69-2211

WILLOUGHBY, Thomas Velbert, 1940-

THE CRYSTAL AND MOLECULAR STRUCTURE OF THE ADDITION COMPLEX OF $CaCl_2$ WITH GLYCYLGLYCYLGLYCINE.

The University of Oklahoma, Ph.D., 1968
Chemistry, physical

University Microfilms, Inc., Ann Arbor, Michigan

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE CRYSTAL AND MOLECULAR STRUCTURE OF THE ADDITION COMPLEX OF $CaCl_2$ WITH GLYCYLGLYCYLGLYCINE

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

THOMAS VELBERT WILLOUGHBY

1968

THE CRYSTAL AND MOLECULAR STRUCTURE OF THE ADDITION COMPLEX OF $CaCl_2$ WITH GLYCYLGLYCYLGLYCINE

APPROVED BY

DISSERTATION COMMITTEE

ACKNOWLEDGEMENT

The author wishes to express his appreciation to Dr. Dick van der Helm for suggesting this problem and for valuable assistance throughout the course of this study.

The author gratefully acknowledges the financial support for this investigation from the National Institutes of Health.

The author would like to thank Dr. Dick van der Helm and Dr. F. R. Ahmed for providing structure factor and Fourier programs for this work, and also his fellow graduate students who provided other necessary computer programs.

Also, the author wishes to thank Mrs. Verlene Yohn, who typed this manuscript and extend his thanks to Ellen Gahan who key punched the initial data cards for the computer programs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

THE CRYSTAL AND MOLECULAR STRUCTURE OF THE ADDITION COMPLEX OF $CaCl_2$ WITH GLYCYLGLYCYLGLYCINE

CHAPTER I

INTRODUCTION

The crystal structure determination of $CaCl_2$ (glycylglycylglycine)$\cdot 3H_2O$ was undertaken as a part of an overall program for studying amino acids and peptides and the interaction of metal ions with amino acids and peptides. This study is being carried out primarily on compounds which have been crystallized from neutral solutions. Freeman and other workers have made extensive crystal structure studies on metal-peptide complexes which have been crystallized from basic solutions (14, 15, 17).

The importance of determining the structures of small amino acids and peptides lies in their use as models from which further knowledge of the structures of proteins, enzymes, and polypeptides can be obtained. Likewise, most of the crystal structure analyses of metal-amino acid and metal-peptide complexes have been carried out on the assumption that such complexes act as models for the metal binding sites on proteins.

Although the x-ray crystallographer is limited, by definition, to the solid state, it is assumed that complexes

found in crystals are also present in the solutions from which crystals grow and, therefore, geometrical information from crystal structure analyses can be transferred to species which exist in solution. Subject to this assumption, the results of crystal structure analyses may be used to establish geometrical criteria according to which models for metal-peptide interaction can be assessed. If a proposed model violates geometrical requirements established for complexes in the solid state, then, it is unlikely to be a valid representation of any species which exists in solution.

The structure of $CaCl_2$ (glycylglycylglycine)$\cdot 3H_2O$ was of particular interest for it afforded an opportunity to study interactions between non-transition metals and peptides. Although the structures of many transition metal complexes of amino acids and peptides have been determined (17), there have been no structure determinations on the corresponding alkali and alkaline earth metal complexes.

One of the objectives of this investigation was to see what effects, if any, the metal bonding to the peptide has on the bond distances and angles in the peptide molecule and on the allowed conformations of the peptide molecules (28).

It was also of interest to compare the nature of the peptide bonding to the Ca ion with that of the peptide bonding to transition metal ions. One would definitely expect the metal bonding to be of a different nature in

the Ca-peptide complex than in the transition metal-peptide complexes, since the Ca bonding is expected to be primarily electrostatic in nature.

As in the structures of other peptides and metal-peptide complexes, there is the possibility of finding configurations related to the helical or pleated-sheet arrangements which are generally observed in proteins and polypeptides.

CHAPTER II

EXPERIMENTAL

Crystals of $CaCl_2$ (glycylglycylglycine)$\cdot 3H_2O$ belonging to the triclinic system were obtained by evaporation of a solution containing equimolar amounts of glycylglycylglycine and $CaCl_2$. The crystals grown in this manner were almost invariably twinned; however, it was found that crystals with a small percentage of twinning could be obtained by adding a large excess of $CaCl_2$ to the solution before evaporation.

Density measurements were made on the crystals using the flotation method. The experimental density was determined to be 1.63 gm/cm^3 $\pm$ .01 gm/cm^3 using solutions of CCl_4 and CH_3I. The calculated density is 1.638 gm/cm^3. Cell dimensions and other unit cell data are given in Table I.

This unit cell conforms to the conventions of Donnay (1) if the axes are relabelled such that $c < a < b$. The Donnay cell is chosen such that the axes are the shortest three non-coplanar translations in the lattice, labelled so as to have $c < a < b$, and oriented so that the angles α and β are non-acute.

The Delaunay reduced cell (2), which is sometimes

Table I

Unit Cell Data

Space Group	$P_{\bar{1}}$	
	$a = 6.961 \pm .002\,\text{Å}$	$a^* = .14408\,\text{Å}^{-1}$
	$b = 8.558 \pm .002\,\text{Å}$	$b^* = .11747\,\text{Å}^{-1}$
	$c = 12.115 \pm .003\,\text{Å}$	$c^* = .082778\,\text{Å}^{-1}$
	$\alpha = 85.82 \pm .03^\circ$	$\alpha^* = 94.08^\circ$
	$\beta = 91.49 \pm .02^\circ$	$\beta^* = 88.81^\circ$
	$\gamma = 94.28 \pm .04^\circ$	$\gamma^* = 85.82^\circ$
Asymmetric Unit	$CaCl_2(ggg)\cdot 3H_2O$	
Observed density	1.63 gm/cm^3	
Calculated density	1.638 gm/cm^3	

used for tabulation of unit cell parameters, is chosen such that α, β, and γ are all equal to or greater than 90° and the angles between the diagonal of the cell and the axes are all less than or equal to 90°. The cell dimensions for the Delaunay cell are

$a = 10.622\,\text{Å}$

$b = 14.315\,\text{Å}$

$$c = 12.115\text{Å}$$

$$\alpha = 143.40^\circ$$

$$\beta = 92.39^\circ$$

$$\gamma = 114.66^\circ$$

The transformation matrix used in converting from the old axes to the Delaunay axes is

$$S = \begin{pmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}$$

This is also the transformation matrix used for converting the indices h, k, and l to indices in the Delaunay system.

In order to transform the old reciprocal cell axes to the reciprocal cell axes of the Delaunay system, one needs another transformation matrix. This matrix is the transpose of the inverse of the matrix S and is given below.

$$(S^{-1})^T = \begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ 1 & -1 & -1 \end{pmatrix}$$

The matrix $(S^{-1})^T$ is also the matrix required to transform the fractional coordinates (x, y, z) of a point in the unit cell to the fractional coordinates (x', y', z') of a point in the Delaunay cell. These transformations for the axes, Miller indices, reciprocal axes, and fractional

coordinates in going from the old system to the Delaunay system can be given by the four matrix equations

$$\begin{array}{rl} \text{axes} & a' = Sa \\ \text{Miller indices} & h' = Sh \\ \text{reciprocal axes} & b' = (S^{-1})^T b \\ \text{coordinates} & x' = (S^{-1})^T x \end{array} \qquad a=\begin{pmatrix} a \\ b \\ c \end{pmatrix}, \; h=\begin{pmatrix} h \\ k \\ l \end{pmatrix}, \; b=\begin{pmatrix} a^* \\ b^* \\ c^* \end{pmatrix}, \; x=\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

where the primed system refers to the Delaunay system. Transformation matrices for the real axes and the reciprocal axes are given in the International Tables for X-Ray Crystallography (3), however, the relationship between these matrices is not shown.

Although these four matrix equations are intended for the specific transformation matrix S in going from the old system to the Delaunay system, they are valid for any transformation of axes governed by a transformation matrix S.

There is some controversy over the uniqueness of the different cell reductions used by Delaunay and Donnay. Of the two, the Delaunay reduction has been studied more thoroughly and is thought to be unique except for a few special uses (4). The reduction used by Donnay is in many cases more convenient; however, its uniqueness is still in question.

From the volume of the unit cell V and the density ρ, the value of nM can be calculated using the equation

$$nM = \rho V N_o$$

where n is the number of molecules in the unit cell, M is the molecular weight, and No is Avagadro's number. The value of nM was calculated to be 712 which would correspond to two molecules of $CaCl_2(ggg)\cdot 3H_2O$(ggg=glycylglycylglycine) in the unit cell. As a precaution and check on the empirical formula of $CaCl_2(ggg)\cdot 3H_2O$ as given by the x-ray data, a quantitive determination of Cl was done on the crystals by using Mohr's method. The percentage of Cl in the compound was determined to be 19.8 which is within experimental error of the theoretical value of 20.0. The compound $CaCl_2(ggg)\cdot 3H_2O$ was first prepared by Pfeiffer and co-workers (5).

The space group was assumed to be $P_{\bar{1}}$ instead of P_1 since a compound will usually crystallize in a centric space group unless it requires a non-centric space group.

Four different crystals were used in measuring the intensity data. The crystals were thin rectangular plates with largest dimensions ranging from .1mm. to .4mm. A total of 2327 intensities were measured using integrated θ- 2θ scans on a General Electric XRD-5 X-Ray Diffractometer equipped with scintillation counter and single crystal orienter. The data were measured up to a 2θ value of 145° using Cu Kα radiation (λ = 1.5418Å). This represents 1800 intensities or all of the data up to a 2θ value of 110° and approximately one-third of the data between 110° and 145°.

Absorption corrections were made on the intensity data using a 3-dimensional numerical integration. The intensity I of a diffracted beam of x-rays can be given by the 3-dimensional integral over the volume of the crystal,

$$I = Io \int e^{-ut}\, dV$$

where u is the linear absorption coefficient of the crystal, t is the total path length traversed by the incident and diffracted beams in the crystal, dV is the diffracting element of volume, and Io is the intensity of the diffracted beam had there been no absorption. The desired quantity Io can therefore be obtained from the measured intensity I by evaluation of the 3-dimensional integral. This was done using a computer program written by Dr. Philip Shapiro which does a 3-dimensional numerical integration using the method of Gauss.

Absorption corrections on the intensity data were considered necessary because of the large value of the linear absorption coefficient, $73.8 cm^{-1}$, and because of the shape of the crystals which were thin plates. The value of the correction factor, $\frac{Io}{I}$, ranged from 1.2 to 15 for the 2327 reflections.

Due to the incapability of distinguishing between the centrosymmetric space group $P_{\bar{1}}$ and the non-centrosymmetric space group P_1 by direct x-ray methods, it was decided to run the $N(z)$ test of Howells, Phillips, and

Rogers on the intensity data (6). The basis of the N(z) test is that the intensities follow two different distribution functions depending on whether the space group is centrosymmetric or not. If the space group is centrosymmetric the distribution function is $P(z) = (2\pi z)^{-\frac{1}{2}}\exp(-\frac{1}{2}z)$ and if the space group is non-centrosymmetric $P(z) = \exp(-z)$ where each intensity is expressed as a fraction $z=I/\langle I\rangle$, $\langle I\rangle$ being the local average intensity. The value of N(z), the fraction of the z values which are less than or equal to a particular z, is given by the integral of the distribution function, P(z), from z = 0 to z.

$$N(z) = \int_0^z P(z)\, dz$$

This reduces to $N(z) = 1-e^{-z}$ for the noncentrosymmetric case and $N(z) = \mathrm{erf}(\frac{1}{2}z)^{\frac{1}{2}}$ for the centrosymmetric case. A plot of these two functions is shown in Figure 1 as well as the values of N(z) which were calculated by a computer program from the experimental intensity data. It can be seen from the graph that the experimental N(z) curve lies very close to the theoretical N(z) curve for the centrosymmetric case and therefore it was concluded that the space group was the centrosymmetric $P_{\bar{1}}$.

Lorentz-polarization corrections were made on the intensities by multiplying them by the factor

$$L.P. = \frac{2 \sin 2\theta}{1 + \cos^2 2\theta}$$

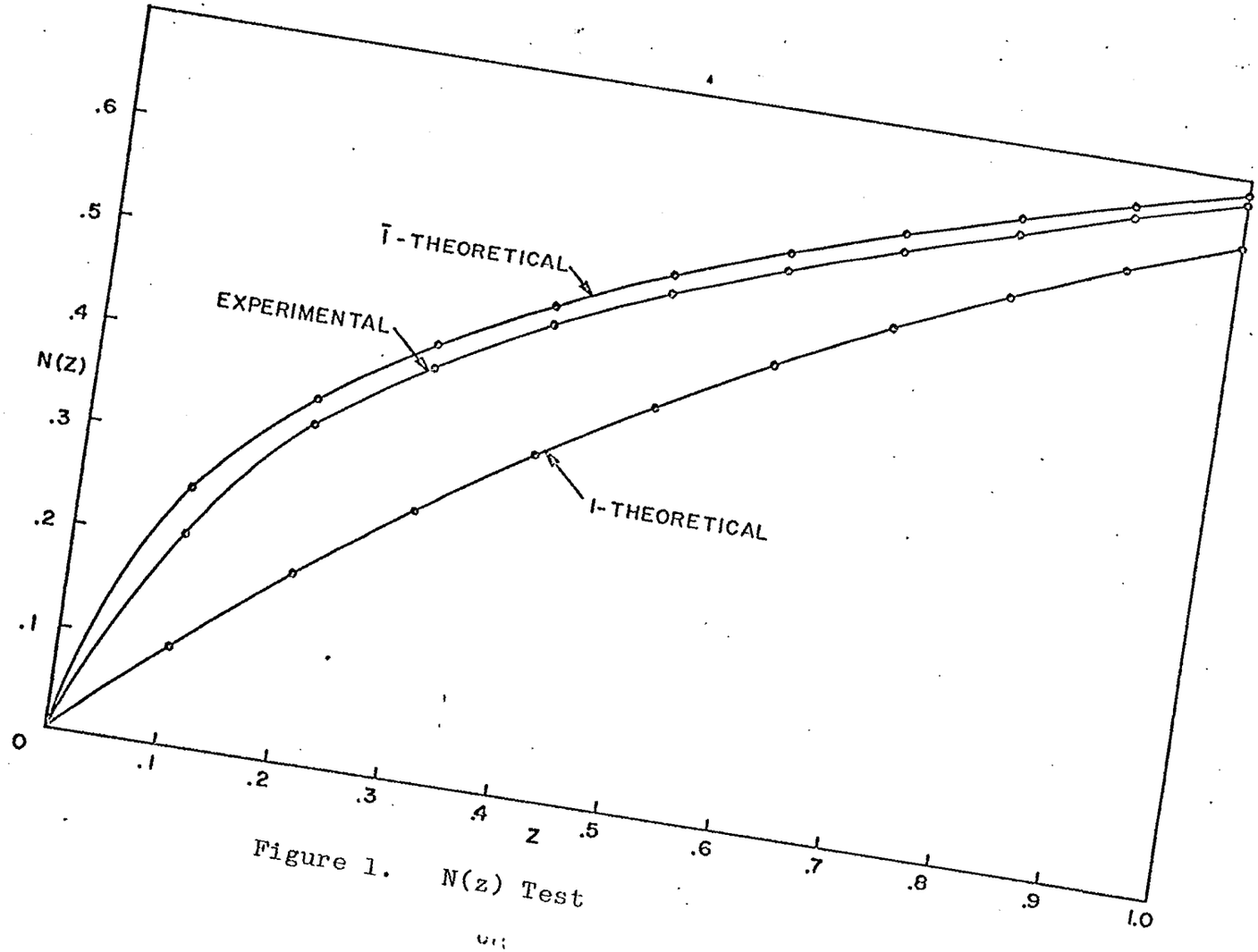


Figure 1. N(z) Test

The intensity data were sharpened using the sharpening function

$$M = \left(\frac{z}{\sum f_i}\right)^2 \exp\left(\frac{P \sin^2\theta}{\lambda^2}\right)$$

where the arbitrary sharpening parameter P was taken to be .5 and where z is the total number of electrons in the atoms and f_i is the scattering factor for the i^{th} atom. With this value of P the exponential part of the sharpening function changes from 1.0-1.2 in going from low order to high order reflections and is, therefore, practically insignificant in its sharpening effect when compared to the term $\left(\frac{z}{\sum f_i}\right)^2$. The sharpening of the intensities by multiplying them by the sharpening function M or by some other suitable sharpening function has been found necessary in most cases in order to get well resolved peaks in the Patterson function.

CHAPTER III

SOLUTION OF THE PATTERSON FUNCTION

After making the necessary corrections on the intensity data the Patterson function is then ready to be calculated. The value of the Patterson function at the point (u, v, w) is defined by the equation

$$P(u,v,w) = V\int_0^1 \int_0^1 \int_0^1 \rho(x,y,z)\ \rho(x+u,\ y+v,\ z+w)dxdydz$$

where $\rho(x,y,z)$ is the electron density at the point (x,y,z) in the unit cell of volume V and is given by the equation

$$\rho(x,y,z) = \frac{1}{V} \sum_h \sum_k \sum_{l=-\infty}^{\infty} Fhkl \exp -2\pi i(hx+ky+lz)$$

Substituting in this expression for the electron density and noting that the integrals of the type

$$\int_0^1 e^{i2\pi(h+h')x}\ dx = 0 \quad \text{if } h' \neq -h$$

one obtains the summation form of the Patterson function

$$P(u,v,w) = \frac{1}{V} \sum_h \sum_k \sum_{l=-\infty}^{\infty} |Fhkl|^2 \exp 2\pi i(hu+kv+lw)$$

which, by applying the condition $\overline{Fhkl} = Fhkl$, can be further simplified to its most common form

$$P(u,v,w) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l=-\infty}^{\infty} |Fhkl|^2 \cos 2\pi(hu+kv+lw)$$

By virtue of its defining equation the Patterson function will give a peak at the point (u,v,w) only when the electron density is finite at both points (x,y,z) and (x+u, y+v, z+w), i.e., only when u,v, and w are the components of an interatomic vector in the unit cell. This function uses as coefficients in the Fourier series the $|F|^2$ values, which are directly derivable from the measured x-ray intensities. Since the Patterson function uses only the measured intensity values, it can be calculated directly at any point (u,v,w) and will give a peak at this point if u,v, and w are the components of an interatomic vector in the unit cell. The magnitude of this peak will be roughly proportional to the product of the number of electrons of the two atoms with this interatomic vector. From this property it follows that peaks between heavy atoms such as Ca and Cl will be much larger than peaks between the light atoms C, O, N. This makes the Patterson function a natural starting point for finding the positions of the heavy atoms present in the structure.

In this particular case there are one Ca^{+2} ion and two Cl^{-1} ions in the asymmetric unit with two molecules present in the unit cell. This gives a total of six heavy

atoms in the unit cell which should give rise to 15 interatomic vectors, however, six of these vectors are repeated by virtue of the center of symmetry in the unit cell and, therefore, only nine different interatomic vectors between heavy atoms will appear in the Patterson map. Since six of these vectors are repeated, leaving only the three vectors between centrosymmetrically related heavy atoms as single vectors, one could expect six heavy atom peaks in the Patterson map to be twice as large as the remaining three. In the subsequent discussion the heavy atom double peaks shall be referred to as H-H(2) peaks and the heavy atom single peaks as H-H(1) peaks.

Complications arise in interpreting the Patterson function due to the fact that heavy atom-light atom interatomic vectors between the heavy atoms Ca and Cl and the light atoms C, O, and N all occur twice and give rise to double peaks in the Patterson map. The peaks which are due to interatomic vectors between centrosymmetrically related heavy atoms are single peaks. Since Ca^{+2} and Cl^{-1} have 18 electrons each, or roughly twice as many electrons as the light atoms C, O, and N, it follows that the single peaks due to heavy atom-heavy atom interatomic vectors, designated H-H(1), are about the same magnitude as the double peaks due to heavy atom-light atom interatomic vectors, designated H-L(2). Because of this fact it is impossible by size alone to distinguish the 3H-H(1) peaks from the

large number, approximately 100, of H-L(2) peaks present in the Patterson map. This leaves only the six double peaks due to heavy atom-heavy atom interatomic vectors, designated H-H(2), as the ones which should be easily found from the Patterson map.

