.

In order to apply the corrections in this manner it was necessary to modify the structure factor program used.[14] To make this change Patterson's equation, which was written in terms of the ratios of $\Delta f'$ and $\Delta f''$ to f, the scattering factors. Patterson's equation is:

$$|F_\pm|^2 = A^2 + B^2 + 2 \sum_r \delta_{1r}(AA_r + BB_r) - 2\sigma \sum_r \delta_{2r}(AB_r - BA_r)$$

$$+ \sum_r \sum_s (\delta_{1r}\delta_{1s} + \delta_{2r}\delta_{2s})(A_rA_s + B_rB_s)$$

$$- \sigma \sum_r \sum_s (\delta_{1r}\delta_{2s} - \delta_{1s}\delta_{2r})(A_rB_s - A_sB_r).$$

A and B are the real and imaginary parts of the structure factor for the non-dispersive parts of the scattering factors for all atoms. A_r and B_r are the contributions from the dispersive parts of the scattering factor for the r^{th} atom. $\delta_{1r} = \Delta f'_r/f_r$ and $\delta_{2r} = \Delta f''_r/f_r$. σ is +1 for $|F_+|^2$ and -1 for $|F_-|^2$. $H_r = A_r/f_r$ and $K_r = B_r/f_r$, that is H_r and K_r are the geometrical and thermal contributions of the r^{th} atom, summed over the symmetry related positions. The equation may be rewritten in terms of $\Delta f'$ and $\Delta f''$ rather than δ_1 and δ_2. That equation is:

$$|F_{\pm}|^2 = A^2 + B^2 + 2A\Sigma_r(\Delta f'_r H_r - \sigma\Delta f''_r K_r) + 2B\Sigma_r(\Delta f' K_r + \sigma\Delta f'' H_r)$$

$$+ \Sigma\Sigma_{rs}(\Delta f'_r \Delta f'_s H_r H_s + \Delta f''_r \Delta f''_s H_r H_s + \Delta f'_r \Delta f'_s K_r K_s$$

$$+ \Delta f''_r \Delta f''_s K_r K_s) - \sigma\Sigma\Sigma_{rs}(\Delta f'_r \Delta f''_s H_r K_s - \Delta f'_s \Delta f''_r H_r K_s$$

$$- \Delta f'_r \Delta f''_s H_s K_r + \Delta f'_s \Delta f''_r H_s K_r).$$

It was this equation which was used to calculate the anomalous dispersion corrections.

The refinement was continued, making the anomalous dispersion corrections on each cycle, until the shifts were once again less than one-third of the estimated standard deviations. When the shifts had diminished to this level, Fourier and difference Fourier maps were calculated to check the correctness of the structure. These maps revealed no significant peaks except those predicted by the structure.

At this point the structure analysis was complete. The observed and calculated structure factors are related by the equation $F_c = 1.166 \times F_o$. The final residual, or R value was 0.065 over all reflections. The final electron densities for each of the non-hydrogen atoms is listed in table 2. The electron and difference densities for the hydrogen atoms are also listed in table 2.

TABLE 2a

Electron Density of the Atoms

Atom	Height	Atom	Height
Zn	60.2	$H_1(N_1)$	1.3
N_1	8.1	$H_2(N_1)$	0.5
$C\alpha_1$	6.7	$H_1(C\alpha_1)$	0.9
C'_1	7.7	$H_2(C\alpha_1)$	1.2
O_1	11.5	H (N_2)	0.9
N_2	9.3	$H_1(C\alpha_2)$	0.9
$C\alpha_2$	7.6	$H_2(C\alpha_2)$	0.7
C'_2	7.8	H (N_3)	1.1
O_2	9.7	$H_1(C\alpha_3)$	0.6
N_3	8.4	$H_2(C\alpha_3)$	0.7
$C\alpha_3$	7.1	$H_1(O^W_1)$	0.7
C'_3	7.2	$H_2(O^W_1)$	0.7
O^1_3	8.1	$H_1(O^W_2)$	1.3
O^2_3	10.1	$H_2(O^W_2)$	1.4
O^W_1	9.0	$H_1(O^W_3)$	0.7
O^W_2	9.9	$H_2(O^W_3)$	0.8
O^W_3	8.0	$H_1(O^W_4)$	1.0
O^W_4	7.4	$H_2(O^W_4)$	0.7
S	23.0		
O^S_1	8.1		
O^S_2	8.6		

TABLE 2b

Electron Density of the Atoms from the Difference Fourier

Atom	Height	Atom	Height
Zn	-0.80	$H_1(N_1)$	0.48
N_1	0.06	$H_2(N_1)$	0.48
$C\alpha_1$	-0.02	$H_1(C\alpha_1)$	0.34
C'_1	-0.14	$H_2(C\alpha_1)$	0.54
O_1	0.09	$H(N_2)$	0.32
N_2	-0.08	$H_1(C\alpha_2)$	0.41
$C\alpha_2$	0.19	$H_2(C\alpha_2)$	0.51
C'_2	-0.04	$H(N_3)$	0.42
O_2	-0.06	$H_1(C\alpha_3)$	0.32
N_3	0.04	$H_2(C\alpha_3)$	0.43
$C\alpha_3$	-0.06	$H_1(O^W_1)$	0.62
C'_3	-0.05	$H_2(O^W_1)$	0.39
O^1_3	0.06	$H_1(O^W_2)$	0.39
O^2_3	-0.13	$H_2(O^W_2)$	0.49
O^W_1	0.31	$H_1(O^W_3)$	0.27
O^W_2	0.31	$H_2(O^W_3)$	0.37
O^W_3	-0.09	$H_1(O^W_4)$	0.47
O^W_4	-0.12	$H_2(O^W_4)$	0.24
S	-0.80		
O^S_1	-0.35		
O^S_2	-0.21		

TABLE 3a

Atomic Parameters of $Zn(ggg) \cdot 1/2SO_4 \cdot 4H_2O$

Atom	X	Y	Z
Zn	.37186(2)	.49433(9)	.06174(4)
N_1	.42408(16)	.3213(5)	.0154(3)
$C\alpha_1$	.4009(2)	.1533(7)	.0028(4)
C'_1	.35280(17)	.1378(6)	.0649(3)
O_1	.33002(12)	.2638(4)	.0949(2)
N_2	.33577(15)	-.0157(5)	.0774(3)
$C\alpha_2$	.2862(2)	-.0512(6)	.1228(4)
C'_2	.28811(17)	-.0766(6)	.2328(3)
O_2	.32249(14)	-.0229(5)	.2851(3)
N_3	.24725(17)	-.1588(5)	.2684(3)
$C\alpha_3$	.23857(18)	-.1875(7)	.3727(4)
C'_3	.19473(18)	-.0910(6)	.4184(5)
O^1_3	.16721(16)	.0014(6)	.3710(3)
O^2_3	.18912(12)	-.1116(4)	.5102(2)
O^W_1	.37096(15)	.5741(7)	.1995(3)
O^W_2	.41891(13)	.7054(5)	.0302(3)
O^W_3	.47572(13)	.3182(6)	.3672(3)
O^W_4	.42189(18)	.1153(5)	.2425(4)
S	.50000	.7411(3)	.2500
O^S_1	.46026(17)	.6379(6)	.2953(3)
O^S_2	.47674(17)	.8443(6)	.1737(3)

TABLE 3a - continued

Atom	X	Y	Z	Isotropic Temperature Value
$H_1(N_1)$	.433	.343	-.044	3.85
$H_2(N_1)$	.444	.310	.059	3.85
$H_1(C\alpha_1)$	.429	.065	.039	4.37
$H_2(C\alpha_1)$	.387	.160	-.067	4.37
$H\ (N_2)$	.355	-.106	.052	3.37
$H_1(C\alpha_2)$	.264	.040	.113	3.47
$H_2(C\alpha_2)$	.274	-.140	.092	3.47
$H\ (N_3)$	.225	-.109	.225	3.64
$H_1(C\alpha_3)$	.267	-.160	.408	3.70
$H_2(C\alpha_3)$	.226	-.300	.385	3.70
$H_1(O^W_1$	.395	.620	.220	4.51
$H_2(O^W_1)$	.353	.550	.255	4.51
$H_1(O^W_2)$	.435	.690	-.030	4.07
$H_2(O^W_2)$	.443	.740	.063	4.07
$H_1(O^W_3)$	.444	.275	.334	5.04
$H_2(O^W_3)$	.485	.485	.337	5.04
$H_1(O^W_4)$	.416	.097	.319	7.10
$H_2(O^W_4)$	.445	.005	.225	7.10

TABLE 3b

Anisotropic Temperature Values

$\exp[-(b_{11}h^2 + b_{22}k^2 + b_{33}l^2 + b_{23}kl + b_{13}lh + b_{12}hk)]$

Atom	b_{11}	b_{22}	b_{33}	b_{23}	b_{13}	b_{12}
Zn	.00120(1)	.01206(9)	.00351(3)	.00070(8)	-.00015(3)	.00021(6)
N_1	.00129(7)	.0155(8)	.0040(2)	.0017(7)	.0003(2)	.0001(4)
C^{α}_1	.00176(10)	.0133(9)	.0055(3)	.0001(9)	.0012(3)	-.0008(5)
C'_1	.00106(7)	.0178(8)	.0040(3)	.0012(8)	.0000(2)	.0001(4)
O_1	.00115(5)	.0101(5)	.0051(2)	-.0002(6)	.0008(2)	.0010(3)
N_2	.00128(7)	.0100(7)	.0043(2)	-.0013(6)	.0013(2)	.0008(4)
C^{α}_2	.00132(9)	.0109(9)	.0044(3)	-.0021(9)	.0001(3)	-.0013(4)
C'_2	.00114(8)	.0106(8)	.0038(3)	-.0015(8)	-.0002(2)	.0002(4)
O_2	.00149(6)	.0191(7)	.0048(2)	-.0016(7)	-.0009(2)	-.0021(4)
N_3	.00133(6)	.0170(8)	.0034(2)	-.0013(7)	.0004(2)	-.0010(4)
C^{α}_3	.00137(9)	.0153(9)	.0036(3)	.0003(8)	.0009(2)	.0004(4)
C'_3	.00118(8)	.0114(8)	.0035(3)	.0006(8)	-.0004(2)	-.0009(4)
$O1_3$	.00222(8)	.0385(10)	.0041(2)	.0032(8)	.0000(2)	.0085(5)
$O2_3$	.00145(6)	.0138(6)	.0035(2)	-.0004(5)	.0003(2)	.0005(3)
O^{w}_1	.00212(8)	.0300(9)	.0032(2)	-.0027(7)	.0004(2)	-.0078(5)
O^{w}_2	.00163(6)	.0168(7)	.0047(2)	-.0014(6)	.0007(2)	-.0028(4)
O^{w}_3	.00189(7)	.0263(9)	.0045(2)	.0000(8)	.0003(2)	-.0006(4)
O^{w}_4	.00235(9)	.0227(10)	.0105(4)	-.0058(10)	-.0013(3)	.0006(5)
S	.00131(3)	.0180(3)	.0047(1)	.0*	.00041(9)	.0*
O^{s}_1	.00211(8)	.0271(10)	.0066(3)	.0026(9)	.0004(2)	-.0029(5)
O^{s}_2	.00266(9)	.0190(8)	.0060(3)	.0014(8)	-.0023(3)	.0010(9)

* The b_{23} and b_{12} elements of sulfur are identically zero, because the sulfur occupies a special position.

TABLE 3c

Principal Axes and Direction Cosines of Anisotropic Ellipsoids

Atom	B	l_1	l_2	l_3
Zn	3.25	.894	.444	-.040
	3.11	-.407	.850	.332
	2.52	.181	-.281	.942
N_1	4.14	.163	.926	.337
	3.48	.943	-.246	.221
	2.82	-.288	-.282	.914
C^{α}_1	5.40	.808	-.148	.569
	3.60	-.372	.620	.690
	3.25	.455	.770	-.446
C'_1	3.29	.123	.722	.679
	2.85	.861	.261	-.435
	2.73	.492	-.639	.590
O_1	4.14	.500	.099	.860
	3.07	.641	.624	-.445
	2.27	.379	-.331	.863
N_2	4.26	.746	.029	.664
	2.94	.397	.781	-.481
	2.03	-.533	.623	.571
C^{α}_2	3.88	.790	-.544	.280
	3.47	-.453	-.212	.865
	2.28	.411	.811	.415
C'_2	3.28	-.643	-.499	.580
	2.88	.765	-.441	.468
	2.44	.022	.745	.665
O_2	5.47	-.532	.845	.035
	4.29	-.547	-.376	.747
	2.75	.645	.378	.663
N_3	4.63	-.404	.896	-.178
	3.40	.888	.432	.156
	2.42	-.217	.095	.971

TABLE 3c- continued

Atom	B	l_1	l_2	l_3
C^{α}_3	4.17	.692	.634	.343
	3.76	-.557	.772	-.302
	2.34	-.457	.018	.889
C'_3	3.59	.764	-.550	-.335
	2.68	.396	.811	-.429
	2.45	.508	.195	.838
O^1_3	11.07	.563	.823	.068
	3.75	.733	-.460	-.499
	2.74	.379	-.331	.863
O^2_3	4.01	.899	.419	.120
	3.46	-.390	.896	-.207
	2.59	-.195	.139	.970
O^w_1	10.16	-.588	.804	-.083
	3.29	.806	.577	-.124
	2.34	.051	.140	.988
O^w_2	5.68	.698	-.667	.257
	3.40	.177	.510	.841
	3.07	.693	.542	-.475
O^w_3	6.80	-.141	.989	-.006
	5.05	.978	.141	.149
	3.30	-.148	-.014	.988
O^w_4	8.77	-.361	-.396	.843
	5.95	.907	-.355	.222
	5.17	.212	.846	.488
S	4.64	.000	1.000	.000
	3.79	-.702	.000	.711
	3.21	.711	.000	.702
O^s_1	7.74	-.487	.865	.117
	5.44	.673	.286	.681
	4.31	-.555	-.411	.722
O^s_2	7.94	.900	.083	-.425
	5.04	.066	.942	.326
	3.48	.428	-.322	.843

The B values correspond to isotropic temperature factors along the three principal axes of the vibration ellipsoid. l_1, l_2, and l_3 are the direction cosines of that axis with respect to the a, b, and c axes respectively.

TABLE 3d

List of Observed and Calculated Structure Factors
Using Final Parameters of All Atoms

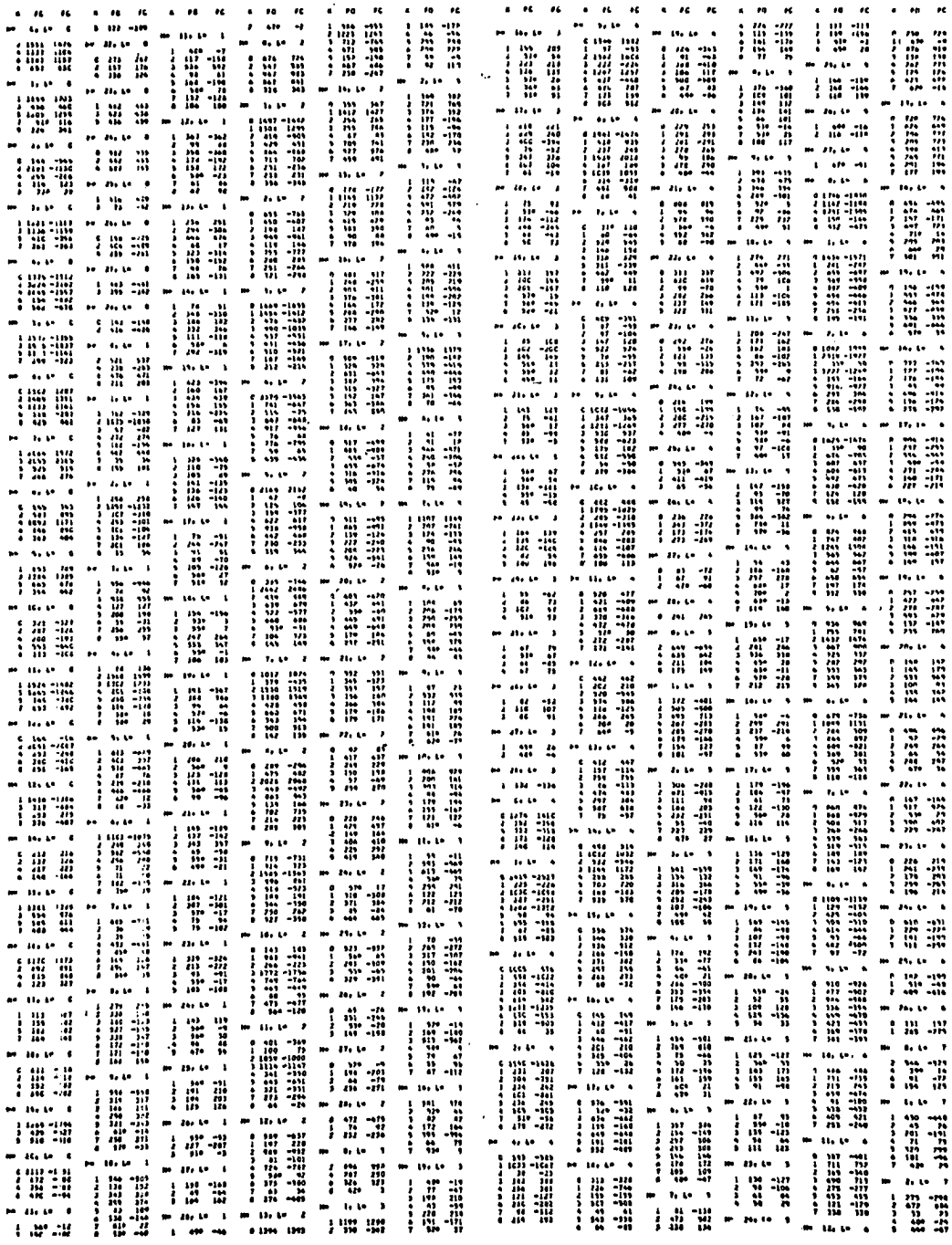
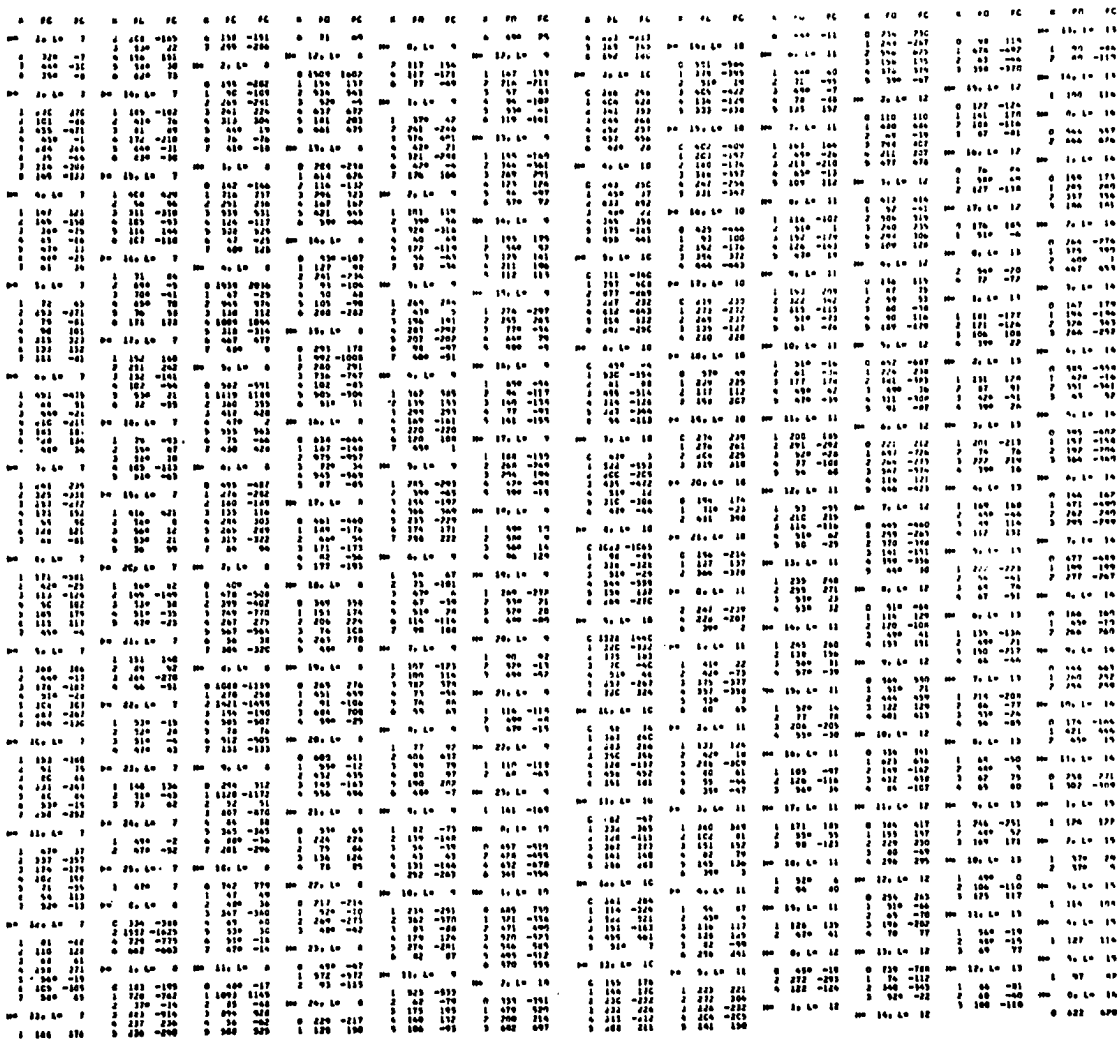


TABLE 3d (Continued)



CHAPTER VI

DISCUSSION OF THE STRUCTURE

The most immediately notable feature of the crystal structure of zinc glycyl glycyl glycine is its segregation into an aqueous region and a protein-like region. With the zinc ion as the link between them, the peptide chains form infinite chains, connected through the bridging action of the zinc ion. The chains form an extended, two stranded, coil around a two-fold screw axis parallel to the c axis. These coils then stack along the glide plane parallel to the bc plane to form an infinite, protein-like sheet of material. This is clearly shown in figures 2 and 5.

The tripeptide molecules are linked through the zinc ions. The zinc ion is chelated by the amine terminal nitrogen and the carbonyl oxygen of the first glycyl residue. The second peptide molecule is bonded to the zinc through the negatively charged oxygen of the carboxylate group of the carboxyl terminal or third glycyl residue. Thus the zinc ion serves in place of a peptide bond in the formation of the infinite protein-like peptide chains. The sulfate ion is located on the two fold rotation axis parallel to the b axis. This sulfate ion and the two water molecules not bonded to the zinc ion form infinite columns of water and sulfate with the two fold rotation axis as the center of the column. These columns do not penetrate the

protein-like sheet. This aqueous column is connected to the peptide sheet through hydrogen bonds from the zinc ligands.

Discussion of Zinc Surroundings

The nature of the bonds between a zinc ion and its ligands is a much avoided subject. Not, however without reason, for the possibilities are many and complex and the means for distinguishing between them are few and inconclusive. This paucity of information is both experimental and theoretical and is a consequence of the fact that zinc is not a transition metal ion--its 3rd subshell is complete.

Thus when an irregular and unexpected zinc coordination was found its proper description was not obvious. There are six potential ligands. The zinc to ligand distances fall into three groups. There are three distances clustered about 2.0 Angstroms, two more near 2.16 Angstroms and the sixth distance is 2.78 Angstroms. The atoms and distances are, N_1 at 2.034 A^o, O_1^w at 1.979 A^o, $^1_{-c}O_3^2$ at 1.964 A^o, O_1 at 2.186 A^o, O_2^w at 2.126 A^o and $^1_{-c}O_3^1$ at 2.782 A^o. The spatial relationships of these atoms may be seen in figure 1 and tables 4a and 4b.

