

A study of bromocortisol acetate and  
other organic molecules by x-ray crystal analysis

Submitted to the University of Glasgow for  
the degree of Doctor of Philosophy in the  
Faculty of Science

by

R.A.L. Miller, B.Sc.

ProQuest Number: 11011840

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 11011840

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346

## CONTENTS

page

Preface

Summary

Chapter 1      Diffraction of x-rays by crystals and  
some methods of crystal structure analysis

1.1	The Laue equations	1
1.2	Bragg's Law	3
1.3	Factors affecting intensity	5
1.4	The structure factor	8
1.5	The Fourier series	10
1.6	The difference Fourier series	12
1.7	Trial and error methods	14
1.8	The Patterson map	16
1.9	The Patterson superposition	19
1.10	The use of heavy atoms	21
1.11	The heavy atom method	22
1.12	Isomorphous replacement	25
1.13	The method of least squares	27
1.14	Computing and temperature factors	29
	References	30

## CONTENTS (Contd.)

	page
<u>Chapter 2</u>	<u>The crystal structure analysis of</u> <u>9-<math>\alpha</math>-bromocortisol acetate</u>
2.1	Introduction 32
2.2	Experimental Data 38
2.3	The Structure Analysis 40
2.4	Tables and Figures
2.5	Results of the Analysis 50
2.6	Discussion 53
	References 59
 <u>Chapter 3</u>	 <u>The Crystal Structure Analysis of the</u> <u>Adduct of Triphenylarsine and</u> <u>Acetylene Dicarboxylic Methyl Ester</u>
3.1	Introduction 61
3.2	Experimental Data 65
3.3	The Structure Analysis 68
3.4	Tables and Figures
3.5	Results of the Analysis 74
3.6	Discussion 76
	References 83

## CONTENTS (Contd.)

	page
<u>Chapter 4</u>	<u>The Crystal Structure Analysis of</u> <u>2-(2-Hydroxyethylamino)-2-Thiazoline</u>
4.1	Introduction 84
4.2	Experimental Data 88
4.3	The Structure Analysis 90
4.4	Tables and Figures
4.5	Results of the Analysis 97
4.6	Discussion 98
	References 108

## PREFACE

I wish to express my thanks to my supervisors Professor J.M. Robertson, Professor G.A. Sim and Dr. G. Ferguson for their encouragement and advice during the course of my work.

I am grateful to Professor D.W.J. Cruickshank, Drs. J.G. Sime, K.W. Muir, D.R. McGregor, J.S. Rollet and W. Oberhansli for making available to me their computer programmes.

I am indebted to Professor J.M. Robertson, Dr. G. Ferguson and Mr. E.W. Macaulay for their critical and helpful reading of this thesis.

My thanks are due to the Department of Scientific and Industrial Research for their financial support during the time of this work.

Finally, my thanks go to my fiance Miss J. Grimshaw, for her help and patience during writing, and to Miss M.F. Butchard for her most excellent work in typing this thesis.

## SUMMARY

X-ray studies have been carried out on organic crystals which contain heavy atoms. Three structures have been successfully solved; the adduct of triphenylarsine and acetylene dicarboxylic methyl ester; the heterocyclic compound 2-(2-hydroxyethylamino)-2-thiazoline formed on reaction of 2-methylthio-2-oxazoline with 2-mercaptoethylamine; and the 9- $\alpha$ -bromo derivative of cortisol acetate.

In order to obtain molecular dimensions and overall shape for the molecule, an x-ray analysis of the 9- $\alpha$ -bromo derivative of cortisol acetate was undertaken. The total structure was evident after examination of the fourth electron density map. Refinement produced a final discrepancy, R, of 13.2% and showed that the expected distortion of the molecule due to the bromine atom, had not occurred.

Elucidation of the structure of the triphenylarsine adduct resulted from straightforward application of the heavy atom technique and the complete structure was assigned with certainty after the fourth three-dimensional Fourier synthesis. Subsequent refinement by Fourier and least squares methods reduced the discrepancy, R, to 14.3%. The structure was found to be other than that first predicted by considering the corresponding reaction

of triphenylphosphine. Two possible resonance forms have been proposed for the adduct and are supported by the results of this analysis.

In the reaction of 2-methylthio-2-oxazolines and 2-mercaptoethylamine, products with unusual properties were isolated. One such product was predicted to be a spiro compound and corroboration of this was sought from x-ray analysis of the crystal. The normal heavy atom technique failed to give a solution to the structure and the Patterson superposition method was applied. The structure was solved in this way and was refined to a discrepancy, R, of 12.1%. The structure was found to contain only one heterocyclic ring and not two as required by the proposed spiro compound.

Molecular dimensions for the three structures are in favourable agreement with the expected values and with values gained from other crystal structure analyses. The various classes of bond lengths and angles have been discussed in some detail in later sections.



# C H A P T E R   1

DIFFRACTION OF X-RAYS BY CRYSTALS AND  
SOME METHODS OF CRYSTAL STRUCTURE ANALYSIS

## 1.1

THE LAUE EQUATIONS

A crystal consists of a three-dimensional array of atoms, molecules or ions, described as a lattice. The smallest repetitive unit of the lattice is a volume called the unit cell having axes  $\underline{a}$ ,  $\underline{b}$  and  $\underline{c}$  and angles, between the axes,  $\alpha$ ,  $\beta$  and  $\gamma$ . The conditions for diffraction by x-rays are determined by this lattice irrespective of the crystal structure.

Supposing some scattering material such as an electron be placed at each lattice point, the positions of these electrons can be defined by a set of vectors  $\underline{r}$  such that

$$\underline{r} = u\underline{a} + v\underline{b} + w\underline{c}$$

where  $u$ ,  $v$  and  $w$  are integers and  $\underline{a}$ ,  $\underline{b}$  and  $\underline{c}$  are the vectors defining the lattice. Two such lattice points, A and B, are shown in figure 1.1.

The Laue equations are generated by consideration of the phase difference between the waves scattered in a particular direction when an x-ray beam impinges on the lattice. This beam of wave length  $\lambda$  is shown in figure 1.1 in a direction defined by a vector  $\underline{S}_0$  with magnitude  $1/\lambda$ . The waves scattered from the two lattice points A and B are in the direction defined by the vector  $\underline{S}$  which is equal in magnitude to  $\underline{S}_0$  and the path difference between the waves is given by

$$\begin{aligned} AD - CB &= \lambda(\underline{r} \cdot \underline{S} - \underline{r} \cdot \underline{S}_0) \\ &= \lambda \underline{r} \cdot \underline{R} \end{aligned}$$

$$\text{where } \underline{R} = \underline{S} - \underline{S}_0$$

The condition for the wave scattered at A to be in phase with the wave scattered at B is that the path difference  $\underline{r} \cdot \underline{R}$  must be an

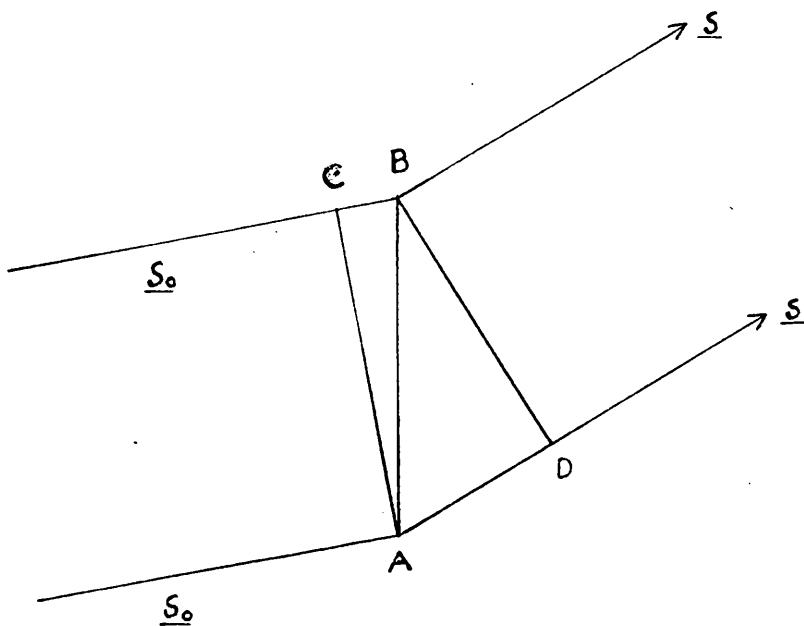


FIGURE 1.1

integer,

$$\text{i.e. } (\underline{u}\underline{a} + \underline{v}\underline{b} + \underline{w}\underline{c}) \cdot \underline{R} = \text{integer}$$

Since this must be true when  $u$ ,  $v$  and  $w$  change by integral values and since  $u$ ,  $v$  and  $w$  are already integral, it follows that

$$\underline{a} \cdot \underline{R} = h$$

$$\underline{b} \cdot \underline{R} = k$$

$$\underline{c} \cdot \underline{R} = l$$

where  $h, k$  and  $l$  are integers. These are the Laue equations.

W.L. Bragg identified the integers  $h$ ,  $k$  and  $l$  of the Laue equations with the Miller indices of the lattice planes and thus made the equations more convenient for the interpretation of x-ray spectra and the determination of crystal structure. In order to demonstrate the relationship between the Laue equations and the Bragg law, the Laue equations may be written in the form

$$\underline{a}/h \cdot \underline{R} = 1$$

$$\underline{b}/k \cdot \underline{R} = 1$$

$$\underline{c}/l \cdot \underline{R} = 1$$

By subtracting these in pairs, the following relationships are gained,

$$(\underline{a}/h - \underline{b}/k) \cdot \underline{R} = 0$$

$$(\underline{a}/h - \underline{c}/l) \cdot \underline{R} = 0$$

$$(\underline{b}/k - \underline{c}/l) \cdot \underline{R} = 0$$

The vector  $\underline{R}$  is therefore at right angles to the vector  $(\underline{a}/h - \underline{b}/k)$  and similarly is at right angles to both  $(\underline{a}/h - \underline{c}/l)$  and  $(\underline{b}/k - \underline{c}/l)$ . In figure 1.2,  $\underline{R}$  is perpendicular to  $PQ$ ,  $QR$  and  $RP$  and is therefore perpendicular to the plane  $PQR$  which has intercepts  $(\underline{a}/h, \underline{b}/k, \underline{c}/l)$  on the unit cell axes. Thus  $\underline{R}$  is in the direction of the normal to the plane with Miller indices  $(hkl)$ .

In figure 1.3  $\underline{S}_o$  makes an angle  $\theta$  with the plane and since  $|\underline{S}_o| = |\underline{S}|$  by definition,  $\underline{R}$  must be the bisector of the incident and diffracted beams. Thus  $\underline{S}$  must also be inclined at an angle  $\theta$  and so diffraction may be regarded as a reflection from the lattice

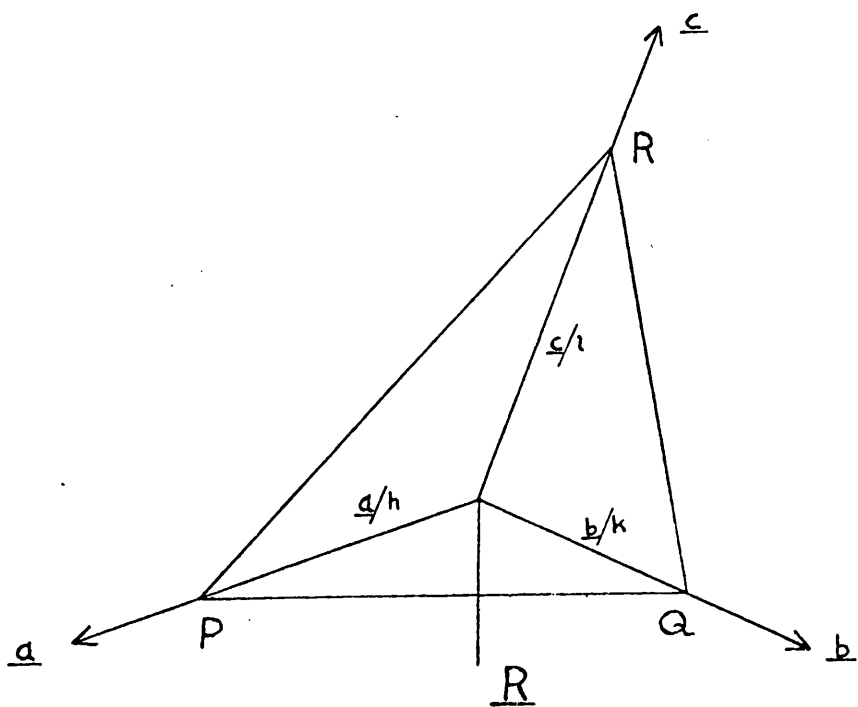


FIGURE 1.2

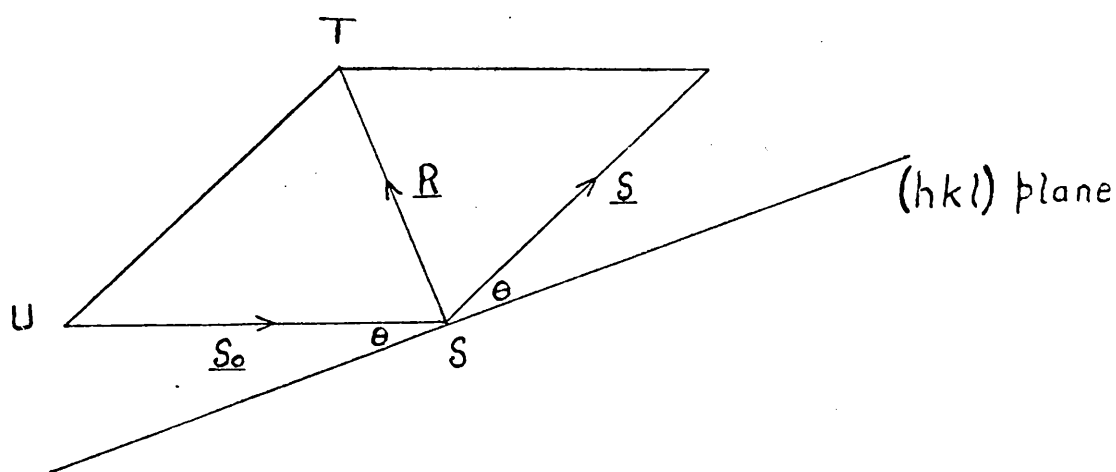


FIGURE 1.3

plane with Miller indices  $(hkl)$ .

If  $\underline{US}$  represents the vector  $\underline{So}$  then,

$$\begin{aligned} |\underline{R}| &= ST = 2US \cos (90^\circ - \Theta) \\ &= 2 \sin \Theta / \lambda \end{aligned}$$

Now  $d$ , the distance from the origin to the plane  $PQR$  is the projection of the vector  $\underline{a}/h$ ,  $\underline{b}/k$ ,  $\underline{c}/l$  on the vector  $\underline{R}$ ,

$$\text{i.e. } d = (\underline{a}/h \cdot \underline{R}) / |\underline{R}| = \lambda / 2 \sin \Theta$$

since  $\underline{a}/h \cdot \underline{R} = 1$  by Laue's equations,

$$\lambda = 2d \sin \Theta$$

This equation is known as Bragg's Law.



## 1.3

FACTORS AFFECTING INTENSITY

Although it is possible to relate a proposed structure to a set of relative intensities, it is necessary in modern crystallography to have a formula which will relate the intensity of the primary beam to that of the various reflected beams. For reflections recorded from a single ideally imperfect crystal on a Weissenberg photograph, the intensity,  $I$ , can be related to the structure amplitude,  $|F|$ , by the equation,

$$I = \frac{N^2 e^4 \lambda^3 v}{4\pi^2 C^4} \cdot \frac{1 + \cos^2 2\theta}{\sin 2\theta} \cdot \frac{1}{2\mu} \cdot F^2.$$

The first group of terms is a constant for the crystal. The factor

$\frac{1}{2\mu}$  corrects for absorption in the crystal but may be neglected where only light atoms are present in the crystal or where the crystal is very small. The expression  $\frac{1 + \cos^2 2\theta}{\sin 2\theta}$  depends on  $\theta$  and allows for the effect of polarisation on the x-ray beam and of the varying time crystal planes spend in the reflecting position. This latter effect is known as the Lorentz factor and is responsible for the  $\frac{1}{\sin 2\theta}$  part of the expression. Further corrections may be applied to account for the effects of primary and secondary extinction.

There are two additional factors which reduce the intensity of the resultant beam and are of great importance in crystal structure analysis. The first is due to the scattering units not being point sources. Since the diameter of the electron cloud of an atom is of the order of one or two Angstroms, rays scattered from different points in this volume are not necessarily in phase and the intensity of the

resultant beam may be reduced. The scattering power of an atom is designated  $f$ , the atomic scattering factor, and is a function of the angle of scattering and of the distribution of electrons in the atoms. For small angles of diffraction, the phase differences are small and therefore the scattered amplitude will approach  $Z$ , the atomic number of the atom.

The second factor which reduces the intensity of the resultant beam is that of the thermal vibrations of the atoms. At ordinary temperatures atoms in crystals vibrate with some atoms being displaced from their mean positions in one direction while others are displaced in other directions. This introduces further phase differences resulting in a reduction of intensity. By using the factor  $f = f_0 \exp. \left( \frac{-B \sin^2 \theta}{\lambda} \right)$ , approximate allowance can be made for thermal motion (1). In the expression above  $\theta$  is the Bragg angle,  $f_0$  is the atomic scattering factor for atoms at rest and  $B$ , the Debye temperature factor, is a constant. The value of  $B$  is given by  $B = 8\pi^2 \bar{U}^2$  where  $\bar{U}^2$  is the mean square displacement of the atoms from their mean positions. In general every crystallographically independent atom in a unit cell will have different thermal vibrations since this depends on the surroundings of an atom as well as on its inertia. The expression given above assumes that the thermal vibrations are equal in all directions. This is not necessarily true and in many molecules vibrations are markedly anisotropic. This is usually ignored at the beginning of a structure analysis and can be allowed for in later stages of refinement.