Actually, instead of the six large peaks predicted, there were eight peaks present in the Patterson function which were large enough to be considered possible H-H(2) peaks. It was suspected that two of these large peaks were due to coincidental overlapping of H-L(2) peaks, this being very common due to the large number of peaks in the Patterson function. The eight largest peaks range in magnitude from 667 to 1351, calculated on an arbitrary scale, and are listed in Table II. The magnitude of the ninth largest peak is 570.

Table II

Eight Largest Peaks in Patterson Function

u	v	w	P(u,v,w)	Peak
.24	.48	.18	1351	A
.40	.20	.44	1295	B
.12	.92	.38	1234	C
.18	.72	.26	1165	D
.92	.18	.12	1049	E
.72	.34	.38	964	F
.88	.46	.18	739	G
.48	.88	.18	667	H

The components of the interatomic vectors expected between six heavy atoms in space group $P\bar{1}$ are given in terms of the coordinates of three of the atoms in Table III.

Table III

Components of Heavy Atom Vectors

u	v	w	
x_2-x_1	y_2-y_1	z_2-z_1	H-H(2)
x_3-x_1	y_3-y_1	z_3-z_1	H-H(2)
x_3-x_2	y_3-y_2	z_3-z_2	H-H(2)
x_1+x_2	y_1+y_2	z_1+z_2	H-H(2)
x_1+x_3	y_1+y_3	z_1+z_3	H-H(2)
x_2+x_3	y_2+y_3	z_2+z_3	H-H(2)
$2x_1$	$2y_1$	$2z_1$	H-H(1)
$2x_2$	$2y_2$	$2z_2$	H-H(1)
$2x_3$	$2y_3$	$2z_3$	H-H(1)

From the previous discussion, the six double peaks, designated H-H(2) in Table III, should be among the eight largest peaks in the Patterson function listed in Table II. Attempts were made to find atomic positions consistent with the components of the H-H(2) vectors and which would generate the components of six of the eight large peaks in the Patterson. Three sets of positions were found by trial and error methods, each of which would generate seven of the eight large peaks listed in Table II. In all three cases

the six H-H(2) vectors generated corresponded to peaks in Table II. The positions of the atoms and the interatomic vectors which they generate for each of the three possible solutions are listed in Tables IV, V, and VI. The vectors are generated in the same order as they are given in Table III and the large peaks which are used from Table II are labeled by letter sequence.

It should be noted that some of the components, u, v, and w have been transformed into their symmetry related forms which were used in calculating the Patterson function. These transformations which consist of adding +1 or -1 to any or all of the u, v, and w values or changing the signs of all of them do not affect the value of the Patterson function. This can be expressed by considering the value of the Patterson function at the point (u,v,w), $P(u,v,w) = P(-u,-v,-w) = P(1+u,v,w) = P(u-1,v,w) =$ etc., and can be shown from the defining equation of the Patterson function.

With the unexpected result of finding three different solutions of the Patterson function, a survey was made of the number of possible solutions that could be obtained. If it is assumed that the eight high Patterson peaks (A-H) listed in Table II contain the six H-H(2) peaks, which is certainly a valid assumption, it is possible to search those eight vectors systematically for solutions.

In the space group $P\bar{1}$ with three heavy atoms in the asymmetric unit, the unit cell contains six heavy atoms at

Table IV

Trial and Error Solution I

$x_1 = .24$	$x_2 = .48$	$x_3 = .64$	
$y_1 = .93$	$y_2 = .41$	$y_3 = .15$	
$z_1 = .10$	$z_2 = .28$	$z_3 = .52$	
	u	v	w	*P(u,v,w)
(A)	.24	.48	.18	1351
(B)	.40	.22	.42	1295
(D)	.16	.74	.24	1165
(F)	.72	.34	.39	964
(C)	.88	.08	.62	1234
(G)	.12	.56	.80	739
(H)	.48	.86	.20	667
	.96	.83	.56	$\sim$240
	.28	.30	.04	$\sim$268

*The symbol $\sim$ is used when peak does not lie on its calculated position but merely close.

Table V

Trial and Error Solution II

$x_1 = .16$	$x_2 = .56$	$x_3 = .68$
$y_1 = .07$	$y_2 = .27$	$y_3 = .21$
$z_1 = .97$	$z_2 = .41$	$z_3 = .77$

	u	v	w	P(u,v,w)
(B)	.40	.20	.42	1295
(H)	.52	.14	.80	667
(C)	.13	.94	.36	1234
(F)	.72	.34	.38	964
(D)	.84	.28	.74	1165
(A)	.24	.48	.18	1351
	.32	.14	.94	~ 530
(G)	.12	.54	.82	739
	.36	.42	.54	~ 200

Table VI

Trial and Error Solution III

$x_1 = .08$	$x_2 = .32$	$x_3 = .80$
$y_1 = .86$	$y_2 = .34$	$y_3 = .20$
$z_1 = .13$	$z_2 = .31$	$z_3 = .51$

	u	v	w	P(u,v,w)
(A)	.24	.48	.18	1351
(F)	.72	.34	.38	964
(H)	.48	.86	.20	667
(B)	.40	.20	.44	1295
(C)	.88	.06	.64	1234
(G)	.12	.54	.82	739
(D)	.16	.72	.26	1165
	.64	.68	.62	468
	.60	.40	.02	320

locations $R_1(x_1,y_1,z_1)$, $R_2(x_2,y_2,z_2)$, $R_3(x_3,y_3,z_3)$, $R_4(=-R_1)$ $R_5(=-R_2)$, and $R_6(=-R_3)$. The positions give rise to three H-H(2) peaks of the type R_2-R_1, R_3-R_2, and R_1-R_3. One can easily see that these three vectors form a closed set, i.e., $(R_2-R_1) + (R_3-R_2) + (R_1-R_3) = 0$ or $(R_2-R_1) + (R_3-R_2) = R_3-R_1$. The first step (I) in the systematic search will therefore be a search for all closed sets of three vectors in the eight A-H vectors. The three other H-H(2) vectors are of the type $R_1 + R_2$, $R_2 + R_3$, and $R_1 + R_3$. Three relations can be obtained by subtraction of these vectors.

$$\begin{aligned}(R_1+R_2) - (R_2+R_3) &= R_1-R_3\\ (R_1+R_2) - (R_1+R_3) &= R_2-R_3 \quad \text{(II)}\\ (R_2+R_3) - (R_1+R_3) &= R_2-R_1\end{aligned}$$

In other words, the vectors R_1+R_2, R_2+R_3, and R_1+R_3 form a closed set in that all possible subtractions will yield the vectors R_2-R_1, R_3-R_2, and R_1-R_3. The second step (II) in the systematic solution is, therefore, once a closed set of three vectors is found in step I, to search the remaining five vectors for a set of three which adhere to the equations (II). The third step (III) of the solution will be the numerical calculation of the solutions found by applying conditions I and II.

It should be pointed out that each independent solution will be found four times following this procedure. Besides solution R_1,R_2,R_3 (identical to $-R_1,R_2,-R_3$) one

will find solutions $-R_1$, R_2, R_3 (identical to $R_1,-R_2,-R_3$), $R_1,-R_2,R_3$ (identical to $-R_1,R_2,-R_3$), and $R_1,R_2,-R_3$ (identical to $-R_1,-R_2,R_3$) as a result of the four possible choices of the asymmetric unit of the unit cell. An example of this will be given in Step III.

To facilitate the procedure, all possible vectorial relationships between the eight vectors A-H are listed in Table VII. Each relationship occurs three times because a calculation is made for each separate vector. The relationships allow for a possible error in the fractional coordinates of the vectors of $\pm$.02.

Step I

The possible answers to Step I are contained in Table VII and are listed in Table VIII. Each independent relationship of Table VII is a solution which adheres to the requirements established for this initial step. There are seven of these relationships.

Step II

Each of the seven closed sets listed in Table VIII should be investigated independently. For example: (1) -A+B = D. It is the purpose to look if one can find in the vectors C, E, F, G, and H one or more sets of three vectors which adhere to the conditions of equations II. Using Table VII, we see, for instance, the relationships

C-G = A
C+F =-D
F+G =-B

Table VII

Vectorial Relationships

A	B	C
- A + B = D	B - A = D	C - A = G
- A + C = G	B + C =-H	C - D = E
- A + H = F	B + F =-G	C + B =-H
		C + D +-F

D	E	F
D + A = B	E + D = C	F + C =-D
D + E = C		F + B =-G
D + F =-C		F + A = H

G	H
G + A = C	H - A = F
G + B =-F	H + B =-C

Table VIII

Independent Vectorial Relationships

- A + B = D	(1)
- A + C = G	(2)
- A + H = F	(3)
B + C =-H	(4)
B + F =-G	(5)
C - D = E	(6)
C + D =-F	(7)

(the assessment of signs to the vectors will be worked out in Step III).

In addition, one finds in Table VIII the relationships

$$C + F = -D$$

$$C + H = -B$$

$$H - F = A$$

For the relationship -A + B = D, we have, therefore, found two solutions adhering to conditions (II)

A B D C F G

and

A B D C F H

Investigating solutions (2) - (6) of Table VIII with the help of Table VII, one finds a total of 12 solutions, four of which involve vectors ABCDFG, four involve

vectors ABCDFH, and four involve the vectors ABCFGH. These solutions are listed in Table IX. There are, therefore, three independent solutions with each independent solution occurring four times.

Table IX

Possible Solutions for Six Vectors

	Condition I	Condition II	Vectors involved
(11)	ABD	CFG	ABCDFG
(12)	ABD	CFH	ABCDFH
(13)	ACG	BDF	ABCDFG
(14)	ACG	BFH	ABCFGH
(15)	AFH	BCD	ABCDFH
(16)	AFH	BCG	ABCFGH
(17)	BCH	ADF	ABCDFH
(18)	BCH	AFG	ABCFGH
(19)	BFG	ACD	ABCDFG
(20)	BFG	ACH	ABCFGH
(21)	CDF	ABG	ABCDFG
(22)	CDF	ABH	ABCDFH

The only relationship in Table VIII which can't be used is relationship (6). This could have been expected, because vector E has only one relationship (see Table VII).

Step III

The assessment of signs alluded to in Step II is in

fact the proper assignment of vectors to R_1+R_2, R_2+R_3, and R_1+R_3. This can be shown, for instance, with solution (11).

From Table VII: A + D = B. This allows the arbitrary assignment

$$\text{(a)} \quad A = R_2-R_1$$
$$D = R_3-R_2$$
$$B = R_3-R_1$$

which fits the equation A + D = B. Also, from Table VII, using two of the vectors C, F, and G, one finds

$$C + F = -D = R_2-R_3$$

This allows for two possible sets of assignments to vectors C and F.

$$\text{(b)} \quad C = R_1 + R_2 \quad \text{or} \quad \text{(c)} \quad F = R_1 + R_2$$
$$F = -R_1 - R_3 \qquad\qquad C = -R_1 - R_3$$

Using another relationship between vectors C, F, and G, a choice between (b) and (c) can be made. From Table VIII: C - G = A or G = C - A. Using equations (a) and (b), one finds $G = R_1 + R_2 - R_2 + R_1 = 2R_1$ which is improper. Using, however, equations (a) and (c), one finds $G = -R_1 - R_3 - R_2 + R_1 = -(R_2 + R_3)$, which is the desired type of relationship. One has, therefore, for solution (11) the following relationships:

$$A = R_2 - R_1$$
$$D = R_3 - R_2$$
$$B = R_3 - R_1$$

$$C = -(R_1 + R_3)$$
$$F = (R_1 + R_2)$$
$$G = -(R_2 + R_3)$$

which follow the relationships derived in Steps I and II, i.e.

$$A + D = R_3 - R_1 = B$$
$$C - G = -R_1 - R_3 + R_2 + R_3 = R_2 - R_1 = A$$
$$C + F = -R_1 - R_3 + R_1 + R_2 = R_2 - R_3 = -D$$
$$G + F = -R_2 - R_3 + R_1 + R_2 = R_1 - R_3 = -B$$

Filling in the numerical values for A, D, B, C, F, and G, one can solve for R_1, R_2, and R_3, i.e.

$$x_2 - x_1 = .24 \qquad x_1 + x_2 = .72 \qquad -x_1 - x_3 = .12$$
$$y_2 - y_1 = .48 \qquad y_1 + y_2 = .34 \qquad -y_1 - y_3 = .92$$
$$z_2 - z_1 = .18 \qquad z_1 + z_2 = .38 \qquad -z_1 - z_3 = .38$$

which yields:

(Solution 31a)

$$x_1 = .24 \qquad x_2 = .48 \qquad x_3 = .64$$
$$y_1 = .93 \qquad y_2 = .41 \qquad y_3 = .15$$
$$z_1 = .10 \qquad z_2 = .28 \qquad z_3 = .52$$

Solution (13) of Table IX is supposed to yield the same solution. Following the procedure as described before, the following assignment is obtained.

$$G = R_2 - R_1 \qquad B = R_1 + R_3$$
$$A = R_3 - R_2 \qquad D = R_1 + R_2$$
$$C = R_3 - R_1 \qquad F = -(R_2 + R_3)$$

Using numerical values for these vectors one obtains the solution 31b.

(Solution 31b)

$x_1 = .65$ $x_2 = .53$ $x_3 = .75$

$y_1 = .12$ $y_2 = .58$ $y_3 = .08$

$z_1 = .04$ $z_2 = .22$ $z_3 = .40$

Changing the origin of the unit cell to $(0, 0, \frac{1}{2})$ this becomes

(Solution 31b')

$x_1 = .65$ $x_2 = .53$ $x_3 = .75$

$y_1 = .12$ $y_2 = .58$ $y_3 = .08$

$z_1 = .54$ $z_2 = .72$ $z_3 = .90$

which is identical to solution 31a with $R_1(31a) = -R_3(31b')$, $R_2(31a) = -R_2(31b')$, and $R_3(31a) = R_1(31b')$ and which is the identity of the solutions R_1, R_2, R_3 and R_1, R_2, $-R_3$ as described before.

Similarly, the solution of (19) in Table IX is related to (31a) as R_1, $-R_2$, R_3, and the solution of (21) in Table IX is related to (31a) as $-R_1$, R_2, R_3.

The solutions arising from (12), (15), (17), and (22) are identical to each other but different from solution 31. This solution is listed as number 32 in Table X. Similarly, (14), (16), (18), and (20) all yield solution 33 of Table X.

In conclusion, there are thus three and only three independent solutions (listed as 31, 32, and 33 in Table X)

Table X

Possible Positions of Heavy Atoms

	$x_1 = .24$	$x_2 = .48$	$x_3 = .64$
(31)	$y_1 = .93$	$y_2 = .41$	$y_3 = .15$
	$z_1 = .10$	$z_2 = .28$	$z_3 = .52$
	$x_1 = .16$	$x_2 = .56$	$x_3 = .68$
(32)	$y_1 = .07$	$y_2 = .27$	$y_3 = .21$
	$z_1 = .97$	$z_2 = .41$	$z_3 = .77$
	$x_1 = .08$	$x_2 = .32$	$x_3 = .80$
(33)	$y_1 = .86$	$y_2 = .34$	$y_3 = .20$
	$z_1 = .13$	$z_2 = .31$	$z_3 = .51$

for which the six H-H(2) vectors fit the eight largest vectors (A-H) in the Patterson. These solutions are identical to the ones listed in Tables IV, V, and VI which were obtained by trial and error methods. In referring to the H-H(2) interatomic vectors, the peak heights will be taken from Tables IV, V, and VI which correspond to solutions 31, 32, and 33, respectively.

Each of the three possible solutions will now be checked for the occurrence of the peaks due to the three H-H(1) vectors. The three H-H(1) peaks for each of the three solutions are listed in Table XI. In both solution

Table XI

H-H(1) Vectors for Three Possible Solutions

u	v	w	P(u,v,w)
	(Solution 31)		
.48	.86	.20	667
.96	.82	.56	~240
.28	.30	.04	~268
	(Solution 32)		
.32	.14	.94	~530
.12	.54	.83	739
.36	.42	.54	~200
	(Solution 33)		
.16	.72	.26	1165
.64	.68	.62	468
.60	.40	.02	320

31 and solution 32 two of the H-H(1) vectors which are calculated do not correspond to peaks in the Patterson map. However, in both solutions there are always peaks of the proper magnitude which are close enough to the calculated positions to not exclude these as possible solutions of the Patterson function. In solution 33 the three H-H(1) vectors all correspond to peaks in the Patterson map.

With three possible solutions of the Patterson function for the positions of the heavy atoms, it was decided to look at each of these in detail in order to see if they conform to the predicted relative peak heights.

In solution 31, if atom 1 is taken to be the Ca^{+2} ion and atoms 2 and 3 the Cl^{-1} ions, the relative peak heights are all reasonably well consistent with what is expected. The H-H(1) peak due to centrosymmetrically related Ca^{+2} ions is larger than the other H-H(1) peaks. This is expected since the electron density of the Ca^{+2} ion should be greater than that of the Cl^{-1} ions, even though these ions both have 18 electrons. The H-H(2) peaks due to the Ca^{+2} - Cl^{-1} vectors are larger than the H-H(2) peaks due to the Cl^{-1} - Cl^{-1} vectors which is also expected.

Looking at solution 32, the heights of the peaks can be reasonably well explained if atom 2 is chosen as the Ca^{+2} ion. This is consistent with the heights of the H-H(1) peaks. The peak due to centrosymmetrically related Ca^{+2} ions is larger than the peaks due to centrosymmetrically

related Cl^{-1} ions. Also, of the six double peaks, H-H(2), the ones due to Ca^{+2}-Cl^{-1} vectors are larger than the ones due to Cl^{-1}-Cl^{-1} vectors as would be expected.

The peaks due to the heavy atom interatomic vectors which are generated in solution 33 also conform well to their predicted peak heights. If atom 1 is taken as the Ca^{+2} ion, the H-H(2) peaks which are due to Ca^{+2}-Cl^{-1} vectors are larger than the ones due to Cl^{-1}-Cl^{-1} vectors. Two of the H-H(1) peaks are of proper magnitude, but the one due to the Ca^{+2}-Ca^{+2} vector is much larger, 1165, which is about the same size as the largest H-H(2) peaks. Although the Ca^{+2}-Ca^{+2} H-H(1) peak is expected to be larger than the Cl^{-1}-Cl^{-1} H-H(1) peaks, since the Ca^{+2} ion has a greater electron density than the Cl^{-1} ion, it is doubtful that this alone could account for the difference in size between this peak and the other two H-H(1) peaks between centrosymmetrically related Cl^{-1} ions.

It can be seen that in each of the three possible solutions of the Patterson function, the relative peak heights could be satisfactorily explained in terms of the electron densities of the Ca^{+2} and Cl^{-1} ions.

Since examination of the relative peak heights did not give a clear-cut choice between the three solutions, an attempt was made to resolve this ambiguity by calculating structure factors for each solution. Structure factor calculations were made for each of the three solutions to

the Patterson function based on the positions of the heavy atoms. For each case the discrepancy index $R = \frac{\Sigma||Fo| - |Fc||}{\Sigma|Fo|}$ dropped to about .55 after three refinements where the $|Fo|$'s are the observed structure factors and the $|Fc|$'s are the calculated structure factors. This was unexpected since there should be better agreement between the observed and calculated structure factors when the correct positions of the heavy atoms are used and, therefore, R should be smaller for the correct solution of the heavy atoms.

With the failure of the structure factor calculations to give any information as to which would be the correct solution of the Patterson function, a further look was taken at the fitting of the heavy atom peaks in the Patterson function.

In each of the possible heavy atom solutions of the Patterson function, six of the eight largest peaks listed in Table II correspond to H-H(2) peaks. For solution 31, these are peaks A, B, C D F, and G, for solution 32, A, B, C, D, F, and H, and for solution 33, A, B C, F, G, and H. If one also considers the three peaks due to the H-H(1) vectors, then seven of the eight peaks listed in Table II are used in each of the three solutions. The only peak in Table II which isn't used in any of the solutions is peak E which has a magnitude of 1049.

Each of the three possible solutions of the Patterson function seemed to be fairly well consistent with the

predicted peak heights. In each solution one of the H-H(1) peaks was found among the eight largest peaks in the Patterson function, and was therefore larger than the expected value for a single peak. This, however, was not considered to be a serious objection since coincidental peak overlapping in the Patterson function is not uncommon due to the large number of peaks. Although the large H-H(1) peak in solution 33 was peak D, the fourth largest peak in the Patterson function, this solution was favored over the other two because a very good fit of the Patterson function was obtained for all of the H-H(1) and H-H(2) vectors. In both solution 31 and solution 32, two of the H-H(1) vectors do not lie exactly on peaks in the Patterson function, but merely close to peaks. This was the primary objection to these two solutions.