The first question which needs to be answered is: Is the sixth, most distant atom, in fact a ligand? There are several circumstances which make it possible that this is a ligand. There are several structures reported,[5] which exhibit zinc-oxygen distances of this magnitude, and which have been proclaimed bonded interactions. Also the presence of the sixth atom has caused distortion of the zinc surrounding by significant amounts from any of the geometries predicted by bonding theory for atoms with five ligands. Further zinc bonding has, for many years, been considered primarily electrostatic in nature,

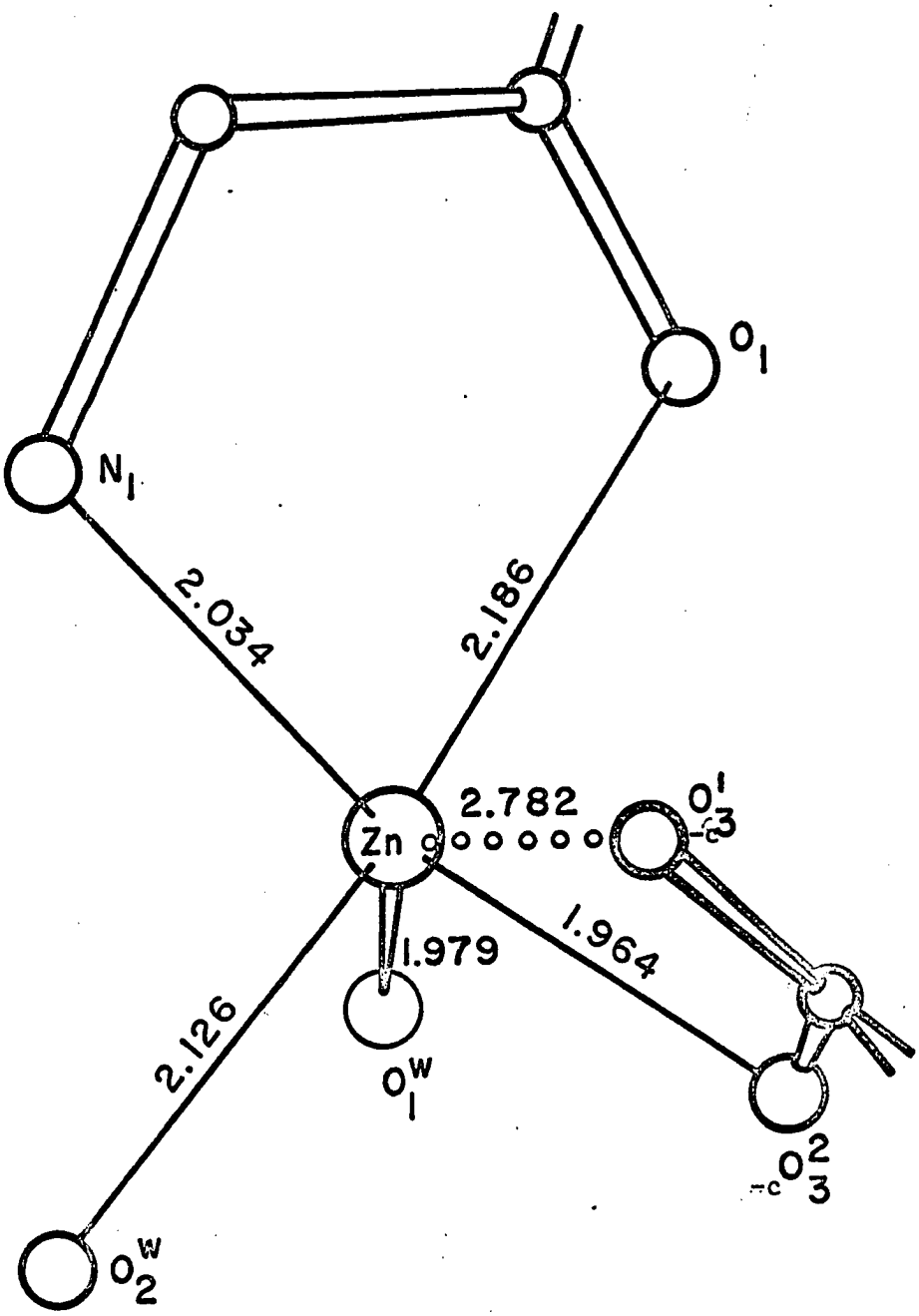


Figure 1. The Zinc Ion and its Ligands.

TABLE 4a

Comparison of Zinc Surrounding with Values for Ideal Trigonal Bipyramid and Tetragonal Pyramid

Angle	Ideal Trigonal Bipyramid	Observed	Ideal Tetragonal Pyramid[55]
O^{W}_{1}-Zn-N_1	120°	121.3(2)°	100°
O^{W}_{1}-Zn-O_1	90	94.1(2)	100
O^{W}_{1}-Zn-O^{W}_{2}	90	86.6(2)	100
O^{W}_{1}-Zn-$^{1}_{-c}O^{2}_{3}$	120	100.0(2)	100
N_1-Zn-O_1	90	79.5(1)	88.4
N_1-Zn-O^{W}_{2}	90	95.7(2)	88.4
O^{W}_{2}-Zn-$^{1}_{-c}O^{2}_{3}$	90	90.4(1)	88.4
O_1-Zn-$^{1}_{-c}O^{2}_{3}$	90	94.6(1)	88.4
N_1-Zn-$^{1}_{-c}O^{2}_{3}$	120	138.4(2)	160
O_1-Zn-O^{W}_{2}	180	174.7(1)	160

Average deviation from trigonal bipyramid is 7.3°.

Average deviation from tetragonal pyramid is 10.1°.

TABLE 4b

Distances of the Ligands from the Zinc

Bond	Length
$Zn-N_1$	2.034(4) Å
$Zn-{}^{1}_{-c}O^{2}_{3}$	1.964(3)
$Zn-O^{w}_{1}$	1.979(4)
$Zn-O_1$	2.186(3)
$Zn-O^{w}_{2}$	2.126(4)
$Zn-{}^{1}_{-c}O^{1}_{3}$	2.782(4)

Standard deviations are in parentheses

TABLE 4c

Known Zinc Complexes with Penta-coordination

Compound	Reference	Configuration
Zinc N-methylsalicyaldimine	56	trigonal bipyramid
bis-acetylacetonate Zinc	57	trigonal bipyramid
monoaquo bis(acetylacetonate) Zinc	58,64	intermediate
$Zn(Zn(S_2CN(CH_3)_2)_2.C_5H_5N$	59	trigonal bipyramid
NN' disalicylidene EDTA Zn	60	tetragonal pyramid
Zn tetraphenyl porphine H_2O	61	tetragonal pyramid
terpyridyldichlorozinc	62,65	tetragonal pyramid[1]
bis l-serinato zinc	63	tetragonal pyramid
Zinc glutamate dihydrate	21	tetragonal pyramid[2]

[1]This structure was initially described as a distorted trigonal bipyramid.

[2]The original author considers this structure to be 5 surrounded, however, others have disagreed.[5]

TABLE 4d

Least Squares Planes of the Zinc Surrounding

Plane	Atoms to which plane was fitted	Equation of Plane*
I	O^{w}_{1} $^{1}_{-c}O^{2}_{3}$ N_1	-15.565x - 6.002y + 3.751z = -8.471
II	O^{w}_{1} $^{1}_{-c}O^{2}_{3}$ N_1 Zn	15.562x + 6.007y - 3.733z = 8.487
III	N_1 O_1 $^{1}_{-c}O^{2}_{3}$ O^{w}_{2}	2.619x + 1.360y + 13.321z = 2.121
IV	Zn N_1 O_1 $^{1}_{-c}O^{2}_{3}$ O^{w}_{2}	2.642x + 1.336y + 13.331z = 2.188

*x, y, and z are expressed in fractional coordinates.

Distances from the Planes to Certain Atoms

Atom	Plane I	Plane II	Plane III	Plane IV
Zn	-0.052 Å	0.039 Å	0.348 Å	0.278 Å
N_1		-0.015	-0.367	-0.432
O_1	2.107	-2.120	0.367	0.302
$^{1}_{-c}O^{1}_{3}$	-0.184	0.167		
$^{1}_{-c}O^{2}_{3}$		-0.013	-0.338	-0.412
O^{w}_{1}		-0.010	2.290	2.219
O^{w}_{2}	-2.168	2.156	0.337	0.264

and one therefore expects irregular co-ordinations. Finally this oxygen lies only 0.167 A^{o} from the least squares plane of the zinc ion and its three nearest ligands. This distance is not an unusually large departure from planarity for the fourth ligand of the basal plane of an octahedral zinc surrounding.

However it does not appear that a zinc oxygen distance of 2.78 A^{o} should be considered bonded, at least for this structure. Pauling[16] has developed an empirical formula for fractional bond lengths. Its results are admittedly only rough approximations and depend on the bond length which is assigned as a single bond. The equation is

$$D(n) = D(1) - 0.60 \log n$$

where D(n) is the length of a bond of bond number n, n less than one and D(1) is the assumed length of a single bond of the same type. If the other zinc oxygen bonds in the possible basal plane are assumed to be single bonds the sixth bond calculates to have a bond order of less than 0.05 which is certainly negligible in view of the accuracy of the equation. Even if the second longest zinc oxygen distance is used the bond order is only 0.10. Further it is probable that this longer assumed distance is not applicable since it is perpendicular to the proposed basal plane, an orientation which is predicted to be of a different bond type by many molecular orbital calculations.[17,22]

Also the position of this atom may be explained more readily in terms of factors other than a bond to the zinc ion. It is in the same carboxylate group as the O_3^2 which is bonded to the zinc. This of course severely limits its possible positions since the peptide chain is fairly extended and therefore has little freedom of movement

remaining. Also this distant oxygen forms a rather strong hydrogen bond with the water molecule in the proposed basal plane of a second zinc ion. These two facts would seem to adequately explain the position of this oxygen. This explanation coupled with the negligible bond order would seem to rule out the possibility of any significant bonding interaction with the zinc ion. Therefore it is concluded that the zinc ion has five ligands rather than six.

Three other crystal structures in which there is a similar approach to the metal ion by the second oxygen have been published. They are monoaquo copper(II) glycylglycylglycinato chloride hemihydrate by Freemen, Robinson, and Schoone;[46,5] copper(II) glutamate dihydrate by Marsh and Gramaccioli;[75] and zinc(II) glutamate dihydrate by Gramaccioli.[21] In all three instances the sixth atom would complete a badly distorted octahedral surrounding by filling the second axial position. The two axial bond lengths and the angle these bonds make with each other are: 2.30 Å, 2.82 Å, and 148° for the copper(II) glycylglycylglycinate chloride; 2.30 Å, 2.58 Å, and 149° for the zinc glutamate.

In the original publications none of these three metal ions were thought to be octahedrally bonded. According to Gramaccioli and Marsh, "The coordination about the copper atom is approximately square planar, the square comprising the two oxygens and a nitrogen atom of glutamate groups and a water molecule. . . . a fifth atom, O(2) occupies an axial position at 2.30 Å from the copper atom; a sixth, O(4), at 2.59 Å and considerably displaced from the other axial position completes a severely distorted octahedron."[75] Whether or

not these axial atoms are bonded to the copper atom is not discussed. In the article on zinc glutamate Gramaccioli refers to this copper surrounding as "(4+1) coordination."[21] In the discussion of zinc glutamate, Gramaccioli,[21] calls the zinc surrounding a tetragonal pyramid and notes that its average deviation from a regular tetragonal pyramid is only 4.0^{o}. In the initial article on copper(II) glycyl glycylglycinato chloride, Freeman describes the bonding of the copper atom as "fivefold coordination."[46] He however notes that this may be the result of access to the sixth ligand position being blocked by the second oxygen of the carboxylate group. The first oxygen of this group is bonded to the copper atom.

However Freeman has changed his point of view. In a review article on metal chelates of peptides and amino acids[5] he says that all three are distorted octrahedra. He views the copper octahedra as a square planar structure with four strong planar bonds and two weaker bonds perpendicular to this plane. The two weaker bonds are allowed to assume fractional bond orders so that the coordination number of the copper ion varies by increments of ¼ from four to six. He then attempts to establish a correlation between the coordination number, the absorption maxima of the compound in solution, and the color of the crystals. The results are reproduced in Table 5, which was taken from "Advances in Protein Chemistry," volume 22.[5]

In doing this he neglects the fact that nearly all cupric ions are thought to be square planar with two more distant ligands when in aqueous solution. Specifically the complex $CuCl_2(OH_2)_2$, and the complex ions $Cu(OH_2)_4^{++}$ and $Cu(NH_3)_4^{++}$ are of this type. These

TABLE 5

Ligand Fields and Coordination Numbers in Cu(II) Complexes

Complex	Four closest ligand atoms				Next-nearest ligand atoms	Coord. no. of Cu	Color	λ max.
$Cu(biu)_2Cl_2$	O=	O=	O=	O=	2 Cl^-(2.96)	6	Blue-green	-
Cu(Gly-Gly-Gly)Cl·1/2H_2O	NH_2	O=	O^-	Cl	OH_2(2.3),O=(2.8)	5 3/4	Blue-green	730
Cu(Glu)·2H_2O	NH_2	O^-	O^-	OH_2	O=(2.3), O=(2.6)	5 3/4	Blue	620
$Cu(Gly)_2$·H_2O	NH_2	O^-	O^-	O^-	OH_2(2.4), O=(2.7)	5 1/2	Blue	630
$Cu(pro)_2$·2H_2O	NHR	NHR	O^-	O^-	2 OH_2(2.5)	5 1/2	Blue	610
$Cu(\beta\text{-Ala})_2$·6H_2O	NH_2	NH_2	O^-	O^-	2 OH_2(2.5)	5 1/2	Blue	-
$Cu(\beta\text{-}NH_2But)_2$·2$H_2O$	NH_2	NH_2	O^-	O^-	2 OH_2(2.5)	5 1/2	Blue	-
Cu(Gly-l-His)·1 1/2 H_2O	NH_2	N	N_{im}	O^-	OH_2(2.5), O=(3.0)	5 1/4	Blue	595
Cu(Gly-Gly)·3 H_2O	NH_2	N	O^-	OH_2	OH_2(2.3)	5	Blue	635
NaCu(Gly-Gly-Gly)·H_2O	NH_2	N	N	O^-	N(2.6)	4 3/4	Violet	555
Cu(β-Ala-l-His).2H_2O	NH_2	N	N_{im}	O^-	OH_2(2.5)	4 3/4	Violet	-
$K_2Cu(Biu)_2$.4H_2O	NH	NH	NH	NH	-	4	Violet-pink	505
Na_2Cu(Gly-Gly-Gly-Gly)·10 H_2O	NH_2	N	N	N	-	4	Violet-pink	520
Na_2Cu(Gly-Gly-Gly-Gly-Gly).4H_2O	NH_2	N	N	N	-	4	Violet-pink	510

Taken from an article by H. C. Freeman in Recent Advances in Protein Chemistry, Vol. 22.[5]

three ions show colors which span most of the color range exhibited by the compounds in Table 5. $CuCl_2(OH_2)_2$ forms green crystals and has, in concentrated solutions, an absorption maxima at 850 millimicrons.[78] In dilute solution the compound hydrolizes to give the $Cu(OH_2)_4^{++}$ ion. The wavelength of the absorption maximum shifts from that of the $Cu(OH_2)_4^{++}$ ion to 850 millimicrons gradually as the concentration of the solution is increased.[78] $Cu(OH_2)_4^{++}$ is found in crystals of $CuSO_4.5H_2O$); the crystals are light blue and the $Cu(OH_2)_4^{++}$ ion has an absorption maximum, in solution, at about 800 millimicrons.[77] The $Cu(NH_3)_4^{++}$ ion is found in crystals of $CuSO_4.4NH_3.H_2O$ which are deep blue; the solution absorption maximum is about 600 millimicrons.[77] Only those compounds in which the ligands are deprotonated amides are outside the range exhibited by the above three complexes. A more complete description of this explanation and more of the experimental evidence which supports it may be found in Advanced Inorganic Chemistry by Cotton and Wilkenson.[77] Examination of Table 5 reveals that the change of color may be accounted for by the change in the types of ligands rather than the change in the number of ligands present. This explanation, in terms of the types of ligands, would seem to be superior to an explanation in terms of the change in the number of ligands because it would change the order if the two compounds Cu(gly.l-His)$1\frac{1}{2}H_2O$ and Cu(Gly.Gly).$3H_2O$ so that they would be in the right order with respect to their absorption maxima. Freeman's explanation, in terms of the total number of ligands, has them out of order with respect to their absorption maxima.

It should be further noted that Freeman's primary point is that those complexes to which he has assigned a coordination number of four in the solid state will also exhibit this low coordination

number in solution. That is these complexes will not bond to two water molecules to form an octahedral complex upon being dissolved in aqueous solution. The arguments in favor of this point are quite reasonable. However it should be pointed out that it is not necessary to accept the formation, in the solid state, of the highly strained and elongated axial bonds which he postulates, in order to accept his primary thesis. In fact Freeman points out that "it is unlikely that these strained Cu-O bonds survive dissolution of the crystal." They are probably replaced with water molecules when in solution.

While this discussion does not prove the non-existence of the postulated bonds it does show that one is not forced to postulate these bonds in order to explain the observed properties of the compounds. If the postulated bonds are indeed present, then the solid state spectra of the copper compounds should show abnormalities caused by the extreme assymetry of the bonding. Thus it should be possible to investigate the existence of these bonds and resolve the question.

The conclusion that the zinc ion has only five ligands is justified only if the bonding is largely covalent and not merely electrostatic. This is because the geometry of electrostatic bonding is determined mainly by space filling requirements. If the bonding is covalent then the zinc ion must use one of the sp^3d hybridization schemes, either a trigonal bipyramid or a square pyramid.[74] Also all of the orbitals must come from the fourth main shell.

There is evidence,[17] both experimental and theoretical, that third series transition and post transition metals use their 4d orbital to form rather strong covalent hybrid bonds in some complexes. The most

relevant experimental evidence is the isolation in solution and in the solid state of the optical isomers of zinc trisethylenediamine chloride and zinc trisethylenediamine sulfate.[18] While these compounds are stable for only a few hours in solution, even this limited stability indicates significant covalent character in the bonds.

Assured, with reasonable certainty that the zinc ion is surrounded by five covalently bound ligands, it is now useful to examine the geometry of the bonding arrangement, (fig. 1). It was noted earlier that three of the ligands were closer to the zinc than the other two. These three are: N_1, O_1^w and $\overset{1}{_{-c}}O_3^2$. These three are essentially co-planar with the zinc ion and the angles between them at the zinc ion are:

angle N_1-Zn^{++} O_1^w is 121 degrees,

angle O_1^w - Zn^{++} $\overset{1}{_{-c}}O_3^2$ is 100 degrees

and angle $\overset{1}{_{-c}}O_3^2$ - Zn^{++} - N_1 is 138 degrees.

The bonds between the zinc ions and the other two ligands, O_2^w and O_1 make angles of 90±5 degrees with the bonds between the zinc and the three close ligands, except for the one involving both ligands of the chelate ring which is 79.5°. This geometry appears to be best described as a moderatley distorted trigonal bipyramid.

A systematic comparison of the angular distortions of the actual bonding from of geometry of either the trigonal bipyramid or the tetragonal pyramid may be found in Table 4a. The angular values used for the tetragonal pyramid are those suggested by Gillespie.[55] This configuration minimizes the electrostatic repulsions among the ligands. Statistical confirmation of the correctness of describing

the geometry as that of a distorted trigonal bipyramid is also found in Table 4a. This arises from the average deviation of the observed angles from those of either idealized geometry. The average deviation from the trigonal bipyramid is 7.3^{o} while the average deviation from the tetragonal pyramid is 10.1^{o}.

Table 4d lists some least squares fitted planes through the zinc surrounding. Planes I and II are planes through the ligands in the basal plane of a trigonal bipyramid surrounding. The relevant ligands and the zinc are very closely co-planar for this geometry. Planes III and IV are the basal planes of a tetragonal pyramid and a square planar, (4+1) pyramid respectively. The atoms used to calculate these planes are obviously not co-planar. In fact the deviations approach the tetrahedral distortion found in bis(l-serinato) zinc.

The major distortion, that of the angles in the base plane of the trigonal bipyramid, can be ascribed to the dislocation of $^{1}_{-c}O^{2}_{3}$. This atom is part of the carboxylate group on the peptide. Therefore its freedom of movement is severely limited by conformational limits of the peptide chain. Further this oxygen cannot move without displacing the other oxygen on the carboxylate group and disturbing its hydrogen bonding and conformational and packing contacts. Thus it is likely that this distortion is the result of the constraints imposed on the structure by packing considerations.

The only other really significant distortion is the 79.5^{o} N_1-Zn-O_1 angle. This is undoubtedly caused by the fact that both atoms are in the first glycyl residue and are the atoms through which the peptide chelates the zinc ion. Freeman[5] points out that chelation

rarely results in a significant distortion of the peptide residue involved. This means that this angle is a function of the bond lengths between the metal and the chelating ligands. This hypothesis is supported by the N-Zn-O angles in the chelate rings of zinc di-l-serine and zinc-l-glutamate. The angle in zinc glutamate is 79.4°.[21] The two angles in zinc di-l-serine are 79.6° and 80.0°.[63]

There have been very few zinc structures determined by diffraction techniques in which the zinc is unequivocably five surrounded.[19,20] A fairly complete listing of these structures will be found in Table 4c. The zinc glutamate dihydrate structure has been described as octahedral.[5] However, as discussed previously it does not seem likely that the sixth ligand is actually bonded to the zinc ion.

There has been little theoretical work done on five surroundings. Further what has been done is often contradictory. Valence bond, molecular orbital, and electrostatic repulsion approaches to pentacoordination all yield the result that the trigonal bipyramid is the most stable configuration for d^0 and d^{10} electronic states. The most successful approach thus far has been through the use of molecular orbital calculations using modified Slater orbitals to calculate overlap integrals.[17,22] Craig et al.[17] have shown that, for a maximized overlap integral, the apex bond are predicted to be about ten per-cent longer than the equatorial bonds. Further the bonds are of approximately the same strength. Cotton[22] assumes all five bonds to be of the same length. He then shows that the overlap is greater for the equatorial bonds than for the axial bonds under this condition.

The three zinc ligand bonds in the equatorial plane are of normal

length. In fact they are the length predicted by the Pauling tetrahedral co-valent radii. This result is in agreement with the calculations of Craig et al..[17] Their results indicate that the equatorial hybrid orbitals each contain twenty-five per cent 4s character, sixty-seven per cent 4p character and only eight per cent 4d character. Thus they have essentially the same composition as tetrahedral orbitals.

The axial, or apex, zinc oxygen bonds are significantly longer than is predicted by the tetrahedral radii, about ten per cent longer. This is the direction and magnitude of the elongation predicted by Craig et al..[17] It is tempting to argue for the correctness of those calculations in view of this experimental result. However, the elongation is not found in all trigonal bipyramidal structures. A suitable explanation of this elongation is not obvious.

It should be pointed out that this is the first trigonal bipyramid surrounding found for zinc with ligands which might be considered biological in nature. Also this is the largest of the biological type ligands thus far used with zinc. The structures previously determined were with ligands which had less freedom of movement than glycyl glycyl glycine. Thus it is possible that packing considerations influenced the zinc surrounding to a greater extent than in this structure. Therefore it is possible that the trigonal bipyramid could have an important role in the biochemistry of zinc.

Since one of the principal reasons for determining the molecular structure of zinc glycylglycylglycine was to investigate the differences between the bonding geometries of zinc and copper ions in biological environments, these differences should be examined at this time. There are

three pairs of well determined copper and zinc chelates; the ligands are, in order of increasing complexity, l-serine, l-glutamic acid, and glycylglycylglycine.

The copper(II) and zinc ions in the serine complexes are both five coordinated.[63,76] The geometries are best described as distorted square planar with a fifth axial ligand. The four ligands in the base plane are all within 0.12 Å of the least square plane fitted to these four atoms in the copper chelate.[76] The corresponding deviations in the zinc chelate range from 0.31 Å to 0.38 Å.[63] Two of the ligands in the zinc chelate are approximately 0.1 Å above the zinc, along the direction perpendicular to the least squares plane through the four ligands. The other two are approximately 0.5 angstroms below the zinc along the same perpendicular axis. Thus it could be said that the addition of the fifth ligand has distorted this surrounding from tetrahedral toward the (4+1) pyramid. Therefore the copper chelate approximates the ideal (4+1) coordination much more closely than does the zinc chelate.

The differences between the bonding geometries of the copper and zinc ions in the glutamate structures are similar but more pronounced than those in the serine chelates. As in the serine chelate the copper bonding is best described as a square planar plus one, (4+1), arrangement.[21,75] The zinc structure, however, is best described as a tetragonal pyramid.[21] All five zinc-ligand bonds are approximately of the same length, while the fifth bond in the copper surrounding is appreciably longer. In both of these compounds the four ligands of the base plane are planar within 0.03 Å. The copper is displaced from this plane by 0.15 Å, the zinc by 0.32 Å.

The copper surrounding in copper(II) chloride glycylglycylglycine more closely resembles that in copper(II) serine than that in copper(II)

glutamate. That is the four base plane ligands deviate appreciably from being coplanar.[46] The bonding geometry however is still best described as a distorted (4+1) coordination. As has been shown the zinc bonding in the glycylglycylglycine chelate most closely approximates a trigonal bipyramid.

Thus in these three chelates the copper(II) ion demonstrates a strong preference for the square planar plus one geometry while the zinc has approximated three different geometries. In all six crystal structures the number of ligands was the same, five. The most important difference between the zinc and copper ion in biological environments seems to be that the zinc bonding is much more labile than is that of the copper ion. This impression is enhanced by the fact that zinc frequently exhibits tetrahedral geometry while copper(II) has been known to assume this geometry only in a few of its complexes such as $CuCl_4^{=}$.