In the anisotropic case the vibrations of an atom are described by a symmetrical tensor,  $\underline{U}$ , with six independent components such that the mean square amplitude of vibration in the direction of a unit vector  $\underline{l} = (l_1, l_2, l_3)$  is

$$\bar{U}^2 = \sum_{i=1}^3 \sum_{j=1}^3 U_{ij} l_i l_j$$

$$\text{or } \bar{U}^2 = U_{11} l_1^2 + U_{22} l_2^2 + U_{33} l_3^2 + 2U_{23} l_2 l_3 + 2U_{31} l_3 l_1 + 2U_{12} l_1 l_2$$

$U$  and  $l$  are defined with respect to the reciprocal axes  $\underline{a}^*$ ,  $\underline{b}^*$ ,  $\underline{c}^*$  so that the components of  $\underline{U}$  in the (100) direction parallel to  $\underline{a}^*$  is

$$\bar{U}^2 = U_{11}$$

If the axes are orthogonal the direction  $30^\circ$ , say, round from  $\underline{a}^*$  towards  $\underline{b}^*$  has  $l = (\frac{\sqrt{3}}{2}, \frac{1}{2}, 0)$  and the component of  $\underline{U}$  in that direction is

$$\bar{U}^2 = U_{11} \left(\frac{\sqrt{3}}{2}\right)^2 + U_{22} \left(\frac{1}{2}\right)^2 + 2U_{12} \left(\frac{\sqrt{3}}{2}\right) \left(\frac{1}{2}\right)$$

$$= \frac{3}{4} U_{11} + \frac{1}{4} U_{22} + \frac{\sqrt{3}}{2} U_{12}$$

## 1.4

THE STRUCTURE FACTOR

A crystal with  $N$  atoms in the unit cell can be regarded as consisting of  $N$  interpenetrating lattices with all the lattice points occupied by atoms. These lattices will obey the Laue and Bragg conditions but the  $N$  lattices will in general be out of phase. The intensities of the various planes will therefore depend upon the atomic arrangement within the unit cell.

The position of the  $j$ th atom situated at the point  $(x_j, y_j, z_j)$  where  $x$ ,  $y$  and  $z$  are fractions of the unit cell vectors can be represented by the vector  $\underline{r}_j$ . The path difference between the waves scattered by the  $j$ th atom and an atom at the origin of the unit cell can be shown to be  $\lambda \underline{r}_j \cdot \underline{S}$  and the corresponding phase difference is  $2\pi \underline{r}_j \cdot \underline{S}$ . By combining these phase differences for all atoms in the unit cell, the expression for the complete wave scattered in the crystal would be

$$\begin{aligned}
 F &= \sum_{j=1}^N f_j \exp. 2\pi i \underline{r}_j \cdot \underline{S} \\
 &= \sum_{j=1}^N f_j \exp. 2\pi i (x_j \cdot \underline{a} \cdot \underline{S} + y_j \cdot \underline{b} \cdot \underline{S} + z_j \cdot \underline{c} \cdot \underline{S}) \\
 F(hkl) &= \sum_{j=1}^N f_j \exp. 2\pi i (hx_j + ky_j + lz_j)
 \end{aligned}$$

In this expression  $f_j$  represents the scattering factor of the atom.

The quantity  $F(hkl)$  is known as the structure factor and it is a complex

quantity which can be represented by a modulus,  $|F(hkl)|$ , known as the structure amplitude and a phase constant  $\alpha(hkl)$ . The structure amplitude can be obtained from the measured intensities but the phase is not an observable quantity. The structure factor can be evaluated by means of the expressions

$$\begin{aligned} |F(hkl)| &= \sqrt{A^2 + B^2} \\ \alpha(hkl) &= \tan^{-1} B/A \end{aligned}$$

$$\text{where } A = \sum_{j=1}^N f_j \cos 2\pi(hx_j + ky_j + lz_j)$$

$$\text{and } B = \sum_{j=1}^N f_j \sin 2\pi(hx_j + ky_j + lz_j)$$

This is valid for crystals of any symmetry but when a centre of symmetry is present at the origin there is no need to calculate the sine terms since they are bound to add to zero. The structure factor can therefore be obtained by summing the cosine terms alone and the possible phase angles are thus limited to 0 or  $\pi$  depending upon whether the expression for A is positive or negative. In terms of the electron density

$\rho(xyz)$  the structure factor can be expressed

$$F(hkl) = \int_0^1 \int_0^1 \int_0^1 \rho(xyz) \exp. 2\pi i(hx + ky + lz) dx.dy.dz.$$

## 1.5

THE FOURIER SERIES

Within reasonable limitations, a periodic function can be represented by an appropriate sum of sines and cosines known as a Fourier series. The three dimensional Fourier series expressed in the terms of the theory of the complex variable is

$$f(xyz) = \sum_n \sum_{-\infty}^{\infty} \sum_p A(nmp) \exp. - 2\pi i \left( \frac{nx}{a} + \frac{my}{b} + \frac{pz}{c} \right)$$

Since a crystal is periodic, Bragg, in 1915, suggested that its electron density could be represented by such a series. In order to evaluate this series and so obtain the electron density at any point in the crystal, it is necessary to calculate the coefficients  $A(nmp)$ . Substituting this series in the general expression for the structure factor gives,

$$F(hkl) = \int_0^1 \int_0^1 \int_0^1 \sum_{-\infty}^{\infty} \sum_p VA(nmp) \exp. 2\pi i(mx + ny + pz) \\ \times \exp. 2\pi i(hx + ky + lz) dx dy dz$$

On integrating, every term is zero except that for which  $m = -h$ ,  $n = -k$  and  $p = -l$  which gives,

$$F(hkl) = \int_0^1 \int_0^1 \int_0^1 VA(\bar{h}\bar{k}\bar{l}) dx dy dz.$$

$$\text{i.e. } F(hkl) = VA(\bar{h}\bar{k}\bar{l})$$

This shows the Fourier coefficients,  $A$ , to be directly related to the corresponding structure factors. The electron density,  $\rho$ , is given by,

$$\rho_{(xyz)} = \sum_{hkl} \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} \frac{F(hkl)}{V} \exp. - 2\pi i(hx + ky + lz)$$

The zero term of the series is a constant,

$$F(000) = \int_0^1 \int_0^1 \int_0^1 \rho(xyz) dx. dy. dz. = Z$$

A more convenient form of this series is

$$\rho_{(xyz)} = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} \frac{|F(hkl)|}{V} \cos [2\pi(hx + ky + lz) - \alpha(hkl)]$$

where  $\alpha(hkl)$  is the phase angle associated with each structure factor.

## 1.6

THE DIFFERENCE FOURIER SERIES

This synthesis is a Fourier summation utilising coefficients  $(F_o - F_c)$  and is sometimes called the error synthesis. It is due to Bunn and has been studied and exploited by Booth (2) and by Cochran (3). The correctness of a proposed structure can be tested by this method since a correct structure will yield a difference map which is more or less flat with only small undulations due to random errors in the observed data and in the proposed model. If in the structure there are wrongly placed atoms, large undulations will appear and the positions of the incorrectly placed atoms will appear in holes while their true positions will be indicated by peaks. The structure must then be modified to give an almost featureless map. In this way the difference synthesis proved useful in establishing the crystal structure of sodium benzyl penicillin (4).

In the refinement of a structure the difference synthesis can be used to gain more accurate coordinates. This follows from the fact that when a proposed structure and actual structure are almost the same, the series termination errors are nearly identical and therefore vanish on subtraction. Also temperature factor corrections can be made for both isotropic and anisotropic thermal motion. These have different but characteristic contour shapes and can therefore be used to detect anisotropic motion. A further discrepancy between  $F_o$  and  $F_c$  is caused by the omission of the hydrogen atoms which are not normally visible on a Fourier map. When the positions of the rest



of the atoms have been determined the hydrogen atoms may show up in the difference map as fairly well defined peaks. This is one of the main advantages of this technique and is of particular importance in the study of hydrogen bonding in smaller molecules.

## 1.7

TRIAL AND ERROR METHODS

This method is of particular use in the elucidation of the structure of certain 'special' molecules. Small molecules having symmetry which is utilised by the space group, flat symmetrical molecules such as analogues of benzene and structures in which orientation of the molecule is the main problem are examples of the type of structure susceptible to a solution by the trial method. Examples of these are dibenzyl(5) and diphenyl (6). The trial structures are arrived at by various methods. Considerations of symmetry, intense spots and unit cell dimensions can give information on molecular orientation and shape. An early example of this was the use of x-rays in the field of steroids. Bernal's unit cell dimensions (7) of various sterols contributed to settling the controversy surrounding the structure of these compounds, and further, showed that the trans configuration was preferred.

Another useful method of arriving at a trial structure is the construction of a weighted reciprocal lattice from which molecular orientation may be found. If a number of structures are possible, the calculation of structure factors for all of them and comparison with the observed values becomes a laborious and time consuming task. A useful way of circumventing this is to build up the Fourier transform of a set of atoms. Holes can be punched representing the atomic positions in projection of the proposed structure, the size of the holes representing the chemical type. This mask is placed in front of parallel light and the diffraction pattern represents the required Fourier transform which can be compared with the relevant weighted

reciprocal lattice. In this way possible structures are quickly evaluated and can be further developed.

## 1.8

THE PATTERSON MAP

The Patterson map is a vector representation of crystal structure. In this approach the unknown phases of the structure factors are temporarily ignored by introducing the structure amplitude,  $|F_o|$ , into a Fourier summation in the form  $|F_o|^2$  and studying the information made available.

In Patterson's treatment of the problem (8) a quantity  $A(u,v,w)$  which can be called the weighted average distribution of density of scattering matter about a point  $(x,y,z)$  in the crystal is defined by the equation,

$$A(u,v,w) = \frac{1}{V} \int_0^a \int_0^b \int_0^c \rho(x,y,z) \rho(x+u, y+v, z+w) dx dy dz.$$

If the two functions of  $\rho$  are expanded by expressing them in terms of the corresponding Fourier series we have,

$$A(u,v,w) = \frac{1}{V^2} \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} |F|^2(hkl) \cos 2\pi \left( \frac{hu}{a} + \frac{kv}{b} + \frac{lw}{c} \right)$$

This equation is simplified for the centre of symmetry at the origin always present in a Patterson map and shows that  $A(u,v,w)$  may be calculated directly from the information furnished by the x-ray spectra.  $A(u,v,w)$  can only attain a large value if both  $\rho(x,y,z)$  and  $\rho(x+u, y+v, z+w)$  are large.

This would appear to be the perfect solution of the phase problem and indeed for very small molecules the atomic positions can often be found from consideration of the vector map. However, the complexity of the vector map increases rapidly with increase in the number of atoms in the unit cell resulting in overlap of peaks. The possibility

of resolving very many peaks then becomes slim since the number of vectors from a unit cell containing  $n$  atoms is  $n(n - 1)/2$ . In general the unravelling of a vector map to give atomic positions is a most difficult task and in the case of complex molecules is virtually impossible. If the atoms were point sources the problem of overlap would be greatly alleviated. This is the basis for the sharpened Patterson function (9, 10) in which the structure factors,  $F(hkl)$ , are converted to sharp structure factors,  $F(s)(hkl)$ , by the relation

$$F(s)(hkl) = F(hkl)/\hat{F}(hkl)$$

where  $\hat{F}(hkl)$  is the average of the scattering curves of the atoms present, scaled to unit power in a given crystal. Using this method of improving resolution and utilising the three-dimensional Patterson function a measure of success has been gained for fairly complex structures (11,12).

There are certain sections in Patterson maps which are of special interest in that they should theoretically contain only vectors from symmetrically related atoms. For example in the space group  $P2_1/C$  with equivalent positions  $(x,y,z)$ ,  $(\bar{x},\bar{y},\bar{z})$ ,  $(x, \frac{1}{2} - y, \frac{1}{2} + z)$ ,  $(\bar{x}, \frac{1}{2} + y, \frac{1}{2} - z)$  the special vectors  $(0, \frac{1}{2} - 2y, \frac{1}{2})$  and  $(2x, \frac{1}{2}, \frac{1}{2} + 2z)$  arise. The former is known as a Harker line and the latter a Harker section. They can provide valuable information without the necessity of a full three dimensional computation although lack of resolution and overlap is still a problem of complex molecules. Chance vectors which are not symmetrically related can occur on both the Harker line and the Harker section especially where a large number of atoms are present in the unit cell and it would appear that the most important outlet for

the Patterson technique lies in the analysis of molecules containing one or more heavy atoms.

## 1.9

THE PATTERSON SUPERPOSITION

This method of solving vector maps was discovered independently and simultaneously by Clastre and Gay (13), by Garrido (14), by Beevers and Robertson (15) and by McLachlan (16). The method used by us and described here is the vector convergence method of Beevers and Robertson. It is of particular use in a crystal of fairly high symmetry with about four or more equivalent positions where a moderately heavy atom is present which does not adequately phase the Fourier synthesis. If the coordinates of the heavy atom are found from a Patterson map then this method is applicable to the three dimensional Patterson synthesis.

Since there are fairly high peaks in the Patterson due to vectors between heavy and light atoms, located at the ends of vectors emanating from the heavy atom, there is an image of the crystal structure from each of the  $n$  heavy atoms. In the Patterson map there are as many images as there are heavy atoms and by placing the origins of the  $n$  Patterson maps on the respective heavy atom positions, the atoms of the crystal structure may be picked out where there are  $n$  densities superposed on each other. No attempt is made at this stage to discern between light atoms of different chemical type, the peak height on the Patterson degenerating to a cross on the map marking a position. Chemical types can usually be assigned at a later stage of refinement from various considerations such as chemical knowledge, bond lengths and peak heights. A number of structures have been solved by the superposition method. One example is the structure of

Limonin, the bitter principle of citrus fruits, an iodoacetate derivative of which was solved by a combination of superposition and Fourier methods (17).

In a more mathematical approach, Buerger (18) has shown that there are certain advantages in forming the minimum functions which is the lowest value for superposed peaks in the Patterson. There is a high probability that peaks in the minimum function will represent true atomic sites.



There are two methods of utilising heavy atoms in crystal structure determination. These are known as the heavy atom method and the isomorphous replacement method. In the heavy atom method only a single form of the crystal is required and the phases used in the first Fourier summation are those calculated for the heavy atom alone. The problem of how heavy an atom is required has been considered by Luzatti (19). For a compound containing between twenty and thirty light atoms a bromine atom should be adequate while an iodine atom should be sufficient to solve a structure containing a much larger number of atoms. With very large molecules, where molecular weights exceed 10,000, an impossibly heavy atom would be required and at present the most productive attack appears to be by the method of isomorphous replacement. This latter method, described in more detail in a subsequent section, requires at least two forms of the crystal which are isomorphous. For proteins the unit cell and the protein molecule must be essentially the same in each crystal but one of them must have one or more heavy atoms at a location where the other has none. The position of the heavy atom must first be found by some Patterson method and in favourable cases the phase of the protein contribution can be found. An example is the structure analysis of horse haemoglobin (molecular weight - 34,000) using silver and mercury atoms for isomorphous replacement (20).

## 1.11

THE HEAVY ATOM METHOD

In the majority of crystal structure analyses of complex molecules, success has depended upon the presence of one or more atoms whose atomic numbers are considerably greater than those of the other atoms. This method involves approximations which in favourable cases are close enough to allow the Fourier method of refinement to be utilised. The structure factor,  $F$ , may be split into two parts

$$F = F(h) + F(l)$$

where  $F(h)$  is the contribution to the structure factor of the heavy atom and  $F(l)$  the contribution of the light atoms.

The centrosymmetric case is the simpler since the phase angle is either 0 or  $\pi$ . A suitable heavy atom derivative would be such that the sign of  $F$  and  $F(h)$  are identical in a large majority of cases. The ability to predict whether a heavy atom will give a sufficient percentage of identical signs for  $F$  and  $F(h)$  has been investigated (21). This percentage is dependent on  $r$  which is

$$r = (\sum F^2(h) / \sum F^2(l))^{1/2}$$

When  $r = 1$ , 80% of the signs are the same while if  $r = 2$ , 90% of the signs are identical. In a typical heavy atom structure analysis  $r$  is about 1 or 2. It would appear that the higher the value of  $r$ , the better the chance of solving the structure. However, there are disadvantages accompanying increase in  $r$ . As  $r$  increases,  $F(h)$  becomes much greater than  $F(l)$  and since details of the light atoms depend upon  $F(l)$ , resolution and accuracy of this part of the molecule decreases.

Further error is caused by absorption which increases with increase of the heavy atom size.

In the non-centrosymmetric case (22) the phase of the structure factor is not given correctly or incorrectly by the heavy atom method but rather a distribution of error dependent on  $r$ , is obtained. For example when  $r = 1$ , 38% of the errors lie in the range  $\pm 20^\circ$  of the correct phase angle while when  $r = 2$  the proportion in the  $\pm 20^\circ$  range rises to 67%. Luzatti (19) has shown that for a given  $r$ , the resolution in the non-centrosymmetric case is much poorer than in the centrosymmetric case.