Instead of doing Fourier syntheses using the signs of the structure factors calculated from the heavy atom positions for each of the three different solutions, it was decided to try to interpret the Patterson function further, assuming that solution number 33 was the correct one, in hopes of finding the atoms surrounding the Ca ion. Since the bonding of the Ca^{+2} ion was expected to be mainly electrostatic in nature, it seemed unlikely that it would be bonded to any atoms other than oxygen atoms. Bonding to Cl ions was excluded because the lengths of the heavy atom vectors found in the Patterson showed no close contacts

between Ca and Cl ions. The possibility of having the Ca ion surrounded by oxygen atoms was also supported by the structure of the cyclic amine complex $CaBr_2 \cdot 10H_2O \cdot 2(CH_2)_6N_4$ where $Ca(H_2O)_6^{2+}$ octahedra are formed and no Ca-Br bonds are formed (7).

The normal length of the Ca-O bond ranges from 2.2 to 2.6Å. A search was made in the Patterson map for peaks due to interatomic vectors that would be included in this length range. Six peaks were found around the origin due to interatomic vectors of this length, five of which were of the proper magnitude expected for H-L(2) peaks, ranging in magnitude from 436 to 571. The sixth peak was about twice as large, 1049, and was the only peak, peak E, listed in Table II which could not be satisfactorily interpreted as a heavy atom-heavy atom peak. The components of these vectors and the heights of the peaks are listed in Table XII.

Table XII

Possible Ca-O Vectors

u	v	w	P(u,v,w)
-.30	-.18	.00	530
.08	.22	-.14	436
.24	.16	-.10	498
.24	-.06	.12	571
.16	-.16	.14	321
-.08	.18	.12	1049

Changing the origin of the unit cell to ($\frac{1}{2}$, 0, 0) and using the coordinates of the Ca ion from solution 33, x=.42, y=.14, z=.87, the positions of the oxygen atoms were calculated from the components of the vectors listed in Table XII. For each of the vectors listed in Table XII the position of the oxygen atom was calculated as is illustrated by using the components of the first vector listed.

$u = x_o - x_{Ca} = -.30$ $\quad x_{Ca} = .42 \quad x_o = .12$

$v = y_o - y_{Ca} = -.18$ $\quad y_{Ca} = .14 \quad y_o = .96$

$w = z_o - z_{Ca} = .00$ $\quad z_{Ca} = .87 \quad z_o = .87$

The components of the interatomic vectors between this oxygen atom and the five other heavy atoms in the unit cell were then calculated. After calculating the components of these five vectors, the Patterson map was then checked to see if peaks were present corresponding to each of these vectors. In this case, all the peaks were present due to heavy atom-oxygen vectors, so it was considered to be highly probable that this actually was the position of an oxygen atom.

Because of the impossibility of knowing initially whether the components of the vector from Ca^{+2} to 0 were (-.30, -.18, .00) or (.30, .18, .00), positions for the oxygen atoms were calculated for both cases. For the second case

$u = x_o - x_{Ca} = .30$ $\quad x_{Ca} = .42 \quad x_o = .72$

$v = y_o - y_{Ca} = .18$ $\quad y_{Ca} = .14 \quad y_o = .32$

$w = z_o - z_{Ca} = .00$ $\quad z_{Ca} = .87 \quad z_o = .87$

In the same manner as before, the Patterson was checked for the vectors between this oxygen and the five other heavy atoms in the unit cell. Not all of these vectors were present in the Patterson map, so the position of this oxygen was considered to be unlikely and was discarded.

For each of the five other vectors listed in Table XII, which were considered to be possible Ca-O vectors, positions were found for oxygen atoms which would generate all of the heavy atom-oxygen peaks. In each case, the vector from the Ca^{+2} at (.42, .14, .87) to an oxygen was given the components of one of the vectors listed in Table XII and the position of the oxygen atom was calculated, then, the vector from the Ca^{+2} to the O was given the components of the vector in the opposite direction and the corresponding position of this oxygen was calculated. The Patterson was then checked to see if all of the heavy atom-O interatomic vectors were present for each of the two positions calculated for the O atom. For the first five peaks listed in Table XII only one of the two positions calculated for the oxygen atom could be justified by finding all of the heavy atom-O vectors present in the Patterson. When the two possible positions were calculated for the oxygen atom corresponding to the vectors (-.08, .18, .12) and (.08, -.18, -.12) which would give rise to the sixth peak listed in Table XII, it was found that all of

the heavy atom-oxygen vectors were present for both possible positions of the oxygen atom. This was very satisfying to find because it explained this peak being approximately twice as large as the other Ca-O peaks and also explained why it could not be satisfactorily interpreted as a heavy atom-heavy atom peak. The unusually large magnitude of the Ca-Ca peak could now be explained also in terms of the Ca-O bonding. With two Ca-O bonds being linear and of the same length, two of the O-O vectors are necessarily parallel to the Ca-Ca vector and have the same length. This would cause two of the peaks due to these two O-O vectors to lie on top of the Ca-Ca peak.

Thus, by assuming heavy atom solution 33 of the Patterson function was the correct one, the positions of seven oxygen atoms surrounding the Ca ion could be found and these positions were verified by finding all of the heavy atom oxygen vectors in the Patterson map. At the same time, by including the oxygen atoms in the interpretation of the Patterson function, the ambiguities in the peak heights for solution 33 were resolved thereby eliminating the major objection to this solution. With this success in further interpreting solution 33 of the Patterson function, it was considered highly probable that this was actually the correct solution.

The positions of the seven oxygens found from the Patterson function that give rise to the six vectors

listed in Table XII are listed in Table XIII.

Table XIII

Positions of Oxygen Atoms

x	y	z
.12	.96	.87
.34	.92	.01
.66	.30	.77
.66	.08	.99
.26	.30	.73
.34	.32	.99
.50	.96	.75

Looking at the positions calculated for the oxygen atoms it is noticed that the second and fourth oxygen atoms listed in Table XIII are related by a center of symmetry. Having two centrosymmetrically related oxygen atoms bonded to the same Ca ion requires that these two oxygen atoms be shared by another Ca ion which is related by a center of symmetry to the first. Finding two atoms being shared in bonding to two metal atoms is a fairly common occurrence.

Trial Structure and Refinement

Structure factors were calculated based on the positions of the Ca ion, 2Cl ions and six oxygen atoms giving an R index value of .42 as compared to the value of

.55 obtained when the positions of the Ca ion and two Cl ions were used. A Fourier series calculation was then made using the signs of the structure factors based on the positions of the Ca ion, 2 Cl ions, and 6 O atoms and the amplitudes of the structure factors obtained from intensity measurements.

This Fourier series, which represents the electron density in the unit cell, gave peaks for all the rest of the atoms in the structure. The positions of 6 C's and 3 N's of the peptide and the O of one water molecule were found from the electron density map. A trial structure was thus obtained since all of the atoms of $CaCl_2(ggg)\cdot 3H_2O$ had been found. The first structure factor least squares cycle with all atoms given isotropic temperature factors of 3.00 resulted in an R value of .24. Structure factor least squares cycles using isotropic temperature factors for all atoms reduced the R-value to .18. At this point the Ca, 2Cl's and 7 O's were given anisotropic temperature factors and the structure was further refined to an R-value of .13. Refining the structure further with all atoms given anisotropic temperature factors did not significantly lower the R-value.

Having reached a point where the structure was no longer refining, it was decided to make anomalous dispersion corrections for the scattering factors of the Ca and Cl ions. The dispersion effects are taken into account by representing

the scattering factors for Ca and Cl as

$$f = fo + \delta_1 + i\delta_2$$

where fo is the ordinarily tabulated scattering factor for the atom. The value of fo is calculated on the assumption that the scattering power of an electron in an atom is the same as that of a free electron. δ_1 and δ_2 are the real and imaginary parts of the dispersion correction for the scattering factor which are due to the tightly bound electrons that are concentrated in a small volume near the nucleus.

The effect of δ_1 and δ_2 on the structure factors for the centrosymmetric case can be seen from the structure factor equation.

$$F_{hkl} = 2 \sum_j f_j \cos 2\pi(hx_j+ky_j+lx_j)\ T_j$$

where T_j is the exponential temperature factor expression for the j^{th} atom. Substituting in $f_j = fo_j + \delta 1_j + i\ \delta 2_j$ for the scattering factor of the j^{th} atom, we get

$$F_{hkl} = 2 \sum_j fo_j \cos 2\pi(hx_j+ky_j+lz_j)\ T_j$$

$$+ \sum_r 2\ \delta 1_r \cos 2\pi(hx_r+ky_r+lz_r)\ T_r$$

$$+ i \sum_r 2\ \delta 2_r \cos 2\pi(hx_r+ky_r+lz_r)\ T_r$$

where the first term gives the calculated value of the structure factor which is due to the non-dispersive parts

of the scattering factors and the last two terms are due to the dispersive parts of the scattering factors, the sum r being over the dispersive atoms. From this expression it is noted that the calculated structure factor F_{hkl} is a complex quantity for a centrosymmetric crystal when there are dispersive atoms present. The structure factor for a given reflection can then be written

$$F = A + \sum_r \delta 1_r A_r + i \sum_r \delta 2_r A_r$$

where

$$A = 2 \sum_j fo_j \cos 2\pi(hx_j+ky_j+lz_j) T_j$$

$$A_r = 2 \cos 2\pi(hx_r+ky_r+lz_r) T_r$$

the sum r being over the dispersive atoms in the structure and the sum j being over all of the atoms. The calculated value of $|F|^2$ is given by

$$|F|^2 = (A + \sum_r \delta 1_r A_r)^2 + (\sum_r \delta 2_r A_r)^2$$

$$|F|^2 = A^2 + 2A \sum_r \delta 1_r A_r + \sum_r \sum_s (\delta 1_r \delta 1_s + \delta 2_r \delta 2_s) A_r A_s$$

The calculated value of $|F|^2$ is then represented by the right hand side of the last equation and not by A^2 which is assumed in most structure factor least squares calculations. It should also be pointed out that the value of $|F|^2$ depends on δ_2 as well as δ_1 and, therefore, dispersion effects cannot be completely corrected for by simply adding

the δ_1 values to the scattering factors of the dispersive atoms.

The derivation given above is mathematically correct but not physically correct. The phase of a structure factor for a centrosymmetric crystal is real and not complex and in addition the diffracted intensity related to $|F|_o$ is affected by the anomalous dispersion and not A. It is therefore physically more meaningful to use A for the calculated structure factor and to correct $|F|_o$ for the anomalous dispersion instead (8).

This can be done by subtracting from $|F|_o^2$ the two terms involving δ_1 and δ_2 in the previous equation, i.e.,

$$|A|_o^2 = |F|_o^2 - 2 A_c \sum_r \delta 1_r A_r - \sum_r \sum_s (\delta 1_r \delta 1_s + \delta 2_r \delta 2_s) A_r A_s$$

This can only be done in the final stages of the refinement because it requires values for A_c and A_r (for all dispersive atoms). The least squares refinement can now proceed using A and A_o for all reflections.

In the structure of $CaCl_2(ggg)\cdot 3H_2O$ there are three dispersive atoms, the Ca and two Cl's. The values of δ_1 and δ_2 for Ca are .3 and 1.4 electrons, respectively, and for Cl .3 and .7 electrons, respectively (9). The new effective $|F|_o$ values which are given by $|A|_o$ were calculated and found to be smaller than the original $|F|_o$ values in nearly all cases, with most corrections in the

neighborhood of 3 or 4% of the $|F|_o$ values. In looking at the calculated values for the two correction terms, the first term is always much larger than the second. The reason that the corrected values are nearly always smaller than the measured values can be seen by looking at the first correction term. In the first term, A_c carries the proper sign of the structure factor and is equal to the calculated structure factor with dispersive effects not included. The term $\sum_r \delta l_r A_r$ is a sum over the dispersive atoms which are the heavy atoms in the structure and because δ_l is positive and has the same value for Ca and Cl, this sum will have the same sign as the structure factor would have based only on the heavy atoms. Since the signs of A_c and $\sum_r \delta l_r A_r$ are nearly always the same, the term $2Ac \sum_r \delta l_r A_r$ will nearly always be positive and will, therefore, make $|A|_o$ less than $|F|_o$.

After making anomalous dispersion corrections on the observed structure factors, a difference Fourier map was calculated in order to try to find the positions of the hydrogen atoms. Peaks were found in the difference Fourier map which corresponded to 11 of the 17 hydrogen atoms present in the structure. The positions of these hydrogen atoms are listed in Table XIV.

The structure was further refined with the hydrogen atoms given isotropic temperature factors and all other atoms given anisotropic temperature factors. The final

Table XIV

Positions of Hydrogen Atoms

Bonded to	x	y	z
N (2)	.31	.37	.34
$O_w(2)$	.25	.27	.66
$O_w(2)$	.21	.41	.72
N (3)	.01	.67	.83
N (1)	.21	.86	.38
$C^{\alpha}(1)$	.08	.57	.40
C'(1)	.29	.54	.46
$C^{\alpha}(2)$	.49	.32	.18
C'(2)	.48	.49	.14
$C^{\alpha}(3)$	.98	.29	.95
C'(3)	.82	.36	.02

weighting scheme used in the structure factor least squares calculations is given by the two formulas

$$\sqrt{w} = |Fo|/P_1 \quad \text{if} \quad |Fo| \leq P_1$$

or

$$\sqrt{w} = P_1/|Fo| \quad \text{if} \quad |Fo| > P_1$$

where P_1 was given the value of 15.0. For the observed data, up to a 2θ value of 145°, the final R value was .094 and all of the positional and temperature parameter shifts were less than 1/5 their estimated standard deviations. The relatively high value of the discrepancy index

R was attributed to the errors introduced by using four different crystals to obtain the intensity data and to the high value for the linear absorption coefficient. The final positional and anisotropic temperature factor parameters along with their estimated standard deviations are given in Tables XV and XVI.

Vibrational ellipsoids were calculated for the anisotropic atoms and the lengths and direction cosines of their principal axes are given in Table XVII.

The final parameters listed in Tables XV, XVI, and XVII were obtained from the refinement using structure factor data up to a 2θ value of 145°. As a matter of interest the refinement was also carried out using structure factor data only to to a 2θ value of 110°. The final positional parameters along with their estimated standard deviations for this limited amount of data are given in Table XVIII. The values of these parameters all lie within the estimated standard deviations of those obtained by using all of the data in the refinement. Bond angles and distances calculated using these positional parameters are given in Table XX. These also lie within the estimated standard deviations of those obtained by using data up to 145° (see Table XXIII).

The final values for the calculated and observed structure factors for all of the data are given in Table XIX. Atomic scattering factors used in the structure factor calculations were taken from the International Tables for X-ray Crystallography (9).

A final Fourier electron density map was calculated and the peak heights of the atoms are listed in Table XXI.

The final difference Fourier map was calculated with the eleven hydrogen atoms not included in the calculated structure factors. These hydrogens reappeared in the difference Fourier with peak heights ranging from .6 to 1.2 electrons/$\mathring{A}^3$. Several spurious peaks were found with heights up to .7 electrons/$\mathring{A}^3$. Four regions of negative density were found in the difference Fourier with values ranging from -.8 to -1.2 electrons/$\mathring{A}^3$. One of the regions was in the vicinity of a Cl ion, however, the other regions of negative electron density seemed to be spurious.

Table XV

Final Atomic Positions

Atom	x	y	z
Ca	.4161(3)*	.1418(2)	.8719(2)
Cl(1)	.2934(6)	.2057(4)	.4999(3)
Cl(2)	.8233(4)	.3331(3)	.3176(2)
N(1)	.2571(17)	.7706(11)	.4327(8)
$C^{\alpha}(1)$	.2246(17)	.6079(12)	.4045(8)
C'(1)	.3008(14)	.5888(12)	.2894(8)
O (1)	.3333(11)	.7001(8)	.2241(6)
N (2)	.3234(12)	.4398(9)	.2685(6)
$C^{\alpha}(2)$	.3950(14)	.4003(11)	.1667(9)
C'(2)	.2577(13)	.3450(10)	.0771(8)
O (2)	.3178(10)	.3192(8)	-.0119(5)
N (3)	.0675(12)	.3267(8)	.0987(6)
$C^{\alpha}(3)$	-.0785(15)	.2821(12)	.0214(8)
C'(3)	-.1744(15)	.1192(11)	.0458(7)
O (3)	-.3273(10)	.0854(8)	-.0074(6)
O (4)	-.1130(10)	.0258(8)	.1197(5)
$O_w(1)$	.1382(11)	.0315(9)	.2930 (7)
$O_w(2)$	.2315(13)	.2939(9)	.7392(6)
$O_w(3)$	.4693(13)	.9633(9)	.7408(7)

*Standard deviations are given in parentheses.

Table XVI

Final Anisotropic Thermal Parameters and Their Standard Deviations (in parentheses)*

Atom	B_{11}	B_{22}	B_{33}	B_{23}	B_{13}	B_{12}
Ca	.0143(4)	.0046(2)	.0030(1)	.0006(3)	-.0007(4)	-.0005(5)
Cl(1)	.0417(12)	.0141(5)	.0049(2)	-.0009(5)	-.0048(8)	.0011(12)
Cl(2)	.0183(6)	.0108(4)	.0046(2)	-.0021(4)	.0023(5)	-.0009(8)
N (1)	.0398(35)	.0086(14)	.0047(7)	-.0009(16)	.0026(25)	.0020(35)
C^{α}(1)	.0210(28)	.0080(15)	.0044(7)	.0015(17)	.0017(23)	-.0046(32)
C'(1)	.0123(22)	.0092(15)	.0041(7)	-.0027(16)	-.0067(20)	-.0042(29)
O (1)	.0189(18)	.0068(10)	.0048(5)	.0040(11)	.0029(15)	-.0040(21)
N (2)	.0166(21)	.0080(13)	.0024(5)	-.0018(13)	.0002(16)	-.0037(25)
C^{α}(2)	.0129(23)	.0066(14)	.0060(8)	.0018(17)	-.0033(22)	.0038(28)
C'(2)	.0117(21)	.0039(12)	.0042(7)	-.0003(14)	-.0033(19)	.0007(24)
O (2)	.0142(16)	.0101(11)	.0041(5)	-.0064(12)	.0060(14)	-.0016(21)
N (3)	.0215(22)	.0028(10)	.0029(5)	.0028(11)	-.0069(17)	-.0050(23)
C^{α}(3)	.0141(24)	.0083(15)	.0048(7)	.0018(17)	-.0041(21)	-.0093(29)
C'(3)	.0187(26)	.0073(14)	.0030(6)	.0002(15)	.0045(21)	-.0035(30)
O (3)	.0127(16)	.0101(11)	.0069(6)	.0056(13)	-.0059(16)	-.0127(21)
O (4)	.0168(17)	.0064(9)	.0046(5)	.0035(11)	-.0038(14)	-.0021(20)
O_W(1)	.0184(19)	.0140(13)	.0071(6)	-.0051(15)	-.0081(18)	.0154(25)
O_W(2)	.0295(24)	.0108(13)	.0052(6)	-.0009(13)	-.0040(19)	.0103(27)
O_W(3)	.0254(22)	.0089(12)	.0080(7)	-.0047(14)	.0052(20)	-.0023(25)

*Temperature factor = $\exp-(h^2B_{11}+k^2B_{22}+l^2B_{33}+hkB_{12}+hlB_{13}+klB_{23})$.