The Hydrogen Bonding Scheme

The reader is referred to tables 6a, 6b, 6c and 6d of a systematic presentation of the distance and angular relationships among the atoms involved in the hydrogen bonds found in this compound. Table 6e is a systematic presentation of the geometry of the bonded contacts of each water molecule.

The most notable thing about the hydrogen bonding arrangement is that it emphasizes the existence of two different regions in the crystal structure, an anhydrous peptide region and an aqueous region around the sulfate ion. The hydrogen bonding between the two regions is weak. The main link is through the zinc ion. Both water molecules which are bonded to the zinc ion, O_1^W and O_2^W, form strong hydrogen bonds with the sulfate ion.

TABLE 6a

Hydrogen Bond Distances and Angles

Bond	Atom X	Atom Y	Distance(X...Y)	Angle(X-H...Y)
I	N_1	$^{7}_{-c}O^{s}_{1}$	3.152(6) Å	171°
II	N_1	$^{3}O^{w}_{3}$	3.052(6)	155
III	N_2	$_{-b}O^{w}_{2}$	3.166(5)	161
IV	N_3	$^{6}_{-b}O_{1}$	3.152(5)	170
V	O^{w}_{1}	$^{6}O^{1}_{3}$	2.597(5)	167
VI	O^{w}_{1}	O^{s}_{1}	2.699(5)	154
VII	O^{w}_{2}	$^{7}_{-c}O^{w}_{3}$	2.658(5)	168
VIII	O^{w}_{2}	O^{s}_{2}	2.698(5)	156
IX	O^{w}_{3}	O^{w}_{4}	2.723(5)	142
X	O^{w}_{3}	O^{s}_{1}	2.768(6)	143
XI	O^{w}_{4}	$_{-b}O^{s}_{2}$	2.765(6)	167
XII	O^{w}_{4}	O_2	2.858(6)	90*

*It seems likely that this interaction is a van der Waal's contact between the oxygen atoms rather than a hydrogen bond. The reasons for this are discussed in the text.

TABLE 6b

Other Relevant Distances and Angles in the Hydrogen Bonds

Bond	Distance(X-H)	Distance(H...Y)	Angle(Z-X...Y)
I	N_1-H_1	$H_1 \cdots {}^{7}_{-c}O^{s}_{1}$	C^{α}_{1}-$N_1 \cdots {}^{7}_{-c}O^{s}_{1}$
	.858 Å	2.30 Å	95.6(3)°
II	N_1-H_2	$H_2 \cdots {}^{3}O^{w}_{3}$	C^{α}_{1}-$N_1 \cdots {}^{3}O^{w}_{3}$
	.789 Å	2.31 Å	113.4(3)°
III	N_2-H	$H \cdots {}_{-b}O^{w}_{2}$	C^{α}_{2}-$N_2 \cdots {}_{-b}O^{w}_{2}$
	.943 Å	2.24 Å	113.8(3)°
IV	N_3-H	$H \cdots {}^{6}_{-b}O_{1}$	C'_{2}-$N_3 \cdots {}^{6}_{-b}O_{1}$
	.861 Å	2.29 Å	109.1(3)
V	O^{w}_{1}-H_2	$H_2 \cdots {}^{6}O^{1}_{3}$	H_1-$O^{w}_{1} \cdots {}^{6}O^{1}_{3}$
	.906 Å	1.31 Å	95°
VI	O^{w}_{1}-H_1	$H_1 \cdots O^{s}_{1}$	H_2-$O^{w}_{1} \cdots O^{s}_{1}$
	.773 Å	1.97 Å	120.4(3)°
VII	O^{w}_{2}-H_1	$H_1 \cdots {}^{7}_{-c}O^{w}_{3}$	H_2-$O^{w}_{2} \cdots {}^{7}_{-c}O^{w}_{3}$
	.926 Å	1.74 Å	93°
VIII	O^{w}_{2}-H_2	$H_2 \cdots O^{s}_{2}$	H_1-$O^{w}_{2} \cdots O^{s}_{2}$
	.814 Å	1.93 Å	116°
IX	O^{w}_{3}-H_1	$H_1 \cdots O^{w}_{4}$	H_2-$O^{w}_{3} \cdots O^{w}_{4}$
	.988 Å	1.87 Å	118°
X	O^{w}_{3}-H_2	$H_2 \cdots O^{s}_{1}$	H_1-$O^{w}_{3} \cdots O^{s}_{1}$
	1.420 Å	1.49 Å	92°
XI	O^{w}_{4}-H_2	$H_2 \cdots {}_{-b}O^{s}_{2}$	H_1-$O^{w}_{4} \cdots {}_{-b}O^{s}_{2}$
	1.093 Å	1.677 Å	107°
XII	O^{w}_{4}-H_1	$H_1 \cdots O_2$	H_2-$O^{w}_{4} \cdots O_2$
	1.060 Å	2.64 Å	67°

*See note in preceding table.

TABLE 6c

Van der Waals' Contacts Less than 3.50 Å

Atom X	Atom Y	Distance
O_1	$^{6}N_3$	3.15 Å
O_2	$^{6}N_3$	3.43
O_2	$^{6}C^{\alpha}{}_3$	3.33
$O^2{}_3$	$^{1}_{-b}N_2$	3.18
$O^2{}_3$	$^{1}_{-b}C^{\alpha}{}_2$	3.16
$C^{\alpha}{}_2$	$^{6}_{-b}O_1$	3.37
$C^{\alpha}{}_2$	$^{7}_{-b,-c}O^2{}_3$	3.21
$O^w{}_2$	$^{1}_{-c}C'{}_3$	2.43

TABLE 6d

Minimum Contact Distances[42]

Contact	Normally Allowed	Outer Limit
C...C	3.20 Å	3.00 Å
C'..C'	2.95	2.90
C...O	2.80	2.70
C...N	2.90	2.80
C...H	2.40	2.20
O...O	2.70	2.60
O...N	2.70	2.60
O...H	2.40	2.20
N...N	2.70	2.60
N...H	2.40	2.20
H...H	2.00	1.90

TABLE 6e

Geometry of the Possible Bonding Contacts of the Waters

A. O^{W}_{1}

Angles at the Water Oxygen

Angle	θ
H_1-O^{W}_{1}-H_2	102°
Zn-O^{W}_{1}-H_1	119
Zn-O^{W}_{1}-H_2	136

Hydrogen Bond Acceptors ($^{6}O^{1}_{3}$, O^{S}_{1})		Hydrogen Bond Donors (none)
Angle	θ	
Zn-O^{W}_{1}-$^{6}O^{1}_{3}$	141.4(3)°	
Zn-O^{W}_{1}-O^{S}_{1}	120.5(2)	
O^{S}_{1}-O^{W}_{1}-$^{6}O^{1}_{3}$	86.3(2)	

B. O^{W}_{2}

Angles at the Water Oxygen

Angle	θ
H_1-O^{W}_{2}-H_2	101°
Zn-O^{W}_{2}-H_1	109
Zn-O^{W}_{2}-H_2	127

Hydrogen Bond Acceptors ($^{7}_{c}O^{W}_{3}$, O^{S}_{2})		Hydrogen Bond Donors ($_{b}N_{2}$)	
Angle	θ	Angle	θ
Zn-O^{W}_{2}-$^{7}_{c}O^{W}_{3}$	115.1(2)°	Zn-O^{W}_{2}-$_{b}N_{2}$	97.6(1)°
Zn-O^{W}_{2}-O^{S}_{2}	119.9(2)	$^{7}_{c}O^{W}_{3}$-O^{W}_{2}-$_{b}N_{2}$	126.3(2)
O^{S}_{2}-O^{W}_{2}-$^{7}_{c}O^{W}_{3}$	109.2(2)	O^{S}_{2}-O^{W}_{2}-$_{b}N_{2}$	86.5(2)

TABLE 6e - continued

C. O_{W_3}

Angle at the water oxygen H_1-O^W_3-H_2 110°

Hydrogen bond acceptors O^W_4, O^S_1

Hydrogen bond donors $^3N_1$, $^7O^W_2$

Angles between the hydrogen bonds

Angle	θ
O^W_4-O^W_3-O^S_1	105.1(2)°
O^W_4-O^W_3-$^3N_1$	139.6(2)
O^S_1-O^W_3-$^3N_1$	107.2(2)
$^7O^W_2$-O^W_3-O^S_1	10616(2)
$^7O^W_2$-O^W_3-$^3N_1$	91.4(2)
$^7O^W_2$-O^W_3-O^W_4	101.7(2)

D. O^W_4

Angle at the water oxygen H_1-O^W_4-H_2 100°

Hydrogen bond acceptors $_bO^S_2$, O_2*

Hydrogen bond donors O^W_3

Angles between the hydrogen bonds

Angle	θ
O^W_3-O^W_4-O_2	124.1(2)°
O^W_3-O^W_4-$_bO^S_2$	114.9(2)
O_2-O^W_4-$_bO^S_2$	103.0(3)

* It seems likely that this interaction is a van der Waals contact rather than a hydrogen bond; the reasons for this are discussed in the text.

One of these oxygens, O_1^w forms the only strong hydrogen bond with the peptide. This hydrogen bond is between O_1^w and[6] O_3^1, one of the atoms in the carboxylate group. The oxygen-oxygen distance is 2.59 Å, the shortest hydrogen bond in the structure. This bond is expected to be especially strong because of the partial positive charge on the donor atom and the partial negative charge on the acceptor atom. O_1^s is the acceptor atom for the second hydrogen bond formed by O_1^w.

The second oxygen bonded to the zinc also donates both of its hydrogen atoms for hydrogen bonding. Both bonds are shorter than the "normal" oxygen-oxygen distance of 2.75 angstroms[31] and therefore are supposed to be stronger than the usual oxygen-oxygen hydrogen bond. Again this is not surprising, upon consideration of the partial charge on the donor and one of the acceptor atoms. The acceptor atoms are O^s_2 at 2.698 Å and $^7_{-c}O_3^w$ at 2.658 Å. It is possible that O_2^w also accepts a hydrogen from $_bN_2$ to form a weak hydrogen bond. The nitrogen-oxygen distance is 3.166 Å which is long but still within the accepted limits for hydrogen bond formation. Marsh and Donohue[40] report that 2.90 angstroms is the typical distance for a peptide hydrogen bond involving the amide hydrogen. Also the oxygen is situated very near to the extended axis of the nitrogen-hydrogen bond. However it is unlikely that the hydrogen bond actually exists. First the fact that this water molecule is one of the zinc ligands makes it a very poor hydrogen bond acceptor. The geometry of this possible hydrogen bond may be seen in figure 5. Secondly the zinc and two hydrogens covalently bonded to the oxygen form a trigonal surrounding rather than the tetrahedral surrounding expected for the formation of four bonds. The angles involved in the surroundings of the water molecules are listed in table 6e.

As mentioned above $^{7}_{-c}O_3^W$ accepts a hydrogen bond from O_2^W. It also accepts a hydrogen bond from N_1 which is 3.052 Å from $^{3}O_3^W$. This distance is considerably longer than the 2.80 to 2.85 angstroms[40] usually reported for ammonium nitrogen to oxygen hydrogen bonds. This is however the shortest hydrogen bond in this structure involving a peptide hydrogen. This water also donates both of its hydrogens to form hydrogen bonds. The hydrogens are donated to O_1^S and O_4^W. The oxygen-oxygen distances are 2.768 Å and 2.733 Å respectively.

O_4^W also donates a hydrogen to the sulfate ion, specifically O_4^W hydrogen bonds to $_{-b}O_2^S$ at a distance of 2.76 angstroms. This completes a bridge formed by O_3^W and O_4^W between translationally related sulfate ions. The translation is along the b axis, the axis of the aqueous column. Thus this aqueous column is a continuous structure held together by hydrogen bonds. It is open to question whether or not the other hydrogen in O_4^W is involved in hydrogen bonding. O_4^W is 2.858 Å from O_2, however the difference Fourier used to locate the hydrogen atoms shows the O_4^W-hydrogen axis as nearly perpendicular to the line between the O_4^W-O_2 axis. There is absolutely no peak between the two oxygen atoms which might be a hydrogen atom. Further the internuclear separation of the two oxygen atoms agrees well with that predicted by the Pauling[16b] van der Waals radii of 1.40 angstroms for oxygen. It should also be noted that when isotropic temperature factors were refined O_4^W had a temperature factor of 6.6 compared to 4.5 for O_3^W, the other water not bonded to the zinc. This indicated more freedom of movement and probably less bonding. Therefore perhaps the hydrogen position shown by the difference Fourier is correct and O_4^W is involved in only two hydrogen bonds.

From the preceding discussion it should be clear that the environment of the water molecules contains two features which could be considered abnormal. These features are the possible acceptance of a hydrogen bond by O^{W}_{2}, which is one of the zinc ligands, and the failure of one of the hydrogen atoms on O^{W}_{4} to enter into the formation of a hydrogen bond.

Examination of the hydrogen bonding angles in Table 6e reveals that the water molecule O^{W}_{2} is surrounded by a distorted trigonal pyramid. This is not usually considered one of the geometries favorable to hydrogen bond formation. However the bipyramid has appeared in the crystal structure of sodium perxenate octahydrate.[79]

It is difficult, if not impossible, to explain the apparent nonformation of a hydrogen bond between O^{W}_{4} and O_{2}. The angles listed in Table 6e indicate that O^{W}_{4} could donate a hydrogen to form a hydrogen bond with O_{2} without disrupting the other hydrogen bonds in which it participates. However the hydrogen atom is not located in a position such that the bond is formed.

There is one other notable feature which involves a water molecule. This feature is the large anisotropy in the thermal parameters of O^{W}_{1} and O^{1}_{3}. These are the two atoms involved in hydrogen bond V. The major axes of thermal ellipsoids of the two atoms in the hydrogen bond are very nearly parallel and almost perpendicular to the bonds which hold these two atoms. There is no readily apparent explanation for these large anisotropies. Frequently such incongruous anisotropies are the result of disorder within the crystal. This should be considered as a possibility even though the exact nature of the disorder is not made clear by examination of the final Fourier maps, the result of which is

given in table 2 or by examination of a three dimensional model.

The amine terminal nitrogen, N_1, is involved in two weak hydrogen bonds. It donates hydrogens to $^{3}O_{3}^{w}$ and to $^{1}_{-c}O_{1}^{s}$. The nitrogen oxygen distances are 3.052 Å and 3.152 Å respectively, significantly longer than the "normal" value of 2.85 angstroms. This is the third hydrogen bond for which O_1^s is the acceptor oxygen. It is not unusual for a sulfate to accept a third hydrogen bond, especially if the donor is an ammonium ion.[41] N_2, as previously discussed, probably does not enter into the formation of a hydrogen bond.

There is only one potential hydrogen bond between the peptide chains. This is from N_3 to $^{6}_{-b}O_1$. This bond, if it exists, is very weak, the nitrogen to oxygen distance is 3.152 Å. The N_3-H axis is coincident with the N_3-O_1 bond. Therefore the geometry is favorable to hydrogen bond formation; the distance, however, is unfavorable. Further O_1 is the carbonyl oxygen involved in the formation of the chelate ring with the zinc ion. This would be expected to reduce its ability to accept a hydrogen bond. Therefore at best this hydrogen bond is very weak. This may be seen in figure 5.

Thus all twelve hydrogens capable of hydrogen bonding could be involved in a hydrogen bond. The sulfate-aqueous column has arranged itself to maximize its internal hydrogen bonding between it and the water molecules which are zinc ligands. The hydrogen bonds formed by the peptide hydrogens are very weak and appear to have little effect on the conformation of the peptide chain. There is only one possible hydrogen bond between the anhydrous peptide region and the aqueous column and its existence is questionable. This lack of influence by hydrogen

bonding forces is in marked contrast to the important role of hydrogen bonding in uncomplexed peptide. Also in the $CaCl_2$ complex of glycyl glycyl glycine the combination of hydrogen bonding and metal-carbonyl oxygen bonding forces the peptide chain into an unfavorable conformation.

The Sulfate Ion

The sulfate ion is situated on a two fold rotation axis, one of the crystallographic symmetry operations present in this structure. Therefore there are only two independent bond lengths and three angles needed to completely describe this sulfate ion. All of the angles are within one and half degrees of the tetrahedral angle of 109.5 degrees. The bond lengths are the same. The experimental value of 1.445 angstroms agrees with the expected value of 1.45 angstroms.[31] From these values it is reasonable to conclude that the negative charge on the sulfate ion is equally distributed among the four oxygens.

TABLE 7

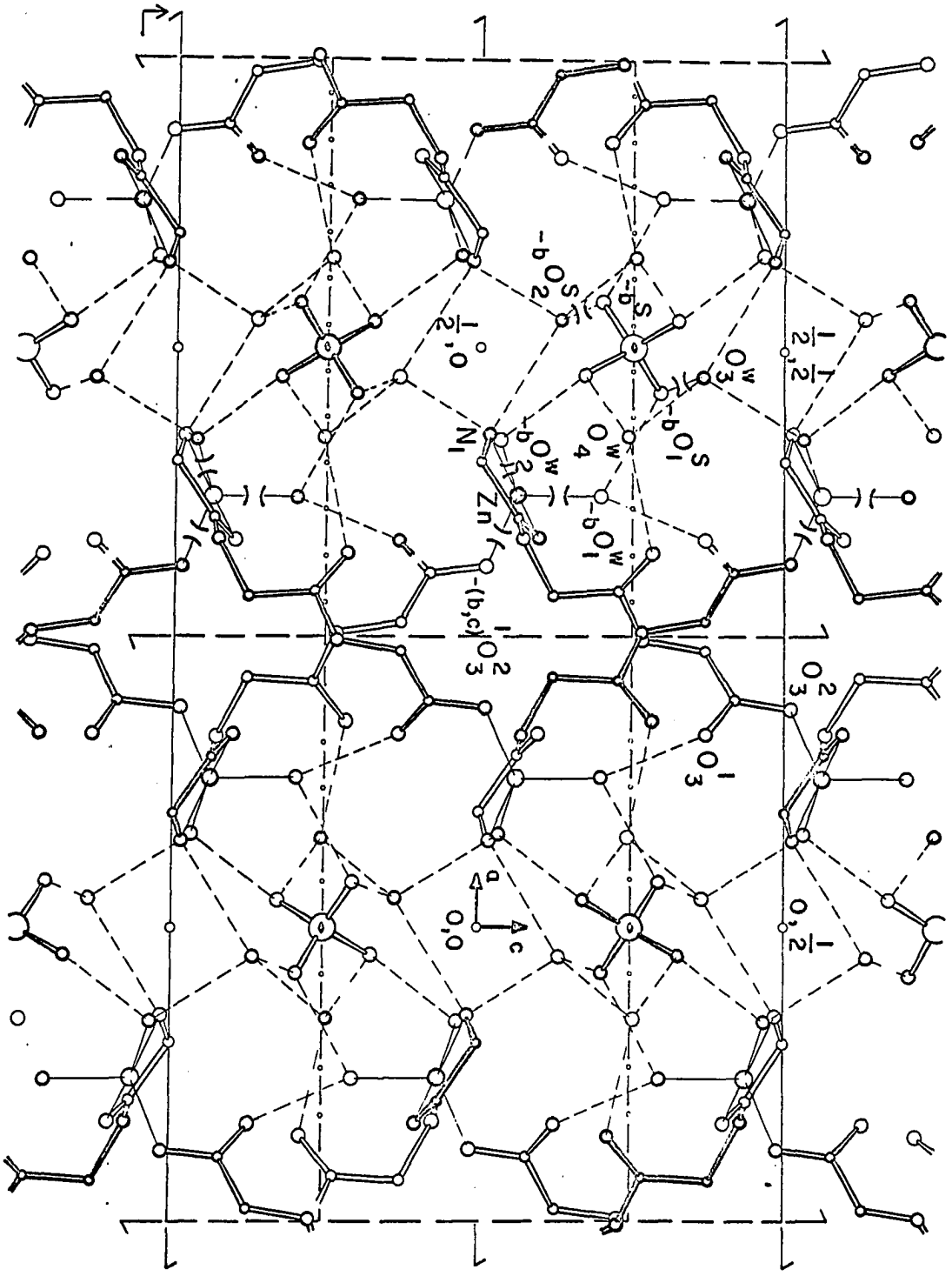
Distances and Angles in Sulfate Ion

Bond	Distance	Angle	Θ
$S-O_1$	1.445(4) Å	O_1-S-O_2	109.4(2)°
$S-O_2$	1.445(4)	$O_1-S-{}^3O_1$	110.8(3)
		$O_2-S-{}^3O_2$	108.2(3)

Geometry of the Peptide

The peptide bond distances and angles are generally explained in terms of the resonance structures A and B, shown on the next page.

Figure 2. The ac Projection of $Zn(ggg).\frac{1}{2}SO_4.4H_2O$



Increasing thickness indicates increasing b, from $-\frac{1}{2}$b to $+\frac{1}{2}$b.

TABLE 8a

Bond lengths and Angles in the Tripeptide Molecule*

Bond	Length	Angle	Θ
$N_1-C^{\alpha}{}_1$	1.482(7) Å	$N_1-C^{\alpha}{}_1-C'_1$	110.1(4)°
$C^{\alpha}{}_1-C'_1$	1.505(7)	$C^{\alpha}{}_1-C'_1-N_2$	115.3(4)
C'_1-O_1	1.240(6)	$O_1-C'_1-C^{\alpha}{}_1$	120.7(5)
C'_1-N_2	1.318(7)	$O_1-C'_1-N_2$	123.8(4)
$N_2-C^{\alpha}{}_2$	1.448(7)	$C'_1-N_2-C^{\alpha}{}_2$	122.3(4)
$C^{\alpha}{}_2-C'_2$	1.510(7)	$N_2-C^{\alpha}{}_2-C'_2$	114.8(4)
C'_2-O_2	1.216(6)	$C^{\alpha}{}_2-C'_2-N_3$	113.6(4)
C'_2-N_3	1.335(7)	$O_2-C'_2-C^{\alpha}{}_2$	123.6(4)
$N_3-C^{\alpha}{}_3$	1.453(7)	$O_2-C'_2-N_3$	123.7(5)
$C^{\alpha}{}_3-C'_3$	1.506(7)	$C'_2-N_3-C^{\alpha}{}_3$	123.5(4)
$C'_3-O^1{}_3$	1.212(6)	$N_3-C^{\alpha}{}_3-C'_3$	115.9(4)
$C'_3-O^2{}_3$	1.267(6)	$C^{\alpha}{}_3-C'_3-O^1{}_3$	122.3(4)
		$C^{\alpha}{}_3-C'_3-O^2{}_3$	115.2(4)
		$O^1{}_3-C'_3-O^2{}_3$	122.4(4)

TABLE 8b

Average Values in Peptides and Metal-Peptide Complexes

	Peptides (Marsh and Donohue)	Metal Complexes (Freeman)
Bond	Length	Length
N-C (terminal)	1.49 Å	1.49 Å
C^{α}-C'	1.51	1.53
C'-O	1.42	1.26
C'-N	1.325	1.30
N-C	1.455	1.46
Angle	Θ	Θ
N-C^{α}-C'	111°	111°
C^{α}-C'-N	116	115
C^{α}-C'-O	120.5	119
O-C'-N	123.5	126
C'-N-C^{α}	122	123

*Standard deviations are in parentheses.

TABLE 8c

Least Squares Planes of the Peptide Chain

Plane	Atoms to which plane was fitted	Equation of the plane*
I	O_1 N_2 C^{α}_1 C'_1	14.132x - .420y + 11.361z = 5.639
II	O_1 N_2 C^{α}_1 C'_1 C^{α}_2	13.365x - .661y + 11.583z = 5.334
III	C^{α}_2 C'_2 O_2 N_3	12.806x - 6.890y - 1.664z = 3.818
IV	C^{α}_2 C'_2 O_2 N_3 C^{α}_3	12.476x - 6.966y - 1.428z = 3.776
V	C^{α}_3 C'_3 O^1_3 O^2_3	16.102x - 6.103y +2.423z = 3.598

* X, y, and z are expressed in fractional coordinates.