The location of the heavy atom in the crystal can be determined by the Patterson method. When a heavy atom is present the vectors between light atoms become second order quantities compared with the heavy atom vectors. Therefore large peaks will show on the Patterson map from which the heavy atom position in the crystal may be deduced. This will provide approximate phasing for application of the Fourier method which in favourable conditions should provide a picture of some or all of the atoms of the structure. Further application of the Fourier synthesis with phases appropriate to the heavy atom and light atoms should reveal the total structure and give more accurate coordinates for the atoms. Figure 1.4(a) portrays the approximate phasing gained from the heavy atom method.  $F(1)$  represents the contribution of the unknown part of the structure and the end of this vector may lie anywhere on the circumference of circle 1. The addition of  $F(1)$  to  $F(h)$ , the calculable contribution of the heavy

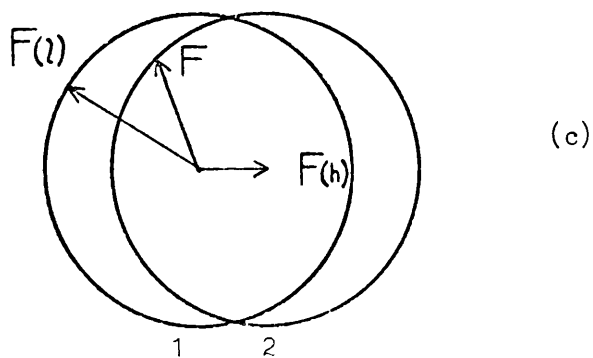
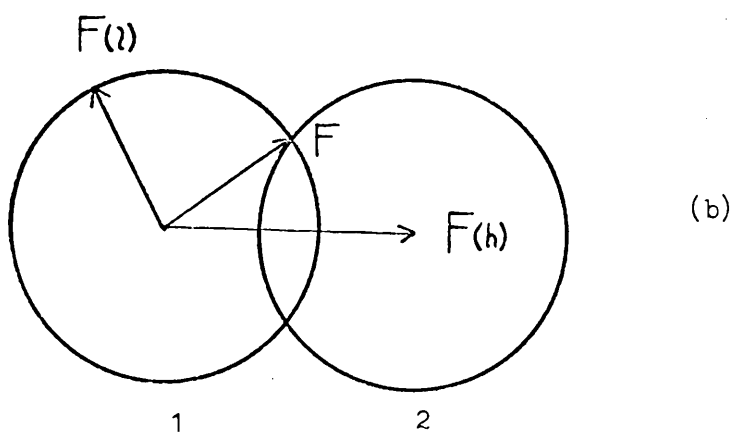
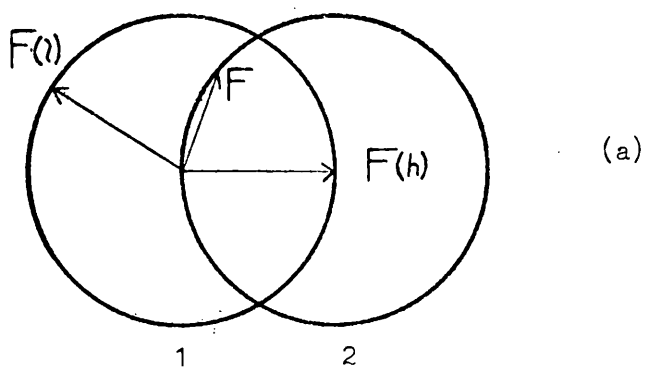


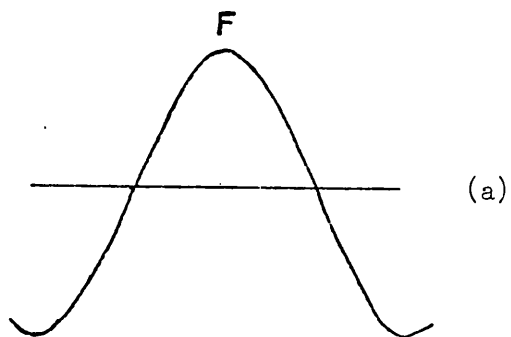
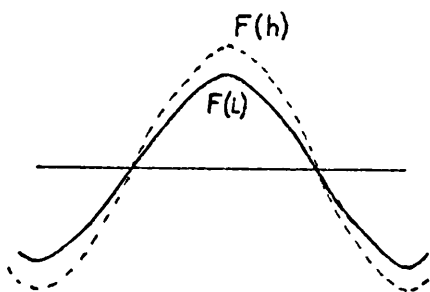
FIGURE 1.4

atom, gives the structure factor of the heavy atom derivative  $F$  and the end of this vector is now restricted to lie somewhere on the circumference of circle 2. The phases of the structure factors of the heavy atom derivative,  $F$ , are therefore restricted to a considerable extent by the phases of the heavy atom  $F(h)$ . By assigning these latter phases to the structure factors, an approximation to the structure of the heavy atom derivative is usually obtained.

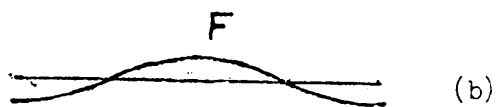
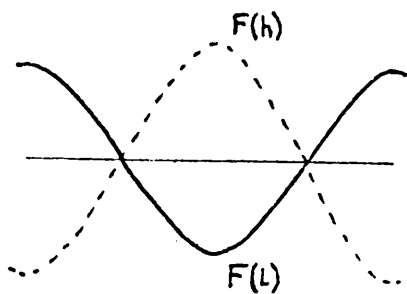
Figure 1.4(a) assumes  $|F(1)| = |F(h)|$ , (b) assumes  $|F(h)| > |F(1)|$  with the phases now more closely restrained to those of  $F(h)$  and (c) shows  $|F(h)| < |F(1)|$  and hence poorer approximation.

A more powerful approach to the phase problem is the isomorphous replacement method. If two different heavy atoms can be substituted successively in the molecule without altering the overall crystal structure the phase relationships can be calculated from a difference effect. With the help of accurate intensity measurements it is not necessary for the heavy atom contribution to outweigh the contribution of the remaining part of the molecule. The principle of the method is shown by figure 1.5(a). for the centrosymmetric case. The waves scattered by the atoms in the unknown structure combine to give a resultant  $F(1)$  of known amplitude but unknown phase. If a centre of symmetry is present the phases may be represented by a peak or by a trough as in the diagram.

If another atom is added containing a large number of electrons and which does not disturb the remainder of the structure to any large extent, the contribution of this atom can be represented by the dotted peak  $F(h)$  and the resultant amplitude of the heavy atom derivative is  $F$ . Thus by noting whether the resultant  $F$  increases or decreases, we can determine whether the original amplitude,  $F(1)$ , was a peak or a trough. This is the principle of the isomorphous replacement method. An almost perfect example was found in the phthalocyanine structures (23). If the heavy atom derivative,  $F$ , is isomorphous with the parent compound and if accurate measurements can be made then this method can be used equally well in the asymmetric case, although complications are introduced. In figure 1.6(a) it can be seen that there are two



(a)



(b)

FIGURE 1.5

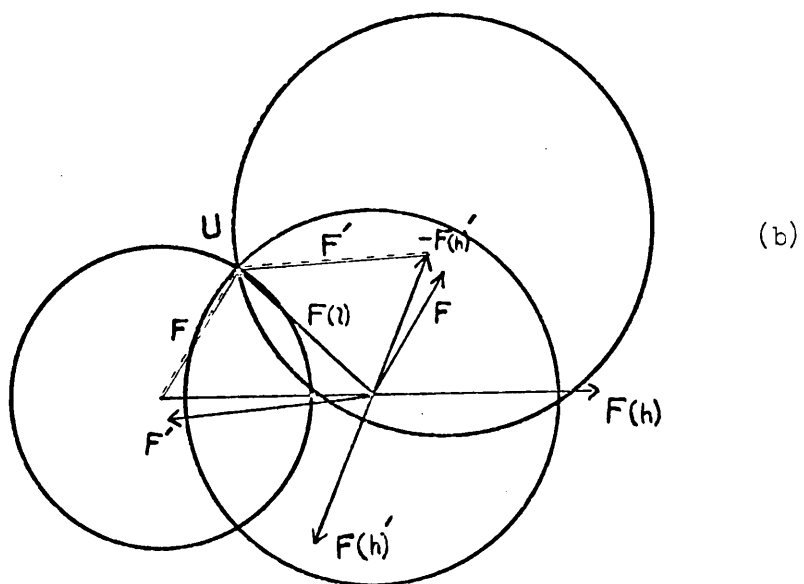
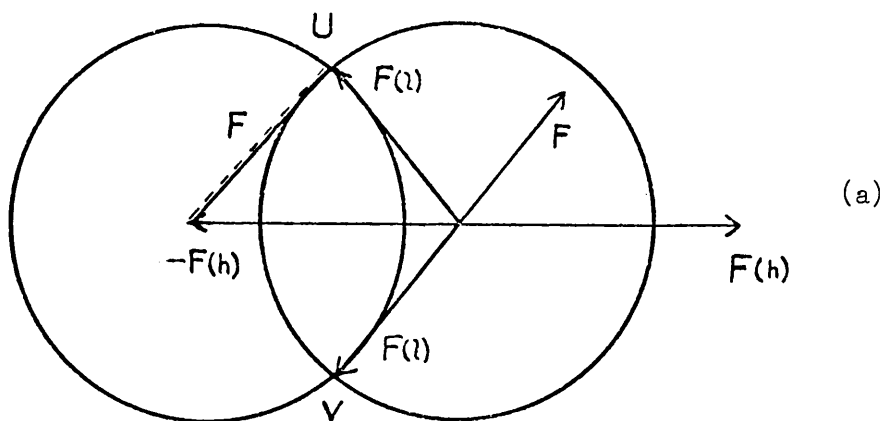


FIGURE 1.6



solutions to the vector equation  $F = F(h) + F(l)$  (24). These are indicated in the figure at the points of intersection, U and V, of the two circles. Each value of  $F(l)$  can be included in a Fourier summation, the resultant electron density map containing a spurious centre of symmetry. This can often be unravelled by consideration of known chemical features such as benzene rings etc.

The complete solution can be obtained if a third isomorphous derivative is available. This multiple isomorphous replacement method has been studied by Bijvoet et al. (24) and by Harker (25). When another isomorphous heavy atom derivative is available with structure factor  $F^1$  and a known heavy atom contribution  $F(h)^1$ , the unambiguous solution for  $F(l)$  is at the point U in figure 1.6(b), the point of intersection of the three circles. If isomorphism were perfect and measurements accurate, the solution would be as complete in the asymmetric case as in the centrosymmetric case of the phthalocyanines. The main outlet for this method is in the study of proteins. In fact, the multiple isomorphous substitution is not much used in the elucidation of smaller natural products, mainly because of the difficulty in obtaining suitable isomorphous derivatives and because of the fact that the simpler heavy atom method is usually adequate. There are of course examples, a recent one being the solution of the structure of aflatoxin  $G_1$  (26) in which three different solvates proved to be closely isomorphous.

The use of least squares was first introduced by Hughes (27) in the structure analysis of melamine. The two main advantages of this method are that the effects of series termination are overcome and the influence of inaccurate coefficients can be decreased.

The best values of the atomic parameters are those which minimise the function,

$$R = \sum_q w(hkl) \left[ |F_o| - |F_c| \right]^2 \quad \dots\dots\dots (1)$$

where  $q$  is the number of independent observations and  $w(hkl)$  is the weight of the plane  $(hkl)$  and should be inversely proportional to the square of the probable error in the corresponding  $|F_o|$ . If  $P_1, P_2, \dots, P_r$  are the  $n$  parameters in the  $|F_c|$  whose values are to be determined then for  $R$  to be a minimum

$$\frac{\partial R}{\partial P_j} = 0 \quad \dots\dots\dots (2)$$

$$\text{i.e.} \quad \sum w \Delta \frac{\partial |F_c|}{\partial P_j} = 0 \quad \dots\dots\dots (3)$$

where  $\Delta = |F_o| - |F_c|$ . For a trial set of  $P_j$  close to the correct values  $\Delta$  may be expanded by Taylor's series thus;

$$\Delta(P + E) = \Delta(P) - \sum_i E_i \frac{\partial |F_c|}{\partial P_i} \quad \dots\dots\dots (4)$$

where  $E_i$  is a small change in a parameter  $P_i$ .

Substituting (4) in (3) produces the normal equations.

$$\sum_{i=1}^n \left( \sum_q w \frac{\partial |Fc|}{\partial P_i} \cdot \frac{\partial |Fc|}{\partial P_j} \right) \cdot E_i = \sum_q w \Delta \frac{\partial |Fc|}{\partial P_j} \dots\dots\dots (5)$$

These may be written in the matrix form;

$$\sum_i a_{ij} E_i = b_j$$

$$\text{where } a_{ij} = \sum_w \frac{\partial |Fc|}{\partial P_i} \cdot \frac{\partial |Fc|}{\partial P_j}$$

$$\text{and } b_j = \sum_w \Delta \frac{\partial |Fc|}{\partial P_j}$$

During the course of this thesis both the D.E.U.C.E. computer and the K.D.F.9 computer were used. Initially the capacity of the K.D.F.9 machine was limited such that fully anisotropic structure refinement was impracticable. This is referred to later and is the reason for the at times confusing use of partial anisotropic, partial isotropic refinement.

Three isotropic temperature factors were used during the course of this thesis. The D.E.U.C.E. computer was programmed to use  $\alpha$  during Fourier refinement and  $B\theta$  during least squares refinement while the K.D.F.9 utilised  $U(\text{iso})$  for both Fourier and least squares refinement. In order to avoid confusion all temperature factors have been converted to  $U(\text{iso})$ .

# REFERENCES - CHAPTER 1

1. J.M. Robertson, Z. Kryst., 112, 1959.
2. A.D. Booth, Nature, 161, 765, 1948.
3. W. Cochran, Acta Cryst., 4, 408, 1951.
4. D. Crowfoot, C.W. Bunn, B.W. Rogers-Low, A. Turner-Jones,  
Chemistry of Penicillin, Princeton University Press, 1949.
5. J.M. Robertson, Proc. Roy. Soc., A146, 473, 1934.
6. G.L. Clark, L.W. Picket, J. Amer. Chem. Soc., 53, 167, 1931.
7. J.D. Bernal, Nature, 129, 277 and 721, 1932.
8. A.L. Patterson, Phys. Rev., 46, 372, 1934.
9. A.L. Patterson, Z. Cryst., 90, 517, 1935.
10. S.M. Yu, Nature, 149, 638, 1942.
11. D.P. Shoemaker, J. Donohue, V. Shomaker, R.B. Corey,  
J. Amer. Chem. Soc., 72, 2328, 1950.
12. T.R.R. McDonald, C.A. Beevers, Acta Cryst., 3, 394, 1950.
13. J. Clastre, R. Gay, Compt. rend., 230, 1876, 1950.
14. J. Garrido, C.R. Acad. Sci., Paris, 230, 1878, 1950.  
231, 297, 1950.
15. C.A. Beevers, J.H. Robertson, Acta Cryst., 4, 270, 1951.
16. D. McLachlan Jr., ibid, 37, 115, 1951.
17. S. Arnott, A.W. Davie, J.M. Robertson, G.A. Sim. D.G. Watson,  
J. Chem. Soc., 4183, 1961.
18. M.J. Buerger, Proc. Nat. Acad. Sci., U.S., 36, 376 and 738, 1950.
19. V. Luzatti, Acta Cryst., 6, 142, 1953.
20. D.W. Green, V.M. Ingram, M.F. Perutz, Proc. Roy. Soc., A225, 237,  
1954.
21. G.A. Sim, Acta Cryst, 10, 177, 1957.

22. G.A. Sim, Acta Cryst., 10, 536, 1957.
23. J.M. Robertson, J. Chem. Soc., 615, 1935.  
1195, 1936.
24. C. Bokhoven, J.C. Schoone, J.M. Bijvoet, Acta Cryst., 4, 275,  
1951.
25. D. Harker, Acta Cryst., 2, 1, 1956.
26. K.K. Cheung, G.A. Sim, Nature, 201, 1185, 1964.
27. E.W. Hughes, J. Amer. Chem. Soc., 63, 1737, 1941.

CHAPTER 2

THE CRYSTAL STRUCTURE ANALYSIS OF

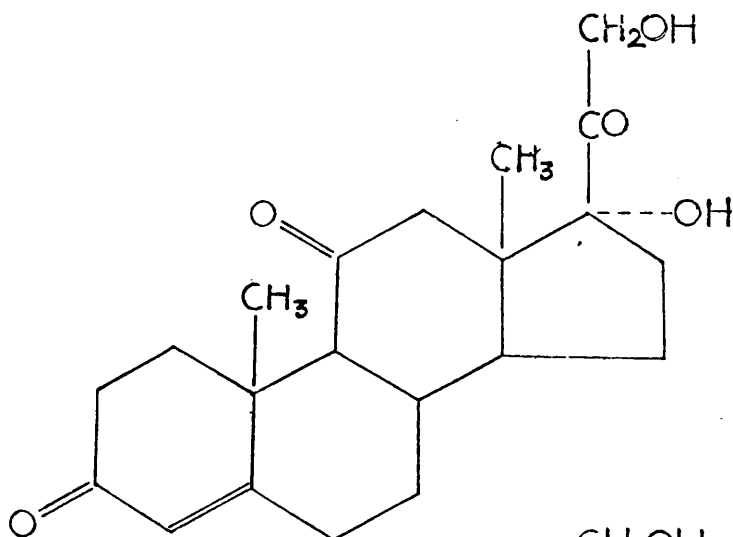
9- $\alpha$ -BROMOCORTISOL ACETATE

## 2.1 INTRODUCTION

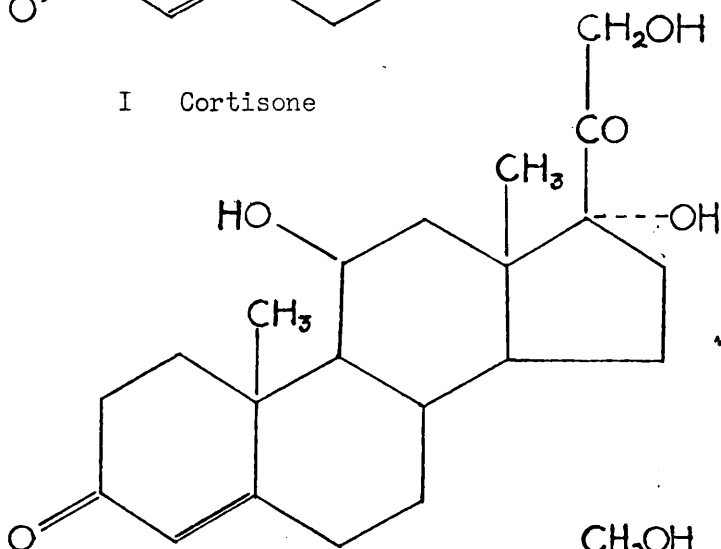


Human adrenal glands are small organs located near the upper end of each kidney. Experiments with animals showed the importance of these glands, since removal of them resulted in death. Further experiments demonstrated that extracts from the cortex of the adrenal glands contain one or more hormones but they did not disclose physiological actions of types later found to be of importance in defining clinical use for adrenocortical hormones. Great effort was put into finding the constituents of the adrenal glands and by 1943 twenty-eight steroids had been isolated from adrenal cortical extracts. Among them were six ketones having one or another type of activity specific to the adrenal function. The ketones, structures I - VI, were found to have electrolyte metabolic activity, i.e. they aid increased retention of  $\text{Na}^+$ ,  $\text{Cl}^-$  and water and excretion of  $\text{K}^+$ . This is normally objectionable, but in the case of Addison's disease which is attended with a high  $\text{K}^+$  and low  $\text{Na}^+$  serum content, this property was beneficial and the steroids, in particular cortexone (VI), were used in this connection.