Table XVII

Lengths and Direction Cosines With Respect to Unit Cell Axes of the Principal Axes of the Thermal Ellipsoids

Atom	Bi($\AA^2$)	l_1	l_2	l_3
Ca	2.86	.970	-.220	-.227
	1.81	.236	.388	.903
	1.23	.052	.895	-.365
Cl(1)	8.27	.986	-.129	-.188
	4.07	.053	.992	.050
	2.70	.159	.000	.981
Cl(2)	3.86	.831	-.504	.281
	2.95	.499	.821	-.052
	2.51	-.244	.268	.958
N (1)	7.70	.997	-.059	.050
	2.72	-.075	.026	.997
	2.50	-.014	.998	.050
$C^{\alpha}(1)$	4.40	.945	-.397	-.030
	2.93	.164	.506	.869
	1.78	.284	.766	-.494
C'(1)	3.71	-.757	.382	.607
	2.89	-.046	.839	-.482
	.93	.652	.388	.631
O (1)	3.94	.949	-.384	-.033
	3.50	.163	.518	.862
	1.18	.268	.764	-.506
N (2)	3.58	.900	-.496	.012
	2.08	.427	.811	-.271
	1.32	.082	.310	.962

Table XVII--Continued

Atom	Bi(Å^2)	l_1	l_2	l_3
$C^{\alpha}(2)$	3.92	-.377	.262	.920
-	2.52	.788	.529	.198
	1.51	-.486	.807	-.339
C'(2)	2.99	-.664	.111	.765
	1.76	.748	.007	.644
	1.12	-.006	.994	-.016
O (2)	4.30	-.529	.609	-.572
	2.44	.741	.614	.015
	1.13	-.414	.502	.820
N (3)	5.05	.896	-.294	-.420
	1.36	.444	.410	.796
	.49	.007	.863	-.436
$C^{\alpha}(3)$	4.61	-.681	.606	.531
	2.29	.320	-.381	.841
	1.33	.649	.698	.101
C'(3)	4.00	.931	-.307	.235
	2.20	.083	.867	.535
	1.36	-.356	-.392	.812
O (3)	6.31	-.513	.620	.682
	2.51	-.460	.465	-.731
	1.09	.724	.632	-.017
O (4)	4.22	.716	-.437	-.627
	2.44	.697	.387	.578
	1.34	-.039	.812	-.523

Table XVII--Continued

Atom	Bi(Å^2)	l_1	l_2	l_3
$O_w(1)$	6.22	.575	.518	-.567
	3.28	.067	.684	.765
	1.94	-.815	.514	-.305
$O_w(2)$	6.06	.928	.226	-.228
	3.11	-.037	.676	.783
	2.59	-.370	.701	-.578
$O_w(3)$	5.78	.724	-.261	.622
	4.00	-.686	-.072	.724
	2.30	.067	.963	.298

Table XVIII

Atomic Positions Using Data up to 110°

Atom	x	y	z
Ca	.4160(3)	.1418(2)	.8719(2)
Cl(1)	.2933(6)	.2054(4)	.5000(3)
Cl(2)	.8236(4)	.3330(3)	.3176(2)
N (1)	.2569(18)	.7706(12)	.4325(8)
$C^{\alpha}(1)$	.2248(17)	.6077(13)	.4045(9)
C'(1)	.3006(14)	.5892(12)	.2891(8)
O (1)	.337(11)	.7004(8)	.2242(6)
N (2)	.3238(13)	.4400(10)	.2683(6)
$C^{\alpha}(2)$	.3947(15)	.3999(12)	.1667(9)
C'(2)	.2579(15)	.3445(11)	.0775(8)
O (2)	.3173(10)	.3191(8)	-.0120(6)
N (3)	.0673(13)	.3271(9)	.0989(7)
$C^{\alpha}(3)$	-.0788(16)	.2822(13)	.0210(9)
C'(3)	-.1757(16)	.1195(12)	.0457(8)
O (3)	-.3270(11)	.0853(9)	-.0076(6)
O (4)	-.1131(11)	.0262(8)	.1197(6)
$O_W(1)$	.1381(12)	.0314(10)	.2933(7)
$O_W(2)$	.2310(13)	.2942(10)	.7390(6)
$O_W(3)$	.4693(13)	.9633(9)	.7406(7)

TABLE XIX

TABLE XIX (Continued)

Table XX

Bond Lengths and Angles in Tripeptide Molecule
Using Data up to 110°

Bond	*Length	Angle	*Degrees
N (1)-C^{α}(1)	1.459(15)Å	N (1) C^{α}(1) C' (1)	110.3(9)°
C^{α}(1)-C'(1)	1.533(15)	C^{α}(1) C'(1) O (1)	121.7(9)
C'(1)-O (1)	1.201(13)	C^{α}(1) C'(1) N (2)	114.1(9)
C'(1)-N (2)	1.342(13)	O (1) C'(1) N (2)	124.2(9)
N (2)-C^{α}(2)	1.416(13)	C'(1) N (2) C^{α}(2)	122.3(8)
C^{α}(2)-C'(2)	1.497(15)	N (2) C^{α}(2) C' (2)	120.2(9)
C'(2)-O (2)	1.213(12)	C^{α}(2) C'(2) O (2)	120.3(9)
C'(2)-N (3)	1.355(14)	C^{α}(2) C'(2) N (3)	118.3(9)
N (3)-C^{α}(3)	1.417(14)	O (2) C'(2) N (3)	121.3(9)
C^{α}(3)-C'(3)	1.515(15)	C'(2) N (3) C^{α}(3)	124.4(8)
C'(3)-O (3)	1.246(13)	N (3) C^{α}(3) C' (3)	114.6(9)
C'(3)-O (4)	1.246(12)	C^{α}(3) C'(3) O (3)	117.0(9)
		C^{α}(3) C'(3) O (4)	121.8(9)
		O (3) C'(3) O (4)	121.0(1.0)

Table XXI

Electron Densities of the Atoms

Atom	Electron Density (electrons/Å^3)
Ca	38.8
Cl(1)	22.3
Cl(2)	28.0
N (1)	7.7
C^{α}(1)	7.6
C'(1)	7.9
O (1)	11.8
N (2)	10.7
C^{α}(2)	8.5
C'(2)	9.3
O (2)	12.5
N (3)	10.4
C^{α}(3)	8.6
C'(3)	8.8
O (3)	11.7
O (4)	12.6
O_w(1)	9.3
O_w(2)	10.6
O_w(3)	10.1

CHAPTER IV

DISCUSSION OF STRUCTURE OF $CaCl_2(ggg)\cdot 3H_2O$

Packing of the Molecules

In order to facilitate the discussion of the structure of $CaCl_2(ggg)\cdot 3H_2O$, the following terms and notation are used. The symbols ggg will be used to denote the tripeptide molecule, glycylglycylglycine. The notation used in the labeling of atoms in the peptide molecule is that proposed during the 1965 Gordon Conference on Proteins. The term "residue" is reserved for the group of atoms -N-C $-\overset{O}{\overset{\|}{C}}'-$. Residues are numbered starting with the terminal N. The number of the residue to which an atom belongs is given in parentheses after the symbol for the atom. With this notation, for example, the atoms of the first peptide residue in ggg are denoted N(1) C^{α}(1) C'(1) O(1). The term "peptide group" is reserved for the sequence of atoms $C^{\alpha}C'$ O N (C^{α}).

The two Cl atoms are labeled Cl(1), Cl(2), and the oxygen atoms of the three water molecules are labeled $O_w(1)$, $O_w(2)$, and $O_w(3)$.

When necessary to distinguish between atoms in

different peptide molecules the subscripts A, B, etc., will be used, i.e., $N(1)_A$ will denote N(1) in molecule A. In referring to symmetry related Cl atoms or water molecules, they will be represented as Cl(1), Cl(1)' or $O_w(1)$, $O_w(1)'$, etc. The coordinates of symmetry equivalent atoms using this notation are given in Table XXII.

Table XXII

Subscript	Atomic Coordinates		
A	x	y	z
B	1-x	1-y	1-z
C	-x	1-y	1-z
D	x+1	y+1	z
E	x+1	y	z
Superscript unprimed	x	y	z
'	1-x	1-y	1-z
''	x	y+1	z
'''	x-1	y	z

In the structure of $CaCl_2(ggg)\cdot 3H_2O$ each Ca ion is bonded to oxygens from four different peptide molecules, and is therefore very important in the packing of the peptide molecules. As shown in Figure 2, the peptide oxygens bonded to Ca(1) are $O(1)_A$, $\bar{O}(3)_C$, $O(3)_D$, $O(4)_D$, and $O(2)_B$.

The peptide molecules are also linked together via hydrogen bonding to the Cl ions. Both N-H groups and the terminal NH_3^+ group of the peptide molecule, which is in the

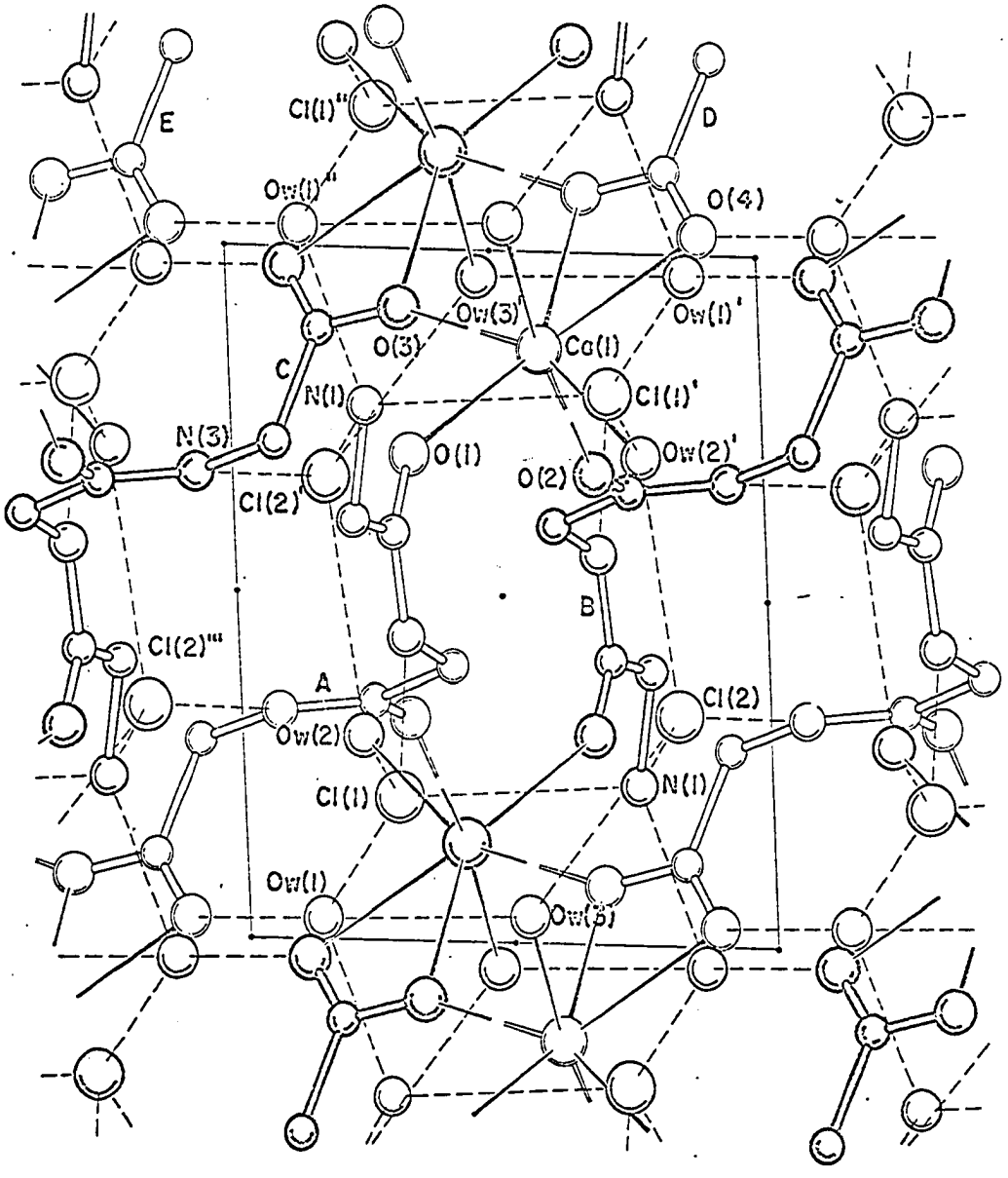


Figure 2

ab Projection of the Structure of

$CaCl_2$ (ggg)· $3H_2O$

form of the zwitterion, are involved in hydrogen bonding to Cl ions. The atom N(2) of molecule A is hydrogen bonded to Cl(2)'''. The terminal NH_3^+ group of molecule A($N(1)_A$) is hydrogen bonded to two Cl ions, Cl(1)' and Cl(2)', which are symmetry related to Cl(1) and Cl(2), respectively. $N(1)_A$ is also hydrogen bonded to the water molecule $O_w(1)''$. The net effect of the hydrogen bonding of the Cl ions is that there is a region in the structure where the peptide molecules are held together chiefly by hydrogen bonding to the Cl ions.

In addition to the N-H----Cl hydrogen bonding the Cl ions are stabilized by hydrogen bonding to the three water molecules in the structure. The atom Cl(1), which is hydrogen bonded to $N(2)_A$ and $N(1)_B$, is also hydrogen bonded to the water molecules $O_w(1)$ and $O_w(2)$. Likewise, Cl(2)' forms hydrogen bonds to $N(1)_A$, $N(3)_C$, $O_w(2)$, and $O_w(3)$'. Each Cl ion, therefore, accepts hydrogen bonds from two water molecules, the terminal NH_3^+ group of one peptide molecule, and the N-H group of another peptide molecule.

The water molecule ($O_w(1)''$) which isn't bonded to the Ca ion forms hydrogen bonds linking two different peptide molecules, A and E. It accepts a hydrogen bond from $N(1)_A$ in addition to one from $O_w(3)''$ and is donor in the hydrogen bonds to $O(4)_E$ and the chloride ion Cl(1)''.

The peptide molecules are held together by means of

this complex hydrogen bond network which makes use of all the available hydrogens for hydrogen bonding, and by means of the bonding of the Ca ion to the oxygens of different peptide molecules. There is no intramolecular or intermolecular hydrogen bonding between the peptide molecules themselves as would be required for the helical or pleated sheet peptide configurations.

Geometry of Peptide Molecule

Bond lengths in the peptide molecule listed in Table XXIII and in Figure 3 show significant deviations from the standard values for free peptides given by Pauling and Corey (1) and the later revised values given by Marsh and Donohue (11). The C^{α}-C' distances of 1.53, 1.51, and 1.51Å are in close agreement with the standard value of 1.51Å given by Marsh and Donohue, however, the N-C^{α}, C'-O, and C'-N distances all seem to be significantly different from the Marsh and Donohue values. Except for the terminal C^{α}-NH_3^+ distance, the N-C^{α} distances of 1.41 and 1.41Å are shorter than the standard value of 1.45Å. The C'-N bond lengths of 1.34 and 1.35Å are consistently longer than the standard value of 1.32Å and the C'-O bond lengths of 1.21 and 1.21Å are consistently shorter than the standard value of 1.24Å. With the estimated standard deviations in the bond lengths of slightly larger than .01Å, these distances are several standard deviations away from the standard

Table XXIII

Bond Lengths and Angles in Tripeptide Molecule

Bond	*Length	Angle	*θ
	Å		
N (1)-C^{α}(1)	1.457(14)	N (1) C^{α}(1) C'(1)	110.5(9)°
C (1)-C'(1)	1.529(14)	C^{α}(1) C'(1) O (1)	121.5(9)
C'(1)-O (1)	1.205(12)	C^{α}(1) C'(1) N (2)	114.3(8)
C'(1)-N (2)	1.339(12)	O (1) C'(1) N (2)	124.2(9)
N (2)-C^{α}(2)	1.414(13)	C'(1) N (2) C^{α}(2)	122.0(8)
C^{α}(2)-C'(2)	1.506(14)	N (2) C^{α}(2) C'(2)	120.1(8)
C'(2)-O (2)	1.208(11)	C^{α}(2) C'(2) O (2)	120.1(9)
C'(2)-N (3)	1.351(13)	C^{α}(2) C'(2) N (3)	118.5(8)
N (3)-C'(3)	1.414(13)	O (2) C'(2) N (3)	121.4(8)
C^{α}(3)-C'(3)	1.513(14)	C'(2) N (3) C^{α}(3)	124.8(8)
C'(3)-O (3)	1.256(12)	N (3) C^{α}(3) C'(3)	114.5(8)
C'(3)-O (4)	1.244(12)	C^{α}(3) C'(3) O (3)	116.7(8)
		C^{α}(3) C'(3) O (4)	122.2(9)
		O (3) C'(3) O (4)	121.0(9)

Average values in peptides and metal-peptide complexes

	Peptides (Marsh and Donohue)	Metal-complexes (Freeman)
Bond	Length	Length
	Å	Å
N -C^{α}(terminal)	1.49	1.49
C^{α}-C'	1.51	1.53
C'-O	1.24	1.26
C'-N	1.325	1.30
N -C^{α}	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 given in parentheses.

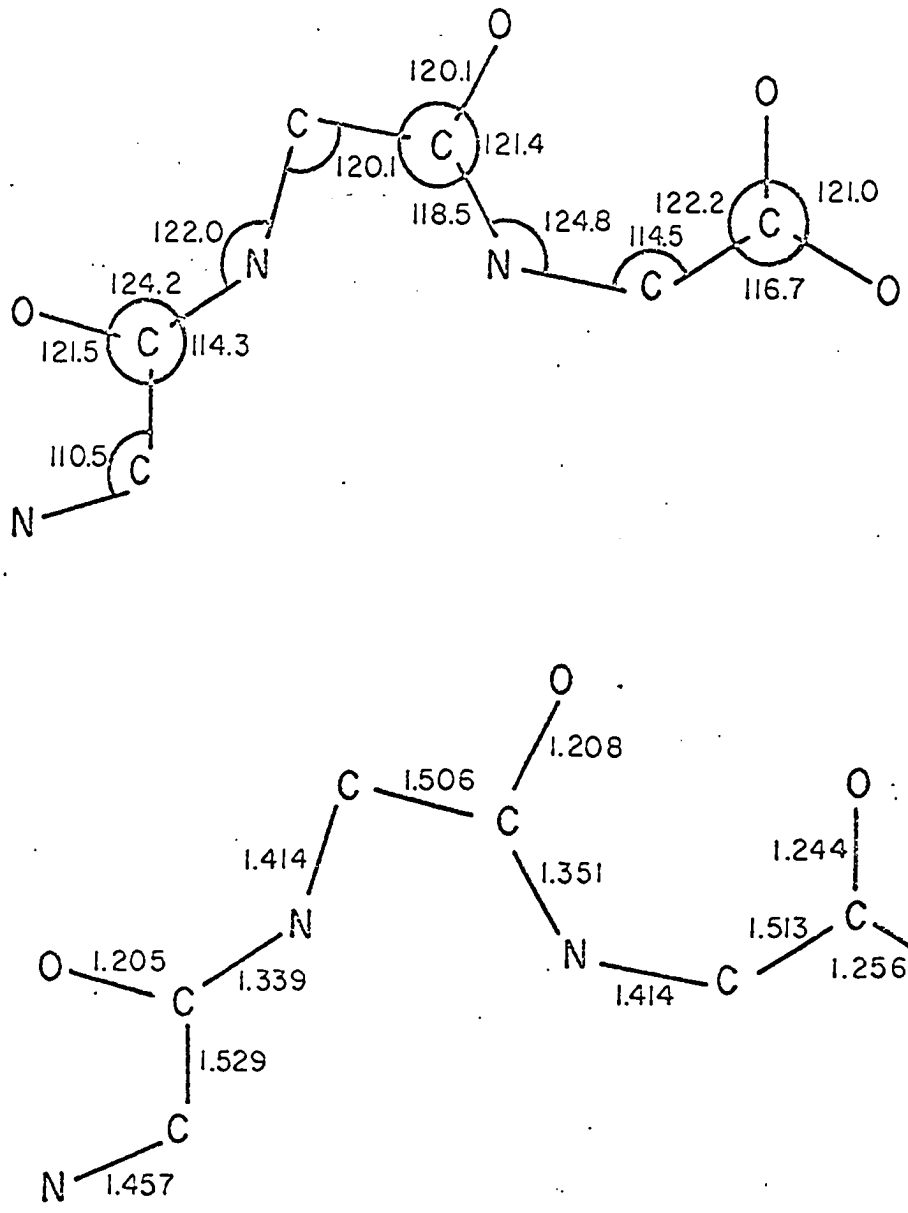


Figure 3. Bond Lengths and Angles in Tripeptide Molecule

values and are considered to be significant.

The shortness of the C'-N bond in peptide linkages, 1.32Å, can be attributed to partial double bond character as can be shown from the two resonance forms in Figure 4.