Distances from the planes to certain atoms

Atom	Plane I	Plane II	Atom	Plane III	Plane IV
O_1	-.008 Å	.001 Å	C^{α}_2	-.003 Å	-.023 Å
N_2	-.007	.060	C'_2	.012	.019
C^{α}_1	-.006	-.045	O_2	-.005	-.001
C'_1	.021	.035	N_3	-.004	.031
C^{α}_2	-.176	-.051	C^{α}_3	-.099	-.026

Atom	Plane V
C^{α}_3	.001 Å
C'_3	-.005
O^1_3	.002
O^2_3	.002

Figure 3. Bond Distances in the Tripeptide Molecule

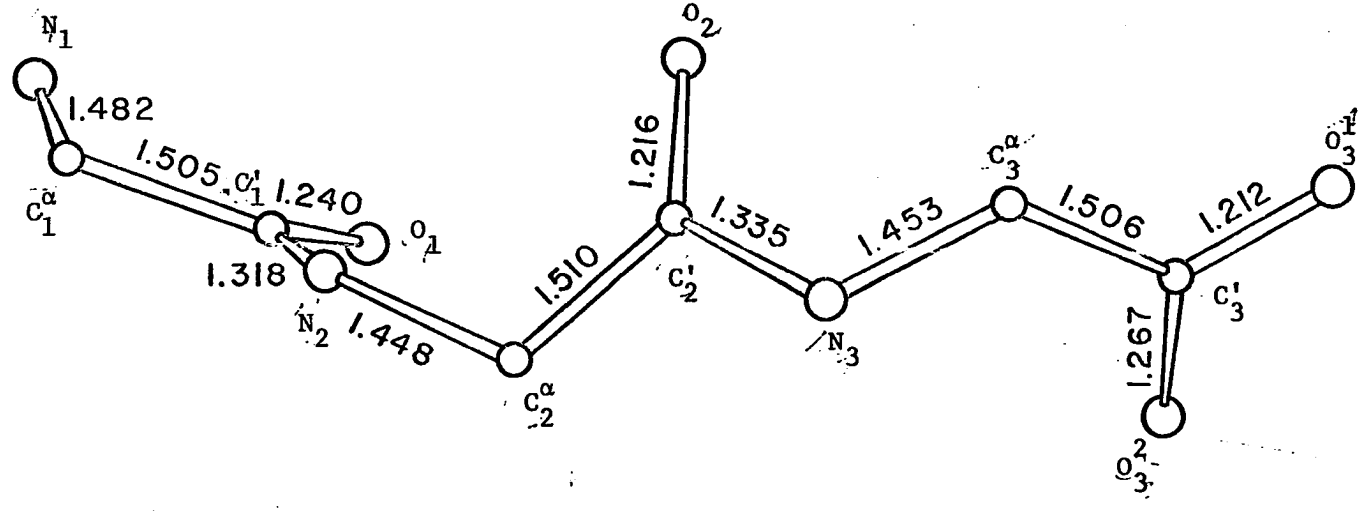
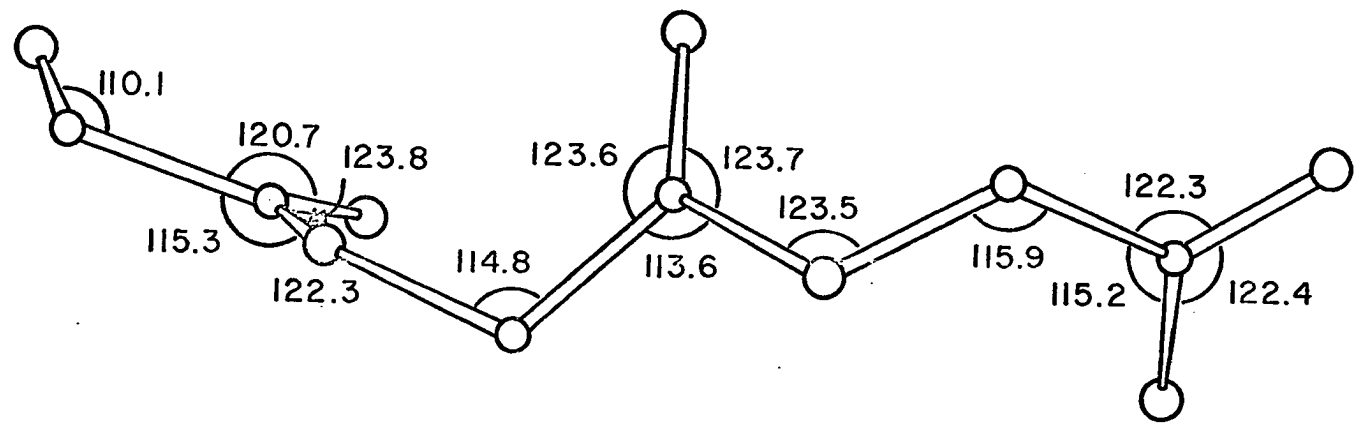
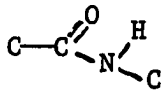
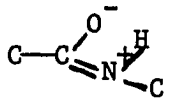


Figure 4. Bond Angles in the Tripeptide Molecule





(A)



(B)

Unfortunately there have been very few peptide structures done with sufficient accuracy to make feasible a quantative discussion of the bonding. Also many of the compounds studies have been obtained under conditions of high pH. However some general trends have been established.[5,40] These trends include confirmation of the general correctness of Pauling's hypotheses concerning peptide group dimensions and allowed peptide configurations. Also the use of carbonyl oxygen as a transitions metals ligands has been established.

The bond distances and angles found in zinc triglycine are all within the experimentally established range of values. The bond distances and angles are in Figures 3 and 4 and Table 8a. The bonding of the peptide molecule is so regular that only a few of the values deviate from the average[5b,40] values found in several compilations. The values from these compilations are found in Table 8b. The bond distances are also in agreement with those published for zinc glycylglycine.[5b]

The N_1-C^{α}_1 distance of 1.482 Å in zinc triglycine is not significantly different from the 1.49 Å found as an average of several complexed peptides by Freeman.[5b] Neither is it much different from the values of 1.48 Å and 1.46 Å found in zinc glycylglycine. The distances for the bonds N_2-C^{α}_2 and N_3-C^{α}_3 are 1.448 Å and 1.453 Å respectively. This is not significantly different from distances of 1.44 and 1.46 angstroms in zinc glycylglycine and 1.46 angstroms as an average of

several complexed or chelated peptides.[5b] The bond angles at the peptide nitrogens N_2 and N_3 are 122.3° and 123.5°. These are within a standard deviation of the average value of 123°. For purposes of comparison the C^{α}-C' bond are generally broken into two groups, those in which C' is part of an amide group and those in which C' is part of a carboxyl group. The two C^{α}-C' bonds which terminate in amide groups, $C^{\alpha}{}_1$-C'_1 and $C^{\alpha}{}_2$-C'_2 have lengths of 1.505 and 1.510 angstroms respectively. The distances found in zinc glycylglycine are 1.50 and 1.51 angstroms. Marsh and Donohue[40a] give 1.51 angstroms as the average for non-chelated peptides. Freeman[5a] gives 1.53 angstroms as an average in chelated structures. The $C^{\alpha}{}_1$-C'_1 bond appears short when compared to Freeman's average value. However this average is heavily weighted with copper chelates which are bonded through a deprotonized peptide nitrogen; this alters the bond lengths appreciably.

The $C^{\alpha}{}_3$-C'_3 bond which terminates in a carboxyl group has a length of 1.506 angstroms. This is definitely shorter than the lengths of 1.52 and 1.54 angstroms found in zinc glycylglycine or 1.52 and 1.527 angstroms average values given by Freemen[5a] and Marsh and Donohue[40a] respectively. However it is within three standard deviations of the average values and one of the lengths found in zinc glycylglycine. Thus it is probably not significant.

The most important mechanism for relieving strain in peptide molecules is the deformation of the N-C^{α}-C' bond angle. The most commonly occuring value is 111° for peptides.[40a] Values of 105° to 115° are common however, and one structure has been reported with a value of 125°.[42] The values in this structure are in the normal range. In the

amine terminal residue, the angle N_1-C^{α}_1-C'_1 is 110.1°, almost exactly the average value. The values for the two peptide group angles, N_2-C^{α}_2-C'_2 and N_3-C^{α}_3-C'_3 are 114.8° and 115.9° respectively. It is not apparent what strains have caused these two angles to deform in this manner. The conformational angles in table 10 indicate that it is not an internal strain. Perhaps it is simply stretching between two zinc ions and over another peptide chain.

The bonding around C'_1 is in agreement with the published averages.[5] The C'_1-O_1 bond length is 1.240 angstroms compared to 1.24Å and 1.26Å in zinc glycylglycine. Freeman[5] gives an average of 1.24 angstroms for C=O bonds when the oxygen atom is complexed with a metal ion. The C'_1-N_2 distance is 1.318 Å, compared to 1.31 angstroms given by Freeman[5] and 1.32 angstroms found in zinc glycylglycine.[5] The three bond angles, C^{α}_1-C'_1-O_1, C^{α}_1-C'_1-N_2, and O_1-C'_1-N_2 are 120.7°, 115.3°, and 123.8° respectively. This agrees with the averages given by Marsh and Donohue[40] rather than those of Freeman.[5] The Marsh and Donohue[40] values are 120.5°, 116°, and 123.5° respectively, those of Freeman are 119°, 115°, and 126°. This discrepancy is probably caused by the fact that Freeman's values are heavily weighted with compounds chelated through deprotonized peptide nitrogen atoms.

The bonding around C'_2 shows significant deviations from the averages of March and Donohue[40b] for uncomplexed peptides, this peptide group is also uncomplexed, and therefore a comparison is valid. The angles C^{α}_2-C'_2-O_2, C^{α}_2-C'_2-N_3, and O_2-C'_2-N_3 are 123.6°, 113.6°, and 123.7°, respectively. These differ significantly from the average values of 120.5°, 116°, and 123.5°, except for the third angle. The C'_2-O_2

bond is 1.216 Å, significantly shorter than the 1.24 angstroms listed by Marsh and Donohue. The 1.335 Å C'_2-N_3 distance is longer than the "average" value of 1.325 angstroms but not significantly. An examination of the conformational angles and packing distances[43] indicates that the deformation, as in the case of the N_2-$C^{\alpha}{}_2$-C'_2 bond angle deformation, is not caused by the necessity to relieve an internal strain.

The stretching proposed to explain the extended N-C^{α}-C' bond angles in the second and third peptide residues fails for this deformation. The stretching would be expected to increase the $C^{\alpha}{}_2$-C'_2-N_3 bond angle. Instead this angle has contracted. Therefore another explanation must be found.

The only feasible explanation is that packing forces and van der Waals' repulsion are responsible for these angular deformations. The particular van der Waals contact is between O_2 and O^{W}_4. The success of this argument of course depends upon the absence of a hydrogen bond between these two atoms. Experimental evidence for the absence of this hydrogen bond was discussed earlier. This repulsion would apply force in the direction required by the deformation of the angles N_2-$C^{\alpha}{}_2$-C'_2, $C^{\alpha}{}_2$-C'_2-N_3, and $C^{\alpha}{}_2$-C'_2-O_2. The absence of this hydrogen bond also gives rise to an explanation for the shortened C'_2-O_2 bond and the possibly lengthened C'_2-N_3 bond. The exceptionally weak hydrogen bond formed by N_3 is also a factor in this deviation of bond length. In terms of the resonance structures, A and B, introduced earlier, these bond length deviations correspond to an increase in the contribution of form A. It should now be noted that the averages[40b] are taken from strongly hydrogen bonded structures. Further hydrogen bonding would

tend to stabilize resonance structure B and therefore increase its contribution at the expense of form A. Thus a non-hydrogen bonded amide would be expected to have a larger contribution from resonance structure A. This is indeed the effect observed and the most probable explanation of the deformation of the second peptide residue.

There is a small but probably real deviation from planarity of the peptide group C^{α}_2-(C'_2-O_2)-N_3-C^{α}_3. The deviations are shown in Table 8c. The deviation corresponds to a two degree rotation about the C'_2-N_3 bond. This is the angle omega two (ω_2) in Table 10. This grouping is generally thought to be planar because of the high contribution of resonance structure B to the bonding in protein-like material. Therefore this small deviation is also consistent with an increased contribution of resonance form A.

There is a similar deformation of the first peptide group, C^{α}_1-(C'_1-O_1)-N_2-C^{α}_2. This deviation from planarity corresponds to a 5.5 degree rotation about the C'_1-N_2 bond. This deviation is a common occurrence in chelated peptide groups.[5c] This rotation would reduce the amount of resonance energy gained by the amide group. Also chelation should increase the contribution of form B. There is therefore no satisfactory explanation of this phenomena.

There is nothing remarkable about the carboxylate ion of the peptide. The four atoms C^{α}_3, C'_3, O^1_3, and O^2 are co-planar within experimental error, as is seen in Table 8c. The bond angles do not differ significantly from the average values for carboxyl groups bonded to metals. The bond lengths from C'_3 to O^1_3 and O^2_3 are 1.212 and 1.267 angstroms respectively. This indicates some localization of the pi

electrons as is predicted as a result of O^2_3 being bonded to the zinc ion. It is meaningless to compare the bond lengths with averages since a large variation is to be expected.

The Chelate Ring

As is typical of five membered rings, the chelate ring formed by the first glycyl residue and the zinc ion is not planar. The ring could be described as having either N_1 or C^{α}_1 puckered out of the plane of the ring. The more satisfactory description is that which has N_1 puckered from the plane of the other four atoms. This is because C^{α}_1, C'_1, O_1, and the zinc ion are more nearly co-planar than are N_1, C'_1, O_1, and the zinc ion, as is seen in Table 9. This is in line with the fact that the linear combination of atomic orbitals approach to either molecular orbital or valence bond theory predict that the atoms C^{α}_1, C'_1, O_1, and the zinc ion should be co-planar. The slight departure from planarity is to accomodate the transannular strain between the zinc ion and C^{α}_1 and C'_1. There is a 19° rotation about the C^{α}_1-C'_1 bond which is necessary in order to accomodate this pucker. The angle at the zinc ion, N_1-Zn-O_1, is 79.5°, typical of this type of chelate. The angles at the ligand atoms, Zn-N_1-C^{α}_1 and Zn-O_1-C'_1, 112.6° and 112.5°, are also quite typical.

Conformation of the Peptide

Because of the uniformness of the dimensions of peptide molecules it is possible to specify completely the conformation of any peptide molecule by specifying the rotation about the bonds along the peptide backbone. In order to facilitate the description of peptide conformations

Figure 5. A Partial ab Projection of $Zn(ggg)\text{-}\frac{1}{2}SO_4.4H_2O$

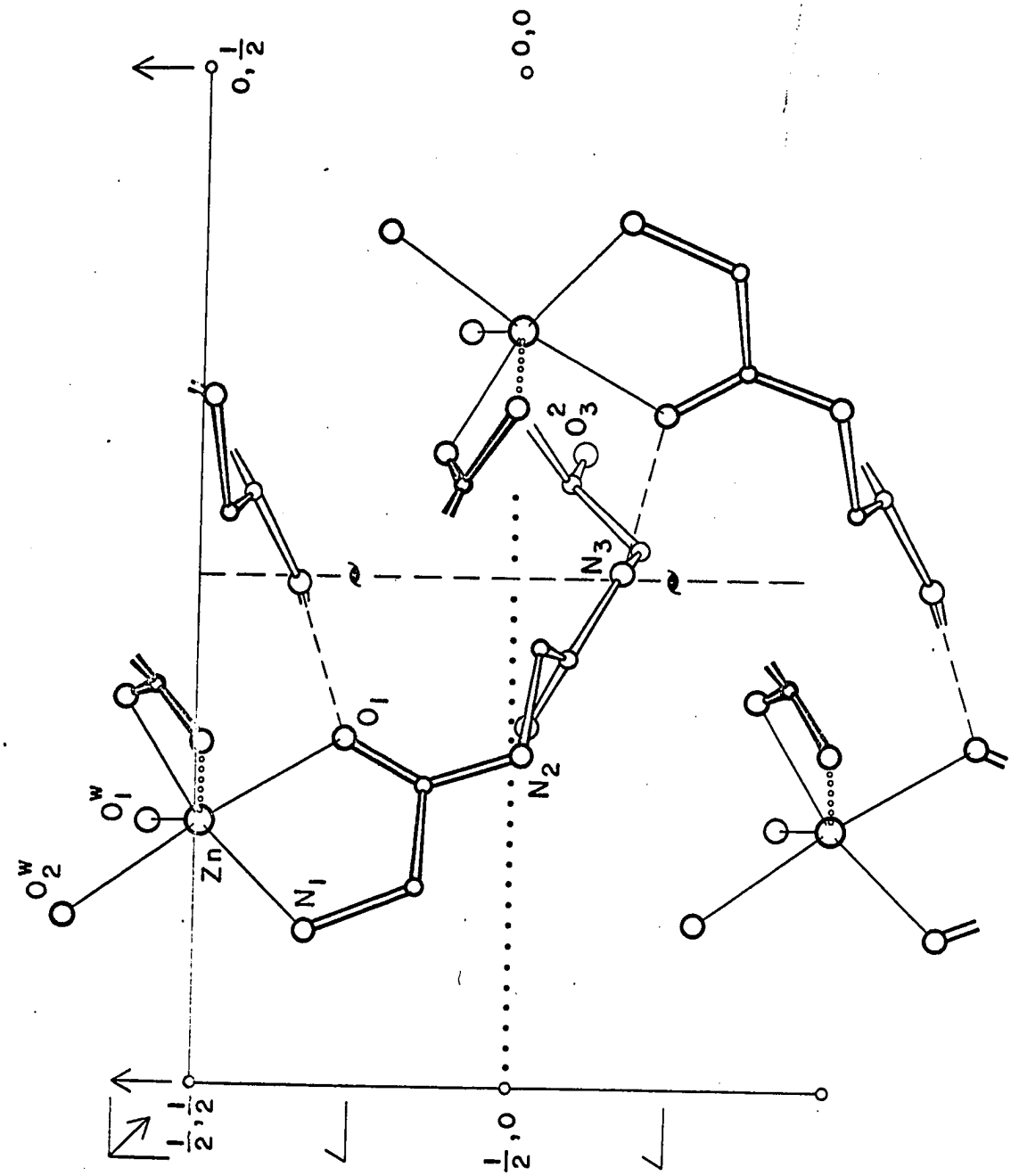


TABLE 9a

Least Squares Planes for the Chelate Ring

Plane	Atoms to which plane was fitted	Equation of the Plane*
I	Zn N_1 C'_1 O_1	10.511x - .432y + 12.398z = 4.487
II	Zn N_1 C'_1 O_1 C^{α}_1	11.794x - .888y + 12.002z = 4.764
III	Zn C^{α}_1 C'_1 O_1	13.864x - .715y + 11.410z = 5.496
IV	C^{α}_1 C'_1 O_1 N_2 Zn	14.052x - .637y + 11.361z = 5.587
V	C^{α}_1 C'_1 N_2 O_1	14.132x - .422y + 11.361z = 5.639

* X,y, and z are expressed in fractional coordinates.

Distances from the planes to certain atoms

Atom	Plane I	Plane II	Plane III	Plane IV	Plane V
Zn	-.027 Å	-.077 Å	.010 Å	.023 Å	.109 Å
N_1	.023	.136	.330	.342	.393
C'_1	-.040	.047	.032	.014	.020
O_1	.044	.032	-.026	-.040	-.008
C^{α}_1	-.304	-.138	-.015	-.019	-.006
N_2				.020	-.007

TABLE 9b

Distances and Angles in the Chelate Ring

Bond	Distance	Angle	θ
$Zn-N_1$	2.034(4) Å	$Zn-N_1-C^{\alpha}_1$	112.6(3) °
$N_1-C^{\alpha}_1$	1.482(6)	$N_1-C^{\alpha}_1-C'_1$	110.1(4)
$C^{\alpha}_1-C'_1$	1.505(7)	$C^{\alpha}_1-C'_1-O_1$	120.7(4)
C'_1-O_1	1.240(5)	C'_1-O_1-Zn	112.5(3)
O_1-Zn	2.186(3)	O_1-Zn-N_1	79.5(1)

in this manner a standard nomenclature, descriptive angles, and reference configuration have been proposed.[8]

There are four angles necessary to describe glycine structures. They are the three rotational angles and the bond angle N-C^{α}-C', which is referred to as $\tau(N_1C^{\alpha}{}_1C'_1)$. This bond angle is necessary because it is subject to wide variation, and its magnitude determines the range of allowed values for the rotational angles.

The first rotational angle is the rotation around the axis of the N_i-$C^{\alpha}{}_i$ bond. It has a value of zero when the N_i-H bond is co-planar with and cis to the $C^{\alpha}{}_i$-C'_i bond of the same peptide residue. The angle is measured as a righthanded rotation, clockwise, when viewed along the vector from N_i toward C^{α}_i. The angle is denoted by the symbol phi, ϕ_i. As a practical matter the angle is calculated using the C'-N bond rather than the N-H bond because the carbon atom is much more accurately located.

The second rotational or conformational angle is the rotation about the $C^{\alpha}{}_i$-C'_i bond. It is assigned a value of zero when N_i-$C^{\alpha}{}_i$ bond is co-planar with and cis to the C'_i-O_i bond. As previously, the angle is measured as a righthanded rotation about the bond axis looking in the forward direction along the peptide chain. In this case the forward direction is from $C^{\alpha}{}_i$ toward C'_i. The angle is referred to as psi, Ψ_i. Some authors have denoted this angle as phi prime, ϕ', where $\Psi = \phi' + 180^{\circ}$. Since some of the tables and conformational maps use the ϕ' instead of both values are given for convenience.

The third conformational angle is the rotation around the C'_i-N_{i+1} bond axis. It is assigned a zero value when the C'_i-O_i bond is co-planar with and trans to the N_{i+1}-H bond. The angle is measured as a right

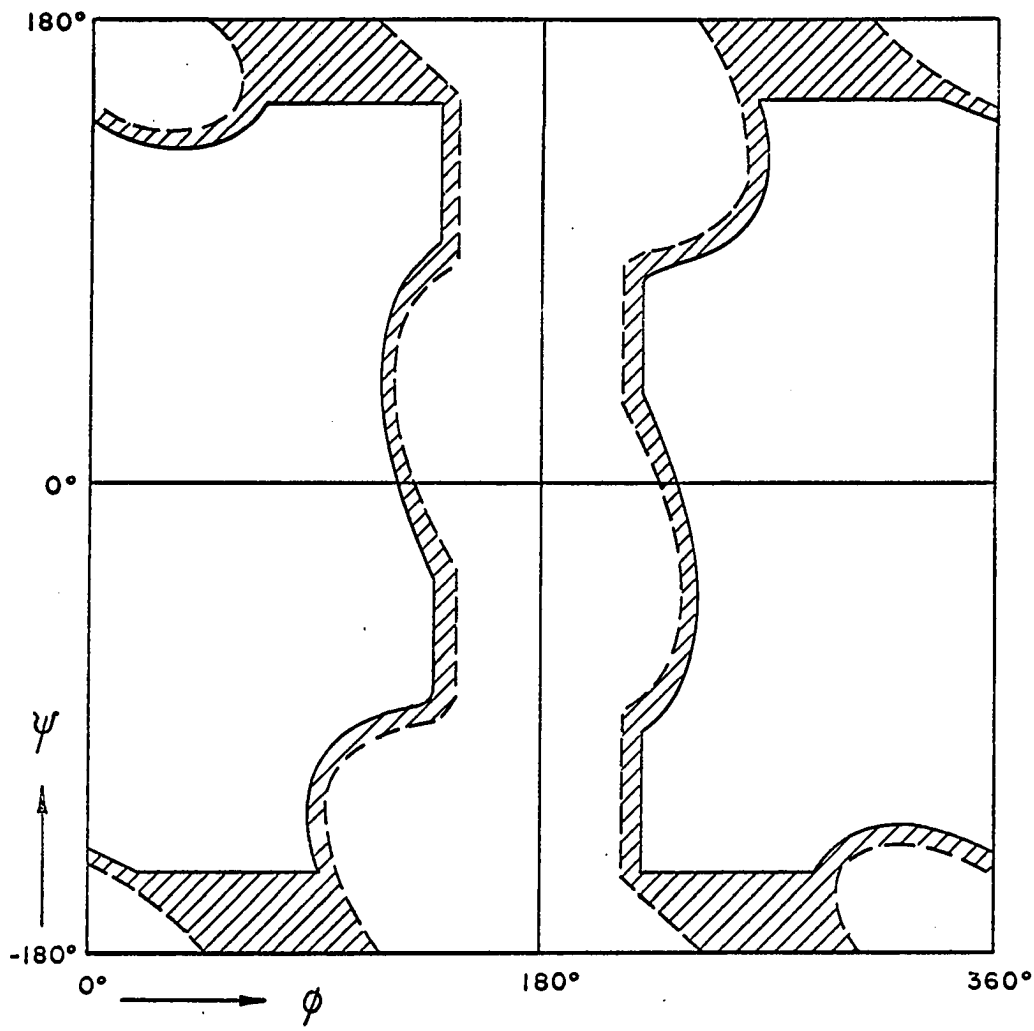
TABLE 10

Conformational Angles of Three Triglycine-Metal Complexes

Angle	$Zn(ggg)\cdot 1/2SO_4 \cdot 4H_2O$	$CaCl_2(ggg) \cdot 3H_2O$	$CuCl(ggg) \cdot 1\ 1/2\ H_2O$
Ψ_1	341°	342°	349°
Φ'_1 *	161	162	169
ω_1	355	2	8
$N_1-C^{\alpha}_1-C'_1$	110.1	110.5	107.9
Φ_2	272	82	294
Ψ_2	338	176	313
Φ'_2	158	356	133
ω_2	358	358	2
$N_2-C^{\alpha}_2-C'_2$	114.8	120.1	111.1
Φ_3	71	291	266
Ψ_3	358	351	353
Φ'_3	178	171	173
$N_3-C^{\alpha}_3-C'_3$	115.9	114.5	110.7