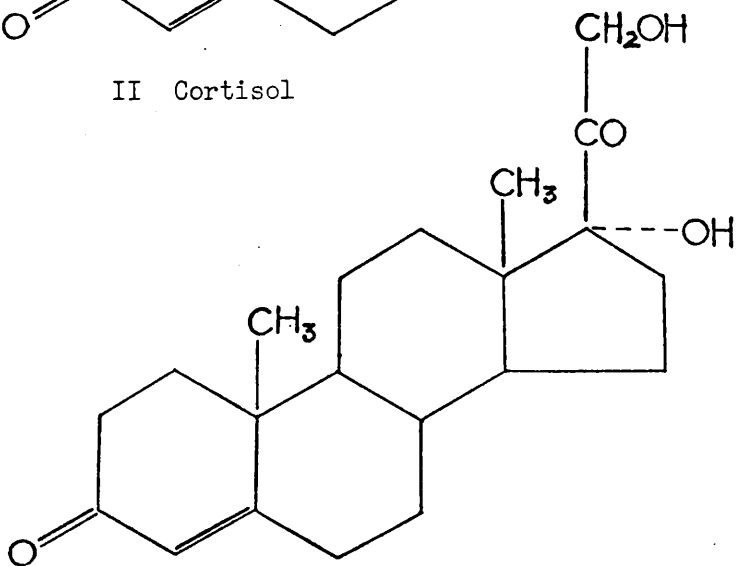
Interest in the adrenal steroids developed greatly when Hench, Kendall and their group at the Mayo clinic found that cortisone (I) would relieve rheumatoid arthritis. To a certain extent early expectations of the steroid as a therapeutic have been disappointing, partly due to the objectionable side effects caused by it. However, the steroids are still of very great importance and in the last decade a great amount of research has been carried out in finding ways of synthesising cortisone and its analogues. In attempting to find



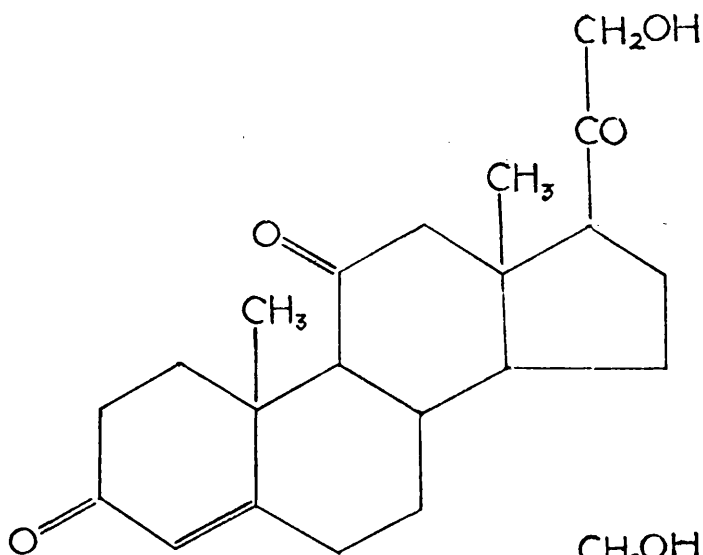
I Cortisone



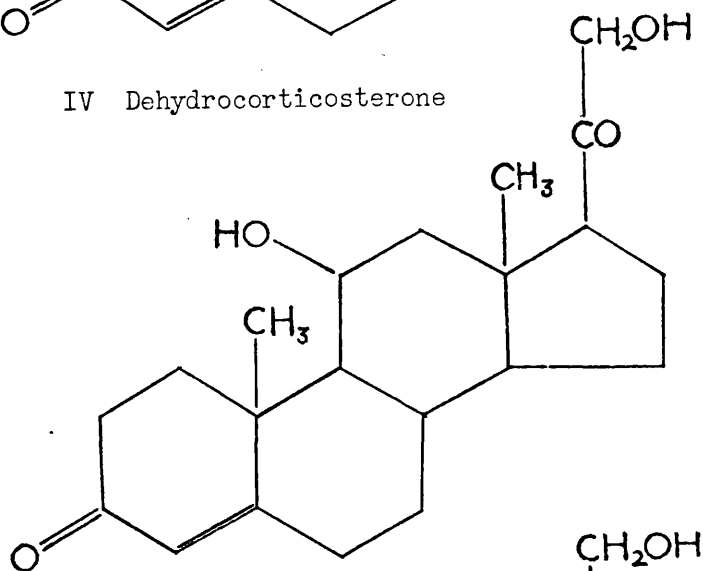
II Cortisol



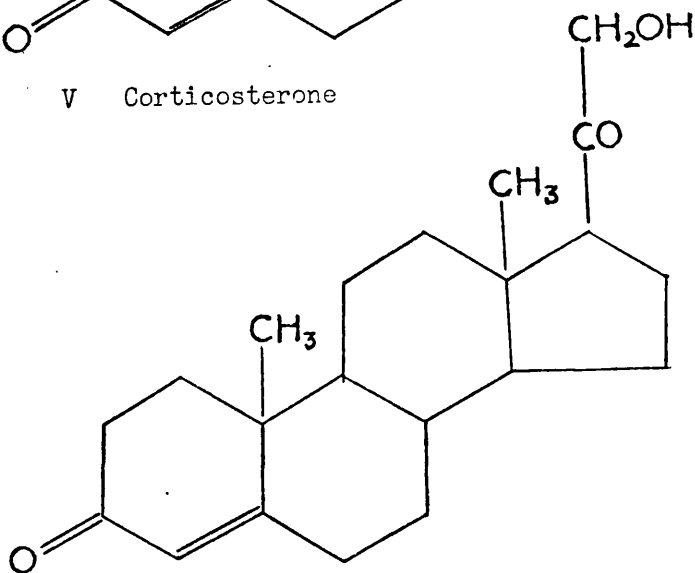
III Cortexolone



IV Dehydrocorticosterone



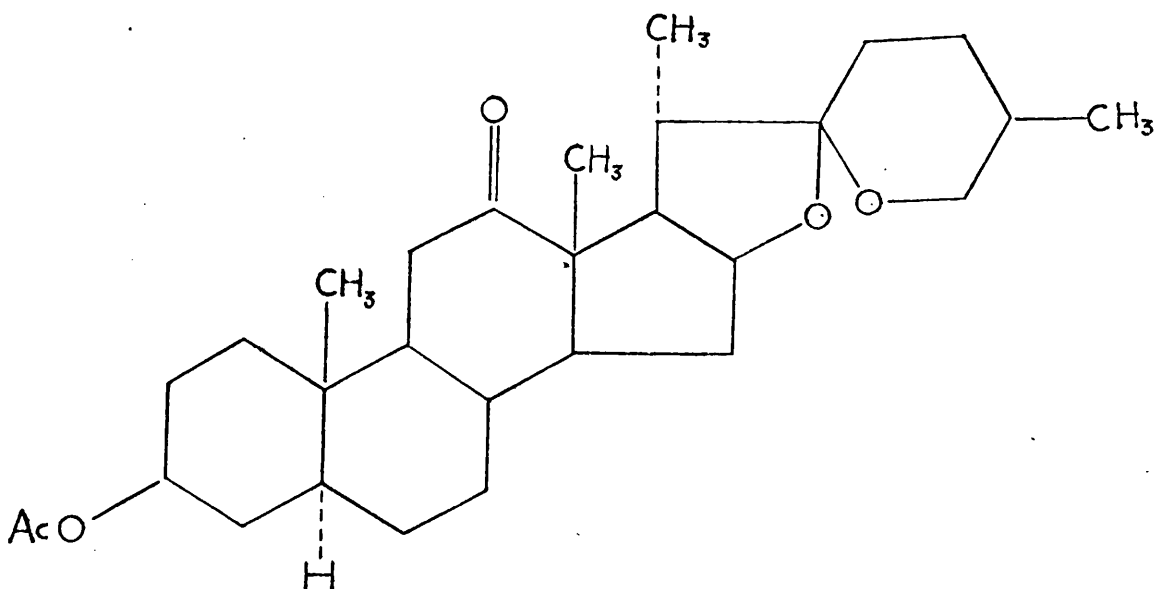
V Corticosterone



VI Cortexone

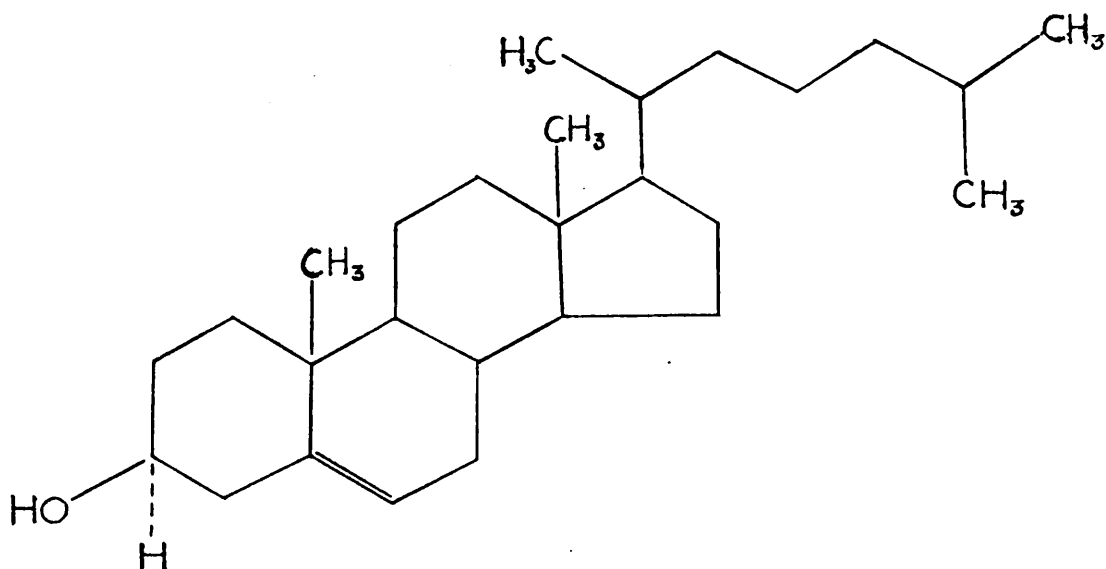
safer and more effective analogues almost every position in the cortisone (I) and cortisol (II) molecules has been modified. The use of cortisol was itself an improvement on cortisone since, unlike the latter, it can be injected directly into, for example, an inflamed rheumatic joint. The therapeutic uses of cortisone and its analogues are not restricted to arthritis but include applications in dermatology, asthma and in the emergency treatment of collapse during surgery.

The elucidation of the structures of the corticoids was carried out mainly by two groups viz. the Reichstein group and the Kendal group. Classical degradation techniques were used in conjunction with partial synthesis. The subsequent research on synthesis has been immense and a large number of starting materials and routes have been found although many were neither practicable nor economic. The only synthesis eventually used on a commercial scale is that utilising hecogenin (VII) as starting material. This compound was isolated by Marker and co-workers (1) and was recognised as a potential starting material for the synthesis of cortisone. In addition, an abundant source was found in the sisal plant which is widely cultivated in East Africa.



VII Hecogenin

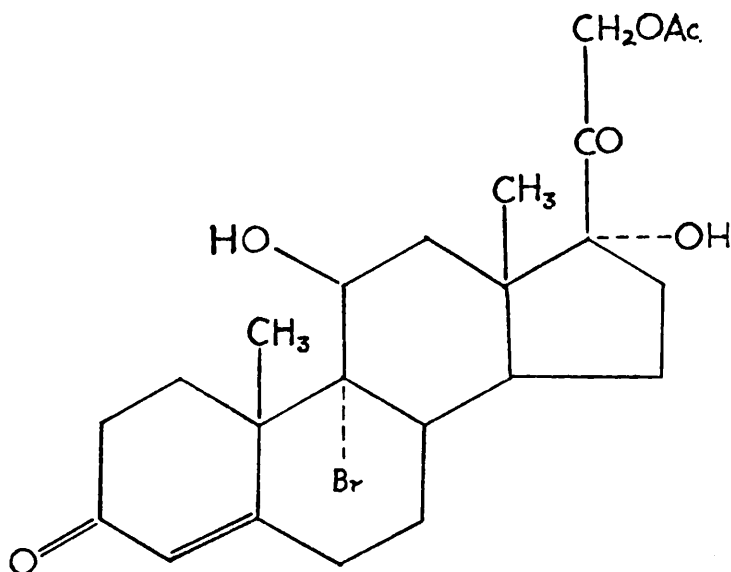
The first total synthesis of cortisone was reported by Woodward and co-workers in 1951 (2) and described in detail the following year (3). The biosynthesis of the corticoids is expected to go through intermediate cholesterol (VIII). The utilisation of cholesterol in man for the



VIII Cholesterol

synthesis of the corticoids has been demonstrated by the administration of labelled cholesterol and the isolation of labelled metabolites of cortisol. The isotope distribution pattern of the cortical steroids biosynthesised from acetate-1-C<sup>14</sup> is in agreement with the postulate that cholesterol is an intermediate although it is not known whether cholesterol is an obligatory intermediate.

Among the more effective variants of cortisone and cortisol are the 9- $\alpha$ -halo compounds prepared originally by Fried and Sabo (4). Included among them is 9- $\alpha$ -Bromocortisol, the 21-acetate of which has been used in this crystal structure analysis and is shown by structure IX. The acetate is biologically equivalent to the parent substance and is generally preferred since the free ketols are subject to oxidation.



IX 9- $\alpha$ -Bromocortisol Acetate

The purpose of this analysis was to gain information on crystal packing, molecular dimensions and overall shape of the molecule. It was expected that the  $\beta$  groups would tend to buckle the molecule but that the large bromine atom in the  $\alpha$  position might compensate for this buckling effect. All these considerations may be of use in explaining the physiological behaviour of the compound and the only method of attacking these problems is by a three-dimensional x-ray analysis.

## 2.2 EXPERIMENTAL DATA



CRYSTAL DATA

9- $\alpha$ -Bromocortisol Acetate -  $C_{23}^{O}H_{31}Br$ .

Molecular Weight = 483.4

Assuming 4 molecules per cell the calculated density = 1.248 g./cc.

Including 4 acetone molecules per cell the calculated density

= 1.285 g/c.c.

The measured density (by flotation method)

= 1.287 g./c.c.

The crystal is orthorhombic with

$$a = 13.92 \pm .04 \text{ \AA}$$

$$b = 24.00 \pm .05 \text{ \AA}$$

$$c = 7.70 \pm .03 \text{ \AA}$$

Volume of the unit cell =  $2572 \text{ \AA}^3$

The number of molecules per cell = 4

Absent spectra hoo when h is odd,

oko when k is odd,

and ool when l is odd.

The space group is  $P 2_1 2_1 2_1$

Total number of electrons per cell =  $F(000) = 1008$

$$\sum f^2 \text{ (light atoms)} = 1243 \text{ (sin } \theta = 0)$$

$$\sum f^2 \text{ (heavy atom)} = 1225 \text{ (sin } \theta = 0)$$

$\mu$  - the absorption coefficient for x-rays ( $\lambda = 1.5418$ ) =  $27.6 \text{ cm}^{-1}$

# INTENSITY DATA

Using copper  $K\alpha$  ( $\lambda = 1.5418 \text{ \AA}$ ) and molybdenum  $K\alpha$  ( $\lambda = 0.7107 \text{ \AA}$ ) radiation, rotation and precession photographs were taken. The unit cell dimensions were obtained from the precession photographs and the space group determined by consideration of the systematically absent spectra. It was found that the  $h00$ ,  $0k0$  and  $00l$  reflexions were absent when  $h, k$  and  $l$  were odd respectively. The space group was therefore determined uniquely as  $P 2_1 2_1 2_1$ .

A small crystal, bathed in a uniform x-ray beam of copper  $K\alpha$  radiation was used for the intensity measurements. No absorption corrections were made. Using a Weissenberg equi-inclination camera the  $hk0$  to  $hk6$  spectra were collected photographically. The multiple film technique (5) was used to achieve correlation of strong and weak reflexions. The intensities were estimated visually by comparison with a calibrated step-wedge and corrected for Lorentz, polarisation and rotation factors (6). Some difficulty was found in estimation of many of the intensities since certain of the reflexions were accompanied by a small satellite spot. This at times came adjacent to or coalesced with the reflexion being estimated. These satellite spots were found to be due to the fact that the crystal used had a few very small crystal fragments attached to it and when an intensity was in doubt an equivalent reflexion from another quadrant was estimated.

The values of the structure amplitudes,  $|F_o|$ , were computed by using the mosaic crystal formula. The various layers,  $hk0 - hk6$ , were not scaled relative to one another, the absolute scale being found later by comparison with the calculated structure factors. In all 2,187 independent structure amplitudes were measured.

## 2.3 THE STRUCTURE ANALYSIS

SOLUTION OF THE PATTERSON MAP FOR THE  
HEAVY ATOM POSITION

For a crystal belonging to the orthorhombic system, the expression for the Patterson function  $P(U, V, W)$  is

$$P(U, V, W) = \frac{8}{v} \sum \sum_0^{\infty} \sum |F(hkl)|^2 \cos 2\pi hu \cos 2\pi kv \cos 2\pi lw.$$

which can be reduced to a two dimensional expression for projections.

The equivalent positions of this space group,  $P 2_1 2_1 2_1$ , are  $x, y, z$  ;  $\frac{1}{2} - x, \bar{y}, \frac{1}{2} + z$  ;  $\frac{1}{2} + x, \frac{1}{2} - y, \bar{z}$  ;  $\bar{x}, \frac{1}{2} + y, \frac{1}{2} - z$  ; and the vectors from symmetry related atoms are  $2x - \frac{1}{2}, 2y, -\frac{1}{2}$  ;  $2x, \frac{1}{2}, 2z - \frac{1}{2}$  ;  $\frac{1}{2}, 2y - \frac{1}{2}, 2z$ .

The two dimensional Patterson projection,  $P(UV)$ , was computed utilising 304 coefficients and the resulting map is shown in figure 2.1. The space group of the Patterson function in the orthorhombic system is  $Pmmm$  and the mirror planes at  $U, V$  and  $W = \frac{1}{2}$  obviate calculation of the complete unit cell. In figure 2.1 and in the quadrant computed, three vectors are expected between bromine atoms. These are the vectors  $2x - \frac{1}{2}, 2y$  ;  $2x, \frac{1}{2}$  ; and  $\frac{1}{2}, 2y - \frac{1}{2}$  ; the latter two being double weight peaks and the first named a single weight peak in a general position. As shown in figure 2.1 these vectors are clearly distinguishable and are labelled A, B and C respectively. The coordinates of the bromine atom in the  $x$  and  $y$  directions were thus calculated.