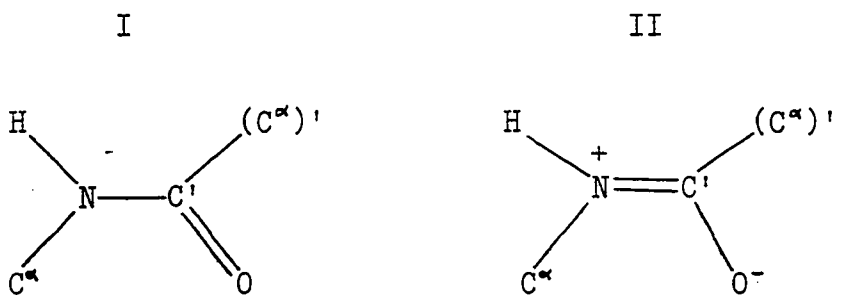


Figure 4

To explain the longer than normal C'-N distances and the shorter than normal C'-O distances in the Ca complex, it is assumed that resonance form I contributes more to the actual structure than is normally the case when the peptide is not complexed. This is the exact opposite of the trend in the Cu-peptide complexes where the average C'-N distance is 1.30Å and the average C'-O distance is 1.26Å indicating a greater than normal contribution of resonance form II (12). Other transition metal complexes of peptides also show this trend. It should be noted that in the Ca-peptide complex both peptide oxygens are bonded to Ca ions. This bonding would be expected to give more ionic character to these oxygens, the reverse of what is observed.

There seems to be no adequate explanation for the

shorter than normal N-C^{α} bond lengths of the second and third residues. However, with the C'-N bond showing more single bond character than it normally does in the free peptide, it might be expected that the C'-N bond and the N-C^{α} bond would be more nearly equivalent than in the free peptide. Although the N-C^{α} distance of 1.46Å involving the terminal NH_3^+ group is appreciably longer than the other N-C^{α} distances, it is also shorter than its average value of 1.49Å found in other structures.

Bond angles in the peptide molecule listed in Table XXIII and shown in Figure 3 are in fairly close agreement with the standard values of Pauling and Corey and the average values of Marsh and Donohue. The only significant differences are in the N-C^{α}-C' angles of the second and third residues. These are 120° and 115°, respectively, and are significantly larger than the Marsh and Donohue value of 111°. The N-C^{α}-C' angle involving the terminal NH_3^+ group has the normal value of 111°.

Since variety in peptide structures is achieved by rotation about the N-C^{α} and C^{α}-C' bonds, the large N-C^{α}-C' angles for the second and third residues give the peptide groups increased rotational freedom to utilize hydrogen bonding and Ca-O bonding more efficiently. Perhaps, as a result of this increased rotational freedom, the second residue has a somewhat unusual configuration in that the nitrogen atom is lying trans to the carbonyl oxygen and is

therefore cis to the nitrogen of the third residue. Marsh and Leung have pointed out that in small peptides there is a strong tendency for the amino or peptide nitrogen atoms to be as close as possible to the carbonyl or carboxyl oxygen atoms of the same residue, and hence, to have a small ψ angle (see discussion on conformational angles) (13). In the present case the peptide nitrogens are involved in H-bonding to the Cl ions and the carbonyl oxygen is bonded to a Ca ion. Although this configuration is unusual, it is thought to have occurred in order to utilize more fully the hydrogen bonding capabilities of both peptide nitrogens (N(2) and N(3)) and leave the carbonyl oxygen (O(2)) more available for bonding to the Ca ion.

The cis configuration for the peptide nitrogens of adjacent residues is also found in $NaCu(ggg)\cdot H_2O$ and in the Na, Cu complexes of the higher peptides gggg and ggggg (14,15). Here the cis configuration is more of a necessity in that the peptide nitrogens of adjacent residues have been deprotonated and are coordinated to the same Cu atom. In these structures, however, there seem to be no significant deviations of the N-C^{α}-C' angles from their normal value of 111°.

Looking at these structures, it appears as if the trans configuration of the nitrogen and carbonyl oxygen of the same residue, and hence, the cis configuration of the nitrogens of adjacent residues, which is unfavored in small

peptides and amino acids is favored in some complexes and chelates where it gives a better geometry for metal coordination or a more efficient use of hydrogen bonding or both.

As was mentioned before, the only significant deviations of the bond angles from those found in the free peptides are in the N-C^{α}-C' angles. It is surprising that smaller C'-N-C^{α} angles are not found in the Ca complex with the C'-N bond showing more single bond character than it normally does in the free peptide. The C'-N-C^{α} angles for the two peptide groups are 122° and 125° as compared to the average Marsh and Donohue value of 122°.

It is interesting to compare the bond lengths and angles in the Ca-peptide complex with those found in the free peptides and in the transition metal peptide complexes. The bond lengths and angles in the transition metal complexes show only two significant deviations from those found in the free peptides. The carbonyl oxygen distance (C'-O), whose average is 1.26Å in transition metal complexes, is longer than the average value of 1.24Å found in free peptides and the average C'-N distance of 1.30Å is shorter than the average value of 1.32Å found in free peptides. In the Ca-ggg complex there are significant deviations in the C'-O, C'-N, and N-C^{α} bond distances and the NC^{α}C' angle. The C'-O distances of 1.21 and 1.21Å are shorter than the average peptide value of 1.24Å and the C'-N distances of

1.34 and 1.35Å are longer than the average peptide value of 1.32Å. These deviations are the exact opposite of those found in the transition metal complexes. The non-terminal N-C^{α} bond distances of 1.41 and 1.41Å are shorter than the average peptide value of 1.45Å and the non-terminal N-C^{α}-C' angles of 120° and 115° are larger than the average peptide value of 111°.

Whether these deviations in the bond lengths and angles of the peptide are characteristic of those in Ca-peptide complexes and in general non-transition metal complexes, or are just a peculiarity of this particular structure will have to await the determination of other structures in this area.

In addition to the constancy of the bond lengths and angles of the peptide, the most important characteristic which is common to the peptides whose structures have been determined is the approximate planarity of the peptide groups. Table XXIV gives the equations of the best least squares planes for the two peptide groups and also the $C^{\alpha}C'ON$ and $C^{\alpha}C'OO$ groups (16).

The two $C^{\alpha}C'ON$ groups (Plane I and Plane III) are certainly planar within the limits of accuracy. For the first peptide group composed of atoms $C^{\alpha}(1)$ C'(1) O(1) N(2) $C^{\alpha}(2)$ (Plane II), the largest deviation of an atom from the best plane is .011Å. With the estimated standard deviations of the coordinates of these atoms being

Table XXIV

Least Squares Planes Through Peptide Molecule

Plane	Atoms to which plane was fitted	*Equation of Plane
I	$C^{\alpha}(1)C'(1)O(1)N(2)$	6.4475x+ .4557y+4.1208z=3.3941
II	$C^{\alpha}(1)C'(1)O(1)N(2)C^{\alpha}(2)$	6.4347x+ .4924y+4.1651z=3.4265
III	$C^{\alpha}(2)C'(2)O(2)N(3)$	-1.1504x+8.0329y-3.1517z=2.2346
IV	$C^{\alpha}(2)C'(2)O(2)N(3)C^{\alpha}(3)$	-1.0611x+8.0529y-3.1153z=2.2718
V	$C^{\alpha}(3)C'(3)O(3)O(4)$	-4.1449x+3.9429y+8.8406z=1.6199

Atom	Deviation from Plane I	Deviation from Plane II	Atom	Deviation from Plane III	Deviation from Plane IV
$C^{\alpha}(1)$	-.002 Å	.003 Å	$C^{\alpha}(2)$	-.001 Å	-.013 Å
C'(1)	.006	.004	C'(2)	.003	.007
O (1)	-.002	-.004	O (2)	-.001	.002
N (2)	-.002	-.011	N (3)	-.001	.020
$C^{\alpha}(2)$	.022	.007	$C^{\alpha}(3)$	-.054	-.016

Atom	Deviation from Plane V
$C^{\alpha}(3)$	-.006 Å
C'(3)	.022
O (3)	-.008
O (4)	-.008

*x, y, and z are fractional coordinates.

approximately .01Å, this deviation is considered to be insignificant; therefore, the first peptide group is planar within the limits of accuracy.

The second peptide group (Plane IV) composed of atoms $C^{\alpha}(2)$ C'(2) O(2) N(3) $C^{\alpha}(3)$ shows deviations from the best plane of up to .02Å. This is about twice the estimated standard deviations of the coordinates and is probably significant. The departure from planarity for this peptide group is also illustrated by looking at the best plane calculated for the atoms $C^{\alpha}(2)$ C'(2) O(2) N(3) (Plane III). It can be seen that the largest deviation of an atom from this plane is .003Å, but the $C^{\alpha}(3)$ atom, which was not included in this plane, lies .054Å out of the plane. Although approximately planar, it can be concluded that the second peptide group is probably not planar within the limits of accuracy. As Freeman has pointed out, the non-coplanarity of the peptide group $C^{\alpha}C'ON(C^{\alpha})'$ is always due to the atom $(C^{\alpha})'$ lying out of the plane of the amide ($C^{\alpha}C'ON$) group (17). Looking at the best plane calculated for the terminal $C^{\alpha}C'OO$ group (Plane V), it can be seen that this group is also significantly non-planar with deviations from the best plane of up to .022Å.

Coordination of Ca Ion

Each Ca ion is bonded to seven oxygen atoms as can be seen from Figure 2 and Figure 5. Two of the oxygen atoms

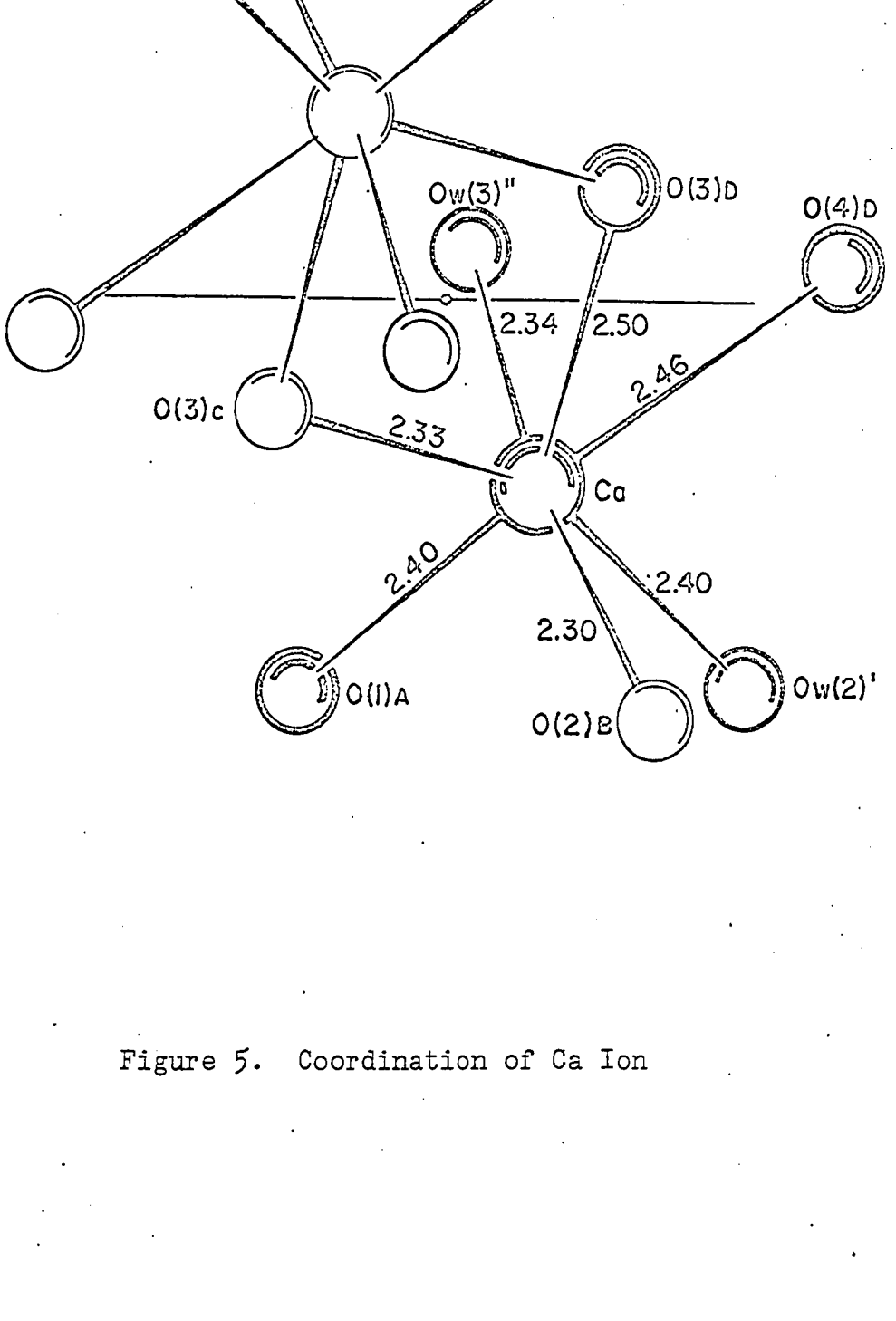


Figure 5. Coordination of Ca Ion

are centrosymmetrically related and are, therefore, shared by a second Ca ion. Seven coordinated Ca ions have been reported elsewhere in the literature, for example, in dicalcium phosphate dihydrate (18), where both seven and eight coordinated Ca ions are present, and in calcium formate (19). Ca ions exhibiting coordination numbers of anywhere from six to nine have been reported in the structures of phosphates and silicates.

In the structure of $CaCl_2(ggg)\cdot 3H_2O$ five of the oxygen atoms which are bonded to the Ca ion form a distorted pentagonal arrangement around the Ca ion, the remaining two are on either side of the plane of this pentagon. The five oxygen atoms which form the pentagonal arrangement consist of a carbonyl oxygen $O(1)_A$ of one peptide molecule at a distance of 2.40Å, an oxygen of a water molecule ($O_W(2)'$) at 2.40Å, both oxygens ($O(3)_D$ and $O(4)_D$) of the carboxylate group of a second peptide molecule at distances of 2.50Å and 2.46Å, and one oxygen ($O(3)_C$) of a carboxylate group of a third peptide molecule at a distance of 2.33Å. Bonded on one side of this plane is a carbonyl oxygen ($O(2)_B$) of a fourth peptide molecule at a distance of 2.30Å and on the other side of the plane an oxygen ($O_W(3)''$) of a water molecule at a distance of 2.34Å. These bond lengths are in the normal range of 2.2 to 2.6Å for Ca-O bond lengths. The various interatomic distances and angles for the metal coordination are given in

Table XXV. The two Ca ions which share in the bonding of two centrosymmetrically related oxygen atoms are 4.00Å apart. It is interesting to note that this is very close to the Ca-Ca bond distance of 3.95Å found in the α-form of Ca metal (20).

There is no bonding of the Ca ions to the Cl ions in the structure of $CaCl_2(ggg)\cdot 3H_2O$. This was noticed in the Patterson function where no heavy atom vectors were found which corresponded to a Ca-Cl bond length of approximately 2.7Å. In this respect the structure is similar to that the cyclic amine complex $CaBr_2\cdot 10H_2O\cdot 2(CH_2)_6N_4$ where $Ca(H_2O)_6^{2+}$ octahedra are formed and no Ca-Br bonds are formed (7). With the Ca ion being surrounded by oxygen atoms of the peptide molecules and water molecules, the Cl ions are used to fill in the many hydrogen bonding sites found at the peptide and amino nitrogen atoms and at the water molecules. In this way the hydrogen bonding of the Cl ions is used to link different peptide molecules together just as the Ca ion does through bonding to the oxygen atoms of different peptide molecules.

Details of Hydrogen Bonding

The hydrogen bonds formed by the terminal NH_3^+ group in $CaCl_2(ggg)\cdot 3H_2O$ are quite similar to the ones formed by the NH_3^+ group in glycyl-L-tyrosine·HCl and DL-glutamic acid·HCl (21,22). In each of these three cases the NH_3^+

Table XXV

Environment of Ca Atom in $CaCl_2(ggg)\cdot 3H_2O$

Bond	Length	Angle	*θ
Ca-O $(4)_D$	2.462(7)Å	O $(4)_D$-Ca-$O_w(2)'$	81.9(3)°
Ca-O $(3)_D$	2.503(8)	$O_w(2)'$-Ca-O $(1)_A$	78.8(3)
Ca-O $(3)_C$	2.329(8)	O $(3)_C$-Ca-O $(1)_A$	81.6(3)
Ca-O $(1)_A$	2.401(7)	O $(3)_C$-Ca-O $(3)_D$	68.5(2)
Ca-$O_w(2)'$	2.395(8)		
Ca-$O_w(3)''$	2.342(8)	O $(2)_B$-Ca-O $(4)_D$	94.8(2)
		O $(2)_B$-Ca-$O_w(2)'$	81.9(3)
Ca(1)-Ca(2)	3.995 (3)	O $(2)_B$-Ca-O $(1)_A$	100.4(2)
		O $(2)_B$-Ca-O $(3)_C$	90.6(3)
		O $(2)_B$-Ca-O $(3)_D$	92.0(2)

Angle	θ
$O_w(3)''$-Ca-O $(4)_D$	79.0(3)°
$O_w(3)''$-Ca-$O_w(2)'$	91.1(3)
$O_w(3)''$-Ca-O $(1)_A$	83.2(3)
$O_w(3)''$-Ca-O $(3)_C$	97.8(3)
$O_w(3)''$-Ca-O $(3)_D$	88.9(3)

*Standard deviations are given in parentheses.

group is hydrogen-bonded to two Cl ions and an oxygen atom. The N-Cl distances for the terminal NH_3^+ group are 3.14Å ($N(1)_A$ to Cl(2)') and 3.21Å ($N(1)_A$ to Cl(1)') in $CaCl_2(ggg)\cdot 3H_2O$. These compare to the N-Cl distances for the terminal NH_3^+ group which are both 3.24Å in glycyl-L-tyrosine·HCl and both 3.18Å in DL-glutamic acid·HCl. The normal range of distances for the NH_3^+ -Cl hydrogen bonds is 3.15 to 3.25Å for a large variety of hydrochloride type structures (23). In $CaCl_2(ggg)\cdot 3H_2O$ the third hydrogen bond formed by the NH_3^+ group is to the oxygen of a water molecule with the N-O distance being 2.86Å ($N(1)_A$ to $O_w(1)''$). This is a normal N-H----O hydrogen bond distance and compares to values of 2.91Å in glycyl-L-tyrosine·HCl and 2.89Å in Dl-glutamic acid·HCl. These are somewhat longer than the amino N-O hydrogen bond distances of 2.83Å in L-leucyl-L-prolyl-glycine and 2.78Å and 2.77Å in glycylphenylalanylglycine (13,24).

The terminal N atom (NH_3^+) is approximately tetrahedrally surrounded by $C^{\alpha}(1)$, two chloride ions, and the oxygen of a water molecule. It is therefore expected that the H atoms are directed towards the Cl ions and the oxygen atom, however, only one of these positions corresponded to a peak in the difference Fourier map. These H-bond distances and angles are given in Table XXVI.