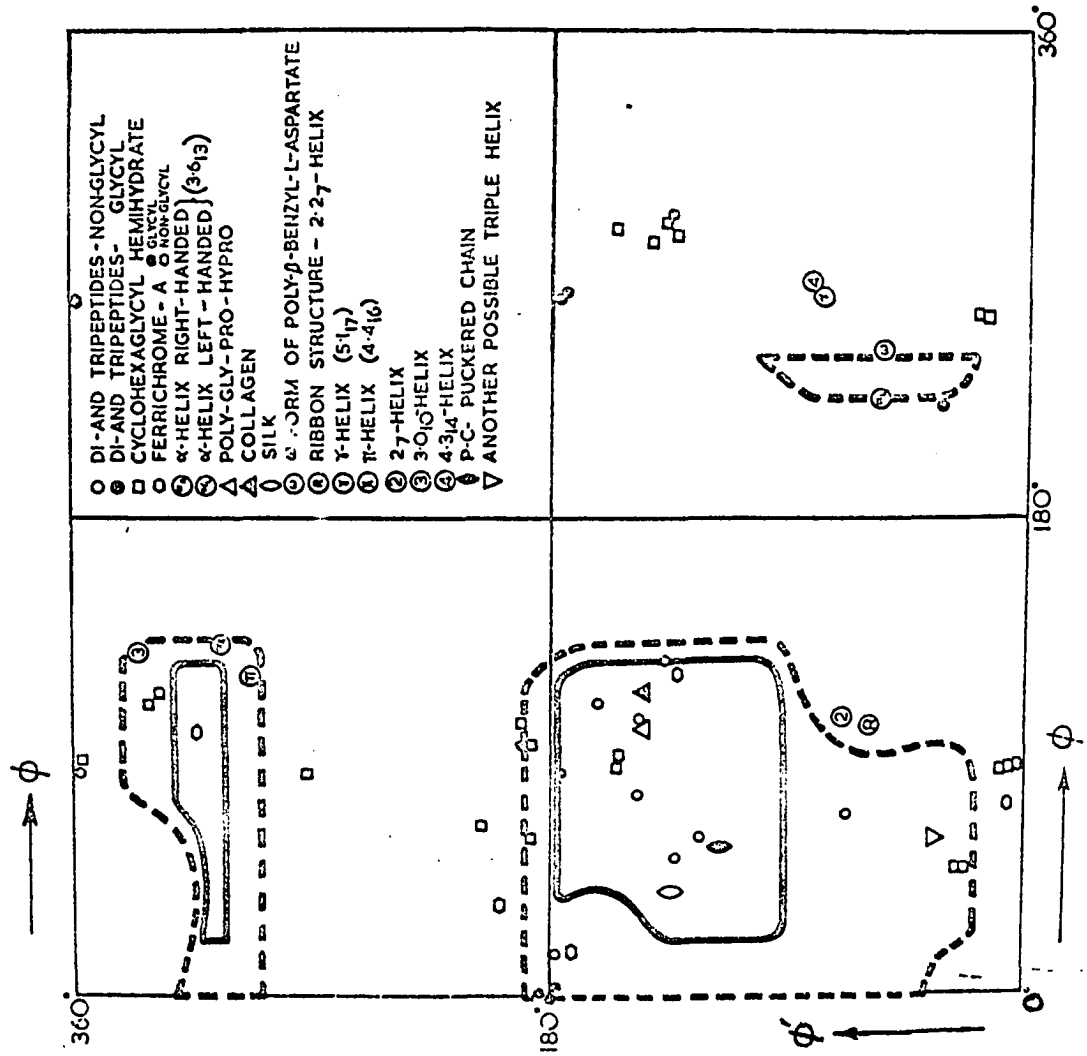
*The angles Φ'_i are the same as the angles Ψ_i. It is from an old nomenclature and has the opposite reference configuration, thus $\Phi'_i = \Psi_i - 180°$. It is included because the angles Φ'_i are plotted in figure 7.

Figure 6. Allowed Conformational Angles for Glycine Residues.



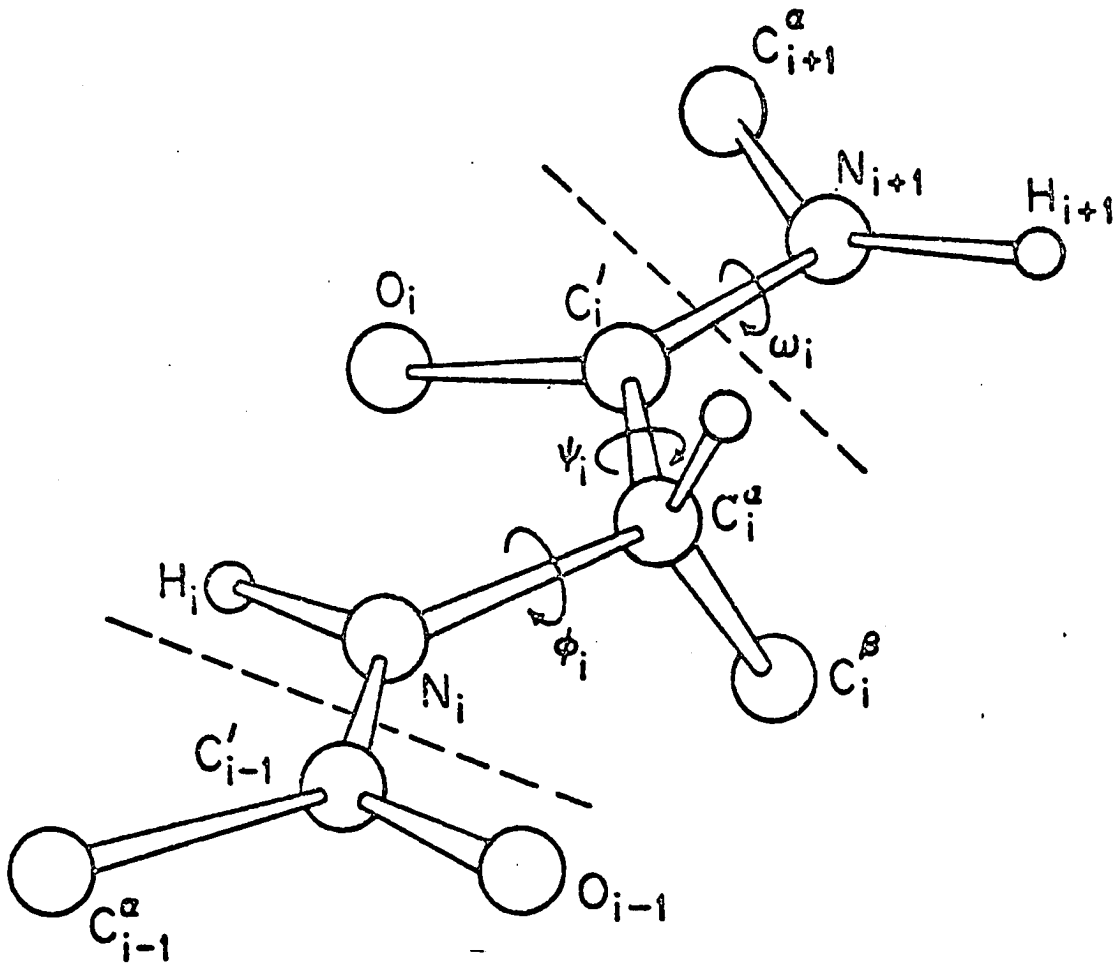
From an article by Ramakrishan and Ramachandran.

Figure 7. Conformational Angles from Known Structures



From an article by Ramakrishan and Ramachandran.[42]

Figure 8. Reference Peptide Configuration



From an article by Edsall et al.[8]

handed rotation about the bond as one looks from C'_i toward N_{i+1}. The angle is referred to as omega, ω_i. As a practical matter the angle is calculated as if the cis configuration of the C'_i-O_i and N_{i+1}-C^{α}_{i+1} bonds were the defined zero. This is because the hydrogen atom is not located as accurately as is the carbon atom. The two methods are equivalent if the nitrogen has trigonal sp^2 bonding. Because of the resonance in the amide this angle is theoretically zero.

If the omega angle is assumed to be zero then only certain values of the phi and psi rotations are possible. These allowed values are a function of the bond angle N-C^{α}-C' and the van der Waal's radii of the atoms. Because of the experimental uncertainty and the importance to theoretical protein models two sets of van der Waal's radii have been used in calculating the theoretically allowed limits of the conformational angles phi and psi. This results in two sets of conformational angles, a normally allowed set corresponding to normally used van der Waal's radii and an "outer limit" set which correspond to a shorter set of van der Waal's radii. Table 6d shows the normal and outer limit van der Waal's radii and Figure 6 shows the allowed ϕ and ψ angles for glycyl residues with a $\tau(NC^{\alpha}C')$ of 115 degrees.

The conformational angle ψ_1 is probably meaningless in terms of allowed peptide or protein conformations. It is expected that this angle will adjust itself to the requirements for the efficient formation of the chelate ring. Its value of 341° is however very similar to that found for uncomplexed amino acids.[40b] This similarity probably results from the requirements of efficient packing for the amino acid structures, rather than any fundamental structural feature. There is no angle ϕ_1 because it

involves the C' atom of the previous peptide residue which does not exist for the amine terminal residue.

A comparison of the calculated conformation angles with the mapping of allowed conformations reveals that all of the angles fall within the normally allowed region. This is confirmed by the relevant intramolecular contact distances which are all larger than those values listed as normally allowed. The only values which might be considered abnormal are the two angles $\tau(N_2C^{\alpha}{}_2C'_2)$ and $\tau(N_3C^{\alpha}{}_3C'_3)$ which are 114.80 and 115.90 respectively. However it is not unusual to find angles of this size.[5,40,42]

It should be noted that it is possible to calculate two values of Ψ_3. This degeneracy results from the fact that residue three is the carboxyl residue and therefore has two oxygens, either of which may be used to calculate Ψ_3. These two values differ from each other by 180°. The value listed in the table is calculated using the oxygen with the shorter C-O bond, the oxygen with the carbonyl character. This is the oxygen not bound to the zinc and the Ψ_3 is 358°. The Ψ_3 calculated using the more ionic of the two oxygens, the one bound to the zinc is 178°.

If the conformational angles for residues two and three are plotted on Figure 7 and compared with the other values plotted there it is obvious that the angles are similar to other structures containing glycine. These include other glycine peptides and the structural proteins. In fact, the conformational angles of residue three are intermediate between those of silk and the currently accepted collagen structure.[45] Both silk and collagen are composed of extended peptide chains and have a high glycine content.

In silk the peptide chains are almost completely extended and planar. A slight degree of pucker along the chain is necessary to accomodate the β-carbon atoms of adjacent peptide chains. The N-H and C=O bonds are approximately perpendicular to the direction of the chain. The adjacent chains in the plane of the amide groups run in opposite directions. The structure is stabilized by an infinite chain of hydrogen bonds between the amide groups of these adjacent chains.

The collagen structure is not so definitely known as is the silk structure. The currently accepted version is based on the polyglycine(II) structure.[45] The basic unit is composed of three extended peptide chains which run in the same direction. The chains are packed together such that the contacts are van der Waals contacts rather than hydrogen bonds. Each individual chain is in an extended left-handed helix that repeats approximately every third residue. These chains then make up the three stranded cable. This three stranded cable is then twisted into a right-handed coil or helix. This macrostructure is referred to as a coiled coil.

The peptide chain in zinc triglycine shows several similarities with the proposed collagen structure. First if the zinc is considered to replace a peptide bond in an infinite chain then the peptide chain makes a continuous coil as is found in polyglycine II[45] and is proposed for collagen. In collagen it is proposed that three strands of coiled peptide coil together to form a supercoil in a structure known as a coiled coil. In zinc triglycine the coil is formed around a two-fold screw axis. This limits the macrostructure to a two stranded coil. The short repeat distance of the crystal and the chain length prevent

the formation of a supercoil. Also the C=O and the N-H bond are oriented more or less perpendicular to the chain as is proposed for collagen and in contrast to other helical structures. Finally, infrared studies[45] indicate that collagen is very poorly arranged for hydrogen bond formation within the three stranded coil. In zinc triglycine only one peptide hydrogen out of four is situated so that it bonds to its own peptide chain. Further this hydrogen bond is rather weak, 3.16 Å. Thus it seems reasonable that the peptide configuration in this and other similar structures could prove useful in building theoretical models of the fibrous proteins.

Comparison of Structural Features with other Triglycine Complexes

Zinc triglycine is the third reported crystal structure of a glycyl glycyl glycine metal complex which was crystalized from a neutral solution. There have been other reported structures containing glycyl glycyl glycine but they were crystalized from a very basic solution and are not indicative of what might be found in a biological system. The other two comparable structures are copper(II) monochloride triglycine sesquihydrate[46] and calcium dichloride triglycine trihydrate.[47] It should be instructive to examine the three structures for similarities and differences.

A comparison of the conformational angles of the three triglycine complexes, Table 10, reveals several similarities. Eight of the nine Ψ angles fall into one range. The exception, Ψ_2 of the calcium complex is distorted so that the oxygen is complexed to the calcium ion. Four of the six ϕ angles fall into the same range. The two exceptions ϕ_2 of the calcium complex and ϕ_3 of the zinc chelate may be explained on

the basis of allowing the complexing atom to reach the position required for bonding to the metal. It is interesting to note that the deviations from the norm are rotations of approximately 180° in all three cases.

It should be noted that the ω_1 angles in the zinc and copper chelates deviate significantly from planarity. This has been noticed in nearly all peptide-metal chelates.[5] This rotation is easier to explain in the basic chelates, where the chelation is through the deprotonized peptide nitrogen. In this case, where the chelation is through the carbonyl oxygen the origin of the distorting force is not clear. The interpretation however is the same for either type of chelate. The bonding to the metal results in a larger contribution of the ionized resonance form in the chelate.

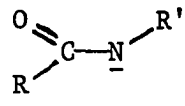
There is one other feature of all three structures which is worth remarking. This is that the $C'_2=O_2$ bond in all three structures is shorter than one expects for peptides. This is also true of the $C'_1=O_1$ bond in the calcium triglycine complex. The bond lengths are for the copper chelate 1.19±0.01 Å, for the calcium complex 1.21±0.01 Å, and for this chelate 1.216±0.006 Å. These distances are significantly shorter than the average distances of 1.24 Å for free peptides[40] and 1.26 Å for carbonyl oxygens not bonded to metal[5] in the compilation of distances in the peptide metal complexes.

It is possible to reconcile these observations with the concept that these bond lengths are determined by the relative contribution of the two resonance forms, H-N-C=O and $H-N^{+}=C-O^{-}$ if one considers that the short bonds have been lengthened less than in other structures rather than actually shortened. Evidence in support of this view is found by examining

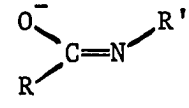
the C=O bond lengths in aldehydes and ketones. There should not be any contribution from similar resonance forms in these structures. The normal value is given as 1.21±0.01 Å,[48] and for compounds with polar substituents such as β-propiolactone and trichloroacetaldehyde, 1.18 Å and 1.15 Å [48] are reported respectively.

If the explanation that these bonds are not shortened, but instead the bonds in the averages have been lengthened to a greater extent, is correct it should be possible to demonstrate the mechanism for this lengthening process. It indeed is possible to postulate reasonable lengthening mechanisms. This is more easily done for the average from metal complex structures.[5]

All of the bond lengths included in this average are from structures in which the amide nitrogen, in the amide group from which the bond length is taken, has been deprotonized. Thus the molecular resonance might be represented as

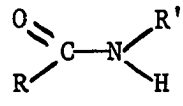


Form A'

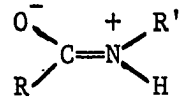


Form B'

rather than the conventional resonance for an amide of



Form A



Form B

An increase in the percentage of B' over B would be expected. This is because increasing B' is dispersing an isolated charge while increasing B is creating opposite charges on nearly adjacent atoms. The structures in the set used for the averages also demonstrate the shortening of the

C-N bond predicted by a larger contribution of the C=N form. The distance for this set of structures is 1.30 Å compared to 1.32 Å for non-chelated peptides.[40] While apparently the matter has not been investigated, it would seem a reasonable supposition that the energy difference between the two deprotonized resonance contributors is less than the energy difference between the two normal resonance contributors.

The bond lengthening mechanism in structures such as the · -helix or pleated sheet is more subtle. It is however equally reasonable and in line with the current chemical thought. The lengthening in this instance is attributed to the induced dipole effect in these structures.

The large magnitude of this effect is the result of two contributing factors. The first and probably most important of these factors is the existence of infinite chains of amide to amide hydrogen bonds. This causes the induced dipole effect to be cumulative. That is the formation of the hydrogen bonds changes the magnitude of the dipole moment of the amide group. However the original magnitude of this dipole was the major factor in determining the strength of the hydrogen bond. Thus this change in the dipole moment causes the hydrogen bond to become stronger. This changed dipole moment and stronger hydrogen bonding then result in an increase of the resonance form $^{-}O-C=N^{+}$.

The second contributing factor to the large induced dipole effect is the initial existence of the resonance structure. It is well known that pi systems, such as this, show large· induced dipole effects.

It should be pointed out that none of these three triglycine-metal complexes forms these infinite chains of amide to amide hydrogen bonds. The bonding in these cases does not present the opportunity for this type of cumulative dipole effect. For these reasons these shorter

C=O bonds seem to be significant and real, not just the normal variation one expects from molecule to molecule.

It was hoped that the compounds used by Marsh and Donohue[40] to compute a set of average dimensions for peptide bonds would yield sufficient data to test this hypothesis. Unfortunately this was not the case. As Marsh and Donohue point out, most of these structures were done before either the data or computing facilities, which are needed to achieve the accuracy necessary to resolve this type of problem, were available. The carbonyl bond lengths of these compounds and the hydrogen bond arrangement of the peptide groups are given in Table 11.

TABLE 11

Carbonyl Bond Lengths of Structures Included

in Marsh and Donohue's Averages

Compound	Reference	C=O Length	Type of Structure
α-glycylglycine	66	1.249(7) Å	anti-parallel pleated sheet
glycyl-l-aspargine	67	1.227(15)	hydrogen bonding not in series
glycyl-l-tryptophane	68	1.226(15)	hydrogen bonding not in series
glutathione	69	1.23(3) 1.24(3)	neither in series
tosyl-l-prolyl-l-hydroxyproline.H_2O	70	1.24(6)	hydrogen bonding not in series
N,N'-diglycyl-l-cystine	71	1.21(3)	hydrogen bonding in series[1]
glycyl-l-phenylalanine glycine	72	1.21(2) 1.23(2)	parallel pleated sheet[2]
l-leucyl-l-prolyl glycine.H_2O	73	1.236(15) 1.272(15)	neither in series

[1] N-H...O distance is 3.31 Å., a very weak hydrogen bond.

[2] These bond lengths should be viewed with extreme skepticism due to systematic factors pointed out by the authors which specifically affect the C=O bonds.

PART II

A PROPOSAL FOR VECTOR REFINEMENT AND SOLUTION OF CRYSTAL STRUCTURES

The advent of electronic computers with moderately large memories and fast computational time has made possible the refinement of structural parameters by the method of least squares. The results of this method depend upon the choice of the origin in the crystal structure. If the origin is not implicit in the mathematical form of the matrix of least squares sums then this matrix will be singular.[23] In a large majority of the possible crystallographic space groups this causes no problem. This is because the symmetry elements are arranged such that the structure factor equations for those space groups assume a particular origin to make efficient use of the symmetry operations. There are however sixty-eight space groups in ten point groups for which their is no symmetry implied origin in one or more dimensions. These are known as the polar space groups and point groups.

The least square matrix in terms of the atomic positions in these polar space groups is singular.[23] This problem is usually overcome by the application of an additional arbitrary constraint for each polar dimension. This constraint takes two usual forms; either one atom is defined as the origin and not allowed to shift or the covariance between all pairs of atoms is set equal to zero. This later method

TABLE 1

Polar Crystallographic Point Groups[23]

Point Group	Number of Space Groups	Number of Polar Dimensions
1, C_1	1	3
m, C_s	4	2
2, C_2	3	1
mm2, C_{2v}	22	1
4, C_4	6	1
4mm, C_{4v}	12	1
3, C_3	4	1
3m, C_{3v}	6	1
6, C_6	6	1
6mm, C_{6v}	4	1

implies that each atom shifts relative to a fixed origin rather than relative to the other atoms. It has been shown that either of these approaches leads to systematic errors in the estimated standard deviations derived from the least squares matrix.[23] The standard deviations derived from these matrices are too large. However there is an alternative approach that does not lead to these errors.

This alternate approach is based on the fact that the intensities which are the measured quantities in the experiment are a function of the atomic composition and the interatomic vectors. Since the set of interatomic vectors contains both vector AB and its inverse BA for all pairs of atoms A,B, it is apparent that the set of interatomic vectors, the Patterson function, is non-polar. This is a direct result of the presence of the inverse operation in the set of symmetry elements present in the symmetry group of the vector set. The inverse, i, operation is defined by the equation, $v \odot i = -v$. Thus the origin must be exactly half way between v and -v and is therefore uniquely defined in all dimensions. Therefore the symmetry group is non-polar.

Development of the Least Squares Equations

If a suitable set of normal equations may be found the method of least squares could be applied to the set of interatomic vectors rather than the atomic positions. Patterson has derived an expression for the absolute magnitude of the intensity, $|F_{hkl}|^2$, in terms of the interatomic vectors and atomic composition of the crystal unit cell.

$$|F_{hkl}|^2 = \sum_{i=1}^{n} f_i^2 + \sum_{\substack{i=1 \\ i\neq j}}^{n} \sum_{j=1}^{n} f_i f_j e^{-2\pi i(hu_{ij}+kv_{ij}+lw_{ij})} \qquad \text{Eq. 1}$$

where f_i is the atomic scattering curve of the ith atom in the unit cell and u_{ij}, v_{ij}, and w_{ij} are the fractional co-ordinates in the a, b, and c directions respectively of the vector between the ith and jth atoms of the unit cell. This equation is of course a complex Fourier series which may be written

$$|F_{hkl}|^2 = \sum_{i=1}^{n} f_i^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} f_i f_j \cos 2\pi(hu_{ij}+kv_{ij}+lw_{ij})$$

$$+ \sum_{i=1}^{n} \sum_{j=1}^{n} f_i f_j i \sin 2\pi(hu_{ij}+kv_{ij}+lw_{ij}) \quad i \neq j \qquad \text{Eq. 2}$$

The expression may be simplified by making use of the centrosymmetric property of the vector set. This yields

$$|F_{hkl}|^2 = \sum_{i=1}^{n} f_i^2 + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} f_i f_j \cos 2\pi(hu_{ij}+kv_{ij}+lw_{ij}) \qquad \text{Eq. 3}$$

These equations are not linear in the unknowns u, v, and w. However, it is well known that the method of least squares may be applied to errors in the parameters u, v, and w if good initial guesses are available for the u, v, and w. This is accomplished by expanding the expression for $|F_{hkl}|^2$ in terms of its Taylor series about the points u_{ij}, v_{ij}, and w_{ij}. The required initial guesses of u, v, and w are available since it has been shown[24] that a Fourier series with $|F_{hkl}|^2$ as coefficients produces a density function in which the high values correspond to the ends of vectors in the vector set.

The Taylor expansion of a function f(x) about the point a is given by:

$$f(x) = f(a) + f'(a)(x-a) + f''(a)(x-a)^2 + \ldots\ldots + f^{n'}(a)(x-a)^n \ldots$$

where x is the "correct" position on the curve, a is the known approximation and $f^{n'}(a)$ represents the nth derivative of the function f(x) with respect to x at the point a.

It is now necessary to introduce the notation needed to apply the above mathematics to the set of observational equations. Let $|F^{I}_{hkl}|^2$ be the ideal or "correct" value of the measured intensity, let $|F^{o}_{hkl}|^2$ be the value actually observed and let $|F^{c}_{hkl}|^2$ be the value calculated by equation 3 above using the approximate u, v, and w. Then

$$|F^{I}_{hkl}|^2 = |F^{o}_{hkl}|^2 + E_{hkl}$$

where E_{hkl} is the error in the experimental observation. The errors in the parameters u_{ij}, v_{ij}, and w_{ij} shall be ε_{ij}, δ_{ij} and τ_{ij} respectively and the required condition for successful least squares approach may then be stated as:

$$\varepsilon_{ij} >> \varepsilon^2_{ij} \quad ; \qquad \delta_{ij} >> \delta^2_{ij} \quad ; \qquad \tau_{ij} >> \tau^2_{ij}$$

With this assumption the Taylor expansion may be applied to $|F_{hkl}|^2$ and all terms past the first derivative ignored. This process yields

$$|F^{I}_{hkl}|^2 = |F^{o}_{hkl}|^2 + E_{hkl} = |F^{c}_{hkl}|^2 + 2 \sum_{i=1}^{n} \sum_{j=1+i}^{n} \left(\frac{\partial(|F^{c}_{hkl}|^2)}{\partial u_{ij}} \varepsilon_{ij} + \frac{\partial(|F^{c}_{hkl}|^2)}{\partial v_{ij}} \delta_{ij} + \frac{\partial(|F^{c}_{hkl}|^2)}{\partial w_{ij}} \tau_{ij} \right) \qquad \text{Eq. 4}$$

where $|F^{c}_{hkl}|^2$ is given by equation 3.