In order to ascertain the  $z$  coordinate of the bromine atom a Harker section was calculated at  $U = \frac{1}{2}$ . Again only a quarter of the section at  $U = \frac{1}{2}$  need be calculated and this is shown in figure 2.2. In this case all the 2,187 terms were included. Only one

bromine-bromine vector per quadrant is expected in a general position and should occur at  $\frac{1}{2}$ ,  $2y - \frac{1}{2}$ ,  $2z$ . This vector is labelled D in figure 2.2 and from it the fractional coordinates of the bromine atom were calculated to be

$$x = 0.36979$$

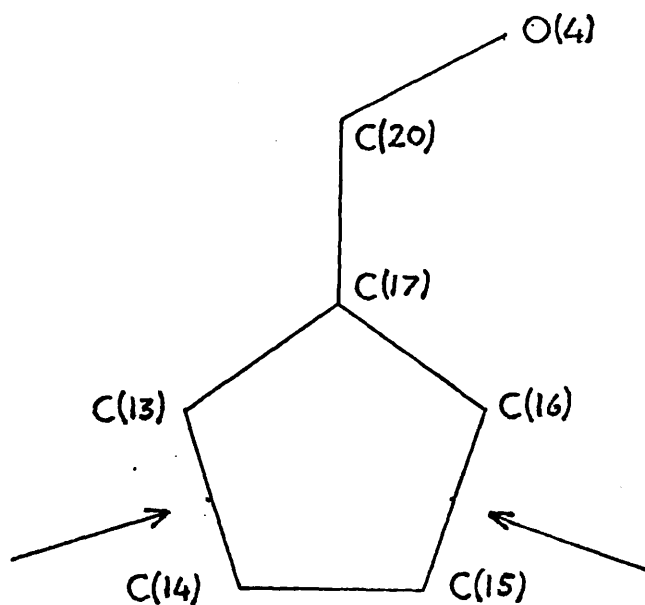
$$y = 0.42116$$

$$z = 0.20170$$

## SOLUTION OF THE STRUCTURE

The value of  $r$  ( $= [\sum f^2(h) / \sum f^2(1)]^{\frac{1}{2}}$ ) was calculated to be 0.99. In the non-centrosymmetric case this corresponds to about 38% of the phase angles being correct to within  $\pm 20^\circ$  (7). The first round of structure factor calculations based on the bromine atom and including an isotropic temperature factor,  $U$ , of 0.052 for the bromine atom, produced an overall  $R$  factor of 38.3%. An electron density map was computed and as expected contained a large number of obviously spurious peaks. Sections through the  $z$  axis were computed at intervals of  $z$  corresponding to approximately  $0.3\text{\AA}$ . The electron density contours were transferred to glass sheets which were stacked one on top of the other affording a three dimensional picture of the unit cell. Apart from the bromine atom only three other peaks could immediately be assigned to atoms with any confidence. These corresponded to C(17), C(20) and O(4) although this was not appreciated at that time. A starting point was therefore established and further study of the electron density map was undertaken in order to elicit as much information as possible. Various groups of peaks were selected and models of them built up and compared with Dreiding models of the known structure. This process was tried a large number of times without success and satisfactory positions for no other atoms could be deduced. At this point a group of peaks, which suggested a distorted but nevertheless credible five membered ring, were located in the vicinity of the three already assigned to atoms. This group included one of the atoms, C(17), already established. These

additional atoms were C(13), C(14), C(15) and C(16) and including C(17) they formed a five-membered ring. Structure X shows the atoms now located, other than the bromine atom.

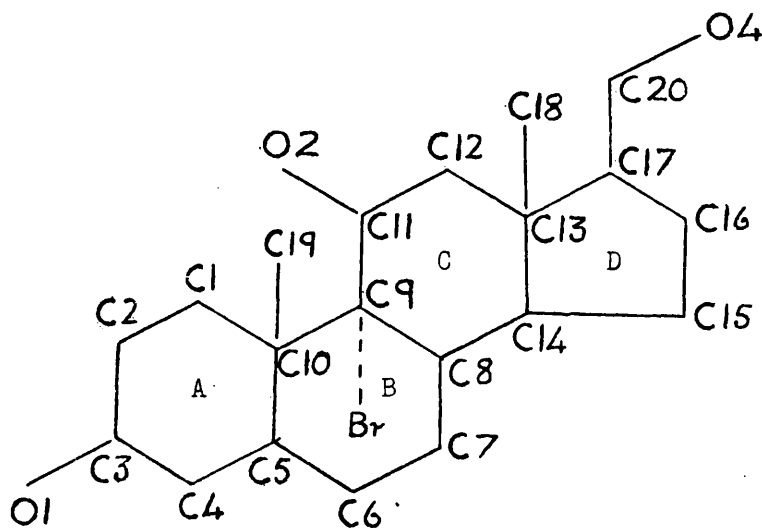


X

This afforded an excellent stepping stone for building up further parts of the structure since the six-membered ring, C, had to be attached to the five membered ring, D, in one of two directions, shown by the arrows in structure X, and furthermore this direction was fixed by the position of the bromine atom in the cell. This ring, C, was located and thence positions for the atoms of rings B and A were also established although the complexity of the Fourier

map made even this seemingly simple operation a hazardous task, and as was found later not all the atoms were unambiguously placed.

Other atoms bonded to the rings were found: the total number of atoms established at this stage was twenty four, being C(1) - C(20), O(1), O(2), O(4) and Br. Structure XI shows the relation of these atoms.



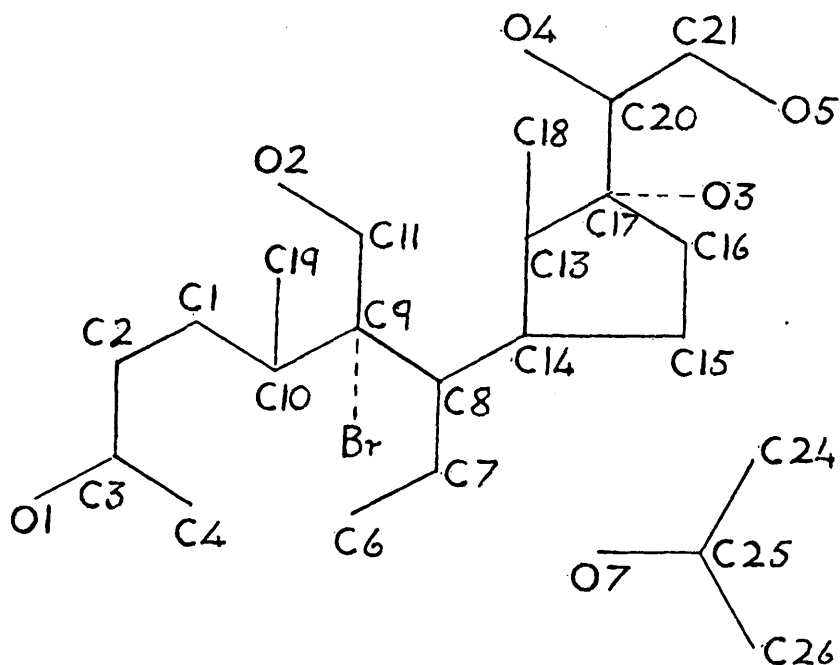
XI



A round of structure factors was calculated based on the twenty four atoms and including an overall isotropic temperature factor,  $U$ , of 0.052. This resulted in a drop in  $R$  of 8%, the value now being 30.5%. An electron density map was computed and drawn up on glass as before. The main interest of this map was to ascertain the positions of the remaining atoms of the molecule, notably the acetate side chain. A number of peaks remained, notably a group of small peaks in the vicinity of ring D, but their size and geometry were considered unfavourable to their being the acetate chain. Nevertheless three pieces of information were gained from the second electron density distribution. Firstly the improved resolution furnished more accurate coordinates for the atoms already established. Also the positions of two of the atoms, C(5) and C(12), located in the first electron density distribution were found to be in doubt. The shape of the peak corresponding to C(5) was very flat with high values of electron density falling off very rapidly to negative values in the  $z$  direction. Since this peak was directly above the bromine atom position and the peak shape was characteristic of a diffraction ripple, it was decided to leave this atom out of the next structure factor calculations. The second peak in doubt, that corresponding to C(12), had become split into two small peaks. The original peak position did not fit very well with the rest of the six-membered ring and although the satellite peak was in a more promising position, this atom was also excluded from the third round of structure factor calculations. Four additional well defined peaks were discovered in the asymmetric unit.

It was thought originally that they corresponded to the acetate side chain but their substantial distance from ring D showed them to be unrelated to the steroid molecule. The resolution and the fact that their geometry was virtually identical to an acetone or acetic acid molecule determined that the peaks were not false, and it was concluded that this was a molecule of crystallisation. This agreed well with the fact that the measured density of the crystal had suggested that an additional molecular weight of around 60 was present. Since the crystal was recrystallised from acetone, this was assumed to be the molecule of crystallisation and this was later verified by consideration of peak heights and bond lengths. The third round of structure factor calculations therefore included twenty two atoms of the steroid molecule plus the four solvent atoms, producing an R factor of 28.5%.

The third electron density distribution was calculated in the region around ring D in order to try and find the position of the side chain. Three well defined peaks in chemically sensible positions were taken as part of the chain and were included in the next structure factor calculations. Before this was done bond angles and lengths were calculated to check whether there were any other suspect atomic positions which might for the better have been excluded. Within the distortion expected at this stage the angles and lengths appeared reasonable and therefore all the atomic coordinates were included, giving an R factor of 25.6%. Structure XII shows the atoms included.



## XII

From the fourth electron density distribution, C(5) and C(12) were located in positions which satisfied the geometry of the system. The remaining atoms of the side chain were still elusive. Tiny peaks were found adjacent to the already located atoms of the chain. These looked far from promising but using their positions for calculation of bond lengths and angles produced results which were very plausible and they were included in the fifth structure factor calculations.

This calculation included all the atoms of the structure plus the atoms of the molecule of crystallisation and furnished an R factor of 21.4%.

## REFINEMENT OF THE STRUCTURE

Refinement of the structure was carried out by the method of least squares. The first cycle furnished coordinates which produced an R factor of 19.5%. A constant temperature factor, U(iso) of 0.0525 had been employed which after the first structure factor least squares calculation ranged from 0.0775 to 0.0443. A further three cycles of block diagonal isotropic least squares refinement reduced the discrepancy to 16.5%. At this stage the  $|F_o|$  values were scaled such that  $\sum_k |F_o| = \sum |F_c|$  for each layer and it was decided to continue refinement with anisotropic temperature factors for the heavy atom and the atoms of the side chain. Two further cycles of this mixed isotropic, anisotropic refinement reduced the discrepancy to 13.2%. Since the parameter shifts, the shift in R and the shift in  $\sum w \Delta^2$  had become small the refinement was concluded. The course of the analysis is shown in table 2.1.

Throughout the structure factor least squares refinement the weighting scheme used was

$$\sqrt{w} = 1 \quad \text{if} \quad |F_o| \leq p_1 ; \quad \sqrt{w} = p_1 / |F_o| \quad \text{if} \quad |F_o| > p_1$$

where  $p_1$  was taken as approximately the average structure amplitude.

## 2.4 TABLES AND FIGURES

TABLE 2.1

		<u>Atoms included</u>	<u>R%</u>
Structure factors	1	1	38.3
	2	24	30.5
	3	26	28.5
	4	29	25.6
	5	34	21.4
		<u><math>\sum w \Delta^2</math></u>	<u>R%</u>
Isotropic least squares cycle	1	20900	19.5
	2	15473	17.3
	3	14642	16.9
	4	14102	16.5
Partially anisotropic least square cycle	5	13757	13.5
	6	13633	13.2

Weighting scheme

$$\sqrt{w} = 1 \quad \text{if} \quad |F_o| \leq p$$

$$\sqrt{w} = p / |F_o| \quad \text{if} \quad |F_o| > p$$

where  $p$  = average structure amplitude

TABLE 2.2

### Measured and Calculated Values of the Structure Factor

[illegible]



TABLE 2.2 (Contd.)

H	K	L	Po	Pe	H	K	L	Po	Pe	H	K	L	Po	Pe	H	K	L	Po	Pe	H	K	L	Po	Pe	H	K	L	Po	Pe
4	7	2	40.3	39.7	9	13	2	22.7	26.1	0	23	3	22.0	19.3	5	4	3	53.5	56.0	10	2	3	25.5	22.1	0	5	4	16.5	16.0
4	7	2	45.6	45.5	9	14	2	18.4	18.4	0	24	3	13.0	11.4	5	5	3	44.0	43.2	10	3	3	25.4	17.1	0	6	5	16.5	16.0
4	9	2	26.0	25.4	9	15	2	16.1	19.4	0	25	3	10.7	7.3	5	6	3	34.6	37.2	10	4	3	13.7	14.1	0	7	6	16.5	16.0
4	10	2	38.6	38.4	9	16	2	14.4	17.6	0	26	3	8.6	5.7	5	7	3	27.7	30.4	10	5	3	10.9	11.5	0	8	7	16.5	16.0
4	11	2	31.9	31.7	9	17	2	12.1	15.3	0	27	3	6.5	4.2	5	8	3	20.1	22.8	10	6	3	8.7	9.3	0	9	8	16.5	16.0
4	12	2	47.4	45.4	9	18	2	6.9	9.9	0	28	3	4.4	2.7	5	9	3	13.6	15.3	10	7	3	7.6	8.2	0	10	9	16.5	16.0
4	13	2	35.2	35.2	9	19	2	13.4	14.0	0	29	3	3.3	1.9	5	10	3	10.4	11.6	10	8	3	6.5	7.1	0	11	10	16.5	16.0
4	14	2	28.1	28.1	9	20	2	6.4	5.7	0	30	3	2.2	1.2	5	11	3	8.3	9.3	10	9	3	5.4	6.0	0	12	11	16.5	16.0
4	15	2	14.5	10.5	10	1	2	10.6	26.2	0	31	3	1.1	0.6	5	12	3	6.8	7.8	10	10	3	4.3	4.9	0	13	12	16.5	16.0
4	16	2	16.4	20.6	10	2	2	30.1	26.9	0	32	3	0.5	0.3	5	13	3	5.7	6.7	10	11	3	3.2	3.8	0	14	13	16.5	16.0
4	17	2	19.5	22.1	10	3	2	39.0	36.7	0	33	3	0.4	0.2	5	14	3	4.6	5.6	10	12	3	2.1	2.7	0	15	14	16.5	16.0
4	18	2	26.6	34.4	10	4	2	43.0	39.7	0	34	3	0.3	0.1	5	15	3	3.5	4.5	10	13	3	1.0	1.6	0	16	15	16.5	16.0
4	19	2	12.4	14.8	10	5	3	13.7	14.0	0	35	3	0.2	0.1	5	16	3	2.4	3.4	10	14	3	0.9	1.0	0	17	16	16.5	16.0
4	20	2	14.4	16.1	10	6	3	31.0	28.1	0	36	3	0.1	0.0	5	17	3	1.3	2.3	10	15	3	0.8	0.9	0	18	17	16.5	16.0
4	21	2	18.4	15.3	10	7	3	25.5	25.7	0	37	3	0.1	0.0	5	18	3	1.2	2.2	10	16	3	0.7	0.8	0	19	18	16.5	16.0
4	22	2	9.3	11.0	10	8	3	25.4	21.5	0	38	3	0.1	0.0	5	19	3	1.1	2.1	10	17	3	0.6	0.7	0	20	19	16.5	16.0
4	23	2	19.3	8.3	10	9	3	25.7	25.7	0	39	3	0.1	0.0	5	20	3	1.0	2.0	10	18	3	0.5	0.6	0	21	20	16.5	16.0
4	24	2	28.6	20.4	10	10	3	32.3	34.9	0	40	3	0.1	0.0	5	21	3	0.9	1.9	10	19	3	0.4	0.5	0	22	21	16.5	16.0
4	25	2	73.5	69.1	10	11	3	14.4	13.2	0	41	3	0.1	0.0	5	22	3	0.8	1.8	10	20	3	0.3	0.4	0	23	22	16.5	16.0
4	26	2	42.9	40.9	10	12	3	14.4	15.3	0	42	3	0.1	0.0	5	23	3	0.7	1.7	10	21	3	0.2	0.3	0	24	23	16.5	16.0
4	27	2	53.2	54.2	10	13	3	20.1	19.1	0	43	3	0.1	0.0	5	24	3	0.6	1.6	10	22	3	0.1	0.2	0	25	24	16.5	16.0
4	28	2	47.6	45.3	10	14	3	18.4	14.4	0	44	3	0.1	0.0	5	25	3	0.5	1.5	10	23	3	0.1	0.1	0	26	25	16.5	16.0
4	29	2	53.2	50.3	10	15	3	6.3	5.2	0	45	3	0.1	0.0	5	26	3	0.4	1.4	10	24	3	0.1	0.1	0	27	26	16.5	16.0
4	30	2	60.1	56.1	10	16	3	6.6	7.0	0	46	3	0.1	0.0	5	27	3	0.3	1.3	10	25	3	0.1	0.1	0	28	27	16.5	16.0
4	31	2	39.6	37.7	10	17	3	10.3	10.3	0	47	3	0.1	0.0	5	28	3	0.2	1.2	10	26	3	0.1	0.1	0	29	28	16.5	16.0
4	32	2	16.3	1.5	10	18	3	6.4	12.2	0	48	3	0.1	0.0	5	29	3	0.1	1.1	10	27	3	0.1	0.1	0	30	29	16.5	16.0
4	33	2	24.0	21.0	11	1	2	14.0	11.5	0	49	3	0.1	0.0	5	30	3	0.1	1.0	10	28	3	0.1	0.1	0	31	30	16.5	16.0
4	34	2	45.5	42.5	11	2	2	9.9	7.7	0	50	3	0.1	0.0	5	31	3	0.1	0.9	10	29	3	0.1	0.1	0	32	31	16.5	16.0
4	35	2	26.0	23.0	11	3	2	13.3	11.5	0	51	3	0.1	0.0	5	32	3	0.1	0.8	10	30	3	0.1	0.1	0	33	32	16.5	16.0
4	36	2	31.2	27.7	11	4	2	24.4	25.2	0	52	3	0.1	0.0	5	33	3	0.1	0.7	10	31	3	0.1	0.1	0	34	33	16.5	16.0
4	37	2	36.2	39.3	11	5	2	26.5	24.2	0	53	3	0.1	0.0	5	34	3	0.1	0.6	10	32	3	0.1	0.1	0	35	34	16.5	16.0
4	38	2	19.4	18.5	11	6	2	20.1	13.4	0	54	3	0.1	0.0	5	35	3	0.1	0.5	10	33	3	0.1	0.1	0	36	35	16.5	16.0
4	39	2	24.5	3.0	11	7	2	10.1	5.4	0	55	3	0.1	0.0	5	36	3	0.1	0.4	10	34	3	0.1	0.1	0	37	36	16.5	16.0
4	40	2	18.2	12.2	11	8	2	12.5	10.0	0	56	3	0.1	0.0	5	37	3	0.1	0.3	10	35	3	0.1	0.1	0	38	37	16.5	16.0
4	41	2	9.9	12.6	11	9	2	12.5	10.0	0	57	3	0.1	0.0	5	38	3	0.1	0.2	10	36	3	0.1	0.1	0	39	38	16.5	16.0
4	42	2	10.1	1.5	11	10	2	10.1	1.5	0	58	3	0.1	0.0	5	39	3	0.1	0.1	10	37	3	0.1	0.1	0	40	39	16.5	16.0
4	43	2	14.4	15.3	11	11	2	22.7	20.2	0	59	3	0.1	0.0	5	40	3	0.1	0.0	10	38	3	0.1	0.1	0	41	40	16.5	16.0
4	44	2	14.4	17.2	11	12	2	14.3	12.1	0	60	3	0.1	0.0	5	41	3	0.1	0.0	10	39	3	0.1	0.1	0	42	41	16.5	16.0
4	45	2	14.2	15.0	11	13	2	10.1	10.1	0	61	3	0.1	0.0	5	42	3	0.1	0.0	10	40	3	0.1	0.1	0	43	42	16.5	16.0
4	46	2	9.3	1.5	11	14	2	9.3	10.3	0	62	3	0.1	0.0	5	43	3	0.1	0.0	10	41	3	0.1	0.1	0	44	43	16.5	16.0
4	47	2	13.4	11.1	11	15	2	10.3	12.2	0	63	3	0.1	0.0	5	44	3	0.1	0.0	10	42	3	0.1	0.1	0	45	44	16.5	16.0
4	48	2	12.7	15.0	11	16	2	6.4	7.6	0	64	3	0.1	0.0	5	45	3	0.1	0.0	10	43	3	0.1	0.1	0	46	45	16.5	16.0
4	49	2	8.4	1.5	11	17	2	6.4	7.6	0	65	3	0.1	0.0	5	46	3	0.1	0.0	10	44	3	0.1	0.1	0	47	46	16.5	16.0
4	50	2	7.7	7.7	12	1	2	22.8	25.1	0	66	3	0.1	0.0	5	47	3	0.1	0.0	10	45	3	0.1	0.1	0	48	47	16.5	16.0
4	51	2	21.2	23.7	12	2	2	20.9	20.7	0	67	3	0.1	0.0	5	48	3	0.1	0.0	10	46	3	0.1	0.1	0	49	48	16.5	16.0
4	52	2	71.6	69.4	12	3	2	10.2	11.5	0	68	3	0.1	0.0	5	49	3	0.1	0.0	10	47	3	0.1	0.1	0	50	49	16.5	16.0
4	53	2	56.6	54.6	12	4	2	14.4	20.0	0	69	3	0.1	0.0	5	50	3	0.1	0.0	10	48	3	0.1	0.1	0	51	50	16.5	16.0
4	54	2	15.1	16.6	12	5	2	14.4	13.2	0	70	3	0.1	0.0	5	51	3	0.1	0.0	10	49	3	0.1	0.1	0	52	51	16.5	16.0
4	55	2	55.6	57.7	12	6	2	17.7	15.4	0	71	3	0.1	0.0	5	52	3	0.1	0.0	10	50	3	0.1	0.1	0	53	52	16.5	16.0
4	56	2	31.2	31.2	12	7	2	17.7	15.4	0	72	3	0.1	0.0	5	53	3	0.1	0.0	10	51	3	0.1	0.1	0	54	53	16.5	16.0
4	57	2	16.5	17.5	12	8	2	14.4	11.5	0	73	3	0.1	0.0	5	54	3	0.1	0.0	10	52	3	0.1	0.1	0	55	54	16.5	16.0
4	58	2	13.1	14.5	12	9	2	7.1	3.9	0	74	3	0.1	0.0	5	55	3	0.1	0.0	10	53	3	0.1	0.1	0	56	55	16.5	16.0
4	59	2	37.5	37.5	12	10	2	9.5	10.5	0	75	3	0.1	0.0	5	56	3	0.1	0.0	10	54	3	0.1	0.1	0	57	56	16.5	16.0
4	60	2	39.2	39.2	12	11	2	10.3	10.3	0	76	3	0.1	0.0	5	57	3	0.1	0.0	10	55	3	0.1	0.1	0	58	57	16.5	16.0
4	61	2	40.1	41.5	12	12	2	9.0	8.2	0	77	3	0.1	0.0	5	58													