Hydrogen bonds formed between the N-H (peptide nitrogen) groups of the second and third peptide residues

Table XXVI

(a) Hydrogen-bonded contacts X....H....Y in $CaCl_2(ggg)\cdot 3H_2O$

Atom X	Atom Y	*dx-y
$N(1)_A$-H . . .	Cl(1)'	3.205(13)Å
$N(1)_A$-H . . .	Cl(2)'	3.144(10)
$N(1)_A$-H . . .	$O_w(1)''$	2.856(13)
Cl(1) . . .	H-$N(2)_A$	3.325(8)
Cl(1) . . .	H-$O_w(1)$	3.134(9)
Cl(1) . . .	H-$O_w(2)$	3.101(8)
Cl(2)' . . .	H-$N(3)_C$	3.202(8)
Cl(2)' . . .	H-$O_w(2)$	3.263(8)
Cl(2)' . . .	H-$O_w(3)'$	3.239(9)
$O_w(1)''$. . .	H-$O_w(3)''$	2.772(12)
$O_w(1)''$-H. . .	$O(4)_E$	2.692(10)

(b) Bond angles at hydrogen-bond donors and acceptors

*Angle		Angle	
Cl(2)'. . .$N(1)_A$. . .Cl(1)'	87.4(3)°	$N(1)_A$. . .Cl(2)'. . .$N(3)_C$	151.7(3)°
Cl(2)'. . .$N(1)_A$. . .$O_w(1)''$	131.4(4)	$N(1)_A$. . .Cl(2)'. . .$O_w(3)'$	86.3(2)
$C^{\alpha}(1)_A$ - $N(1)_A$. . .Cl(2)'	90.3(6)	$N(1)_A$. . .Cl(2)'. . .$O_w(2)$	111.8(2)
$C^{\alpha}(1)_A$ - $N(1)_A$. . .Cl(1)'	102.2(7)	$N(3)_C$. . .Cl(2)'. . .$O_w(3)'$	94.2(2)
$C^{\alpha}(1)_A$ - $N(1)_A$. . .$O_w(1)''$	123.4(7)	$N(3)_C$. . .Cl(2)'. . .$O_w(2)$	90.1(2)
Cl(1)'. . .$N(1)_A$. . .$O_w(1)''$	114.0(4)	$O_w(3)'$. . .Cl(2)'. . .$O_w(2)$	128.0(2)

Table XXVI--Continued

*Angle		Angle	
$O_W(2)$ . . .Cl(1). . .N $(1)_B$	85.3(3)	$O_W(3)''$. . . $O_W(1)''$. . ,Cl(1)''	80.4(3)
$O_W(2)$ . . .Cl(1). . .N $(2)_A$	128.6(2)	$O_W(3)''$. . . $O_W(1)''$. . .N $(1)_A$	76.4(4)
$O_W(2)$ . . .Cl(1). . .$O_W(1)$	148.8(3)	$O_W(3)''$. . . $O_W(3)''$. . .O $(4)_E$	120.2(4)
N $(1)_B$. . .Cl(1). . .N $(2)_A$	97.3(2)	Cl(1)''. . . $O_W(1)''$. . .N $(1)_A$	80.1(3)
N $(1)_B$. . .Cl(1). . .$O_W(1)$	121.3(3)	Cl(1)''. . . $O_W(1)''$. . .O $(4)_E$	146.9(3)
N $(2)_A$. . .Cl(1). . .$O_W(1)$	68.2(2)	N $(1)_A$ . . . $O_W(1)''$. . .O $(4)_E$	127.6(4)
Cl(1) . . .$O_W(2)$. . .Cl(2)'	97.8(2)	Cl(2)' . . . $O_W(3)'$ . . .$O_W(1)'$	122.0(3)
Cl(1) . . .$O_W(2)$. . .Ca	111.5(3)	Cl(2)' . . . $O_W(3)'$ . . .Ca	125.2(30
Cl(2) . . .$O_W(2)$. . .Ca	135.7(3)	$O_W(1)$ . . . $O_W(3)'$ . . .Ca	108.8(4)
C'$(1)_A$. . .N $(2)_A$ . .Cl(1)	110.7(7)	C'$(2)_A$ . . . N $(3)_A$ . . .Cl(2)	133.9(7)
$C^{\alpha}(2)_A$. . .N $(2)_A$ . .Cl(1)	122.0(7)	$C^{\alpha}(3)_A$ . . . N $(3)_A$ . . .Cl(2)	101.1(7)

*Standard deviations are given in parentheses.

and Cl ions have lengths of 3.33Å and 3.20Å, respectively. The normal range for the N-Cl distance in this type of H-bonding is 3.20 to 3.35Å, which is somewhat longer than the hydrogen bonds involving the NH_3^+ group (21).

All of the hydrogens of the three water molecules present in the structure are involved in hydrogen bonding. One of the water molecules ($O_w(2)$) bonded to the Ca ion is hydrogen bonded to two Cl ions (Cl(1) and Cl(2)'), with the Cl-O distances being 3.10 and 3.26Å, respectively. The second water molecule ($O_w(2)'$) bonded to the Ca ion forms hydrogen bonds to a Cl ion (Cl(2)') and to the third water molecule ($O_w(1)'$) of respective lengths 3.24 and 2.77Å. The Cl-O hydrogen bond distances of 3.26Å and 3.24Å at Cl(2)' indicate the formation of weak hydrogen bonds. Normal Cl-O hydrogen bond distances range from 3.00 to 3.20Å (22,25,26). Hydrogen bonds formed by the two water molecules ($O_w(2)$ and $O_w(3)$) bonded to the Ca ion are directed in such a way as to give approximate trigonal coordination of these water molecules. The various hydrogen bond lengths and angles involving the water molecules are listed in Table XXVI.

The water molecule ($O_w(1)''$) which isn't bonded to the Ca ion forms four hydrogen bonds involving another water molecule ($O_w(3)''$), a chloride ion (Cl(1)''), and two different peptide molecules ($N(1)_A$ and $O(4)_E$). Hydrogens are donated by the water molecule to an oxygen of the terminal carboxylate group ($O(4)_E$) and to the Cl ion (Cl(1)''), the

respective H-bond distances between these atoms are 2.69 and 3.13Å. The oxygen of this water molecule accepts H-bonds from the terminal NH_3^+ group ($N(1)_A$) and from a water molecule ($O_W(3)''$) which is bonded to the Ca ion. These H-bond distances are 2.86 and 2.76Å, respectively.

One would expect the water molecule $O_W(1)''$ to have a tetrahedral surrounding of hydrogen bonds, however, three of the atoms involved in the hydrogen bonding form an approximately trigonal arrangement around $O_W(1)''$ (the sum of the hydrogen bonding angles is 355°). This trigonal arrangement consists of the atoms $Cl(1)''$, $N(1)_A$, and $O(4)_E$. The fourth hydrogen bond is formed to the water molecule $O_W(3)''$ in such a way as to give a distorted trigonal pyrimidal arrangement about $O_W(1)''$. The hydrogen bond lengths and angles involving $O_W(1)''$ are listed in Table XXVI.

Conformation of Peptide Chain

The conformation of a peptide chain can be completely specified by giving the relative orientations of the two linked peptide groups at each α-carbon atom. This has been done in a variety of ways using different standard conformations for the peptide chain. The standard conformation of a polypeptide chain as proposed during the 1965 Gordon Conference on Proteins is shown in Figure 6 (27). Dashed lines give the limits of the i^{th} residue;

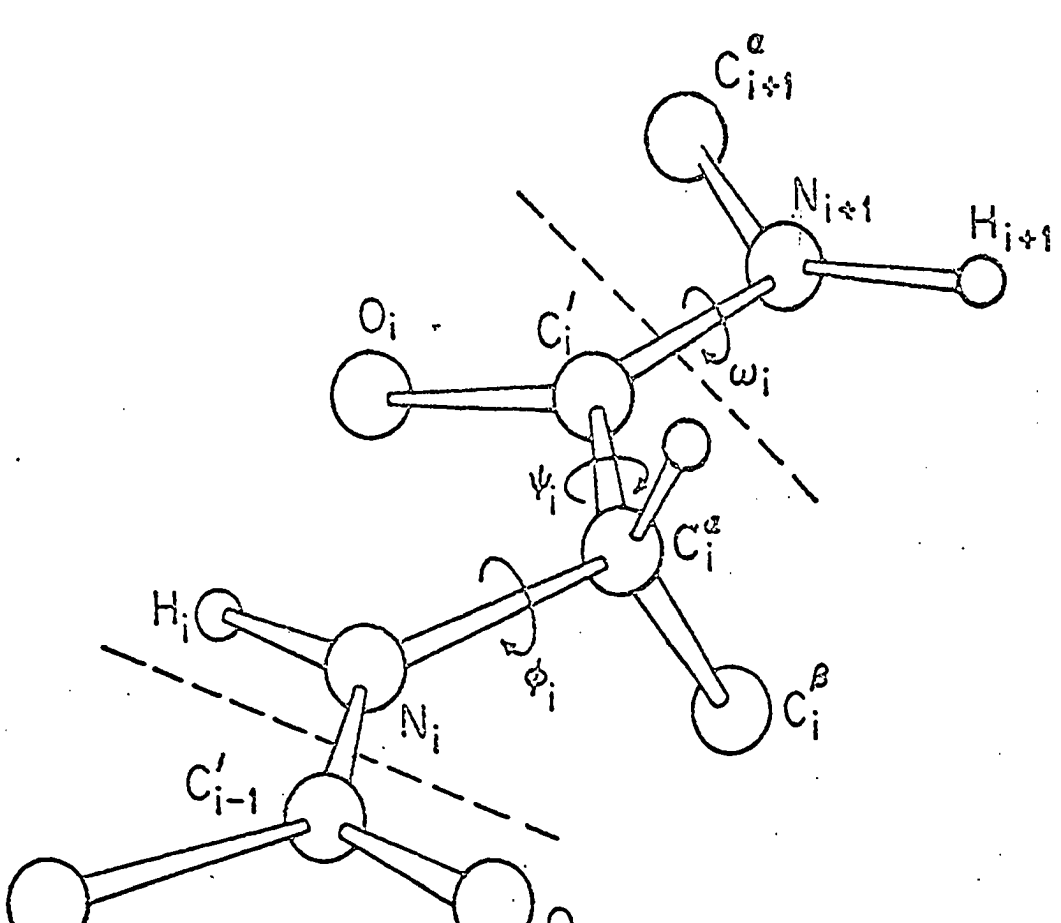


Figure 6. Standard Conformation of Polypeptide Chain

the term residue being reserved for the group of atoms $N\text{-}C^{\alpha}R\text{-}C'O$. Deviations from the standard conformation are given by right handed rotations ϕ, which is about the $N\text{-}C^{\alpha}$ bond, ψ which is about the $C^{\alpha}\text{-}C'$ bond, and w which is about the C'-N bond. As can be seen from Figure 6, the standard conformation represents the fully stretched polypeptide chain which corresponds to $\phi = \psi = w = 0$. This conformation can also be described by the following relationships: $C^{\alpha}\text{-}C'$ bond cis to N-H bond with respect to rotation around the $N\text{-}C^{\alpha}$ bond (ϕ); $N\text{-}C^{\alpha}$ bond cis to C-O bond with respect to rotation around $C^{\alpha}\text{-}C'$ bond (ψ); trans conformation of the peptide bond, i.e., C'-O bond trans to N-H bond (w). Rotation about the peptide (C'-N) bond, which is denoted by w, is not ordinarily used due to the fact that the peptide group $C^{\alpha}C'ON(C^{\alpha})$ is always very closely planar. By giving the rotations ϕ and ψ for each residue, the conformation of the peptide chain can be completely specified.

Values of ϕ and ψ for any residue can be plotted on a conformational map such as the one shown in Figure 7. Not all values of ϕ and ψ are allowed due to the fact that all possible orientations will not be stereochemically allowed because of short contacts between the atoms of adjacent residues. For peptides composed of glycyl residues Figure 7 gives the allowed regions of ϕ and ψ (28). The areas within the full and dashed lines represent the allowed

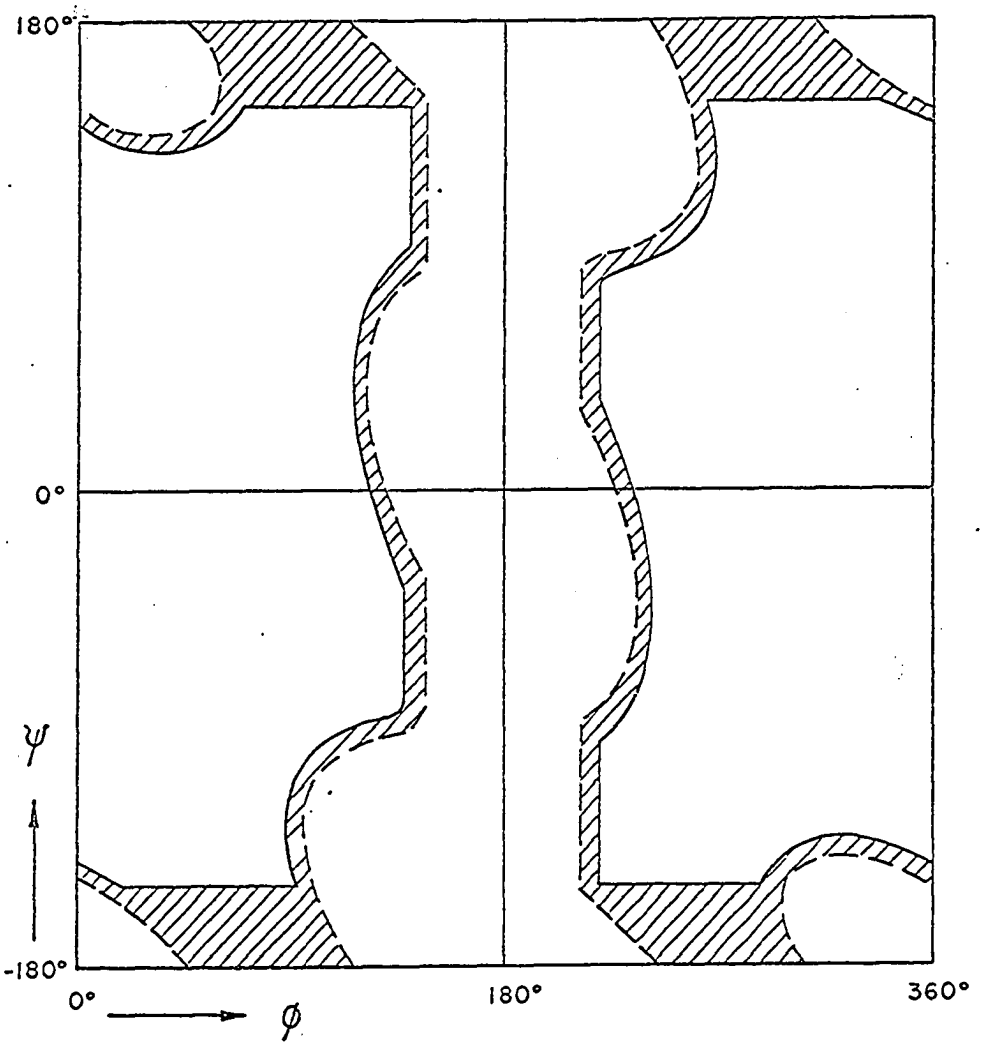


Figure 7. Conformational Map for Glycyl Residues

regions of ϕ and ψ calculated on the basis of two different sets of values for van der Waal's minimum contact distances. These two sets of contact distances termed "normally allowed" and "outer limit" are given in Table XXVII. The allowed regions are calculated for the angle $N\text{-}C^{\alpha}\text{-}C'$ = 115°. The area of the allowed region decreases slightly when the angle $N\text{-}C^{\alpha}\text{-}C'$ is decreased.

Table XXVII

Minimum Contact Distances
(from Edsall et.al., (27))

Contact	Normally allowed	Outer limit
C ... C	3.20Å	3.00Å
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

The values of ϕ , Ψ , and w are given in Table XXVIII for the glycyl residues in both $Cu(ggg)Cl\cdot 1\frac{1}{2}H_2O$, the structure by Freeman, Robinson, and Schoone (29), and

Table XXVIII

Observed Conformations (ϕ, ψ, w) and Corresponding N-C^{α}-C' Angle

Peptide	Residue	ϕ	ψ	w	N-C^{α}-C'
$CaCl_2(ggg)\cdot 3H_2O$	1	--	342°	2°	110.5°
	2	82°	176	358	120.1
	3	291	351	-	114.5
$CuCl(ggg)\cdot 1\frac{1}{2}H_2O$	1	--	349	8	107.9
	2	294	313	2	111.1
	3	266	353	-	110.7

$CaCl_2(ggg)\cdot 3H_2O$. In calculating the angles ϕ and ψ the backbone of the peptide (the bonds C'-N and N-C^{α}, respectively) was used instead of the N-H bond because the hydrogens are not accurately located. It should be noted that there is only one non-terminal residue in ggg and that the angle of rotation ϕ, about the N-C^{α} bond, is undefined for the N-terminal residue. In looking at the conformational map for glycyl residues, Figure 7, it can be seen that the angles ϕ and ψ for each residue in $Cu(ggg)Cl\cdot 1\frac{1}{2}H_2O$ lie in the "normally allowed" region.

For $CaCl_2(ggg)\cdot 3H_2O$ the values of ϕ and ψ for each residue lie in the allowed regions, however, the values of ϕ and ψ for the second residue, which is the only non-terminal residue, lie in the dashed line region which is

allowed only if the "outer limit" contact distances are used.

It should be recalled that this residue has the somewhat unusual configuration of having the nitrogen and carbonyl oxygen lying trans to each other. This residue also has the unusually large value of 120° for the angle N-C^{α}-C' which would increase the allowed region slightly above that given in Figure 7, since this was calculated for the angle N-C^{α}-C' equal to 115°. Even if the allowed region were calculated using the angle N-C^{α}-C' equal to 120°, the ϕ and ψ values for the second residue would still probably lie in the "outer limit" region.

Summary and Comparison of Structural Features

In the structure of $CaCl_2(ggg)\cdot 3H_2O$ the peptide molecules are linked together through hydrogen bonding to the water molecules and Cl ions and by bonding of the metal ions to the peptide oxygen atoms. The peptide molecules are not in the highly extended conformation which seems to be favored in small non-complexed peptides and which is found in the structures of glycylglycylglycine and glycylphenylalanylglycine (11). In these structures intermolecular hydrogen bonding gives rise to packing which is closely related to the so-called pleated-sheet arrangements which are typical of β-proteins with extended polypeptide chains. There is no direct intermolecular or intramolecular hydrogen bonding between the peptide mole-

cules themselves in the Ca complex. This seems to be the case in most metal complexes of small peptides, however, in the structure of the copper complex of glycyl-L-leucyl-L-tyrosine, the peptide molecules are in an extended configuration with intermolecular hydrogen bonding between the peptide molecules (30). The structure shows packing of the molecules which is closely related to the anti parallel pleated-sheet arrangement, and, indeed, the conformational angles approximate those predicted for this arrangement.

The coordination of the metal atom is much different in the Ca-peptide complex than in the transition metal-peptide complexes. In the Ca complex the metal atom is coordinated only to oxygen atoms but in the other metal complexes nitrogen atoms as well as oxygen are used in coordination to the metal. In the metal-peptide complexes whose crystal structures have been determined, except for the Ca complex, the terminal nitrogen atom is always used as a metal bonding site.

Bonding of the Ca ion to the carbonyl oxygen of a peptide group seems to cause significant changes in the bond lengths of the peptide group. The C'-O distance becomes shorter and the C'-N distance becomes longer than is normally found in free peptide. The changes correspond to an increase in the contribution of the resonance form N-C=O. This is in direct contrast to the deviations found in transition metal peptide complexes. The opposite effect

occurs in transition metal complexes when there is metal binding at a peptide nitrogen atom. The C'-O distance becomes longer and the C'-N distance becomes shorter than is found in the free peptide. This indicates an increase in the contribution of the resonance form $N{=}C{-}O^{-}$. Whether the deviations in the C'-O and C'-N bond lengths in the Ca complex are characteristic of changes that occur in Ca-peptide complexes and in general non-transition metal complexes will have to await the determination of other structures of this type.

CHAPTER V

SUMMARY AND CONCLUSIONS

The trial structure of $CaCl_2(ggg)\cdot 3H_2O$ was determined by Patterson-Fourier methods. The positions of the Ca, two Cl's, and six oxygen atoms were located from the Patterson map. A Fourier electron density map then gave the positions of the other atoms in the structure.

The peptide molecules are held together by means of Ca-O bonding and hydrogen bonding to the Cl ions and water molecules. There is no direct hydrogen bonding between the peptide molecules themselves.

Each Ca ion was found to be seven-coordinated with the coordination sphere being made up of five oxygens from four different peptide molecules and oxygens from two water molecules.

The Cl ions serve as hydrogen bonding links between peptide molecules and also as hydrogen bond acceptors from water molecules. There are no Ca-Cl bonds.

Deviations were found in the C'-O and C'-N distances in the peptide molecule corresponding to a greater than normal contribution of the resonance form N-C = O. The

N-C^{α}-C' angle of the second and third residues were found to be larger than their normal value of 111°.

The conformational angles ϕ and ψ for the non-terminal residue of the peptide lie in the "outer limit" region.