The theory of the least squares states that for a set of observational equations the "best" values of u_{ij}, v_{ij}, and w_{ij} are found

when $\sum_{hkl} \epsilon^2_{hkl}$ is a minimum. This quantity is a minimum when the derivatives

$$\frac{\partial(\sum_{hkl} E^2_{hkl})}{\partial \epsilon_{ij}} \quad , \quad \frac{\partial(\sum_{hkl} E^2_{hkl})}{\partial \delta_{ij}} \quad , \quad \frac{\partial(\sum_{hkl} E^2_{hkl})}{\partial \tau_{ij}}$$

are zero. The sum of the errors squared and the necessary derivatives may then be calculated using equations 3 and 4.

$$(\sum_{hkl} E^2_{hkl}) = \sum_{hkl} \left[|F^c_{hkl}|^2 - |F^o_{hkl}|^2 + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \left(\frac{\partial(|F^c_{hkl}|^2)}{\partial u_{ij}} \epsilon_{ij} + \frac{\partial(|F^c_{hkl}|^2)}{\partial v_{ij}} \delta_{ij} + \frac{\partial(|F^c_{hkl}|^2)}{\partial w_{ij}} \tau_{ij} \right) \right]^2 \qquad \text{Eq. 5}$$

and the derivatives

$$\frac{\partial(\sum_{hkl} E^2_{hkl})_{ij}}{\partial \phi_{rs}} = 2 \sum_{hkl} \left[|F^c_{hkl}|^2 - |F^o_{hkl}|^2 + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \left(\frac{\partial(|F^c_{hkl}|^2)}{\partial u_{ij}} \epsilon_{ij} + \frac{\partial(|F^c_{hkl}|^2)}{\partial v_{ij}} \delta_{ij} + \frac{\partial(|F^c_{hkl}|^2)}{\partial w_{ij}} \tau_{ij} \right) \right]$$

$$\frac{\partial(|F^c_{hkl}|^2)}{\partial \Psi_{rs}} \qquad \text{Eq. 6}$$

where ϕ_{rs} is any error term, ϵ_{rs}, δ_{rs} or τ_{rs} and Ψ_{rs} is its associated variable u_{rs}, v_{rs}, and w_{rs} respectively. Then setting these derivatives to zero yields the equations

$$\sum_{hkl} (|F^o_{hkl}|^2 - |F^c_{hkl}|^2) \frac{\partial(|F^c_{hkl}|^2)}{\partial \Psi_{rs}} = \sum_{hkl} 2 \left[\frac{\partial^2(|F^c_{hkl}|^2)}{\partial \Psi_{ij} \partial \Psi_{rs}} \right] \phi_{ij} \qquad \text{Eq. 7}$$

which are the normal equations needed for the least squares method.

Application to the Refinement of Structures

Thermal parameters, either isotropic or anisotropic, could be associated with each vector in the same manner in which they are associated with individual atoms in the usual refinement technique. The details of including these parameters in the conventional least squares on atomic parameters have been worked out previously[25] and the application of analogous terms to vector refinement is identical. Buerger[26] has shown that equation 1 is the product of two complex Fourier series over the atomic positions of the molecule in its symmetry group. From this it is obvious that the isotropic thermal parameters and the diagonal elements of the anisotropic thermal parameters of the vectors are the sums of the corresponding thermal parameters of the atoms which define the vector. These sums are of course weighted by the atomic numbers of the atoms involved. The off-diagonal elements of the anisotropic thermal parameters tensor reflect the alignment of the thermal ellipsiod of the vector.

It will now be demonstrated that these same normal equations arise from the minimization of the residual, $R = \sum_{hkl} (|F^o_{hkl}|^2 - |F^c_{hkl}|^2)^2$ with respect to the errors in the interatomic vectors, ε, δ and τ. This function has also been minimized with respect to the errors in the atomic positions and used for least squares. First the function $\sum_{hkl}(|F^o_{hkl}|^2 - |F^c_{hkl}|^2)^2$ which will yield the smallest value is expressed in terms of $|F^o|^2$, $|F^c|^2$, and corrections to $|F^c|^2$ which will then improve the fit. It is necessary to assume that these corrections are small. Then

$$\sum_{hkl}(|F^{o}hkl|^2 - |F^{c}hkl|^2)^2 = \sum_{hkl}\left[|F^{o}hkl|^2 - |F^{c}hkl|^2 + 2\left(\frac{\partial(|F^{c}|^2)}{\partial u_{ij}}\varepsilon_{ij} + \frac{\partial(|F^{c}|^2)}{\partial v_{ij}}\delta_{ij} + \frac{\partial(|F^{c}|^2)}{\partial w_{ij}}\tau_{ij}\right)\right].$$

Next the derivatives

$$\frac{\partial(\sum_{hkl}(|F^{o}hkl|^2 - |F^{c}hkl|^2)^2}{\partial\phi_{rs}}$$

are computed and set equal to zero. This yields the normal equations in equations 7.

D. W. J. Cruickshank[27] has shown that if the quantity $|F^{o}|^2 - |F^{c}|^2$ is minimized with respect to the errors in the atomic positions, rather than the interatomic vectors, the least squares corrections yield shifts identical to the shifts which result from the refinement of atomic positions by the method of steepest descent applied to the Patterson density. It will now be shown that the least squares corrections of the interatomic vectors are related to the differential Fourier shifts from the Patterson if the least squares is weighted by the function $1/f_i f_j$.

The refinement criteria for the method of steepest descent is that the slope in the function calculated from the observed amplitudes and the slope in the function calculated from the amplitudes calculated by use of equation 3 be identical. This may be expressed by the equations

$$\left(\frac{\partial P^{o}}{\partial \Psi_{ij}{}^{k}}\right)_{ij} = \left(\frac{\partial P^{c}}{\partial \Psi_{ij}{}^{k}}\right) \qquad k = 1,2,3.$$

where P^o and P^c are the Patterson functions calculated from observed and calculated amplitudes respectively and ψ^k_{ij} is one of the co-ordinates of a Patterson peak, u_{ij}, v_{ij}, or w_{ij}, such that $\psi^1_{ij} = u_{ij}$, $\psi^2_{ij} = v_{ij}$, and $\psi^3_{ij} = w_{ij}$. Both sides of the above equation may be expanded in terms of a full multivariate Taylor series and if the corrections are small only the first derivatives are considered. This yields

$$\left(\frac{\partial P^o}{\partial \psi_{ij}{}^k}\right)_{ij} + \sum_k \phi_{ij}{}^k \left(\frac{\partial^2 P^o}{\partial \psi_{ij}{}^k \, \partial \psi_{ij}{}^m}\right)_{ij} = \left(\frac{\partial P^c}{\partial \psi_{ij}{}^k}\right)_{ij} + \sum_k \phi_{ij}{}^k \left(\frac{\partial^2 P^c}{\partial \psi_{ij}{}^k \, \partial \psi_{ij}{}^m}\right)_{ij}$$

$$+ \sum_{m,r,s} \phi_{rs}{}^m \left(\frac{\partial}{\partial \psi_{rs}{}^m}\right)\left(\frac{\partial P^c}{\partial \psi_{ij}{}^k}\right)_{ij}.$$

Application of the criteria that the slope Patterson density be the same for both the observed and calculated functions yields the equations

$$\left(\frac{\partial P^o}{\partial \psi_{ij}{}^k}\right)_{ij} = \left(\frac{\partial P^c}{\partial \psi_{ij}{}^k}\right)_{ij} + \sum_{m,r,s} \phi_{rs}{}^m \frac{\partial}{\partial \psi_{rs}{}^m} \left(\frac{\partial P^c}{\partial \psi_{ij}{}^k}\right)_{ij} .$$

The calculated Patterson density, P^c, is of course a function of the parameters, ψ^k_{ij}. Therefore, $P_{ij}{}^c$ shall represent the calculated Patterson density due to the trial vector ij and its symmetry related vectors. Thus the previously derived equations may be rewritten as

$$\left(\frac{\partial P^o}{\partial \psi_{ij}{}^k}\right)_{ij} = \left(\frac{\partial P^c}{\partial \psi_{ij}{}^k}\right)_{ij} + \sum_{m,r,s} \phi rs^m \frac{\partial}{\partial \psi_{rs}{}^m} \left(\frac{\partial P^c{}_{rs}}{\partial \psi_{rs}{}^k}\right)_{ij}$$

Then if h^k is any index, h, k or l, the coefficient for the error terms may be written as,

$$\frac{-2\pi}{V} \sum_{h,k} h^k \frac{\partial(|F^c|_{ij}^2)}{\partial \psi_{ij}{}^k} \sin 2\pi \left(\sum_{k=1}^{3} h^k \psi_{ij}{}^k\right)$$

where $|F_{ij}{}^c|^2$ is the component of $|F^c|^2$ resulting from the peak at ij. They by inspection of Eq. 7 and Eq. 3 it may be seen that the coefficient of the errors in the least squares method is given by

$$-8\pi \sum_{h,k} h^k f_i f_j \frac{\partial(|F^c{}_{ij}|^2)}{\partial \psi_{ij}{}^k} \sin 2\pi \left(\sum_{k=1}^{3} h^k \psi_{ij}{}^k\right)$$

Further the column vector associated with the least squares method is

$$-8\pi \sum_{h,k} h^k f_i f_j (|F^o|^2 - |F^c|^2) \sin 2\pi \left(\sum_{k=1}^{3} h^k \psi_{ij}{}^k\right)$$

A modified Patterson density P'_{ij} is now introduced, defined such that

$$P'_{ij} = \frac{1}{V} \sum_{h,k,l} f_i f_j |F^o|^2 \cos 2\pi (hu + kv + lw) \ldots$$

By taking advantage of the center of symmetry the following equations may be derived.

$$\left(\frac{\partial P'^o}{\partial \psi_{ij}{}^k}\right) = \frac{-2\pi}{V} \sum_{h} h^k f_i f_j |F^o|^2 \sin 2\pi \left(\sum_{k=1}^{3} h^k \psi_{ij}{}^k\right)$$

$$\left(\frac{\partial P'^c}{\partial \psi_{ij}{}^k}\right) = \frac{-2\pi}{V} \sum_{h} h^k f_i f_j |F^c|^2 \sin 2\pi \left(\sum_{k=1}^{3} h^k \psi_{ij}{}^k\right)$$

$$\frac{\partial}{\partial \psi_{ij}{}^k} \left(\frac{\partial P'^c{}_{ij}}{\partial \psi_{rs}^m}\right)_{rs} = \frac{-2\pi}{V} \sum_{h} h^k f_i f_j \frac{\partial(|F^c|^2)}{\partial \psi_{ij}{}^k} \sin 2\pi \left(\sum_{k=1}^{3} h^k \psi_{rs}{}^k\right)_{rs}$$

Thus the normal equations of the least squares method may be written

$$\left(\frac{\partial P'^{o}}{\partial \Psi_{ij}{}^{k}}\right)_{ij} - \left(\frac{\partial P'^{c}}{\partial \Psi_{ij}{}^{k}}\right)_{ij} = \sum_{r,s,m} \phi^{m}_{rs} \frac{\partial}{\partial \Psi_{rs}{}^{m}} \left(\frac{\partial P'^{c}{}_{rs}}{\partial \Psi_{ij}{}^{k}}\right)_{ij} \quad .$$

The above derivation shows that for small corrections the method of least squares applied to interatomic vectors, as derived in equations three through seven is equivalent to a weighted differential Patterson or steepest descents method in Patterson space. The required weighting function for the steepest descents or differential Patterson function is $f_i f_j$. This is equivalent to weighting the least squares by a factor of $1/f_i f_j$.

The difficulties which arise in the practical application of the method result from the large number of terms to be refined. For a structure with n independent atoms there are nx(n-1)/2 independent interatomic vectors to be refined. Thus for a compound with thirty atoms, there would be 1305 positional parameters to refine. It would therefore be necessary to take a complete set of data to have sufficient data to make least squares valid. However because of the development of automated diffractometers this is no longer such a formidable or tedious task.

The second major practical problem is the inversion of a matrix of this size. The recent advances in computer technology would seem to make the time required feasible. However the possibilities for round off or truncation error are almost limitless for a matrix of this size. This problem could be overcome by careful planning. Therefore there would not seem to be any insurmountable problems to the application of the vector refinement method to crystal structures.

Application of the Vector Refinement to the Solution of Crystal Structures

As discussed previously it is necessary to have initial estimates of the interatomic vectors before the refinement can be undertaken. It is perhaps useful at this point to examine the means available for obtaining these initial vectors and the difficulties associated with the various means.

The most obvious means of obtaining initial estimates of the interatomic vectors is through the use of the Patterson vector density map. If this can be done, and the estimates refined into a set of discrete point vectors by the least squares or the nearly equivalent differential Fourier method, it is then a simple and routine matter to solve[28,26b] for the co-ordinates of the atoms which give rise to this vector set. Thus the refinement becomes a general method for the solution of crystal structures from the Patterson function.

The major difficulty in using this method of structure analysis is of course the overlap in the Patterson function. A practical effect of this overlap is that the method will be of little use in the solution of structures containing a heavy atom. This is because these structures are readily solved using only the easily recognizable subset of interatomic vectors between the heavy atom and the other atoms or by Fourier methods which directly reveal the positons of the atoms in the unit cell. Therefore it is expected that the method will find its widest application in the solution of structures composed of atoms of approximately equal atomic number. That is, the so called "equal atom" structures.

It has been established[27] for the least squares corrections which are realted to corrections by the differential Patterson method in the above manner that the cross terms in the least square matrix between Fourier density peaks which are resolved are negligibly small compared to the terms in the matrix block which relate to only one atom. However the cross terms between those Fourier peaks which are overlapped with one another are not negligible. In this instance the peaks in the Fourier density are the locations of the ends of the interatomic vectors. Also it is well known that the major difficulty in solving crystal structures from the Fourier map of the interatomic vector, the Patterson function, is the fact that these peaks are overlapped[26c] for structures of even moderate complexity. Thus while the common block diagonal approach, often used in the refinement of atomic positions, would be inadequate, the use of the complete least squares matrix would not be required. What would be required is a matrix composed of blocks each of which would contain the complete matrix of elements for a group of overlapped peaks.

There are several steps which may be taken to overcome the overlap problem in an "equal atom" structure. The most basic is to take as complete a set of data as is possible under a given experimental arrangement. This increases the resolving power of the Patterson function as well as decreases the error due to series termination effects. Also this step will reduce the error introduced by some of the other means of resolving overlap.

Other means of resolving the overlap involve mathematical manipulation of either the Fourier amplitudes, $|F^o|^2$, used to calculate

the Patterson density or a manipulation of the density map itself. The most obvious of these techniques is to use "sharpened" Fourier amplitudes. This sharpening is accomplished with a weighting scheme which gives heavier weights to the high order amplitudes which have a shorter periodicity. Many different sharpening schemes have been proposed[29] and one could be picked which offered the most promise for a particular compound. For an equal atom problem it is possible to sharpen to the extent that the coefficients correspond to those of a structure of point atoms at rest.[30] Sharpening the coefficients has the regrettable effect of also increasing the error in the vector density map caused by series termination effects. Extreme sharpening may even result in false peaks appearing in the vector map.

In general, extension of the range of the data and sharpening of the Patterson coefficients will not resolve all of the overlap in the Patterson density map. In order for the least squares method to work, all of the vectors must be resolved and refined. Thus further steps must be taken to locate, with the necessary accuracy, the vectors contained in the overlapped regions of the vector map. It would seem possible to approach this in a systematic manner.

If the data are placed on the absolute scale then the numbers produced in the Patterson map have physical significance. These numbers represent the average of the products of the electron densities of all points in the crystalline unit cell separated by that vector. The integral of this function over the volume of the peak is then the sum of the products of the atomic numbers of the atoms in the crystalline unit cell which are separated by vectors which fall within the volume of

the peak integrated. This then yields the number of vectors under the overlapped peak, assuming again the "equal atom" compound. Further it is possible to check this number of vectors since it is possible to determine, from density, unit cell and molecular weight measurements, the number of atoms within a crystal unit cell and hence the total number of vectors that will be generated.

Furthermore the theoretical density of a Patterson peak may be calculated using simple functions, since spherically symmetrical atoms might be assumed. Then use could be made of the fact that the overlapped peak is the sum of the previously determined number of single vectors. The use of this fact should make it quite straight forward to develop methods for placing single peaks to build the composite peak. The most obvious means of initial placement is through extension of the techniques used in conventional absorption spectroscopy. If suitable criteria could be developed these placements could be initially refined to give the best fit of the overlapped region before the vectors were used in the least square scheme. The criteria which are most obvious are those of the differential Fourier refinement method discussed earlier. Namely that the derivatives of the observed and calculated densities with respect to the co-ordinates u, v, and w be the same at the vector locations. Because of the known overlap further criteria such as the equality of the second derivatives with respect to the coordinates probably should be employed. Once this had been done then the approximate vectors could be used as input for the least squares refinement.

Finally as a result of the progress made in recent years in the programming of solution of nonlinear equations it is possible to extend the range of initial error allowed in the least square and differential Fourier methods. This is done by including second and third derivative terms in the Taylor series expansion of the function. This results in a polynomial in the errors rather than the usual normal equations. However by requiring that the solutions to these equations be small and real the right set of corrections may be found. Eichhorn[44] has shown how these terms may be included in the usual least squares and differential Fourier methods of X-ray diffraction. The extension to the vector space refinement producures proposed here is obvious. However this expedient should not prove necessary if the initial fit of the vectors to the overlapped peaks is sufficiently accurate.

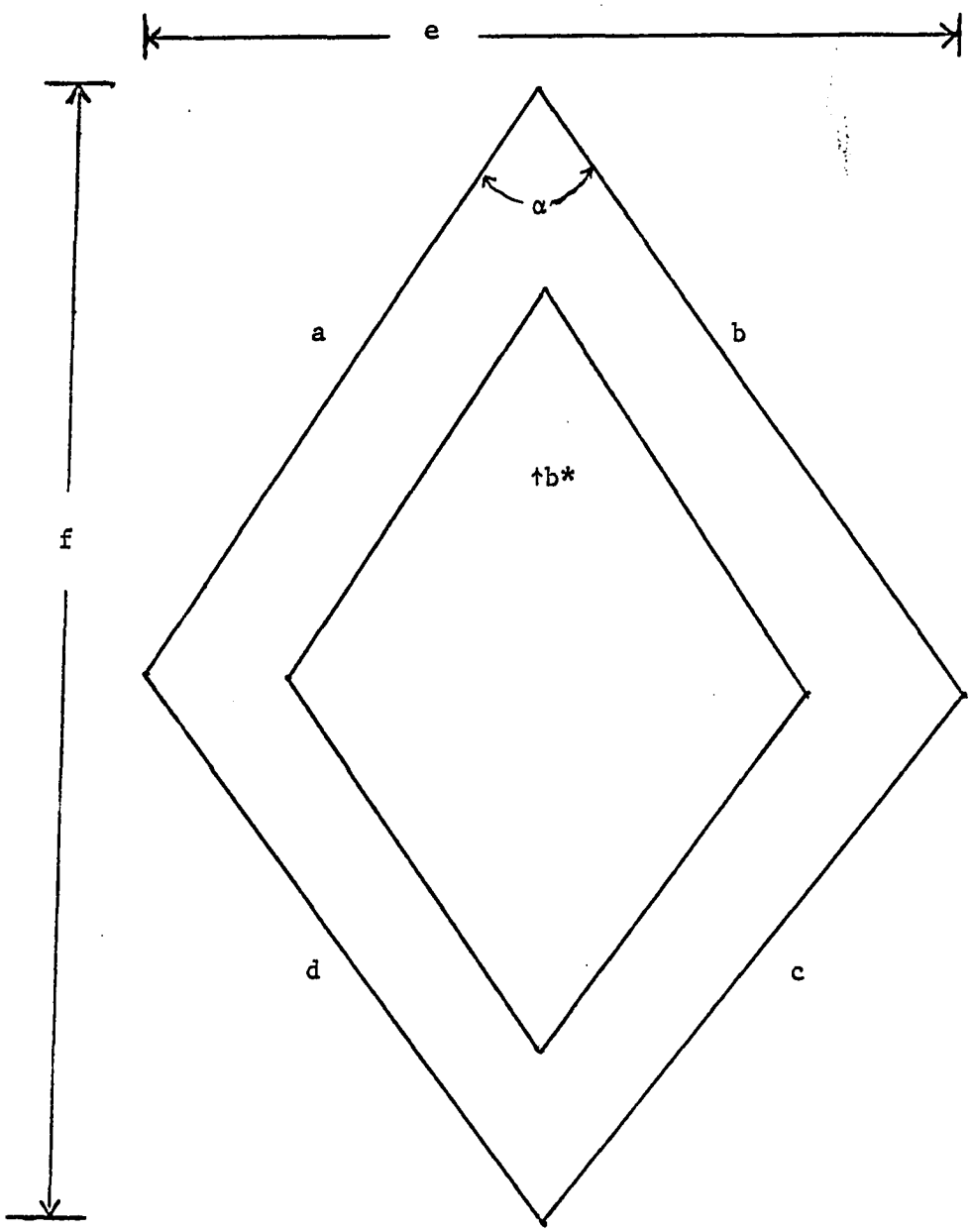
PART III

ATTEMPTED SOLUTIONS OF THE CRYSTAL STRUCTURE OF N-ACETYL PROLINE MONOHYDRATE

The introduction of an l-proline residue into a protein chain causes a gross distortion of many of the commonly occurring secondary and tertiary structures of these molecules. The introduction of an l-proline residue into a right-handed α-helix forces the helix to change direction by at least 35 degrees and disrupts two of the stabilizing hydrogen bonds.[43] Neither can a proline residue be introduced into either the parallel or anti-parallel pleated sheet arrangements. This is because the tertiary amide cannot enter into hydrogen bonding. Also the substitution of $-CH_2-$ for the hydrogen of the amide group results in severe distortion caused by van der Waal's forces. Finally none of the attempts to build a model for the structural protein, collagen have been completely successful. This may be attributed, in large part, to the necessity to include a high percentage of proline and hydroxy-proline resides[45] in the protein sequence.

For these reasons it was decided to investigate the geometry of the l-proline molecule with the nitrogen in the amide form. Acetyl-l-proline was chosen rather than glycyl-l-proline to eliminate the effects of the polar ammonium ion on the conformation of the molecule.

Figure 1. N-Acetyl-l-Proline-Monohydrate Data Crystal



Main Face = $\bar{1}$ 0 1

a = 0.328 millimeters d = 0.293 millimeters

b = 0.345 " e = 0.345 "

c = 0.345 " f = 0.517 "

α = 70 degrees

Large crystals of N-acetyl-l-proline monohydrate were grown by evaporation of an aqueous solution. The data crystal is shown in Figure 1. The only systematic absences were for 0k0 with k odd. This indicated either P_{2_1} or $P_{2_{1/m}}$. Since $P_{2_{1/m}}$ cannot accomodate an l absolute configuration without the simultaneous presence of the d absolute configuration, P_{2_1} was chosen as the space group. The unit cell dimensions are:

$$a = 6.61 \text{ Å}$$

$$b = 10.69 \text{ Å}$$

$$c = 6.66 \text{ Å}$$

$$\beta = 108.89^{\circ}$$

The experimental density was determined to be 1.301 gm/ml, while that calculated for $C_7O_3NH_{11}.H_2O$ N-acetyl-l-proline monohydrate is 1.306 gm/ml. There are two molecules per unit cell.

The intensities were measured with nickel filtered copper $K_{-\alpha}$ radiation using the theta-two theta scan technique. 860 independent reflections were measured out to a two theta of one hundred forty degrees. Of these only 23 were too weak to be observed. Lorentz-polarization and absorption corrections were applied to the data. The linear adsorption coefficient was 4.106 cm^{-1}.

A wilson plot was prepared. A visual fit yielded values of 192 for K, the scale constant, and 4.50 for B, the temperature factor. These values were used to compute the sharpened structure factor coefficients U and E. (See appendix B.)

Identity of the Compound

As previously noted the observed density agreed excellently with the density calculated from the unit cell constants and the molecular

weight of N-acetyl-l-proline monohydrate. However proline is more difficult to acylate than are other amino acids. Therefore it was necessary to investigate the possibility that the crystals were actually l-proline acetate, the acetic acid salt of l-proline.