TABLE 2.2 (Contd.)

H	K	L	Po	Pa	H	K	L	Po	Pa	H	K	L	Po	Pa	H	K	L	Po	Pa	H	K	L	Po	Pa	H	K	L	Po	Pa
4	25	4	15.2	14.4	10	6	4	13.7	14.5	1	25	5	9	12.5	7	11	5	8.0	6.0	0	4	6	6.4	0.5	5	10	6	2.8	6.1
4	26	4	8.6	4.6	10	7	4	8.7	10.2	2	0	5	56.9	57.2	7	12	5	11.4	10.9	0	5	6	20.6	0.8	5	11	6	12.6	11.1
4	27	4	52.2	52.9	10	8	4	20.5	19.2	2	1	5	15.5	15.1	7	13	5	19.7	24.6	0	6	6	32.4	31.0	5	12	6	16.8	15.3
4	28	4	31.2	27.6	10	9	4	6.1	7.5	2	2	5	35.2	35.1	7	14	5	11.2	3.9	0	7	6	19.7	19.4	5	13	6	14.2	13.0
4	29	4	37.1	28.3	10	10	4	21.3	21.1	2	3	5	11.3	11.3	7	15	5	11.9	11.9	0	8	6	7.1	1.4	5	14	6	15.2	14.1
4	30	4	25.9	25.8	10	11	4	12.2	13.1	2	4	5	25.0	20.0	7	16	5	15.5	12.2	0	9	6	20.0	19.9	5	15	6	6.1	9.1
4	31	4	31.8	31.9	10	12	4	6.0	10.2	2	5	5	26.9	29.4	7	17	5	10.6	8.1	0	10	6	31.2	31.0	5	16	6	5.4	12.7
4	32	4	31.3	31.1	10	13	4	5.6	5.6	2	6	5	61.5	54.0	7	18	5	11.9	9.2	0	11	6	11.9	5.3	5	17	6	9.8	19.1
4	33	4	13.7	10.7	10	14	4	10.0	11.7	2	7	5	19.7	19.9	7	19	5	11.9	14.3	0	12	6	11.0	14.3	5	18	6	12.7	11.9
4	34	4	23.3	17.9	10	15	4	10.0	8.5	2	8	5	28.7	22.6	7	20	5	11.9	5.9	0	13	6	14.6	15.1	5	19	6	7.0	6.0
4	35	4	24.0	27.4	10	16	4	9.9	8.7	2	9	5	10.0	17.2	7	21	5	12.2	18.1	0	14	6	5.7	5.4	5	20	6	4.5	4.0
4	36	4	21.4	18.2	10	17	4	7.9	8.7	2	10	5	13.1	7.2	7	22	5	20.2	15.9	0	15	6	10.6	8.7	5	21	6	12.3	11.5
4	37	4	20.7	24.0	10	18	4	6.4	8.7	2	11	5	4.4	42.3	8	1	5	33.6	27.2	0	16	6	23.2	19.4	5	22	6	23.7	23.0
4	38	4	29.9	29.1	10	19	4	8.0	5.4	2	12	5	17.7	14.9	8	2	5	17.7	6.5	0	17	6	11.5	13.1	5	23	6	14.3	15.9
4	39	4	16.5	19.5	10	20	4	23.0	25.6	2	13	5	24.8	20.4	8	3	5	34.1	34.1	0	18	6	26.0	24.7	5	24	6	27.0	22.9
4	40	4	12.0	7.2	10	21	4	8.7	6.3	2	14	5	20.2	15.5	8	4	5	19.3	22.6	0	19	6	33.3	27.5	5	25	6	6.5	9.1
4	41	4	23.0	24.6	10	22	4	12.3	15.9	2	15	5	12.7	15.8	8	5	5	15.9	24.4	0	20	6	17.6	15.6	5	26	6	14.9	13.8
4	42	4	16.5	13.0	10	23	4	17.4	19.3	2	16	5	15.2	17.2	8	6	5	16.5	23.1	0	21	6	22.5	15.3	5	27	6	4.3	7.4
4	43	4	16.7	17.8	10	24	4	8.6	6.5	2	17	5	25.0	16.2	8	7	5	21.3	21.9	0	22	6	14.6	12.5	5	28	6	16.7	19.9
4	44	4	13.9	13.3	10	25	4	17.1	17.2	2	18	5	10.3	11.2	8	8	5	11.4	13.1	0	23	6	12.7	12.8	5	29	6	19.1	17.3
4	45	4	17.6	6.5	10	26	4	5.0	5.3	2	19	5	12.5	13.6	8	9	5	13.6	13.6	0	24	6	12.2	5.7	5	30	6	6.4	5.1
4	46	4	12.3	11.1	10	27	4	11.4	8.3	2	20	5	15.0	16.2	8	10	5	15.3	15.2	0	25	6	9.9	9.8	5	31	6	14.2	13.0
4	47	4	14.6	11.9	10	28	4	10.0	12.6	2	21	5	47.3	41.5	8	11	5	16.6	7.4	0	26	6	24.2	24.2	5	32	6	12.6	9.3
4	48	4	14.1	11.9	10	29	4	7.6	9.6	2	22	5	21.0	16.9	8	12	5	17.2	14.4	0	27	6	21.6	20.5	5	33	6	12.3	11.7
4	49	4	6.5	2.0	10	30	4	11.0	11.8	2	23	5	10.2	16.3	8	13	5	18.2	6.3	0	28	6	10.1	9.7	5	34	6	9.1	9.1
4	50	4	6.6	0.7	10	31	4	4.9	6.4	2	24	5	44.6	42.6	8	14	5	9.1	9.7	0	29	6	15.5	10.3	5	35	6	13.3	14.9
4	51	4	17.8	15.0	10	32	4	11.9	4.8	2	25	5	36.2	31.2	8	15	5	12.6	11.1	0	30	6	16.9	15.9	5	36	6	14.0	4.9
4	52	4	25.9	30.2	10	33	4	3.8	5.6	2	26	5	33.5	35.0	8	16	5	43.9	41.6	0	31	6	12.3	15.7	5	37	6	6.2	5.2
4	53	4	15.6	13.1	10	34	4	16.7	13.2	2	27	5	20.4	22.4	8	17	5	21.1	23.2	0	32	6	4.2	5.7	5	38	6	4.6	3.0
4	54	4	8.1	1.0	10	35	4	4.4	5.5	2	28	5	17.6	19.4	8	18	5	19.5	17.3	0	33	6	10.5	12.1	5	39	6	10.5	7.0
4	55	4	10.1	8.6	10	36	4	16.7	13.2	2	29	5	19.7	20.9	8	19	5	15.0	19.2	0	34	6	9.1	9.8	5	40	6	10.4	9.7
4	56	4	37.8	40.8	10	37	4	8.6	10.8	2	30	5	16.7	18.7	8	20	5	16.7	3.0	0	35	6	6.0	11.9	5	41	6	10.6	13.0
4	57	4	17.3	9.3	10	38	4	10.2	10.5	2	31	5	20.9	23.0	8	21	5	17.9	19.8	0	36	6	3.4	5.0	5	42	6	10.6	13.2
4	58	4	10.9	21.6	10	39	4	15.8	17.2	2	32	5	16.0	13.5	8	22	5	15.6	4.7	0	37	6	7.1	5.2	5	43	6	22.0	15.9
4	59	4	28.8	32.9	10	40	4	11.9	13.2	2	33	5	17.0	20.6	8	23	5	15.6	5.2	0	38	6	35.9	35.9	5	44	6	16.6	13.1
4	60	4	26.3	14.8	10	41	4	8.2	9.3	2	34	5	16.6	16.6	8	24	5	17.2	6.0	0	39	6	12.6	12.6	5	45	6	12.7	12.9
4	61	4	8.7	6.6	10	42	4	8.2	9.3	2	35	5	16.0	13.5	8	25	5	15.6	4.7	0	40	6	7.1	5.2	5	46	6	11.0	11.0
4	62	4	31.0	27.9	10	43	4	12.1	14.4	2	36	5	19.2	25.6	8	26	5	14.8	15.9	0	41	6	46.7	46.7	5	47	6	25.5	16.7
4	63	4	8.7	6.6	10	44	4	12.5	11.9	2	37	5	20.7	7.5	8	27	5	16.6	9.1	0	42	6	26.0	26.0	5	48	6	12.7	12.9
4	64	4	8.4	8.9	10	45	4	12.5	8.5	2	38	5	16.6	16.6	8	28	5	16.6	9.1	0	43	6	7.6	2.9	5	49	6	9.6	6.2
4	65	4	12.8	11.2	10	46	4	4.3	5.1	2	39	5	35.0	9.2	8	29	5	24.2	22.9	0	44	6	8.1	6.6	5	50	6	11.7	13.5
4	66	4	17.7	22.3	10	47	4	4.6	4.6	2	40	5	31.3	31.8	8	30	5	35.6	31.5	0	45	6	35.6	31.5	5	51	6	12.0	14.9
4	67	4	10.6	9.2	10	48	4	6.8	4.6	2	41	5	35.7	25.6	8	31	5	24.9	23.8	0	46	6	6.0	6.2	5	52	6	9.3	9.6
4	68	4	6.1	3.4	10	49	4	12.9	4.5	2	42	5	13.5	5.7	8	32	5	14.0	11.4	0	47	6	10.6	7.4	5	53	6	3.6	2.6
4	69	4	40.0	39.0	10	50	4	14.1	17.4	2	43	5	54.6	45.3	8	33	5	12.2	12.2	0	48	6	13.1	13.1	5	54	6	3.2	2.6
4	70	4	17.6	13.1	10	51	4	11.5	11.3	2	44	5	45.5	44.2	8	34	5	11.0	10.1	0	49	6	12.7	9.9	5	55	6	20.5	16.3
4	71	4	19.1	19.1	10	52	4	11.6	11.6	2	45	5	22.0	22.7	8	35	5	11.2	10.1	0	50	6	16.2	14.7	5	56	6	4.5	7.1
4	72	4	24.3	27.3	10	53	4	7.1	7.1	2	46	5	12.6	11.3	8	36	5	10.8	9.2	0	51	6	17.9	17.9	5	57	6	11.0	11.0
4	73	4	22.6	25.3	10	54	4	7.5	6.2	2	47	5	12.6	11.3	8	37	5	10.8	11.6	0	52	6	11.4	11.4	5	58	6	6.4	2.1
4	74	4	22.6	25.3	10	55	4	5.1	5.8	2	48	5	27.2	24.3	8	38	5	10.8	11.6	0	53	6	5.2	6.2	5	59	6	13.5	12.6
4	75	4	23.0	26.5	10	56	4	4.9	6.4	2	49	5	7.2	7.2	8	39	5	10.6	9.1	0	54	6	5.6	2.8	5	60	6	16.5	16.9
4	76	4	15.9	12.1	10	57	4	4.4	4.9	2	50	5	25.0	23.9	8	40	5	10.6	9.1	0	55	6	5.6	2.8	5	61	6	11.1	11.1
4	77	4	26.7	24.5	10	58	4	7.1	8.8	2	51	5	11.1	10.2	8	41	5	9.9	3.8	0	56	6	2.3	2.7	5	62	6	1.5	12.2
4	78	4	3.4	8.6	10	59	4	13.6	4.7	2	52	5	12.0	12.0	8	42	5	9.9	3.8	0	57	6	13.9	12.2	5	63	6	2.9	2.4
4	79	4	17.1	15.0	10	60	4	7.4																					

TABLE 2.3

9-  $\alpha$ -Bromocortisol Acetate

Final Atomic Coordinates

<u>Atom</u>	<u>x/a</u>	<u>y/b</u>	<u>z/c</u>
C(1)	0.51520	0.38096	0.49803
C(2)	0.57359	0.41019	0.63647
C(3)	0.53166	0.46809	0.69315
C(4)	0.42785	0.47430	0.67123
C(5)	0.37438	0.43142	0.60547
C(6)	0.26265	0.43792	0.63740
C(7)	0.20465	0.41314	0.49366
C(8)	0.24186	0.35863	0.42114
C(9)	0.34887	0.35956	0.37949
C(10)	0.40702	0.37675	0.54073
C(11)	0.38625	0.30910	0.28269
C(12)	0.31815	0.28826	0.13087
C(13)	0.21363	0.28392	0.18941
C(14)	0.18346	0.34115	0.26099
C(15)	0.07493	0.33647	0.27677
C(16)	0.04780	0.29928	0.12778
C(17)	0.13725	0.27988	0.03704
C(18)	0.19877	0.23569	0.31315

TABLE 2.3 (contd.)

<u>Atom</u>	<u>x/a</u>	<u>y/b</u>	<u>z/c</u>
C(19)	0.39479	0.33526	0.69988
C(20)	0.13285	0.22089	-0.03580
C(21)	0.20063	0.20823	-0.18169
C(22)	0.23379	0.11377	-0.10745
C(23)	0.20881	0.05189	-0.12618
C(24)	0.48158	0.06696	0.18449
C(25)	0.45163	0.11597	0.30022
C(26)	0.37170	0.10591	0.43321
O(1)	0.58330	0.50387	0.74520
O(2)	0.40312	0.26506	0.40653
O(3)	0.16696	0.31480	-0.10394
O(4)	0.07803	0.18583	0.01304
O(5)	0.19607	0.14915	-0.22618
O(6)	0.28322	0.13213	0.01055
O(7)	0.48794	0.16021	0.28328
Br(1)	0.36998	0.42177	0.20133

TABLE 2.4

9- $\alpha$ -Bromocortisol Acetate

Standard Deviations of the Final Atomic Coordinates ( $\text{\AA}$ )

<u>Atom</u>	<u>x</u>	<u>y</u>	<u>z</u>
C(1)	0.017	0.016	0.019
C(2)	0.021	0.020	0.023
C(3)	0.023	0.022	0.027
C(4)	0.024	0.023	0.027
C(5)	0.024	0.021	0.027
C(6)	0.018	0.017	0.020
C(7)	0.020	0.020	0.023
C(8)	0.017	0.016	0.019
C(9)	0.013	0.014	0.017
C(10)	0.016	0.015	0.018
C(11)	0.015	0.015	0.019
C(12)	0.016	0.015	0.018
C(13)	0.014	0.014	0.018
C(14)	0.015	0.015	0.017
C(15)	0.017	0.017	0.020
C(16)	0.019	0.018	0.021
C(17)	0.020	0.019	0.021
C(18)	0.018	0.017	0.018

TABLE 2.4 (Contd.)