REFERENCES

1. J.D.H. Donnay and G. Donnay, Crystal Data, 2nd ed., Williams and Heintz Map Corporation, Washington, D.C., 1963.

2. International Tables for X-Ray Crystallography, Vol. I, Kynoch Press, Birmingham (1952).

3. International Tables for X-Ray Crystallography, Vol. II, Kynoch Press, Birmingham (1952).

4. A. L. Patterson and E. Love, Acta Cryst., 10, 111 (1957).

5. P. Pfeiffer, Organishe Molekulverbindungen, F. Enke, Stuttgart, 1927.

6. E. R. Howells, D. C. Phillips, and D. Rogers, Acta Cryst., 3, 210 (1950).

7. L. Mazarella, A. L. Kovacs, P. DeSantis, and A. M. Liquori, Acta Cryst., 22, 65 (1967).

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

9. International Tables for X-Ray Crystallography, Vol. III, Kynoch Press, Birmingham (1952).

10. L. Pauling and R. B. Corey, Nature Lond., 171, 59 (1953).

11. R. E. Marsh and J. Donohue, Advances in Protein Chemistry, 22, 235 (1967).

12. H. C. Freeman, The Biochemistry of Copper, Academic Press, New York, 1966.

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

14. H. C. Freeman, J. C. Schoone, and J. G. Sime, Acta Cryst., 18, 381 (1965).

15. H. C. Freeman and M. R. Taylor, Acta Cryst., 18, 939 (1965).

16. V. Schomaker, J Waser, R. E. Marsh, and G. Bergman, Acta Cryst., 12, 600 (1959).

17. H. C. Freeman, Advances in Protein Chemistry, 22, 257, (1967).

18. G. MacLennan and C. A. Beevers, Acta Cryst., 8, 579 (1955).

19. I. Nitta and K. Osaki, X-Sen Kondankai, 5, 37 (1948).

20. L. E. Sutton, Table of Interatomic Distances and Configurations in Molecules and Ions, The Chemical Society, London, 1965.

21. D. W. Smits and E. H. Wiebenga, Acta Cryst., 6, 531 (1953).

22. B. Dawson, Acta Cryst., 6, 81 (1953).

23. C. J. Brown, Acta Cryst., 2, 228 (1949).

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

25. J. M. Broomhead, Acta Cryst., 4, 92 (1951).

26. R. Parthasarathy, Acta Cryst., 21, 422 (1966).

27. J. T. Edsall, P. J. Flory, J. C. Kendrew, A. M. Liquori, G. Nemethy, G. N. Ramachandran, and H. A. Scheraga, Biopolymers, 4, 121 (1966).

28. C. Ramakrishnan and G. N. Ramachandran, Biophysical Journal, 5, 909 (1965).

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

30. W. A. Franks, Ph.D. Thesis, The University of Oklahoma, Norman, Oklahoma, 1968.

31. R. A. Frazer, W. J. Duncan, and A. R. Collar, Elementary Matrices, The University Press, Cambridge, 1938, p. 142.

APPENDIX A

Least Squares Plane

In finding the least squares plane it is desired to minimize the function $S = \sum_k W_k D_k^2$, which is the weighted sum of the squares of the distances D_k of the k points from the plane. The derivation and minimizing of S follows that given by Schomaker et. al. (16).

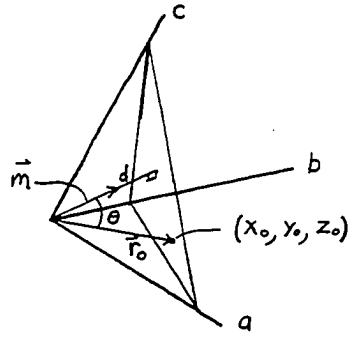
In the subsequent derivation two types of vectors are used. The position vector to a point of fractional coordinates (x, y, z) is always expressed in terms of the unit cell axis

$$\vec{r} = x\vec{a} + y\vec{b} + z\vec{c}$$

but the normal to a plane is always for convenience expressed in terms of the reciprocal vectors, $\vec{a}^*$, $\vec{b}^*$, and $\vec{c}^*$.

$$\vec{m} = m_1\vec{a}^* + m_2\vec{b}^* + m_3\vec{c}^*$$

Before deriving the equation for the distance D of a point from a plane it is useful to have an expression for d the distance of the plane from the origin.



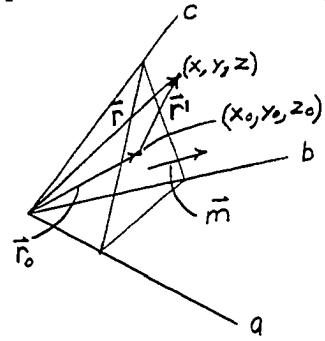
The distance d from the origin to the plane can be derived in terms of the components of the unit normal $\vec{m}$ and a vector $\vec{r}_o$ to an arbitrary point on the plane. The distance d is found by projecting the vector $\vec{r}_o$ onto the unit normal $\vec{m}$

$$\vec{r}_o \cdot \vec{m} = |\vec{r}_o| \cdot |1| \cos\theta = d$$

$$(x_o\vec{a} + y_o\vec{b} + z_o\vec{c}) \cdot (m_1\vec{a}^* + m_2\vec{b}^* + m_3\vec{c}^*) = d$$

$$m_1x_o + m_2y_o + m_3z_o = d$$

Now the equation can be derived for the distance D of a point from the plane in terms of the components m_1, m_2, and m_3 of the unit normal to that plane and d the distance of the plane from the origin.



The distance D of the point of fractional coordinates (x, y, z) from the plane is given by the projection of the vector $\vec{r}^1$ onto the unit normal to the plane m, where $\vec{r}^1$ is a vector from an arbitrary point (x_o, y_o, z_o) in the plane to the point (x, y, z).

$$\vec{r}^1 \cdot \vec{m} = (\vec{r} - \vec{r}_o) \cdot \vec{m} = D$$

$$(x - x_o)\vec{a} + (y - y_o)\vec{b} + (z - z_o)\vec{c} \cdot (m_1\vec{a}^* + m_2\vec{b}^* + m_3\vec{c}^*) = D$$

$$m_1(x - x_o) + m_2(y - y_o) + m_3(z - z_o) = D$$

$$m_1x + m_2y + m_3z - (m_1x_o + m_2y_o + m_3z_o) = D$$

from the equation derived before

$$m_1x_o + m_2y_o + m_3z_o = d$$

therefore,

$$D = m_1x + m_2y + m_3z - d$$

Now that D is expressed in terms of m_1, m_2, m_3, and d an equation can be written for S, the weighted sum of the squares of the distances of the points from the plane

$$S = \sum_k W_k D_k^2 = \sum_k W_k(m_1x_k + m_2y_k + m_3z_k - d)^2$$

where it is remembered that this equation was derived subject to the constraining condition that $\vec{m}$ be a unit vector.

The equation of constraint can be written $\vec{m} \cdot \vec{m} = 1$ where $\vec{m} = m_1\vec{a}^* + m_2\vec{b}^* + m_3\vec{c}^*$ or

$$a^{*2}m_1^2 + b^{*2}m_2^2 + c^{*2}m_3^2 + 2a^*b^* \cos\gamma^* m_1m_2 + 2 a^*c^* \cos\beta^* m_1m_3 +$$

$$2b^*c^* \cos\alpha^* m_2m_3 = 1$$

the last equation can be written

$$\sum_{i,j=1}^{3} g_{ij}m_im_j = 1$$

where $(g_{ij}) = \begin{pmatrix} a^{*2} & a^*b^* \cos\gamma^* & a^*c^* \cos\beta^* \\ a^*b^* \cos\gamma^* & b^{*2} & b^*c^* \cos\alpha^* \\ a^*c^* \cos\beta^* & b^*c^* \cos\alpha^* & c^{*2} \end{pmatrix}$

In order to minimize S subject to an equation of constraint the method of Lagrangian Multipliers is used. The method of Lagrangian Multipliers states that to minimize a function $S(a_1, a_2, \ldots, a_n)$ of several variables $a_1, a_2, \ldots, a_n$ subject to an equation of constraint $G(a_1, a_2, \ldots, a_n) = 0$, one forms the auxiliary function

$$F(a_1 a_2, \ldots, a_n) = S(a_1, a_2, \ldots, a_n) - \lambda G(a_1 a_2, \ldots, a_n)$$

where λ is a Lagrangian Multiplier independent of the a_i. Then one solves the n simultaneous equations for the a_i

$$\frac{\partial F}{\partial a_i} = 0 \quad i = 1, 2, \ldots, n$$

In the present case the function to be minimized is $S = S(m_1, m_2, m_3, d) = \sum_k W_k(m_1 x_k + m_2 y_k + m_3 z_k - d)^2$ and the equation of constraint is the equation that requires $\vec{m}$ to be a unit vector

$$G = G(m_1 m_2, m_3, d) = g_{11} m_1^2 + g_{22} m_2^2 + g_{33} m_3^2 + 2g_{12} m_1 m_2 + 2g_{13} m_1 m_3 + 2g_{23} m_2 m_3 - 1 = 0$$

The auxiliary function can then be written

$$F = \sum_k W_k(m_1 x_k + m_2 y_k + m_3 z_k - d)^2 - \lambda \sum_{i,j=1}^{3} g_{ij} m_i m_j - \lambda$$

remembering that $g_{ij} = g_{ji}$.

Now it is desired to solve the four simultaneous normal equations

$$\frac{\partial F}{\partial m_i} = 0 \quad i=1,2,3 \quad \frac{\partial F}{\partial d} = 0$$

The four normal equations are

$$\tfrac{1}{2}\frac{\partial F}{\partial m_1} = \sum_k W_k x_k (m_1 x_k + m_2 y_k + m_3 z_k - d) - \lambda(g_{11}m_1 + g_{12}m_2 + g_{13}m_3) = 0$$

$$\tfrac{1}{2}\frac{\partial F}{\partial m_2} = \sum_k W_k y_k (m_1 x_k + m_2 y_k + m_3 z_k - d) - \lambda(g_{21}m_1 + g_{22}m_2 + g_{23}m_3) = 0$$

similarly for $\tfrac{1}{2}\dfrac{\partial F}{\partial m_3} = 0$

$$-\tfrac{1}{2}\frac{\partial F}{\partial d} = \sum_k W_k (m_1 x_k + m_2 y_k + m_3 z_k - d) = 0$$

From the last equation d can be solved for in terms m_1, m_2, and m_3

$$\left(\frac{\sum_k W_k x_k}{\sum_k W_k}\right) m_1 + \left(\frac{\sum_k W_k y_k}{\sum_k W_k}\right) m_2 + \left(\frac{\sum_k W_k z_k}{\sum_k W_k}\right) m_3 = d$$

or

$$m_1\bar{x} + m_2\bar{y} + m_3\bar{z} = d$$

The first three equations then become

$$\sum_k W_k x_k [m_1(x_k - \bar{x}) + m_2(y_k - \bar{y}) + m_3(z_k - \bar{z})] = \lambda(g_{11}m_1 + g_{12}m_2 + g_{13}m_3)$$

$$\sum_k W_k y_k [m_1(x_k - \bar{x}) + m_2(y_k - \bar{y}) + m_3(z_k - \bar{z})] = \lambda(g_{21}m_1 + g_{22}m_2 + g_{23}m_3)$$

$$\sum_k W_k z_k [m_1(x_k - \bar{x}) + m_2(y_k - \bar{y}) + m_3(z_k - \bar{z})] = \lambda(g_{31}m_1 + g_{32}m_2 + g_{33}m_3)$$

rewriting these equations in matrix form

$$\begin{pmatrix} \sum_k W_k x_k (x_k - \bar{x}) & \sum_k W_k x_k (y_k - \bar{y}) & \sum_k W_k x_k (z_k - \bar{z}) \\ \sum_k W_k y_k (x_k - \bar{x}) & \sum_k W_k y_k (y_k - \bar{y}) & \sum_k W_k y_k (z_k - \bar{z}) \\ \sum_k W_k z_k (x_k - \bar{x}) & \sum_k W_k z_k (y_k - \bar{y}) & \sum_k W_k z_k (z_k - \bar{z}) \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} = \lambda \begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix}$$

This is a matrix equation of the form

$$Am = \lambda\, gm$$

Now it is desired to transform this equation into a form which is easily solved for the vector m. This can be done by multiplying both sides of the equation on the left by A^{-1}

$$A^{-1}Am = \lambda\, A^{-1}gm$$

$$m = \lambda\, A^{-1}gm$$

$$A^{-1}gm = \frac{1}{\lambda}\, m$$

$$Bm = \frac{1}{\lambda}\, m \text{ where } B = A^{-1}g$$

This equation is of the form $Bm = \mu m$ which is the so-called eigenvalue equation of matrix algebra where μ is a constant. The nontrivial solution vectors m are termed eigenvectors and the corresponding values of μ are termed eigenvalues of the equation.

The set of three homogeneous linear equations in m_1, m_2, and m_3 can be written

$$(B - \frac{1}{\lambda}\, I)m = 0$$

These have nontrivial solutions for m_1, m_2, and m_3 only when the determinantal equation $\left|(B - \frac{1}{\lambda}\, I)\right| = 0$ is satisfied.

$$\left|B - \frac{1}{\lambda}\, I\right| = \begin{vmatrix} B_{11} - \frac{1}{\lambda} & B_{12} & B_{13} \\ B_{21} & B_{22} - \frac{1}{\lambda} & B_{23} \\ B_{31} & B_{32} & B_{33} - \frac{1}{\lambda} \end{vmatrix} = 0$$

This is a cubic equation in λ with roots $\lambda^{(1)} << \lambda^{(2)} \leq \lambda^{(3)}$. The roots of this equation are the reciprocals of the eigenvalues of the matrix equation and the three different solution vectors for m obtained by substituting $\lambda^{(1)}$, $\lambda^{(2)}$, and $\lambda^{(3)}$ into the equation $Bm = \frac{1}{\lambda}m$ are the so-called eigenvectors of the matrix equation. It has been shown that the three eigenvectors are mutually perpendicular (16) and that the eigenvector perpendicular to the best plane is the one that corresponds to the smallest root $\lambda^{(1)}$ or to the largest eigenvalue $1/\lambda^{(1)}$.

Iterative Numerical Solution

It is possible to solve the cubic equation in for its smallest root $\lambda^{(1)}$ and then substitute back in the equation $Bm = 1/\lambda\, m$ in order to find the eigenvector associated with this root. An alternate approach was favored, however, that allowed the smallest root (or largest eigenvalue) and the eigenvector associated with it to be determined simultaneously with very little computation. The method used is a standard iterative numerical method for solving the matrix equation $Bm = \mu m$ for the eigenvector m associated with the largest eigenvalue μ (31).

The procedure is as follows. Multiply an arbitrarily chosen vector m by B. The vector m is arbitrary but in its most convenient form it contains unity in one row and zeros in the other rows. Extract a constant scalar quantity

from the resulting matrix product so as to keep one of its components equal to 1. Multiply the new one-column matrix again by the matrix B. This multiplication process is repeated until the components of the resulting vector and the scalar quantity extracted become constant to any desired degree of accuracy.

The scalar quantity extracted is the largest eigenvalue $\mu^{(1)}$ $(=1/\lambda^{(1)})$ and the resulting vector is the eigenvector associated with it. This vector can then be normalized by dividing each of its components by the magnitude of the vector.

Once the components m_i of the unit normal to the best plane are found it is possible to solve for the distance d of the plane from the origin by the equation

$$d = m_1\bar{x} + m_2\bar{y} + m_3\bar{z}$$

The equation of the best plane is now uniquely determined by the four parameters m_1, m_2, m_3, and d and can be expressed in terms of the fractional coordinates x, y, and z

$$m_1x + m_2y + m_3z = d$$

where the unit vector normal to the plane expressed as a reciprocal vector is given by

$$\vec{m} = m_1\vec{a}^* + m_2\vec{b}^* + m_3\vec{c}^*$$

The equation for the distance D of an arbitrary point of fractional coordinates (x,y,z) from the plane is given by the equation

$$D = m_1x + m_2y + m_3z - d$$

Computer Program for Least Squares Plane

```
 DOUBLE PRECISION ALPHS,BETS,GAMS
 DOUBLE PRECISION WTSUM,CMAGM,D
 DOUBLE PRECISION A(4,4),COEFI,PID(3,3),V(3),DIST
 DOUBLE PRECISION ASTAR, BSTAR,CSTAR,COSA,COSB,COSG,FLAMB
 DOUBLE PRECISION XBAR(3),W(3),CMI(3),CM2(3),CM3(3),U(3)
 DOUBLE PRECISION G(3,3),B(3,3),AINV(3,3),AA(3,3),X(3,30)
 READ (1,90)ASTAR,BSTAR,CSTAR,ALPHS,BETS,GAMS,NPL
 DO 101 MPL=1,NPL
 READ(1,140)N,ITER,MM,NPTS,NWTS
 COSA=DCOS(ALPHS*3.141593/180.0)
 COSB=DCOS(BETS*3.141593/180.0)
 COSG=DCOS(GAMS*3.141593/180.0)
 G(1,1)=ASTAR*ASTAR
 G(1,2)=ASTAR*BSTAR*COSG
 G(1,3)=ASTAR*CSTAR*COSB
 G(2,2)=BSTAR*BSTAR
 G(2,3)=BSTAR*CSTAR*COSA
 G(3,3)=CSTAR*CSTAR
 G(2,1)=G(1,2)
 G(3,1)=G(1,3)
 G(3,2)=G(2,3)
  WRITE (3,62)
WRITE (3,91)((G(I,J),J=1,3),I=1,3)
WTSUM=0.0
```

```
      READ (1,92)((X(I,J),I=1,3),J=1,N)
      WRITE (3,70)
      WRITE (3,117)((X(I,J),I=1,3),J=1,N)
      DO 1 I=1,N
      GO TO (72,71),NWTS
   72 W(I)=1.0
      GO TO 1
   71 READ(1,92)W(I)
    1 WTSUM=WTSUM+W(I)
      WRITE (3,65)
      DO 3 I=I,3
      XBAR(I)=0.0
      DO 2 J=1,N
    2 XBAR(I)=XBAR(I)+W(J)*X(I,J)
      XBAR(I)=XBAR(I)/WTSUM
    3 WRITE (3.94)XBAR(I)
      IF (N-3)106,103,106
  106 DO 45 I=1,3
      DO 45 J=1,3
      A(I,J)=0.0
      DO 5 K=1,N
    5 A(I,J)=A(I,J)+W(K)*X(I,K)*(X(J,K)-XBAR(J))
   45 AA(I,J)=A(I,J)
      WRITE (3,63)
      WRITE (3.141)((A(I,J),J=1,3),I=1,3)
      A(1,4)=1.0
```

```
      A(2,4)=0.0
      A(3,4)=0.0
      DO 55 K=1,3
      DO 32 J=1,3
   32 A(4,J)=A(1,J+1)/A(1,1)
      DO 33 I=2,3
      COEFI=A(I,1)
      DO 33 J=1,3
   33 A(I,J)=A(I,J+1)-COEF1*A(4,J)
      DO 34 I=1,3
      DO 34 J=1,3
   34 A(I,J)=A(I+1,J)
   55 CONTINUE
      WRITE (3,64)
      WRITE (3,141)((A(I,J), J=1,3),I=1,3)
      DO 6 I=1,3
      DO 6 J=1,3
      PID(I,J)=0.0
      DO 6 K=1,3
    6 PID(I,J)=PID(I,J)+AA(I,K)*A(K,J)
      WRITE (3,66)
      WRITE (3,91)((PID(I,J),J=1,3),I=1,3)
      DO 8 I=1,3
      DO 8 J=1,3
    8 AINV(I,J)=A(I,J)
      DO 120 I=1,3
```

```
      DO 120 J=1,3
      B(I,J)=0.0
      DO 120 K=1,3
  120 B(I,J)=B(I,J)+AINV(I,K)*G(K,J)
      WRITE (3,67)
      WRITE (3,91)((B(I,J),J=1,3),I=1,3)
      IF(DABS(B(1,1))-DABS(B(1,2)))121,121,13
  121 IP=2
      GO TO 14
   13 IP=1
   14 EO 15 I=1,3
   15 CM1(I)=B(I,IP)/B(1,IP)
      DO 20 K=1,ITER
      WRITE (3,80)
      DO 18 I=1,3
      CM2(I)=0.0
   19 CM2(I)=CM2(I)+B(I,J)*CM1(J)
      CM3(I)=CM2(I)/CM2(1)
   18 WRITE (3,94)CM3(I)
      WRITE (3,81)
      WRITE (3,94)CM2(1)
      DO 17 I=1,3
   17 CM1(I)=CM3(I)
   20 CONTINUE
      GO TO 108
```

```
103 DO 102 I=1,3
    U(I)=X(I,2)=X(I,1)
102 V(I)=X(I,3)=X(I,1)
    CM1(1)=U(2)*V(3)=U(3)*V(2)
    CM1(2)=U(3)*V(1)=U(1)*V(3)
    CM1(3)-U(1)*V(2)-U(2)*V(1)
108 CMAGM=0.0
    DO 21 I=1,3
    DO 21 J=1,3
 21 CMAGM=CMAGM+G(I,J)*CM1(i)*CM1(J)
    CMAGM=DSQRT(CMAGM)
    WRITE (3,82)
    DO 22 I=1,3
    CM1(I)=CM1(I)/CMAGM
_22 WRITE (3,94)CM1(I)
    IF (N-3)104,105,104
104 WRITE (3,83)
    ELAMB=1.0/CM2(1)
    WRITE (3,94)ELAMB
105 DIST=0.0
    DO 110 I=1,3
110 DIST=DIST+CM1(I)*XBAR(I)
    WRITE (3,68)
    WRITE (3,94)DIST
    GO TO (111,112),MM
```

```
111 READ (1,92)((X(I,J),I=1,3),J=1,NPTS)
    WRITE (3,115)
    DO 114 J=1,NPTS
    D=0.0
    DO 113 I=1,3
113 D=D+CM1(I)*X(I,J)
    D=D-DIST
114 WRITE (3,116)X(1,J),X(2,J),X(3,J),D
112 CONTINUE
101 CONTINUE
 68 FORMAT (1X,20HDIST FROM O TO PLANE//)
 90 FORMAT (3F8.6,3F7.2,I3)
 62 FORMAT (1H1,10X,8HMATRIX G)
 92 FORMAT (20X,3F10.5)
 93 FORMAT (F10.5)
 63 FORMAT (1H1,10X,8HMATRIX A//)
 65 FORMAT (1X,4HXBAR//)
 64 FORMAT (1H1,10X,8HAINVERSE//)
 66 FORMAT (10X,10HA*AINVERSE//)
 67 FORMAT (1H1,10X,12HB=AINVERSE*G//)
 80 FORMAT (1X,13HEIGENVECTOR M//)
 81 FORMAT (1X,10HEIGENVALUE//)
 82 FORMAT (1X,13HUNIT VECTOR M//)
 83 FORMAT (1X,6HLAMBDA//)
115 FORMAT (1H1,10X,6HCOORD.,15X,5HDIST.//)
```

```
116 FORMAT (1X,3F9.5,D14.6//)
140 FORMAT(513)
 70 FORMAT (1X,23HPOINTS USED IN LS PLANE//)
117 FORMAT (1X,3F9.5//)
141 FORMAT (1X,3D15.7//)
 91 FORMAT (1X,3D15.7//)
 94 FORMAT (1X,D15.7//)
    STOP
    END
```