The only reference to the melting point of N-acetyl-l-proline did not indicate whether the hydrated or unhydrated form was reported.[49] The reported compound had been precipitated from an aqueous solution by acid and recrystallized from acetone. The reported melting point is 118°. The crystals on which the data was taken melted at 82°. Recrystallization from acetone did not change the melting point.

A search of the literature on identification of organic compounds failed to yield any compound with a suitable molecular weight and melting point. The melting point of l-proline acetate could not be found.

A proton magnetic resonance spectrum was taken on this compound. A drawing of this spectrum is shown in Figure 2a. The number in parentheses above each peak is the relative area under that peak. If the smallest peak is assumed to be caused by the proton attached to the α- carbon atom of the proline, then the numbers also represent the number of protons which resonate at the frequency at which the peak is found. The peaks have also been lettered. The peak a, as stated above, had been assigned to the single proton on the α-carbon. The peak b has been assigned to the two protons on the δ-carbon atom of the proline residue. Because proline is a cyclic imino acid the δ-carbon is adjacent to the nitrogen atom. This fact explains the location of this peak. The peak c has been assigned to the three hydrogens of the acetyl

Figure 2a. Proton Magnetic Resonance Spectrum of N-Acetyl-l-Proline Monohydrate.

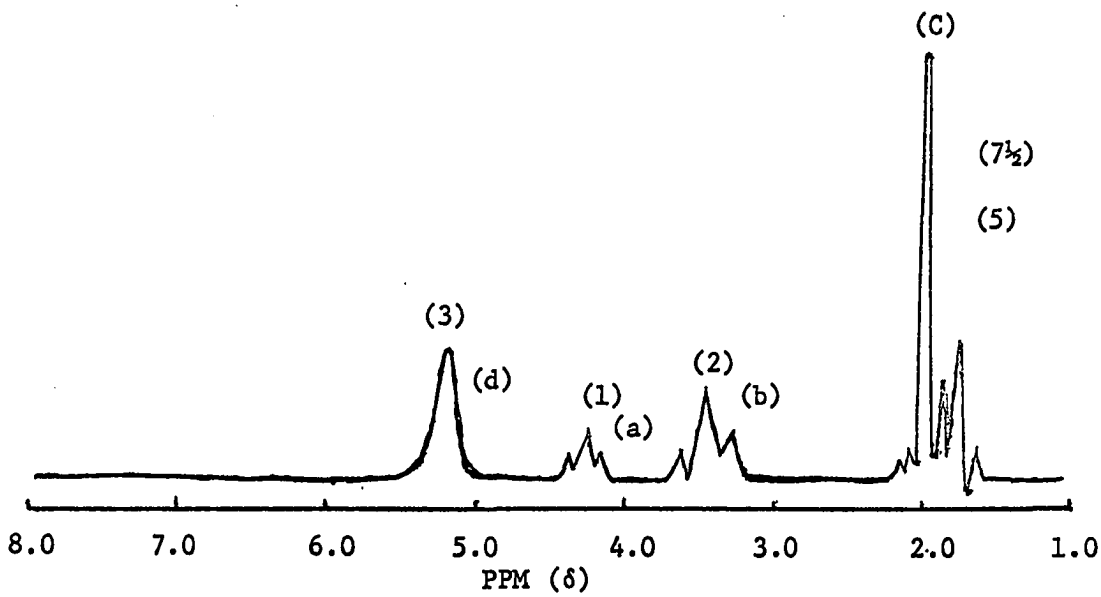
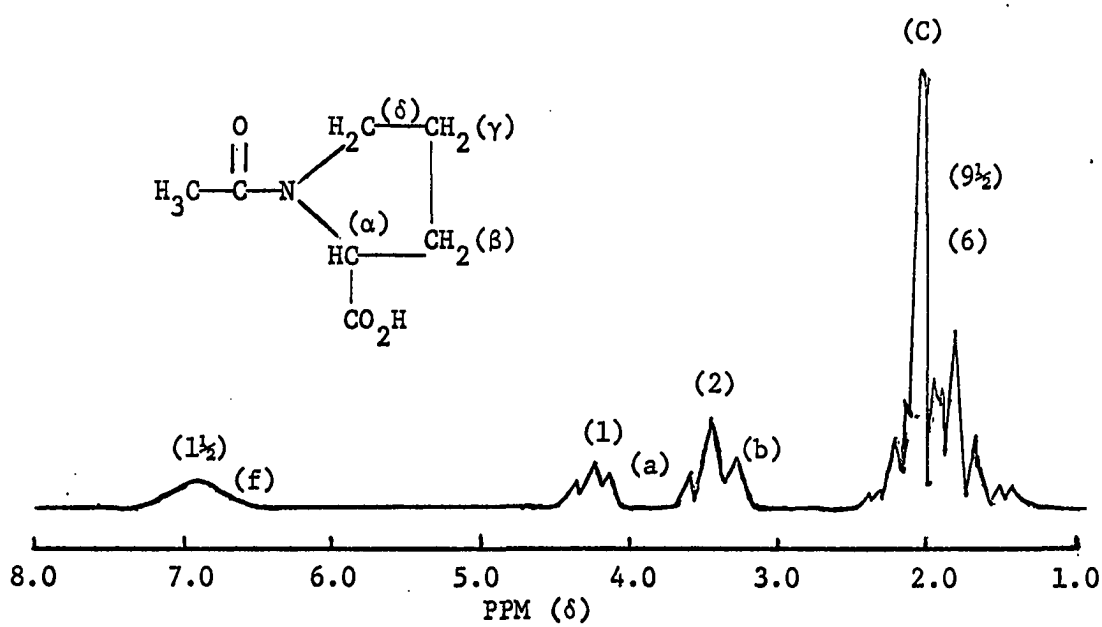


Figure 2b. Proton Magnetic Resonance Spectrum of N-Acetyl-l-Proline after drying.



group and the four hydrogens of the β and γ carbons of the proline residue. The sharp single spike in peak c has been assigned to the three hydrogens of the acetyl group. The second number in parentheses by peak c is the integral, starting from the left, of peak c up to and including this spike. The peak d has been assigned to the acid hydrogen of the acetylproline molecule and to the two hydrogens from the water of hydration. The addition of deuterium oxide to the sample caused this peak to disappear. It was replaced by a sharp single peak which overlapped part of peak b. This clearly showed these three hydrogens to be exchangeable.

However it is also possible to assign the peaks to the protons in proline acetate. In order to resolve this ambiguity 0.0346 grams of the compound were dried in a vacuum desicator over P_2O_5 at 60°C for twenty hours. This yielded 0.0335 grams of dried compound. This product melted at 117-118°C, the recorded melting point of N-acetyl-l-proline. A proton magnetic resonance spectrum of this dried compound was taken. It is shown in Figure 2b.

Peaks a, b, and c in this spectrum have been assigned to the same protons as were discussed above. Peak f has been assigned to the acid hydrogen of acetylproline. The extra protons under peaks f and c have been attributed to contamination of the solvent by water and non-deutrated aceto-nitrile. The solvent was deutrated aceto-nitrile for both spectra, however the solvent was from a different source in each case, which is the reason for the different number of extra protons in the two spectra.

To test the possibility of solvent contamination for the

deutrated aceto-nitrile used to run spectrum 2b, a blank was run. This was done by mixing some of the solvent with carbon tetrachloride. The region in peak c showed the presence of protons in this spectrum also. The addition of a small amount of water to this sample caused part of the protons in this region of the blank spectrum to shift. This indicates that part of these protons were water protons. The rest were methyl protons from non-deutrated aceto-nitrile.

The dried sample on which the spectrum was run was recovered and its melting point checked. The sample melted over a lower and wider range than previously. This is further evidence for the postulated contamination of the solvent by water.

The proton magnetic resonance spectra clearly indicates that the drying process removes water rather than acetic acid from the compound. If acetic acid were being removed the spectra should reveal the loss of one proton from the d peak and three protons from the c peak. Further the three protons lost from the c peak should be the protons represented by the large, sharp spike so that the loss would be obvious, even in the presence of impurities in the solvent. If water were lost the spectra would be expected to show the loss of two of the active protons of peak d to form peak f. There does not appear to be any loss from peak c. It is, therefore, fairly certain that water rather than acetic acid is being lost. Thus these spectra, combined with the close agreement of the observed and calculated density and the melting point of the dried compound, show the identity of the compound to be N-acetyl-l-proline monohydrate.

Attempted Solutions of the Structure

The first attempt to solve the structure was by the use of the Patterson function. This attempt was based on the fact that for space group P_{2_1} the section x, 1/2, z is a Harker[9] section of the Patterson function. This Harker section results from the symmetry operation of a two fold screw axis, which may be written x,y,z → -x, 1/2+y,-z. Therefore the vector between an atom and its symmetry related atom has coordinates 2x, 1/2, 2z. Thus the Harker section shows a projection of the molecule down the b axis. The interatomic distances in this projection are, as can be seen from the form of the Harker vectors, twice those of the real molecule.

Several orientations of the five membered ring of the proline molecule that would fit the Harker section could be found. Only a few of these could be extended to the complete N-acetyl-l-proline molecule. These trial solutions were compared with the complete Patterson function. Most of them did not fit the Patterson function. Structure factors and Fourier electron density maps were computed for those which most nearly matched the Patterson function. The structure factor and Fourier calculations showed these trial solutions to be wrong.

The failure of the Harker section to reveal the structure may be attributed to the main difficulty in solving all structures from their Patterson maps, the severe overlap of the vectors. In this particular Harker section only nine distinct peaks can be located. There should be twelve, one for each atom in the asymmetric unit. Of these nine, five are overlapped with each other in two groups, one of two peaks which may be identified and a second group with three peaks.

The height of the entire region of this second group is greater than the maximum height of any single peak. This makes it hopeless to try to use these peaks to place atoms, therefore the attempts to solve the structure from the Patterson function were abandoned.

An attempt was made to solve the centrosymmetric ac projection of the crystal structure by use of the Harker-Kasper[34] inequalities. However the unitary structure factors, U, associated with the h0l reflections were not large enough to determine a sufficient number of phases. A Fourier calculated by including the phases which almost but not quite satisfied an inequality yielded no useful information.

All further attempts to solve the crystal structure of N-acetyl-l-proline monohydrate were based on the symbolic addition procedure of Hauptman and Karle.[30,39,50,51] The use of this method requires that the origin be specified by assigning phases to appropriate structure factors.[50] In space group P_{2_1} the appropriate structure factors must include two with indicies h0l and one with indicies hkl with $k \neq 0$. Further the two h0l structure factors must be of different parity with respect to the eveness or oddness of h and l. Neither can they be of the type even, 0, even.

The necessity of the parity requirement may be seen by examining the structure factor equation for the h0l reflections. This equation may be written

$$F_{h0l} = 2 \cos 2\pi(hx+lz).$$

The origin, in this projection is on one of the four points; 0,0,0; 1/2, 0; 0, 1/2; or 1/2, 1/2. Thus the specification of the origin on this projection is a matter of picking one of these four points. The

different parity groups change sign in different manners in going from one point to another. It can be seen that the parity group even, 0, even does not change sign while changing origin. Thus it cannot be used to specify the origin. The sign changes are given in the following table.

Parity Group	0,0	1/2,0	0,1/2	1/2,1/2
e,0,e	+	+	+	+
e,0,e	+	+	-	-
o,0,e	+	-	+	-
o,0,o	+	-	-	+

Thus any combination of relative sign changes corresponds to a permissible origin in the hOl projection.

Examination of the general structure factor equations for P_{2_1} reveals that the arbitrary assignment of a phase angle to a subsequent general reflection corresponds physically to choosing arbitrarily the y coordinate of the center of gravity of the contents of the asymmetric unit. The equations are for k=even

$$A = 2 \cos 2\pi (hx + lz) \cos 2\pi ky$$

$$B = 2 \cos 2\pi (hx + lz) \sin 2\pi ky$$

for k = odd

$$A = -2 \sin 2\pi (hx + lz) \sin 2\pi ky$$

$$B = 2 \sin 2\pi (hx + lz) \cos 2\pi ky$$

where $F = A + iB$.

The final choice of the symbolic addition procedure to space group P_{2_1} is the choice of the enantiomorph. That is the choice between a right or left handed structure. Of course when the choice is made

it is not known whether the left or right handed structure has been chosen. Karle and Hauptman[50] state that this choice is made by arbitrarily choosing the sign of any structure invariant whose vectorial value is not either zero or pi. The choice of the positive sign for the phase angle α is equivalent to the choice of an enantiomorph with coordinates (x, y, z) the choice of minus α is equivalent to the choice of the opposite enantiomorph with coordinates (-x, -y, -z). A structure invariant is a structure factor whose vectorial value is independent of the choice of origin. Hauptman and Karle[50] have shown that the structure invariants in P_{2_1} have the indicies of the parity even, 0, even. Thus in P_{2_1} there are no structure invariants which do not have a value of either zero or pi.

In practice in P_{2_1} the enantiomorph is chosen by assigning to a fourth structure factor a symbolic value, represented by a letter. This structure factor is a general structure factor so that it is expected that its phase is not either zero or pi. Phase determination may now proceed using the three arbitrarily assigned phases and the symbol.

Phase determination[39] is initially carried out using what is known as the Σ_2 relationship, which is,

$$\phi_h = \phi_k + \phi_{h-k}.$$

ϕ_h is the phase of the structure factor associated with hkl indicies h. This process is continued until an initial five to ten per cent of the data have been assigned phases in terms of the initially assigned phases and the symbol. This symbol is then given several values between zero and pi. The value which leads to the most self-consistent set is

chosen as the correct value. The negative values between zero and minus pi are not considered because they correspond to the choice of the opposite enantiomorph.[51] Three different initial sets were used in the attempts to solve the structure of N-acetyl-l-proline monohydrate. They are listed below.

SET I

1 0 1 = o(+)

$\bar{1}$ 0 4 = 0(+)

1 9 $\bar{2}$ = 0(+)

2 1 2 = a

SET II

1 0 1 = 0(+)

$\bar{1}$ 0 4 = 0(+)

2 1 2 = $\pi/2$

$\bar{1}$ 10 1 = a

SET III

$\bar{7}$ 0 2 = 0(+)

1 0 1 = 0(+)

$\bar{2}$ 9 1 + 0(+)

4 6 2 = a

$\bar{3}$ 8 3 = b

It will be noticed that two symbols were used in the third set of initial phases. This was necessary because it was not possible to reach all of the structure factors with the Σ_2 relationship using only the four initially assigned phases. This symbol must be allowed to assume all values between minus pi and plus pi. In this case however

the application of the Σ_2 relationship revealed that b = 3a. Thus this relationship was used in assigning values to the symbols.

After the phases are assigned they are refined using what is known as the tangent formula.[39] This formula is:

$$\tan \phi_h = \frac{\sum_k |E_k E_{h-k}| \sin (\phi_k + \phi_{h-k})}{\sum_k |E_k E_{h-k}| \cos (\phi_k + \phi_{h-k})}$$

This equation is the equivalent of the Σ_2 relationship in terms of the vectorial components. This formula was then used, as described in appendix B, to assign further phases. This was continued until the number of assigned phases was ten to fifteen times the number of atoms present in the structure.

At this point Fourier maps were calculated using the assigned phases and the sharpened structure factors, E, as coefficients. These maps were investigated for features of the N-acetyl-l-proline molecule. Several of these features were found in each of the maps. Attempts were made to complete these structures.

These attempts fell into two catagories. The first was the traditional structure factor-Fourier cycle. The atoms found in the E maps, the Fouriers computed using the results of the symbolic addition process, were used as input for a structure factor calculation. Those structure factors which calculated in good agreement with their observed values were used as input for a Fourier calculation. New atoms were located in several of these Fourier maps. However it was impossible to obtain a trial solution which would refine.

The second method used in attempts to correct and complete the initial set of atoms involves the further use of the symbolic

addition procedure.[52,53] The first step involves the calculation of the structure factors using the atoms located in the initial E map. Then phases are assigned to the structure factors associated with high E values. The phases are assigned only if $X*|F_c|>|F_o|$ where X is the fraction of the scattering matter used in the structure factor calculation $|F_c|$ is the absolute value of the calculated structure factor and $|F_o|$ is the absolute value of the observed structure factor. These assigned phases are then used to initiate the phase generation using the tangent formula. When sufficient phases are generated another E map is calculated. This process also resulted in changes and additions to the initial structures. Once again these trial structures could not be refined. In a few cases the process was repeated several times.

Reasons for the Failure of the Symbolic Addition Procedure

The most plausible reason for the failure of the symbolic addition procedure to solve the structure is found in its statistical nature. As explained in Appendix B the general inequality from which the tangent formula and the $\Sigma 2$ relationship are derived places no absolute restriction on the phase of a structure factor in a complex structure. However an analysis of the probability relations between the phase and the structure factor magnitude indicates that the averages or summations used in the tangent formula and the $\Sigma 2$ relation give the most probable phase.[39]

This inequality may be expressed in terms of the unitary structure factors, U, as

$$\left|U_h - \frac{U_k U_{h-k}}{1}\right| \leq \frac{\begin{vmatrix} 1 & U_k^* \\ U_k & 1 \end{vmatrix}^{1/2} \begin{vmatrix} 1 & U^*_{h-k} \\ U_{h-k} & 1 \end{vmatrix}^{1/2}}{1}$$

The three largest unitary structure factors are the -6, 6, 3 with U = 0.495; -4,5,2, U = 0.455; and -7, 0, 2, U = 0.391. Even if these three structure factors were related to each other in the manner required for the application of this inequality they are not large enough to constrain the phase. The result of this is that the acceptance of any phase is dependent only on its consistency with the set from which it was generated and others generated by that set.

When the first phases are generated the process is not statistically proper. That is there are not enough known phases so that new phases are the result of an averaging process. They are determined by the indication from a single pair. This means that there is a high probability that the phase is incorrect. If several of the first phases generated are incorrect by a large amount then the tangent formula will propagate this error and it is likely that the phases will converge to a self-consistent incorrect set. This leads to a failure of this method.

A second source of trouble in P_{2_1} is the fact that it is a polar space group. Because of this the origin is not uniquely defined in the b direction. It is defined only by the phases generated by the symbolic addition procedure. Each of these phases will be incorrect by some amount. This deviation is small and random for sets of phases which lead to correct structures. However in a polar direction each of

these erroneous phases has defined a different origin. This problem is compounded by the small number of structure factors used in calculating the E map. This could lead to a disordering of the structure in the polar direction. This disordering, if systematic, could cause the convergence of some atoms into a single peak and the divergence of others so that bonded pairs would not be recognized.

This lack of a well defined origin, in the polar direction, could offer an explanation of a disconcerting phenomenon which was observed in many of the calculated E maps. This was the occurrence of a peak whose height was at least three times that of the next highest peak. In all but one of these cases this peak occurred at the x and z coordinates predicted by the highest peak in the Harker section of the Patterson function. The y coordinate of a single peak is, of course, not obtainable from the Patterson function for P_{2_1}.

These phenomena would ordinarily be interpreted by the presence of a heavy atom. Potassium or chlorine were the most likely from the apparent electron density. Chemical tests failed to disclose the presence of either of these species. Least square refinement of this proposed atom tended to disperse the high electron density. The Fourier based on phases generated by this "atom" failed to reveal the structure. At least part of this structure should have been apparent if indeed the heavy atom were present. Thus it was concluded that this was not a heavy atom.

The high Patterson peak may be explained either by the presence of several atoms with almost the same x and z coordinates or by the presence of non-symmetry related atoms with a difference of 1/2 in the y coordinate. It is probably caused by a combination of both. If there

are several atoms with the same x and z coordinate then the convergence phenomena postulated above could explain the occurrence of the high peak in the E maps.

It is unfortunate that this hypothesis cannot be tested until the structure is solved. It is, however, worth noting that space group P_{2_1} has a hearsay reputation for causing more difficulty with the symbolic addition procedure than other non-centrosymmetric space groups. Perhaps this is because most of the other non-centrosymmetric space groups, in which symbolic addition has been tried have been non-polar, such as $P_{2_12_12_1}$ or $P_{22_12_1}$.

The Occurrence of False Mirror Planes

There is yet one type of incident which occurred during these attempts which should be recorded. This is that the E maps calculated from the phases generated by two initial sets contained mirror planes. These sets are listed below and are distinct from the initial sets listed earlier.

SET IV	SET V
1 0 1 = 0	1 0 1 = 0
-1 0 4 = 0	-1 0 4 = 0
-1 9 2 = 0	-1 9 2 = 0
-1 10 1 = a	-2 10 2 = a

This mirror plane was parallel to the xz plane and had the same coordinate as the large peak which was discussed earlier. However the mirror plane was not present in all of the E maps which contained the large peak. In these E maps the large peak had a regular octahedral surrounding of nearest neighbors.

At the time this occurred it was thought to be caused by the choice of a reflection with k even for the selection of the enantiomorph. However subsequent events have proved this hypothesis incorrect. The first of these events is the publication of the crystal structure of the alkaloid panamine.[51] This molecule also crystallizes in space group P_{2_1} and the enantiomorph was chosen with the reflection 5, 2, 8 which has k even. Further some of the sets which yielded trial solutions for these attempts on N-acetyl-l-proline used reflections with k even to select the enantiomorph. Set II even used the reflection -1, 10, 1. None of these E maps contained the mirror plane. Therefore it does not appear that the use of reflections with k even caused the mirror planes. These false mirror planes were not investigated further, and their cause is, at present, unknown.

PART IV

SUMMARY AND CONCLUSIONS

The crystal and molecular structure of diaquo zinc(II) glycylglycylglycinato hemisulfate dihydrate was solved using classical Patterson-heavy atom methods. The zinc ion and two light atoms were located from the Patterson map. These three atoms were their own mirror image across the false mirror plane at y=0. This mirror plane was a natural consequence of the symmetry of the Patterson function. Successive cycles of structure factor-Fourier electron density map calculations revealed the crystal and molecular structure.

The zinc ion was revealed to be one of a small but growing group of metal ions with a well established pentaco-ordination. The co-ordination geometry is best described as a moderately distorted trigonal bipyramid. The zinc is chelated through the amine terminal nitrogen and the first carbonyl oxygen of the peptide. Also the zinc ion is complexed to the carbonyl group of another peptide molecule.

This compound is the first trigonal bipyramid surrounding which has been found for a zinc ion with ligands with a biological nature. This geometry offers a possible structural explanation of the inactivity of several zinc containing enzymes.

The bond distances and angles of the peptide molecule are generally within the range of normally reported values. The conformation of the peptide chain is also well within the normally allowed range. Probably the only important deviation in the peptide chain is the slight but significant shortness of the $C'_2=O_2$ bond. This shortness

has also been noted in the structures of the copper chelate[46] and the calcium complex[47] of glycyl glycyl glycine. This shortness has been attributed to the weak and non-cumulative interactions which are found in these structures. If this explanation is then valid then these different dimensions may be useful in constructing theoretical models of the structural protein collagen. This is because infrared studies[45] indicate that this protein does not form the almost infinite chain of hydrogen bonds found in other structural proteins. In fact collagen is apparently a very poor hydrogen bond former, a characteristic shared by zinc triglycine.

It is known that the process of least squares on atomic coordinates on polar space groups leads to errors in the estimated standard deviations of these coordinates. This problem may be overcome by the least squares refinement of the interatomic vectors rather than the atomic coordinates. This process also has the advantage that it yields directly estimated standard deviations of the bond lengths in a structure. Since it is possible to obtain approximate interatomic vectors directly from the Patterson function, the refinement technique also may be used to solve unknown structures. The method should be most useful in solving "equal atom" problems, the type of structure which is most difficult to solve using current methods.

Finally the various attempts to solve the crystal structure of N-acetyl-l-proline monohydrate have been presented. Possible reasons for the failure of methods employed to solve this structure have been proposed. The presentation of these unsuccessful attempts serves to emphasize the difficulty in solving the above "equal atom" structures with current techniques.

REFERENCES

1. Malmström, B. G., Nyman, P. O., Strandberg, B., Triander, B., The Structure and Activity of Enzymes, Goodwin, T. W., Harris, J. I., Hartley, B. S., eds., (Academic Press, New York, 1964), p. 121.

2. Vallee, B. L., Riordan, J. F., Colman, J. E., Proc. Natl. Acad. Sci., 99, 109 (1963).

3. Linskog, S., Nyman, P. O., Biochim. et Biophys. Acta, 85, 462 (1964).

4. Plocke, D. J., Vallee, B. L., Biochemistry, 1, 1039 (1962).

5. Freeman, H. C., Advances in Protein Chemistry, Vol. 22 Anfinsen, C. B., Anson, M. L., Edsall, J. T., Richards, F. M., eds., (Academic Press, New York, 1967) p. 258. (b) Ibid., p. 337. (c) Ibid., p. 341.

6. Flascha, H. A., EDTA Titrations, (Permagon, New York, 1959).

7. Ayres, G. H., Quantitative Chemical Analysis, (Harper, New York, 1958).

8. Edsall, J. T., Flory, P. J., Liquori, A. M., Nemethy, G., Ramachandran, G. N., Scherage, H. A., Biopolymers, 4, 121 (1966).