<u>Atom</u>	<u>x</u>	<u>y</u>	<u>z</u>
C(19)	0.017	0.017	0.017
C(20)	0.017	0.017	0.016
C(21)	0.023	0.025	0.021
C(22)	0.022	0.027	0.030
C(23)	0.035	0.031	0.051
C(24)	0.027	0.025	0.030
C(25)	0.018	0.018	0.023
C(26)	0.026	0.025	0.029
O(1)	0.014	0.017	0.018
O(2)	0.014	0.011	0.013
O(3)	0.012	0.012	0.011
O(4)	0.012	0.016	0.015
O(5)	0.015	0.015	0.015
O(6)	0.018	0.018	0.016
O(7)	0.012	0.012	0.013
Br(1)	0.002	0.002	0.002

TABLE 2.5

9- $\alpha$ -Bromocortisol Acetate  
Isotropic Temperature Factors

<u>Atom</u>	<u>U(iso) x 10<sup>4</sup></u>	<u>Atom</u>	<u>U(iso) x 10<sup>4</sup></u>
C(1)	529	C(12)	484
C(2)	675	C(13)	489
C(3)	878	C(14)	553
C(4)	744	C(15)	634
C(5)	600	C(16)	649
C(6)	627	C(17)	681
C(7)	626	C(24)	989
C(8)	460	C(25)	651
C(9)	398	C(26)	857
C(10)	429	O(7)	617
C(11)	478		

TABLE 2.6

9- $\alpha$ -Bromocortisol Acetate

Anisotropic Temperature Factors x 10<sup>4</sup>

<u>Atom</u>	<u>U(11)</u>	<u>U(22)</u>	<u>U(33)</u>	<u>2U(23)</u>	<u>2U(31)</u>	<u>2U(12)</u>
C(18)	617	755	476	402	-301	-651
C(19)	640	640	306	63	-220	-422
C(20)	673	804	349	250	-457	-380
C(21)	809	939	413	-217	46	-222
C(22)	658	1073	916	-418	104	162
C(23)	1084	873	391	-129	111	275
O(1)	734	1040	1072	-274	-503	-514
O(2)	797	512	603	157	-311	401
O(3)	637	881	266	686	-284	-282
O(4)	617	983	539	-92	163	-684
O(5)	858	903	625	-180	-210	-223
O(6)	996	1278	771	-588	-904	95
Br(1)	613	623	683	451	78	34



TABLE 2.7

9- $\alpha$ -Bromocortisol Acetate

Bonded Distances ( $\text{\AA}$ )

<u>Bond</u>		<u>Distance</u>	<u>Bond</u>		<u>Distance</u>
C(1)	C(2)	1.51	C(12)	C(13)	1.53
C(1)	C(10)	1.54	C(13)	C(18)	1.51
C(2)	C(3)	1.57	C(13)	C(14)	1.54
C(3)	C(4)	1.46	C(13)	C(17)	1.59
C(3)	O(1)	1.19	C(14)	C(15)	1.52
C(4)	C(5)	1.37	C(15)	C(16)	1.50
C(5)	C(6)	1.58	C(16)	C(17)	1.50
C(5)	C(10)	1.48	C(17)	O(3)	1.43
C(6)	C(7)	1.49	C(17)	C(20)	1.52
C(7)	C(8)	1.51	C(20)	O(4)	1.20
C(8)	C(9)	1.52	C(20)	C(21)	1.50
C(8)	C(14)	1.54	C(21)	O(5)	1.46
C(9)	C(10)	1.54	C(22)	O(5)	1.35
C(9)	Br(1)	2.05	C(22)	C(23)	1.53
C(9)	C(11)	1.51	C(22)	O(6)	1.22
C(10)	C(19)	1.59	C(24)	C(25)	1.53
C(11)	C(12)	1.59	C(25)	C(26)	1.53
C(11)	O(2)	1.44	C(25)	O(7)	1.18

TABLE 2.8

9- $\alpha$ -Bromocortisol AcetateInterbond Angles

C(2)	C(1)	C(10)	114°	C(9)	C(10)	C(19)	113°
C(1)	C(2)	C(3)	114	C(9)	C(11)	C(12)	114
C(2)	C(3)	C(4)	115	C(9)	C(11)	O(2)	108
C(2)	C(3)	O(1)	121	C(12)	C(11)	O(2)	111
C(4)	C(3)	O(1)	124	C(11)	C(12)	C(13)	112
C(3)	C(4)	C(5)	120	C(12)	C(13)	C(14)	108
C(4)	C(5)	C(10)	129	C(12)	C(13)	C(17)	115
C(4)	C(5)	C(6)	114	C(14)	C(13)	C(17)	98
C(6)	C(5)	C(10)	116	C(14)	C(13)	C(18)	115
C(5)	C(6)	C(7)	112	C(12)	C(13)	C(18)	112
C(6)	C(7)	C(8)	116	C(17)	C(13)	C(18)	109
C(7)	C(8)	C(9)	114	C(13)	C(14)	C(8)	113
C(7)	C(8)	C(14)	111	C(13)	C(14)	C(15)	104
C(9)	C(8)	C(14)	111	C(8)	C(14)	C(15)	119
C(8)	C(9)	C(10)	110	C(14)	C(15)	C(16)	104
C(8)	C(9)	C(11)	115	C(15)	C(16)	C(17)	109
C(10)	C(9)	C(11)	116	C(16)	C(17)	C(13)	101
C(11)	C(9)	Br(1)	102	C(16)	C(17)	C(20)	115
C(10)	C(9)	Br(1)	106	C(13)	C(17)	C(20)	111
C(8)	C(9)	Br(1)	107	C(16)	C(17)	O(3)	114
C(9)	C(10)	C(1)	111	C(13)	C(17)	O(3)	109

TABLE 2.8 (Contd.)

C(9)	C(10)	C(5)	110°	C(20)	C(17)	O(3)	106°
C(1)	C(10)	C(5)	108	C(17)	C(20)	C(21)	116
C(5)	C(10)	C(19)	105	C(17)	C(20)	O(4)	124
C(1)	C(10)	C(19)	108	C(21)	C(20)	O(4)	120
C(20)	C(21)	O(5)	110	C(24)	C(25)	C(26)	118
C(23)	C(22)	O(5)	117	C(24)	C(25)	O(7)	121
C(23)	C(22)	O(6)	123	C(26)	C(25)	O(7)	122
O(5)	C(22)	O(6)	120	C(21)	O(5)	C(22)	116

TABLE 2.9

9- $\alpha$ -Bromocortisol Acetate

Some Intramolecular Non-Bonded Distances ( $\overset{\circ}{\text{A}}$ )

C(1)	C(19)	2.54	O(2)	C(18)	3.02
C(5)	C(19)	2.44	O(2)	C(19)	2.82
C(9)	C(19)	2.61	C(18)	C(20)	2.86
C(12)	C(18)	2.51	C(1)	Br(1)	3.20
C(14)	C(18)	2.57	C(5)	Br(1)	3.13
C(17)	C(18)	2.53	C(7)	Br(1)	3.22
C(6)	C(19)	3.11	C(12)	Br(1)	3.33
C(8)	C(19)	3.07	C(14)	Br(1)	3.27
C(11)	C(19)	3.28	O(2)	C(13)	3.16
C(2)	C(19)	3.63	O(2)	C(8)	3.18
C(8)	C(18)	3.12	O(2)	C(10)	2.85
C(11)	C(18)	3.16			

TABLE 2.10

9- $\alpha$ -Bromocortisol Acetate

Some Intermolecular Distances  $< 4 \text{ \AA}$

C(6)	O(3) <sub>1</sub>	3.80	O(1)	C(23) <sub>3</sub>	3.67
C(19)	O(3) <sub>1</sub>	3.55	O(3)	C(25) <sub>3</sub>	3.75
C(19)	C(12) <sub>1</sub>	3.66	O(7)	O(3) <sub>3</sub>	2.91
C(26)	O(5) <sub>1</sub>	3.73	O(7)	C(16) <sub>3</sub>	3.41
C(6)	Br(1) <sub>2</sub>	3.87	O(7)	C(17) <sub>3</sub>	3.53
C(1)	O(5) <sub>3</sub>	3.35	O(7)	C(15) <sub>3</sub>	3.60
C(2)	O(4) <sub>3</sub>	3.55	C(3)	C(24) <sub>4</sub>	3.76
C(2)	O(5) <sub>3</sub>	3.86	C(4)	C(24) <sub>4</sub>	3.75
C(7)	O(7) <sub>3</sub>	3.89	O(1)	C(23) <sub>4</sub>	3.25
C(11)	O(4) <sub>3</sub>	3.51	O(1)	C(26) <sub>4</sub>	3.54
C(12)	O(4) <sub>3</sub>	3.83	O(1)	C(24) <sub>4</sub>	3.75
C(19)	O(4) <sub>3</sub>	3.41	O(1)	C(22) <sub>4</sub>	3.82
C(24)	O(3) <sub>3</sub>	3.89	O(7)	O(2) <sub>1</sub>	2.94
C(25)	C(15) <sub>3</sub>	3.85			
C(26)	C(15) <sub>3</sub>	3.86			

The subscripts refer to the following equivalent positions;

1.  $x, y, z$
2.  $\frac{1}{2} - x, -y, \frac{1}{2} + z$
3.  $\frac{1}{2} + x, \frac{1}{2} - y, -z$
4.  $-x, \frac{1}{2} + y, \frac{1}{2} - z$

TABLE 2.11

9- $\alpha$ -Bromocortisol Acetate

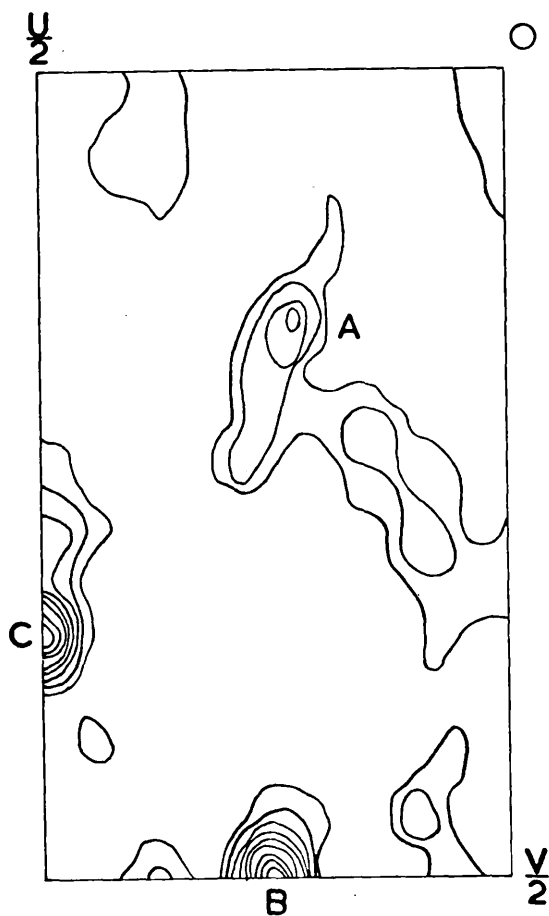
Deviations ( $\overset{\circ}{\text{\AA}}$ ) of the atoms from planes through the five-membered ring

- Planes -    a) C(13) - C(17)  
               b) excludes C(17)  
               c) excludes C(16)  
               d) excludes C(15)  
               e) excludes C(14)  
               f) excludes C(13)

	a	b	c	d	e	f
C(13)	-0.3127	-0.1198	-0.2832	-0.2672	-0.0872	(-0.7539)
C(14)	0.2736	0.1946	0.3010	0.1679	( 0.6506)	0.0137
C(15)	-0.1158	-0.1978	-0.1800	(-0.3173)	0.0971	-0.0217
C(16)	-0.0745	0.1230	(-0.1979)	-0.1771	-0.1553	0.0223
C(17)	0.2294	(0.5890)	0.1621	0.2764	0.1454	-0.0143

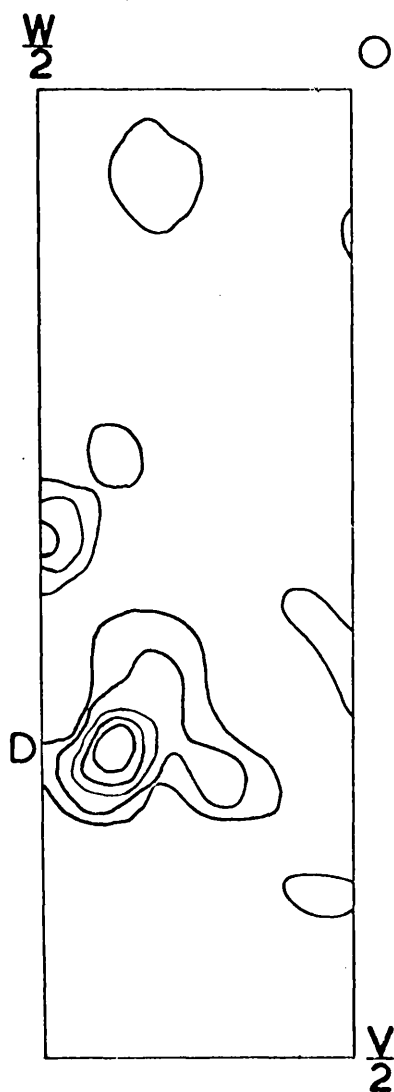
$$\chi^2 (f) = 4.0$$

$$p = 0.05$$



9- $\alpha$ -Bromocortisol Acetate

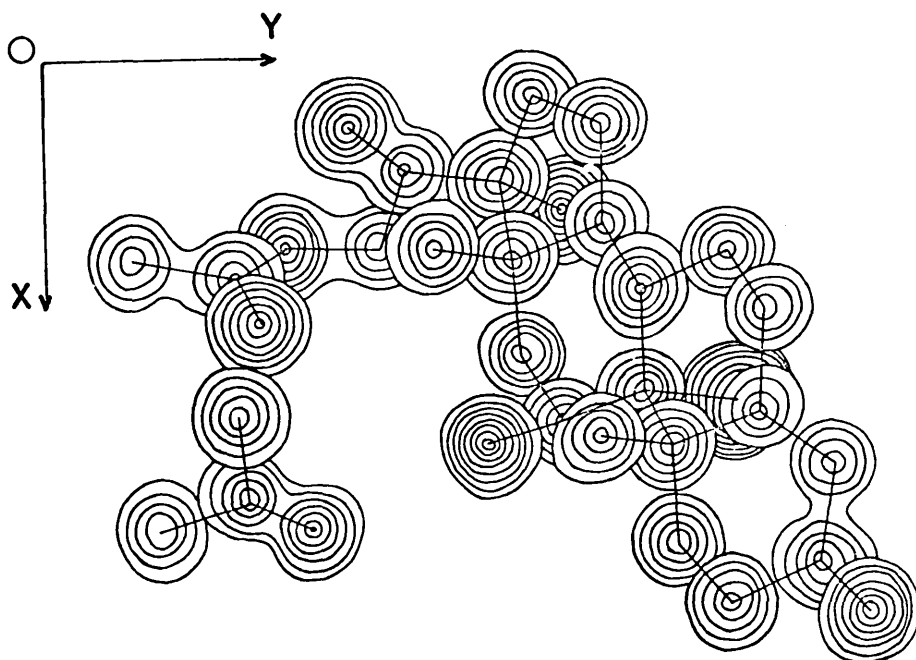
Figure 2.1 Patterson projection along the  $c$  axis. A, B and C denote the bromine-bromine vector peaks. The contour scale is arbitrary.



9- $\alpha$ -Bromocortisol Acetate

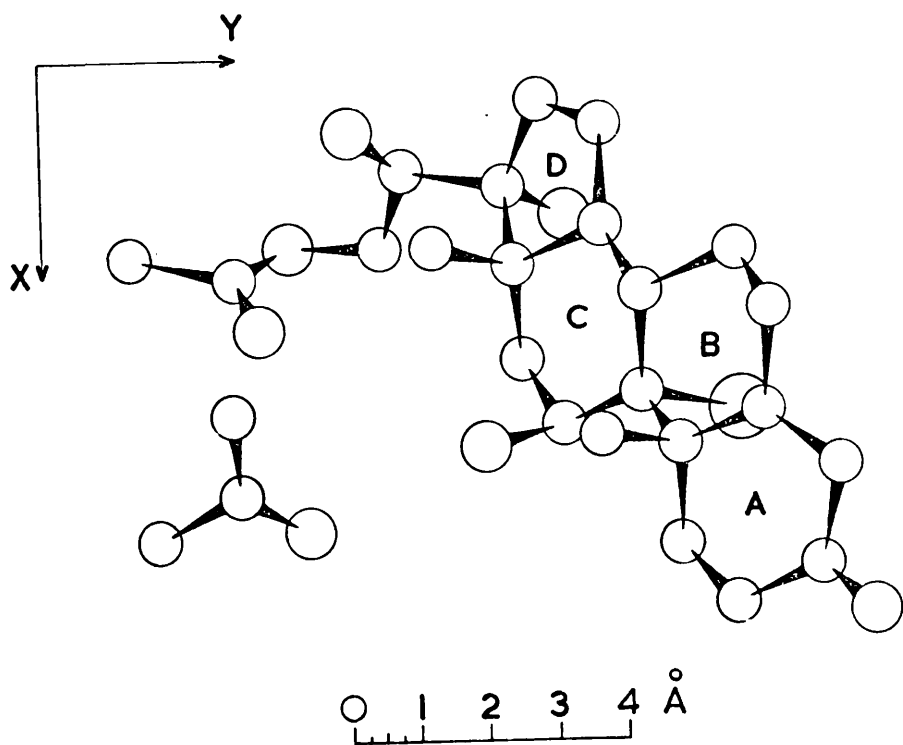
Figure 2.2 The three dimensional Patterson function, the section at  $U = \frac{1}{2}$ . The bromine-bromine vector is marked D. The contour scale is arbitrary.