APPENDIX B

Least Squares Cell Parameters

The best least squares cell dimensions are obtained by minimizing the sum of the deviations squared of some function R of the calculated and observed 2θ values with respect to the cell parameters.

$$(1) \quad \sum_{hkl} E^2_{hkl} = \sum_{hkl} [R^{obs.}_{hkl}(\theta) - R^{calc.}_{hkl}(\theta)]^2$$

$$(2) \quad R_{hkl}(\theta) = \sin^2\theta_{hkl}$$

In this particular case R was chosen as $\sin^2\theta$, however, it could have been chosen as any other function of θ. $R_{hkl}(\theta) = \sin^2\theta_{hkl}$ can be expressed in terms of the unit cell parameters a*, b*, c*, α^*, β^*, and γ^*, for each reflection whose Miller indices are (hkl).

$$(3) \quad \sin^2\theta_{hkl} = \frac{\lambda^2}{4}(h^2a^{*2} + k^2b^{*2} + l^2c^{*2} + 2hka^*b^*\cos\gamma^* + 2hla^*c^*\cos\beta^* + 2klb^*c^*\cos\alpha^*)$$

The problem of minimizing the sum of the deviations squared for the function $\sin^2\theta$ does not lend itself directly to the method of least squares. Since there are cross terms in the parameters a*, b*, c*, α^*, β^*, γ^* the desired linear equations in these parameters cannot be

obtained by the usual least squares method.

A different technique can be used in obtaining the best values in the least squares sense for the cell parameters a^*, b^*, c^*, α^*, β^*, and γ^*, if approximate values of these parameters are known. The function R_{hkl} can be expanded in a Taylor series about the point $(a^*, b^*, c^*, \alpha^*, \beta^*, \gamma^*)$ in six dimensional space. The correct value of the function R_{hkl} can be expressed by the Taylor series

$$(4)\ R_{hkl}\ (a^* + E_{a^*},\ b^* + E_{b^*},\ c^* + E_{c^*}, \ldots,\ \gamma^* + E_{\gamma^*}\ )$$
$$= R_{hkl}(a^*, b^*, c^*, \alpha^*, \beta^*, \gamma^*) + \left(\frac{\partial R_{hkl}}{\partial a^*}\right) E_{a^*} + \ldots + \left(\frac{\partial R_{hkl}}{\partial \gamma^*}\right) E_{\gamma^*}$$

where a^*, b^*, c^*, α^*, β^*, γ^* are the approximate cell parameters and $a^* + E_{a^*}$, $b^* + E_{b^*}$, etc., are the correct values for the cell parameters. All second and higher order terms in the Taylor series are sufficiently small to be neglected if the errors in a^*, b^*, c^*, α^*, β^*, γ^* are small.

The left hand side of this equation represents the true value of R_{hkl} which is equal to the observed value of the function $R^o_{hkl}(\theta)$, denoted by the superscript 0, plus an observational error E_{hkl}. The right hand side of the equation represents the calculated value of the function.

$$(5)\ E_{hkl} + R^o_{hkl} = R^c_{hkl} + \left(\frac{\partial R_{hkl}}{\partial a^*}\right)^c E_{a^*} + \ldots + \left(\frac{\partial R_{hkl}}{\partial \gamma^*}\right)^c E_{\gamma^*}$$

By changing the notation for the i^{th} cell parameter

to q_i with E_i equal to the error in q_i, the equation can be rewritten

$$(6)\quad E_{hkl} + R^{o}_{hkl} = R^{c}_{hkl} + \sum_{i=1}^{6} \left(\frac{\partial R_{hkl}}{\partial q_i}\right)^{c} E_i$$

where it is remembered that R^{c}_{hkl} and $\left(\frac{\partial R_{hkl}}{\partial q_i}\right)^{c}$ are evaluated at the approximate point (a^*, b^*, c^*, α^*, β^*, γ^*). Now it is possible to write the expression for the sum of the deviations squared, where the deviations are the differences in the observed and calculated values for $R_{hkl}(\theta)$.

$$(7)\quad \sum_{hkl} E^2_{hkl} = \sum_{hkl} \left[R^{obs.}_{hkl}(\theta) - R^{cal.}_{hkl}(\theta)\right]^2$$

$R^{calc.}_{hkl}(\theta)$ is represented by the right hand side of equation 6.

$$(8)\quad \sum_{hkl} E^2_{hkl} = \sum_{hkl} \left[R^{o}_{hkl} - \left[R^{c}_{hkl} + \sum_{i} \left(\frac{\partial R_{hkl}}{\partial q_i}\right)^{c} E_i\right]\right]^2$$

or

$$(9)\quad S = \sum_{hkl} E^2_{hkl} = \sum_{hkl} \left[\sum_{i} \left(\frac{\partial R_{hkl}}{\partial q_i}\right)^{c} E_i - \left(R^{o}_{hkl} - R^{c}_{hkl}\right)\right]^2$$

The desired linear normal equations can now be obtained by differentiating the sum of the deviations squared S with respect to each of the error parameters E_i and setting each of the partial derivatives equal to zero.

$$(10)\quad \frac{\partial S}{\partial E_j} = 2 \sum_{hkl} \left[\sum_i \left(\frac{\partial R_{hkl}}{\partial q_i}\right)^c E_i - (R^o_{hkl} - R^c_{hkl}) \right] \left(\frac{\partial R_{hkl}}{\partial q_j}\right)^c = 0$$

$$j = 1, \ldots, 6$$

rewriting these equations

$$(11)\quad \left(\sum_{hkl} \frac{\partial R_{hkl}}{\partial q_1} \frac{\partial R_{hkl}}{\partial q_j}\right) E_1 + \left(\sum_{hkl} \frac{\partial R_{hkl}}{\partial q_2} \frac{\partial R_{hkl}}{\partial q_j}\right) E_2$$

$$+ \left(\sum_{hkl} \frac{\partial R_{hkl}}{\partial q_3} \frac{\partial R_{hkl}}{\partial q_j}\right) E_3 + \ldots + \left(\sum_{hkl} \frac{\partial R_{hkl}}{\partial q_6} \frac{\partial R_{hkl}}{\partial q_j}\right) E_6$$

$$= \sum_{hkl} (R^o_{hkl} - R^c_{hkl}) \frac{\partial R_{hkl}}{\partial q_j}$$

$$j = 1, \ldots, 6$$

It can be seen from the last equation that you get a set of linear equations in the error parameters by this procedure. Now the expressions for the coefficients in these linear equations will be derived using the function $R_{hkl}(\theta) = \sin^2 \theta_{hkl}$ where the cell parameters q_4, q_5, q_6 are taken for convenience as $\cos \alpha^*$, $\cos \beta^*$, $\cos \gamma^*$, respectively, rather than α^*, β^*, γ^*.

$$(12)\quad R_{hkl}(\theta) = \sin^2 \theta_{hkl} = \frac{\lambda^2}{4} (h^2a^{*2} + k^2b^{*2} + l^2c^{*2}$$

$$+ 2hka^*b^* \cos \gamma^* + 2hla^*c^* \cos \beta^* + 2klb^*c^* \cos \alpha^*)$$

$$(13)\quad \frac{\partial R_{hkl}}{\partial q_1} = \frac{\partial R_{hkl}}{\partial a^*} = \frac{\lambda^2}{2} h(ha^* + kb^* \cos \gamma^* + lc^* \cos \beta^*)$$

$$(14)\quad \frac{\partial R_{hkl}}{\partial q_2} = \frac{\partial R_{hkl}}{\partial b^*} = \frac{\lambda^2}{2}\, k(kb^* + ha^* \cos\gamma^* + lc^* \cos\alpha^*)$$

$$(15)\quad \frac{\partial R_{hkl}}{\partial q_3} = \frac{\partial R_{hkl}}{\partial c^*} = \frac{\lambda^2}{2}\, l(lc^* + ha^* \cos\beta^* + kb^* \cos\alpha^*)$$

$$(16)\quad \frac{\partial R_{hkl}}{\partial q_4} = \frac{\partial R_{hkl}}{\partial(\cos\alpha^*)} = \frac{\lambda^2}{2}\, klb^*c^*$$

$$(17)\quad \frac{\partial R_{hkl}}{\partial q_5} = \frac{\partial R_{hkl}}{\partial(\cos\beta^*)} = \frac{\lambda^2}{2}\, hla^*c^*$$

$$(18)\quad \frac{\partial R_{hkl}}{\partial q_6} = \frac{\partial R_{hkl}}{\partial(\cos\gamma^*)} = \frac{\lambda^2}{2}\, hka^*b^*$$

The linear equations in the error parameters can be written as a matrix equation

$$(19)\quad AE = B$$

The elements of the A matrix are

$$(20)\quad A_{ij} = \sum_{hkl} \frac{\partial R_{hkl}}{\partial q_i} \frac{\partial R_{hkl}}{\partial q_j}$$

The components of the vector B are

$$(21)\quad B_i = \sum_{hkl} (R^o_{hkl} - R^c_{hkl}) \frac{\partial R_{hkl}}{\partial q_i}$$

R^o_{hkl} is the value of $\sin^2\theta$ calculated from the 2θ measurements. R^c_{hkl} is the value of $\sin^2\theta$ calculated from the approximate values of a^*, b^*, c^*, α^*, β^*, and γ^*. The derivatives $\frac{\partial R_{hkl}}{\partial q_i}$ are calculated from the equations derived before, using the approximate values of the cell

parameters.

The solution vector E is found by taking the inverse of matrix A

(22) $$E = A^{-1} B$$

The E_i are then added to the approximate cell parameters giving the correct least squares values for the cell parameters.

The standard deviations (31) for the cell parameters q_i are obtained by the equation

(23) $$\sigma_i = \sqrt{\frac{\sum_{hkl} (R^o_{hkl} - R^c_{hkl})^2}{n - p}} \; A^{-1}_{ii}$$

where R^o_{hkl} is the observed value of the function $\sin^2\theta_{hkl}$, R^c_{hkl} is the calculated value of the function $\sin^2\theta_{hkl}$ using the least squares values for a^*, b^*, c^*, α^*, β^*, and γ^*, A^{-1}_{ii} is the i^{th} diagonal element of the matrix A^{-1} calculated using the least squares values of the cell parameters, n is the number of reflections used, and p is the number of cell parameters fitted in the least squares.

Computer Program for Least Squares Cell Parameters

```
      DOUBLE PRECISION UH(100),UH(100),UL(100),R1OBS(100),HH,HK
      DOUBLE PRECISION ALPHS,BETS,GAMS,WLAMB,SDEV,TWOTH,R2OBS
      DOUBLE PRECISION RICAL,DELRI,FLAMB,S,SS,COEFI,HL
      DOUBLE PRECISION A(7,7),D1(6,6),A1(6),B1(6),X(6),E(6,6)
      DOUBLE PRECISION ASTAR,BSTAR,CSTAR,COSA,COSB,COSG
      DOUBLE PRECISION DEV(6)
      READ(1,1)ASTAR,BSTAR,CSTAR,ALPHS,BETS,GAMS,WLAMB,N,ISYS,ITER
      COSA=DCOS(ALPHS*3.141593/180.0)
      COSB=DCOS(BETS*3.141593/180.0)
      COSG=DCOS(GAMS*3.141593/180.0)
    1 FORMAT(3F9.6,3F7.2,F12.6,3I4)
      DO 10 I=1,N
      READ(1,2)IH,IK,IL,TWOTH
      R2OBS=DSIN(TWOTH*3.141593/360.0)
      R1OBS(I)=R2OBS*R2OBS
      UH(I)=IH
      UK(I)=IK
   10 UL(I)=IL
      DO 101 MITER=1,ITER
      DO 5 I=1,6
      B1(I)=0.0
      DO 5 J=1,6
    5 DI(I,J)=0.0
      SDEV=0.0
      WRITE (3,44)
```

```
      DO 99 M=1,N
      FLAMB=WLAMB*WLAMB/2.0
      HH=UH(M)
      HK=UK(M)
      HL=UL(M)
      IH=HH
      IL=HL
      IK=HK
      RICAL=FLAMB*((HH*ASTAR)**2+(HK*BSTAR)**2+(HL*CSTAR)**2+2.0*
     1(HH*HK*ASTAR*BSTAR*COSG+HH*HL*ASTAR*CSTAR*COSB+HK*HL*
     2BSTAR*CSTAR*COSA))/2
      DELRI=R1OBS(M)=RICAL
      SDEV=DELRI*DELRI+SDEV
      WRITE (3,50)IH,IK,IL,RICAL,DELRI
      A1(1)=FLAMB*HH*(HH*ASTAR+HK*BSTAR*COSG+HL*CSTAR*COSB)
      A1(2)=FLAMB*HK*(HK*BSTAR+HH*ASTAR*COSG+HL*CSTAR*COSA)
      A1(3)=FLAMB*HL*(HL*CSTAR+HH*ASTAR*COSB+HK*BSTAR*COSA)
      A1(5)=FLAMB*HK*HL*BSTAR*CSTAR
      A1(4)=FLAMB*HH*HL*ASTAR*CSTAR
      A1(6)=FLAMB*HH*HK*ASTAR*BSTAR
      DO 7 I=1,6
      B1(I)=B1(I)+(R1OBS(M)=RICAL)*A1(I)
      DO 7 J=1,6
      D1(I,J)=A1(I)*A1(J)+D1(I,J)
    7 A(I,J)=D1(I,J)
   99 CONTINUE
```

```
      S=N
      SS=ISYS
      SDEV=SDEV/(S-SS)
      WRITE (3,43)
      WRITE (3,15)SDEV
   15 FORMAT (1X,D15.7)
      WRITE (3,42)
      WRITE (3,47)ASTAR,BSTAR,CSTAR,COSA,COSB,COSG
      A(1,ISYS+1)=1.0
      DO 31 I=2,ISYS
   31 A(I,ISYS+1)=0.0
      DO 55 K=1,ISYS
      DO 32 J=1,ISYS
   32 A(ISYS+1,J)=A(1,J+1)/A(1,1)
      DO 33 I=2,ISYS
      COEFI=A(I,1)
      DO 33 J=1,ISYS
   33 A(I,J)=A(I,J+1)=COEFI*A(ISYS+1,J)
      DO 34 I=1,ISYS
      DO 34 J=1,ISYS
   34 A(I,J)=A(I+1,J)
   55 CONTINUE
      WRITE (3,70)
      DO 72 I=1, ISYS
   72 WRITE (3,80)(D1(I,J),J=1,ISYS)
      WRITE (3,73)
```

```
      WRITE (3,80)(B1(I),I=1,ISYS)
      WRITE (3,82)
      DO 81 I=1,ISYS
   81 WRITE (3,80)(A(I,J),J=1,ISYS)
      DO 45 I=1, ISYS
      DO 45 J=1,ISYS
      E(I,J)=0.0
      DO 45 K=1, ISYS
   45 E(I,J)=E(I,J)+D1(I,K)*A(K,J)
      WRITE (3,48)
      DO 46 I=1,ISYS
   46 WRITE (3,47)(E(I,J),J=1,ISYS)
      DO 75 I=1,6
   75 X(I)=0.0
      DO 40 I=1,ISYS
      DO 40 J=1,ISYS
   40 X(I)=X(I)+A(K,J)*B1(J)
      WRITE (3,49)
      WRITE (3,47)(X(I),I=1,ISYS)
      DO 41 I=1,ISYS
      B1(I)=0.0
      DO 41 J=1,ISYS
   41 B1(I)=B1(I)+D1(I,J)*X(J)
      WRITE (3,74)
```

```
      WRITE (3,47)(B1(I),I=1,ISYS)
      ASTAR=ASTAR+X(1)
      BSTAR=BSTAR+X(2)
      CSTAR=CSTAR+X(3)
      COSB=COSB+X(4)
      COSA=COSA+X(4)
      COSG=COSG+X(6)
      ALPHS=(DATAN(DSQRT(1.0-COSA*COSA)/COSA))*180.0/3.141593
      BETS=(DATAN(DSQRT(1.0-COSB*COSB)/COSB))*180.0/3.141593
      GAMS=(DATAN(DSQRT(1.0-COSG*COSG)/COSG))*180.0/3.141593
      IF(ALPHS)11,11,12
   11 ALPHS=ALPHS+180.0
   12 IF(BETS)13,13,14
   13 BETS=BETS+180.0
   14 IF(GAMS)16,16,17
   16 GAMS=GAMS+180.0
   17 WRITE(3,76)
      WRITE (3,77)ASTAR,BSTAR,CSTAR,ALPHS,BETS,GAMS
      DO 61 I=1,6
   61 DEV(I)=0.0
      DO 60 I=1,ISYS
   60 DEV(I)=DSQRT(SDEV*A(I,I))
      DEV(4)=DEV(4)*180.0/(3.141598*DSQRT(1.0-COSB*COSB))
      DEV(5)=DEV(5)*180.0/(3.141598*DSQRT(1.0-COSA*COSA))
      DEV(6)=DEV(6)*180.0/(3.141593*DSQRT(1.0-COSG*COSG))
      WRITE (3,103)
      WRITE (3,47)(DEV(I),I=1,ISYS)
```

```
101 CONTINUE
103 FORMAT (1X,30HSTAND. DEV. OF CELL PARAMETERS//)
 48 FORMAT (1X,10HA*AINVERSE//)
 80 FORMAT (1X,6D15.7//)
 82 FORMAT (1H1,8HAINVERSE//)
 49 FORMAT (1H1,25HSHIFTS IN CELL PARAMETERS//)
  2 FORMAT (3I4,F7.2)
 50 FORMAT (1X,3I4,2D15.7)
 44 FORMAT (1H1,2X,1HH,4X,1HK,4X,1HL,4X,10HSIN SQUARE,7X,5HDELTA//)
 43 FORMAT (1H1,15HSTAND. DEV. SQ.//)
 42 FORMAT (1H1,5HASTAR,10X,5HBSTAR,10X,5HCSTAR,10X,4HCOSA,11X,
     14HCOSB,11X,4HCOSG//)
 77 FORMAT (1X,3F12.7,3F11.3)
 76 FORMAT (1X,19HNEW CELL PARAMETERS//)
 74 FORMAT (1X,5HB=A*X//)
 47 FORMAT (1X,6D15.7//)
 73 FORMAT (1X,8HVECTOR B//)
 70 FORMAT (1H1,15HCOEFF. MATRIX A//)
    STOP
    END
```