9. Harker, David, J. Chem. Phys., 4, 381 (1936).

10. Wilson, A. J. C., Acta Cryst., 3, 397 (1950).

11. Patterson, A. L., Zietschrift fur Krist. (A), 90, 517 (1935).

12. Pauling, L., Shappell, M. D., Zietschrift fur Krist. (A), 75, 128 (1930).

13. Patterson, A. L., Nature, 143, 939 (1939).

14. Ahmed, F. R., NRC Crystallographic Programs for IBM/360 System.

15. Patterson, A. L., Acta Cryst., 16, 1255 (1963).

16. (a) Pauling, L., The Nature of the Chemical Bond, (Cornell Press, Ithaca, 1960) p. 255. (b) Ibid., p. 260.

17. Craig, D. P., Maccol, A., Nyholm, R. S. Orgel, L. E., Sutton, L. E., J. Chem. Soc., 157, 349 (1954).

18. Neosi, P., Mukherjee, G. K., J. Indian Chem. Soc., 11, 686 (1934).

19. Wood, J. S., Coordination Chemistry Reviews, 2, 417 (1967).

20. Ibers, J. A., Annual Review of Physical Chemistry, 16, 383 (1966).

21. Gramaccioli, C. M., Acta Cryst., 21, 600 (1966).

22. Cotton, F. A., J. Chem. Phys., 35, 228 (1961).

23. Templeton, D. H., Zietschrift fur Krist., 113, 234 (1960).

24. Patterson, A. L., Physics Reviews, 46, 372 (1934).

25. Cruickshank, D. J. W., Pilling, D. E., Bujosa, A., Lovel, F. M., Truter, M. R., Computing Methods and the Phase Problem, (Permagon Press, Oxford, 1961).

26. (a) Buerger, M. J., Vector Space, (John Wiley, New York, 1959), p. 25. (b) Ibid., p. 181. (c) Ibid., p. 13.

27. Cruickshank, D. J. W., Acta Cryst., 5, 511 (1952).

28. Wrinch, D. W., Philosophy Magazine, 27, 198 (1939).

29. Lipson, H., Cochran, W., The Determination of Crystal Structures, Vol. III, (G. Bell and Sons, London, 1966), p. 164.

30. Hauptman, H., Karle, J., Solution of the Phase Problem I: The Centrosymmetric Crystal, (American Crystallographic Association, 1953), p. 34.

31. International Tables for X-ray Crystallography, Vol. II, (Kynock Press, Birmingham, England, 1952).

32. Kuo, S. S., Numerical Methods and Computers, (Addison Wesley, New York, 1965), p. 154.

33. I.B.M. Scientific Subroutine Package, 2nd Edition.

34. Harker, D., Kasper, J., Acta Cryst., 1, 70 (1948).

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

36. Karle, J., Hauptman, H., Acta Cryst., 3, 181 (1950).

37. MacGillavry, C., Acta Cryst., 3, 214 (1950).

38. Woolfson, M. M., Direct Methods in Crystallography, (Oxford Press, Oxford, 1961).

39. Karle, J., Karle, I., Acta Cryst., 21, 849 (1966).

40. Marsh, R. E., Donohue, J., Advances in Protein Chemistry, Vol. 22, Anfinsen, C. B., Anson, M. L., Edsall, J. T., Richards, F. M., (Academic Press, New York, 1967), p. 249.

41. Montgomery, H., Chastain, R. V., Lingafelter, G. C., Acta Cryst., 20, 728 (1966).

42. Ramakrishan, C., Ramachandran, G. N., Biophysical Journal, 5, 909 (1965).

43. Ramachandran, G. N., Ramakrishan, C., Sasiekharan, V., J. Mole. Bio., 7, 95 (1963).

44. Eichhorn, E. L., Acta Cryst., 15, 1215 (1962).

45. Dickerson, R. E., The Proteins, Vol. 2, Nuerath, H., ed., (Academic Press, New York, 1964), p. 742.

46. Freemen, H. C., Robinson, G., and Schoone, J. C., Acta Cryst., 17, 719 (1964).

47. Willoughby, T. V., Ph.D. Dissertation, University of Oklahoma (1968).

48. Interatomic Distances Supplement (The Chemical Society, London, 1965).

49. Birnbaum, S. M., Levintow, L., Kingsley, R. B., Greenstein, J. P., J. of Biological Chemistry, 194, 455 (1952).

50. Hauptman, H., Karle, J., Acta Cryst., 9, 45 (1956).

51. Karle, I., Karle, J., Acta Cryst., 21, 860 (1966).

52. Karle, J., Acta Cryst., B24, 182 (1968).

53. Karle, I. L., Karle, J., Acta Cryst., B24, 81 (1968).

54. van der Helm, D., 1620 Programs from I.C.R.: Structure Factor with Anisotropic B.

55. Gillespie, R. J., J. Chem. Soc., 4672 (1963).

56. Orioli, P. L., Divaira, M., Sacconi, L., Inorg. Chem., 5, 400 (1966).

57. Bennett, M. J., Cotton, F. A., Eiss, R., Acta Cryst., B24, 904 (1968).

58. Lippert. L. L., Truter, M. R., J. Chem. Soc., 4996 (1960).

59. Fraser, K. A., Harding, M. M., Acta Cryst., 22, 77 (1967).

60. Hall, D., Moore, F. H., Proc. Chem. Soc., 259 (1960).

61. Glick, M. D., Cohen, G. H., Hoard, J. L., J. Am. Chem. Soc., 89, 1992 (1967).

62. Corbrdige, D. E. C., Cox, E. G., J. Chem. Soc., 159, 594 (1956).

63. van der Helm, D., Nicholas, A. F., to be published.

64. Montgomery, H., Lingafelder, E. C., Acta Cryst., 16, 748 (1963).

65. Einstein, F. W. B., Penfold, B., Acta Cryst., 20, 924 (1966).

66. Hughes, E. W., Acta Cryst., B24, 1128 (1968).

67. Pasternak, R. A., Katz, L., Corey, R. B., Acta Cryst., 7, 225 (1954).

68. Pasternak, R. A., Acta Cryst., 9, 341 (1956).

69. Wright, W. B., Acta Cryst., 11, 632 (1958).

70. Fridrichsons, J., Mathieson, A. M., Acta Cryst., 15, 569 (1962).

71. Yakel, H. L., Hughes, E. W., Acta Cryst., 7, 291 (1954).

72. Marsh, R. E., Glusker, J. P., Acta Cryst., 14, 1110 (1961).

73. Leung, Y. C., Marsh, R. E., Acta Cryst., 11, 17 (1958).

74. Cotton, F. A., Chemical Applications of Group Theory, (John Wiley, New York, 1963), p. 97.

75. Gramaccioli, C. M., Marsh, R. E., Acta Cryst., 21, 594 (1966).

76. Franks, W. A., Ph.D. Dissertation, University of Oklahoma, (1968).

77. Cotton, F. A. and Wilkinson, G., Advanced Inorganic Chemistry (Interscience, New York, 1962) p. 757.

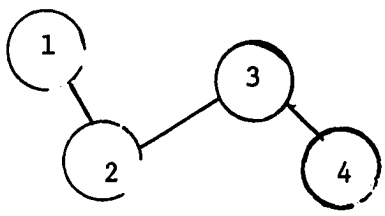
78. Vainshtein, E. E., Antipova-Karataeva, I. I., Russian Journal of Inorganic Chemistry (English Translation), 4, 355 (1959).

79. Ibers, J. A., Hamilton, W. C., MacKenzie, D. R., Inorganic Chem., 3, 1412 (1964).

APPENDIX A

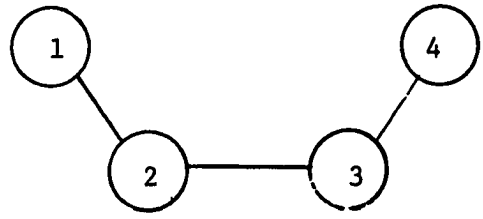
Mathematical Analysis of Conformational Angles

The program calculates the dihedral angle between the two planes determined by the sequences 1,2,3 and 2,3,4 in a list of 4 atomic positions. The zero angle may be specified as either a cis or trans configuration.



Trans configuration

zero angle



Cis configuration

zero angle

This dihedral angle is the right handed rotation about line 2-3 needed to produce the existing configuration from the planar cis or trans reference configuration. The right handed rotation is defined as clockwise motion of atom 4 when one looks in the direction from 2 to 3.

The calculations are performed in Euclidian three space, that is a three dimensional orthonormal co-ordinate system. This results in appreciable simplification of the formulas used in the computations and therefore saves computation time if several angles are to be computed.

In the following discussion A, B, and C will represent the axes of the orthonormal system and a, b, and c will represent the crystal co-ordinate system axes, a*, b* and c* will represent the dual co-ordinate axes of the crystal system. The contravarient components of the atomic co-ordinates in the crystal system will be x, y, and z and their counterparts in Euclidian space will be X, Y, and Z. The fractional atomic co-ordinates of the atoms in the crystal must be given in a right handed system.

The orthonormal co-ordinate system is chosen such that A is coincident with a and of unit length. B is in the ab plane perpendicular to A and of unit length. C is of unit length and perpendicular to both A and B, this causes C to be coincident with c*. These relationships may be expressed in terms of the following equations.

$$A = a/|a|$$

$$C = c^*/|c^*|$$

$$B = C \times A$$

The transformation of atomic co-ordinates is written in matrix form as:

$$(X,Y,Z) = (x,y,z)(\beta).$$

For an orthonormal crystal system the axes A,B,C are coincident with the axes a,b,c. This allows one to immediately write the matrix in terms of the reciprocal cell parameters or dual axes.[31]

$$\beta = \begin{pmatrix} 1/a^* & 0 & 0 \\ 0 & 1/b^* & 0 \\ 0 & 0 & 1/c^* \end{pmatrix}$$

For a monoclinic crystal system A and B are coincident with a and b. The matrix may again be easily written in terms of the reciprocal cell axes and angles.

$$\beta = \begin{pmatrix} 1/a^* \sin\beta^* & 0 & 0 \\ 0 & 1/b^* & 0 \\ -1/c^* \tan\beta^* & 0 & 1/c^* \end{pmatrix}$$

For a triclinic system the fact that all of the reciprocal angles differ from ninety degrees complicates the cross terms of the matrix so that it is not easily expressed in terms of the reciprocal cell dimensions. In this case the β matrix is best expressed in terms of the covariant and contravariant components, g_{ij} and g^{ij} of the metric tensor of the crystal co-ordinate system. The contravariant components of the metric tensor may be directly expressed in terms of the reciprocal cell parameters. If $a^1 = a^*$, $a^2 = b^*$, $a^3 = c^*$, $\alpha^1 = \alpha^*$, $\alpha^2 = \beta^*$, $\alpha^3 = \gamma^*$ then the contravariant components, g^{ij}, of the metric tensor are given by the following formulae.

$$g^{ii} = (a^i)^2$$

$$g^{ij} = a^i a^j \cos\alpha^k, \quad i \neq j \neq k$$

Since $(g^{ij}) = (g_{ij})^{-1}$, the covariant components, g_{ij}, of the metric tensor may be found by inverting the matrix of the components of the contravariant metric tensor. The β matrix of the components of the metric tensor is,[31]

$$\begin{pmatrix} \sqrt{g_{11}} & 0 & 0 \\ g_{12}/\sqrt{g_{11}} & \sqrt{|g| g_{33}/g_{11}} & 0 \\ g_{31}/\sqrt{g_{11}} & \sqrt{|g|}\, g_{32}/\sqrt{g_{11} g_{33}} & 1/\sqrt{g_{33}} \end{pmatrix}$$

where $|g|$ is the determinent of the covariant components of the metric tensor. The formulae for the elements of the metric tensor and the β matrix in terms of its components are valid for all crystal co-ordinate systems. The metric tensor is inverted using the Gauss-Jordan Pivot method[32] programmed in Fortran IV language.[33]

After the atomic co-ordinates are transformed to the orthonormal system the three vectors representing the bonds between the four atoms are formed. For the trans zero conformation the vectors are

$$V_1 = atom_2 - atom_1$$
$$V_2 = atom_3 - atom_2$$
$$V_3 = atom_4 - atom_3$$

For the cis zero conformation the vectors are

$$V_1 = atom_2 - atom_1$$
$$V_2 = atom_3 - atom_2$$
$$V_3 = atom_3 - atom_4$$

The calculation of V_3 in this manner arranges the vectors so that the zero configuration of the vectors is trans while that of the atoms is cis. This allows the angles of rotation to be calculated from the vectors of either the cis or trans arrangement of atoms by the same alogrithms.

Next the unit normals to the planes formed by atoms 1, 2, and 3 and 2, 3, and 4 are calculated.

$$N_{1,2,3} = N_1 = V_1 x V_2 / |V_1 x V_2|$$
$$N_{2,3,4} = N_2 = V_3 x V_2 / |V_3 x V_2|$$

Then the sine and cosine of the angle of rotation are calculated as,

$$\cos \phi = N_2 . N_1$$

$$|\sin \phi| = |N_2 x N_1|$$

The sign of sin ϕ may then be calculated using the trignometric identity

$$\cos(\alpha \pm 90) = \pm \sin \alpha$$

and the fact that V_3 and N_2 are orthogonal. More relevantly, the projection of V_3 on the plane which has V_2 as its normal makes an angle $90+\phi$ with N_1. Then if ψ is the angle between N_1 and V_3 and ϕ the angle between N_1 and N_2, for

$$0 < \phi < \pi \quad \text{then} \quad \pi/2 < \psi < 3\pi/2$$

and for

$$\pi < \phi < 2\pi \quad \text{then} \quad -\pi/2 < \psi < \pi/2.$$

Therefore the sign of cosine of the angle between V_3 and N_1 is the opposite of the sign of sin ϕ, or

$$\text{sign of } [|\sin \phi|] = \text{sign of } [-V_3 . N_1].$$

Knowledge of the sine and cosine of an angle is of course equivalent to knowledge of the angle itself.

APPENDIX B

The Symbolic Addition Procedure

In 1948 Harker and Kasper[34] derived a set of inequalities which would reveal the phase of the large diffraction amplitudes. The inequalities were good only for centrosymmetric space groups. They were developed in terms of normalized structure factor amplitudes, U, where $U = F_o/K \sum_{i=1}^{n} f_i e^{-2BS^2}$.

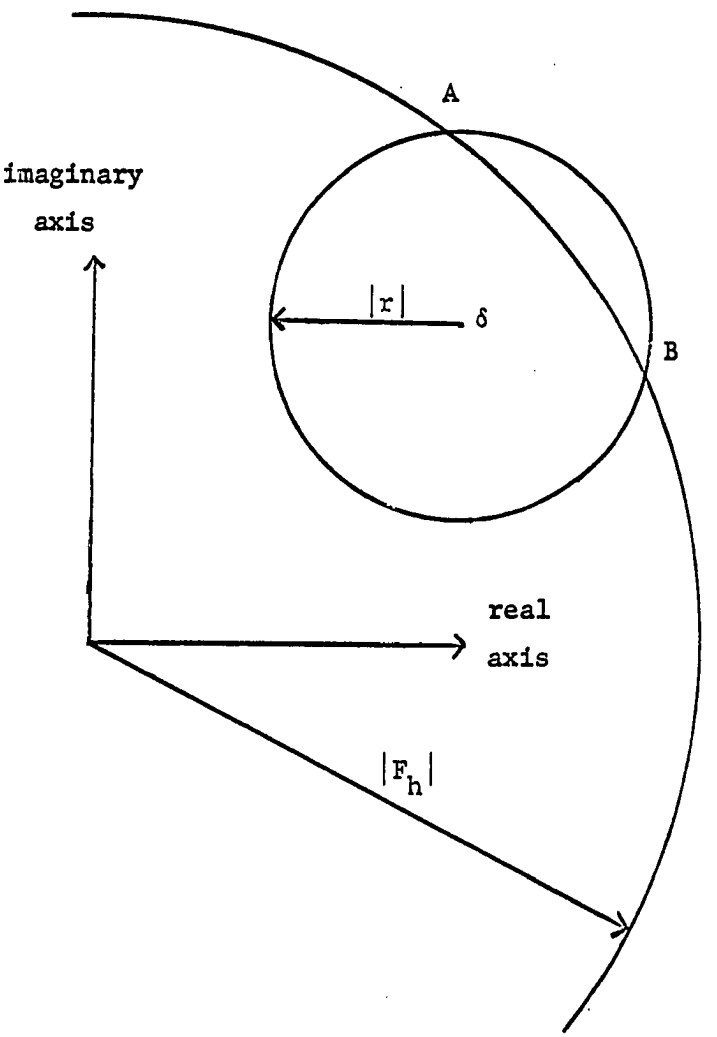
F_o is the observed structure factor amplitude, f_i is the scattering factor for the ith atom in the structure, S is the sine of theta, the diffraction angle, divided by the wave length of the X-rays, and K and B are the scale factor and temperature co-efficient derived from a Wilson plot.[35]

These original inequalities were derived by applying the Schwartz and Cauchy inequalities to the Fourier transform used in X-ray diffraction. Others[36,37] later showed that these inequalities were a result of the non-negativity of the results of the Fourier transform, the electron density. Hauptman and Karle showed that these inequalities were best written as determinants. Also by the introduction of the symmetry constraints on the electron density it is possible to write more powerful phase determining inequalities.

The above mentioned inequalities are severely limited in the complexity of the structure which may be solved by their use. This

Figure 1.

Geometrical Interpretation of the Inequality.



arises from the fact that experience has shown that it is usually necessary for about ten per-cent of the U values to be 0.40 or greater if the structure is to be solved by direct application of the inequalities. If the U's have a normal distribution[38] which is given by the equation

$$P(U)\ dU = (2\pi\varepsilon)^{-1}\ e^{-(u^2/2\varepsilon)}\ dU$$

where

$$\varepsilon = \sum_i \left(f_i / \sum_j f_j \right)^2 ,$$

for a structure composed of only one type of atom, this corresponds to a maximum of sixteen atoms. Of course larger structures have been solved by this method, but in most cases the U distribution was made abnormal by a molecular symmetry which has a systematic effect.

It has also been observed that U's which have values almost large enough to determine their phases usually have the phase indicated by the inequality which nearly applies. Further if several of these inequalities almost apply then the phase is more probably correct than if only one equality relationship which is almost met can be found. Hauptman and Karle have developed a method which systematically exploits these two phenomena. It is known as symbolic addition[30] and is the most successful and widely used of the so-called direct methods.

The original formulation for centrosymmetric crystal structures is a sophistication of the Sayre[38] equation, which is

$$F_h = \frac{\Theta}{V} \sum_k F_k F_{h-k}$$

Θ is a constant which is a function of the particular structure and V is the volume of the unit cell. This equation was derived by applying

the self convolution theorem to the Fourier co-efficients for a crystal structure and the structure which has the square of the electron density of the first structure at all points. Using this equation one is able to establish relationships of the type

$$s(h) \times s(k) \times s(h-k) = +1$$

where s(h) is the sign of F_h.

Karle and Hauptman used normalized structure factor amplitudes, such that the F's are normalized using the equation

$$E_h^2 = \frac{F_h^2}{\varepsilon \sum_i f_i^2 \exp(-2Bs)}$$

where ε is the inverse of the fraction of reflections absent from a particular set of reflections because of space group extinctions. They found that if the structure factor amplitudes are normalized in this manner then it is possible to derive a simple relationship between the size of the summation $\sum_k E_h E_{h-k}$ over these E values with known phases and the probability that the phase of E_h is positive. The probability that the phase of E_h is positive is given by

$$P_+(h) = \tfrac{1}{2} + \tfrac{1}{2}\tanh\alpha_3\alpha_2^{-3/2} E_h \sum_k E_k E_{h-k}$$

$$\alpha_n = \sum_i^m Z_i^n$$

where M is the number of atoms in the unit cell. Since one may arbitrarily choose up to three phases in order to specify the origin it is possible to proceed and compute the probability that each phase is correct. Thus while one is not certain of each phase, those phases most likely to be wrong may be discarded.

However the most significant advancement achieved by Hauptman and Karle was the extension of direct methods to non-centrosymmetric

as well as centrosymmetric structures. This was achieved by use of the inequality,

$$\left| F_h - \frac{F_{h-k}F_k}{F_{ooo}} \right| \leqslant \frac{\begin{vmatrix} F_{ooo} & F_{-(h-k)} \\ F_{h-k} & F_{ooo} \end{vmatrix}^{\frac{1}{2}} \begin{vmatrix} F_{ooo} & F_{-k} \\ F_k & F_{ooo} \end{vmatrix}^{\frac{1}{2}}}{F_{ooo}}$$

where F_{ooo} is the structure factor amplitude of the radiation scattered in the same direction as the incident radiation. This inequality may be rewritten as

$$|F_h - \delta| \leq r.$$

This inequality has the following geometrical interpretation which is illustrated in Fig. 1. The interpretation is that the complex structure factor F_h may not take any value but must be in the area of the complex plane bounded by a circle of radius r whose center is at δ. Further if the magnitude of F_h is known then the complex quantity F_h must be located on the circle swept out by its magnitude and be between points A and B on that circle.

Since the vector k may be varied at will, there are many such limitations on the phase of any amplitude F_h. If only the large F_{h-k} and F_k values are considered then the value of r is small and $F_h \approx \langle \delta_k \rangle_k$ or $F_h \approx \langle F_{h-k}F_k \rangle_k$, where $\langle X_k \rangle_k$ implies the average of the large values of X_k as k varies at will. Since the quantities F_h are complex vectors the averages are vector averages. Thus each F_h is associated with a phase angle by the equation

$$F_h = F_h \exp(i\phi_h)$$

which leads to

$$|F_h| \exp(i\phi_h) = \left\langle |F_{h-k}| \; |F_k| \exp(i\phi_{h-k}) \exp(i\phi_k) \right\rangle_k .$$

Recalling the Sayre equation one concludes

$$\phi_h \tilde{=} \left\langle \phi_{h-k} + \phi_k \right\rangle_k .$$

The limitation to the method is the fact that as the crystal becomes complex, the quantity r becomes so large that the general inequality itself imposes no restrictions on the phase of F_h. However Karle and Karle[39] have shown that the derived phase determining formula, which use the location of δ rather than the size of r, remain valid. During the course of the analysis showing that the phase relationship was valid for complex structures a more accurate formula was derived. This is the well known tangent formula.

$$\tan\phi_h = \frac{\left\langle |E_k \; E_{h-k}| \; \sin(\phi_k + \phi_{h-k}) \right\rangle}{\left\langle |E_k \; E_{h-k}| \; \cos(\phi_k + \phi_{h-k}) \right\rangle}$$

It is this formula that is generally used in the phase assignment procedure leading to a crystal structure.

This formula and the criteria for rejecting unreliable phases has been implemented in a pair of computer programs written for the I.B.M. 360/40 in Fortran. The first program in the pair find all combinations of the type

$$h = k + (h-k)$$

where h and k are the crystallographic indices associated with the large E values. All pairs k and h-k for each h are then stored on magnetic tape for use by the second program. Also stored are the E values the indices and the parity group.

The second program takes an initial set of assigned phases and generates more using the tangent formula. After generating the

phases the program cycles them through the tangent formula to obtain the best self consistent set. On the last refinement cycle the various indicators for accepting or rejecting the assigned phase are calculated and checked against user specified values. The poorly defined phases are removed from the list of assigned phases and the program then repeats the procedure, starting with the generation of new phases.

APPENDIX C

A List of Computer Programs Used during these Investigations

Type of Calculation	Computer and Language	Author
Least Square Unit Cell	I.B.M. 360, FORTRAN IV	T. V. Willoughby
Least Square Plane	I.B.M. 360, FORTRAN IV	T. V. Willoughby
Bond Distances and Angles	I.B.M. 360, FORTRAN IV	M. B. Hossain
Packing Distances	I.B.M. 360, FORTRAN IV	G. Shepherd
Thermal Ellipsoids	I.B.M. 360, FORTRAN IV	W. A. Franks
Structure Factor Least Squares	I.B.M. 360, FORTRAN IV	F. A. Ahmed
Fourier	I.B.M. 360, FORTRAN IV	F. A. Ahmed
Data Reduction	I.B.M. 360, FORTRAN IV	F. A. Ahmed
Data Listing	I.B.M. 360, FORTRAN IV	F. A. Ahmed
Absorption Correction	I.B.M. 1410, FORTRAN II	P. J. Shapiro
Goniostat Settings	I.B.M. 1410, FORTRAN II	P. J. Shapiro
Lorentz-Polarization Corrections	I.B.M. 1410, FORTRAN II	A. F. Nicholas
Absorption Corrections	I.B.M. 360, FORTRAN IV	P. Coppens
Structure Factor Least Squares	I.B.M. 1620, Assembly	D. van der Helm
Fourier	I.B.M. 1620, Assembly	D. van der Helm