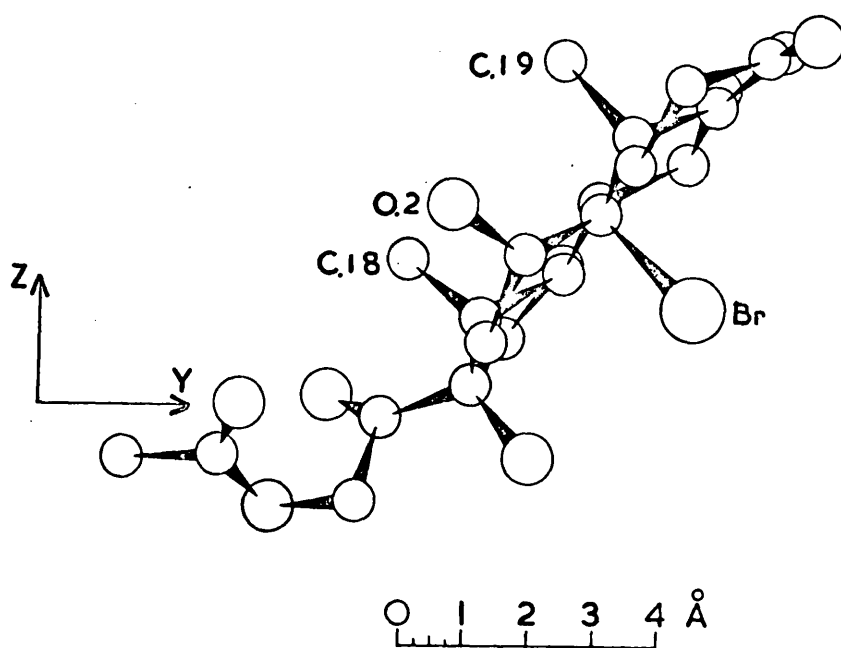
9- $\alpha$ -Bromocortisol Acetate

Figure 2.3 The final three-dimensional electron density distribution shown by means of superimposed contour sections drawn parallel to (001). The contour interval is  $1e/\text{\AA}^3$  except around the bromine atom where it is  $3e/\text{\AA}^3$



9- $\alpha$ -Bromocortisol Acetate

Figure 2.4    The atomic arrangement corresponding to figure 2.3



9- $\alpha$ -Bromocortisol Acetate

Figure 2.5 The atomic arrangement of the steroid as viewed along the  $a$ -axis.

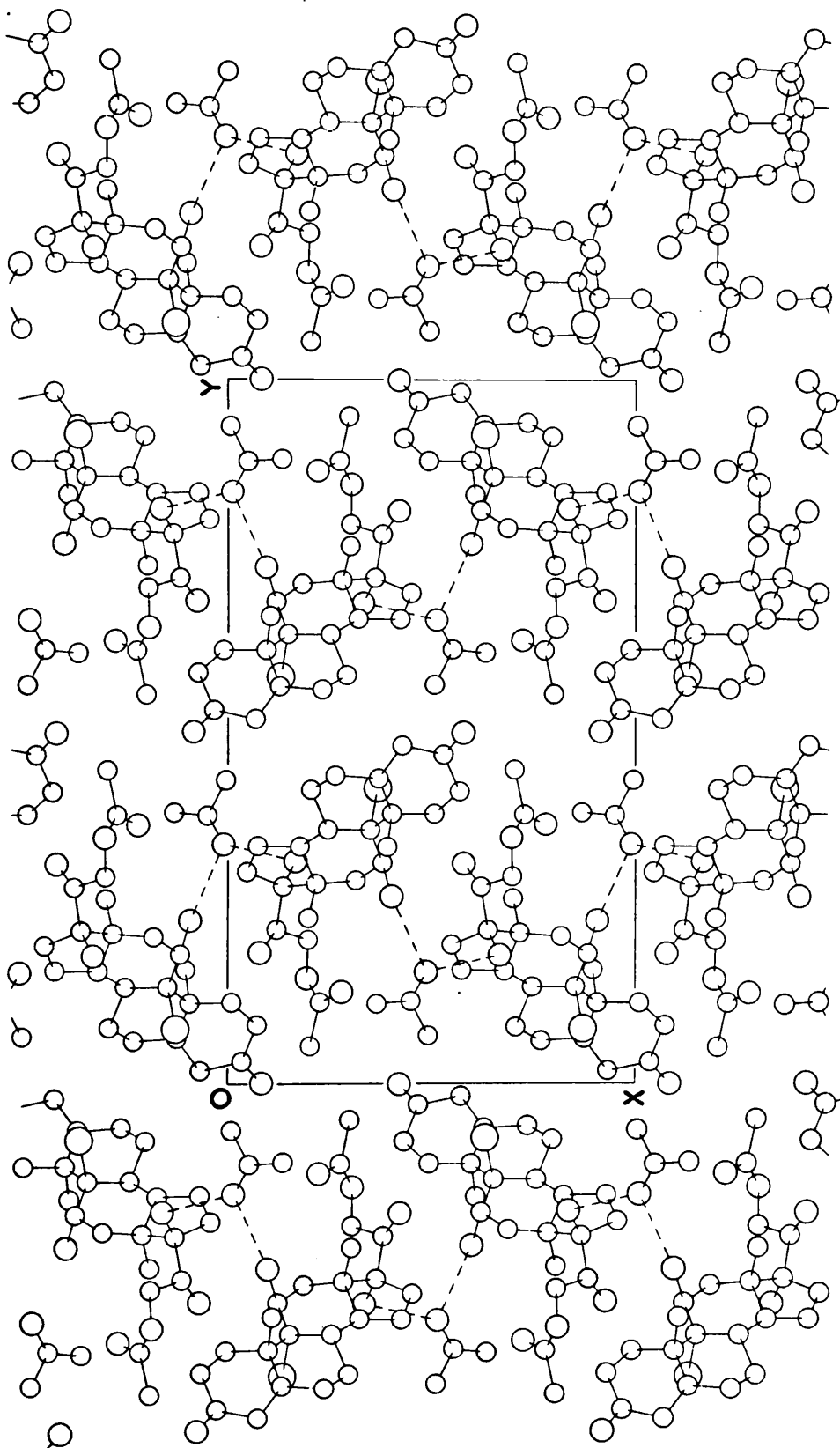


Figure 2.6 The arrangement of molecules in the crystal as viewed along the  $c$ -axis.

## 2.5 RESULTS OF THE ANALYSIS

The final atomic coordinates for one molecule are listed in table 2.3. The standard deviations of the coordinates, derived from the least squares residuals by application of the equation,

$$\sigma^2(U_j) = \sum w \Delta^2 / (m - n) \sum w \left( \frac{\partial \Delta}{\partial U_j} \right)^2$$

where  $m$  is the number of independent observations and  $n$  is the number of parameters to be determined, are listed in table 2.4.

Tables 2.5 and 2.6 show the isotropic and anisotropic temperature factors respectively, from the final least squares cycle. Table 2.2 lists the observed and calculated structure factors while the final three dimensional electron density map is shown in figure 2.3 by means of superimposed contour sections drawn parallel to (001). The atomic arrangement corresponding to this is shown in figure 2.4. Figure 2.5 portrays the atomic arrangement viewed along the  $a$  axis.

The interatomic bond lengths are listed in table 2.7 while the interbond angles are given in table 2.8. The standard deviation  $\sigma(AB)$  of a bond between atoms A and B is given by the formula,

$$\sigma(AB) = \left[ \sigma^2(A) + \sigma^2(B) \right]^{\frac{1}{2}}$$

where  $\sigma(A)$  and  $\sigma(B)$  are the standard deviations of the positions of the atoms A and B given by,

$$\sigma^2(A) = 1/3 \left[ \sigma^2(x) + \sigma^2(y) + \sigma^2(z) \right]$$

The standard deviation,  $\sigma(\beta)$  in radians, for an angle formed at atom B between the bonds AB and BC is given by the formula,

$$\sigma^2(\beta) = \frac{\sigma^2(A)}{AB^2} + \sigma^2(B) \left[ \frac{1}{AB^2} - \frac{2 \cos \beta}{AB \cdot BC} + \frac{1}{BC^2} \right] + \frac{\sigma^2(C)}{BC^2}$$

The average estimated standard deviations for the various bond types is shown below.

<u>Bond type</u>	<u>e.s.d. (Å)</u>
C - C	0.027
C - O	0.025
C - Br.	0.014

The average estimated standard deviation of a valency angle is 1.5°.

Some intramolecular and intermolecular non-bonded distances are shown in tables 2.9 and 2.10 respectively.

Table 2.11 shows the deviation of atoms from various planes calculated through the five membered ring D. The method of calculation is that of Schomaker et al (8). The atoms in groups (a) to (e) are obviously not planar, the deviations from the best planes being large. Plane (f) is the most likely to be a good fit for the four atoms included. The method of determining whether this is true is by application of the significance test based on the  $\chi^2$  distribution. This leads to tests of correspondence between experimental and theoretical data and has led to it being described as a test of "goodness of fit". Values of  $\chi^2$  and their correspondence to the probability of a good agreement having been achieved, are available (9).

In this application  $\chi^2 = \sum \Delta^2 / \sigma^2$  where  $\Delta$  is the deviation in angstroms of an atom from the calculated plane and  $\sigma$  is the mean standard deviation in angstroms of the positional parameters of the atoms. By calculating  $\chi^2$  from the above equation and knowing the number of degrees of freedom ( $n - 3$ ), the probability that no atoms deviate significantly from the calculated plane is found from the tables. If the probability ( $p$ ) is less than one per cent, then it is probable that one or more of the atoms included in the mean plane calculation does deviate significantly.

The equation of the best plane through the four atoms C(14) - C(17) is,

$$0.0844x - 0.7813y + 0.6185z + 4.9246 = 0$$

The value of  $\chi^2$  is 4.0 and the value of  $p$  is approximately 0.05. This means that none of the four atoms deviate significantly from the plane. The distance of the fifth atom, C(13), of the ring D is 0.75<sup>o</sup>Å from the plane and the five-membered ring assumes an envelope conformation.

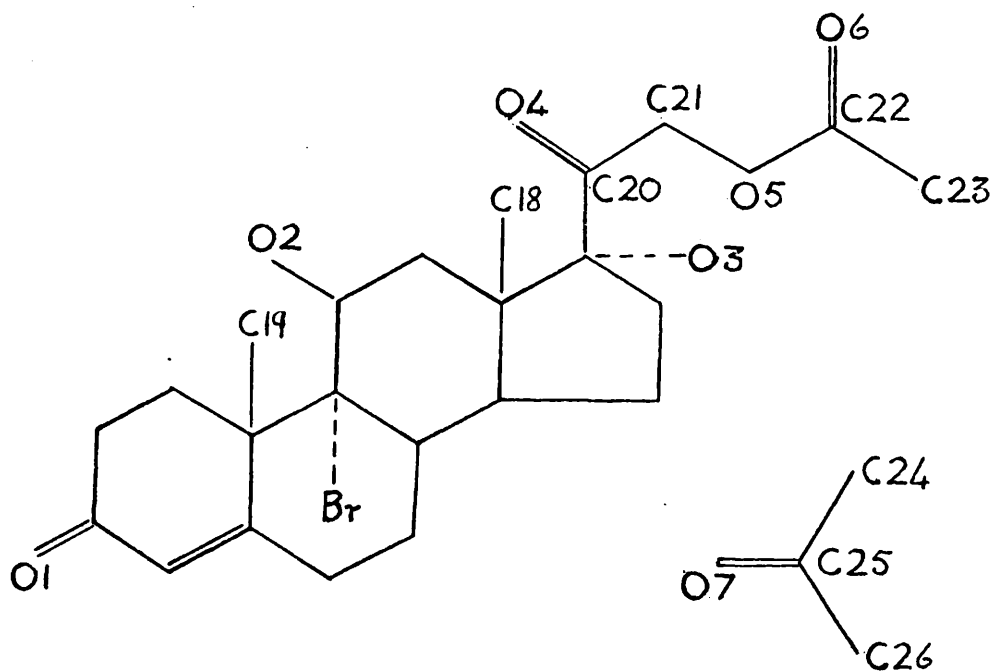
The packing of the molecules viewed along the  $c$  axis is shown in figure 2.6 and also in this figure is shown by means of dotted lines the system of possible hydrogen bonds.



2.6

DISCUSSION

The results of this analysis establish the structure and constitution of 9- $\alpha$ -Bromocortisol acetate and furnish information on crystal packing and molecular dimensions. The structure and the numbering system employed is shown in structure XIII



XIII

It might be expected that the steric repulsion between the  $\beta$ substituted methyl groups C(18) and C(19) and the  $\beta$  substituted hydroxyl group at O(2) would buckle the molecule somewhat but that.

the substitution of the bromine in the  $9\alpha$  position would offset this buckling tendency. Figure 2.5 shows the molecule in the (100) axis projection in which the relation of these groups is clearly exhibited. This figure shows the very flat conformation which the molecule assumes and there is not visual evidence of any overall buckling due to the steric repulsion of the  $\beta$  groups. It would therefore appear that steric repulsion between the bromine atom and carbons, C(1), C(5), C(7), C(11) and C(14), has nullified any such tendency and that the repulsive forces may be alleviated by the molecule assuming a very flat conformation. The intramolecular non-bonded contacts involving C(19) suggest some deformation in this part of the molecule. In a perfectly regular molecule the distances between C(19) and C(1), C(5) and C(9) should be approximately equal. In table 2.9 these distances are shown and it can be seen that the distance C(9) - C(19) is greater than the other two, suggesting that the C(19) atom is being repelled by the O(2) atom. Furthermore the C(11) - C(19) contact of  $3.28\overset{\circ}{\text{\AA}}$  is longer than the distances between C(19) and similarly place carbon atoms. The non-bonded contacts of C(18) are similar to this, the distance C(12) - C(18) of  $2.51\overset{\circ}{\text{\AA}}$  being shorter than both the C(14) - C(18) and C(17) - C(18) contacts. The distances of the bromine atom to C(1), C(5), C(7), C(12) and C(14) are very similar indicating the lack of distortion in the molecule.

The bonded distances are in good agreement with the expected values and with other organic compounds. The average carbon-carbon

single bond is  $1.53\overset{\circ}{\text{\AA}}$  in favourable agreement with the value of  $1.545\overset{\circ}{\text{\AA}}$  in diamond. The average lengths of similar bonds in other heavy atom derivatives viz. cedrelone iodoacetate (10) and bromogeigerin acetate (11) are  $1.55\overset{\circ}{\text{\AA}}$  and  $1.54\overset{\circ}{\text{\AA}}$  respectively. The average  $\text{sp}^3 - \text{sp}^2$  carbon-carbon bond is  $1.53\overset{\circ}{\text{\AA}}$  which is not significantly greater than the accepted value of  $1.525\overset{\circ}{\text{\AA}}$ . An example of this type of bond is the carbon-carbon bond of methyl acetate reported by Sutton et al. (12) as  $1.52\overset{\circ}{\text{\AA}}$ . The only carbon-carbon double bond present in the structure is the C(4) - C(5) bond. This distance is  $1.37\overset{\circ}{\text{\AA}}$  which is not significantly longer than the value of  $1.33\overset{\circ}{\text{\AA}}$  in ethylene (12) and agrees well with the value of  $1.36\overset{\circ}{\text{\AA}}$  in acraldehyde, a double bond in a more comparable environment (12). One further type of carbon-carbon bond remains. This is the  $\text{sp}^2 - \text{sp}^2$  non double bonded distance of C(3) - C(4) and has a length of  $1.46\overset{\circ}{\text{\AA}}$ , identical to that of the central bond in acraldehyde.

The carbon-oxygen bonds fall into three categories, the  $\text{C}(\text{sp}^3) - \text{O}$  single bond, the  $\text{C}(\text{sp}^2) - \text{O}$  single bond and the carbon-oxygen double bond. There are three bonds in the first category. These are C(21) - O(5), C(11) - O(2) and C(17) - O(3) and have lengths of  $1.46\overset{\circ}{\text{\AA}}$ ,  $1.44\overset{\circ}{\text{\AA}}$  and  $1.43\overset{\circ}{\text{\AA}}$  respectively which are in good agreement with the value of  $1.46\overset{\circ}{\text{\AA}}$  given by Sutton et al. (12) for the similar bond of methyl acetate. The  $\text{C}(\text{sp}^2) - \text{O}$  distance in methyl acetate is  $1.36\overset{\circ}{\text{\AA}}$ , a value which agrees well with the length of  $1.35\overset{\circ}{\text{\AA}}$  of the one such similar bond C(22) - O(5) in the cortisol derivative. The final type of bond viz. the double bond is found at C(3), C(20), C(22) and in the acetone molecule of crystallisation. The average bond length is  $1.20\overset{\circ}{\text{\AA}}$  which is comparable to the  $1.21\overset{\circ}{\text{\AA}}$  predicted by

Pauling (13) and the values of  $1.19\text{\AA}$  and  $1.22\text{\AA}$  reported by Asher (14) and Trotter (15) respectively for this type of bond.

The carbon-bromine distance is  $2.05\text{\AA}$ . Examples of values found for such bonds are  $1.99\text{\AA}$  by Hamilton, McPhail and Sim in bromogeigerin acetate (11),  $1.93\text{\AA}$  and  $2.01\text{\AA}$  in the two isomers of 1 - 4 dibromocyclohexane reported by Sutton et al. (12).

Various angle types are present in the structure and these will be considered separately. The average angle round an  $sp^3$  hybridised carbon atom is  $111^\circ$  which is higher than the tetrahedral angle of  $109^\circ 28'$  but not significantly so. Around the  $sp^2$  hybridised carbon atoms, the average angle is  $120^\circ$  which is identical with the expected value. There are six such carbon atoms, C(3), C(4), C(5), C(20), C(22) and C(25). The individual averages at these carbons are all  $120^\circ$ . The angle C(21) - O(5) - C(22) is  $116^\circ$  which is significantly higher than the tetrahedral angle although the presence of the lone pairs on the oxygen atom is expected to cause distortion from a tetrahedral arrangement. The value of the similar angle in methyl acetate is quoted by Sutton et al. (12) as  $113^\circ$ . The average angle of ring A (fig. 2.4) is  $116^\circ$ . This ring contains a double bond C(4) - C(5) and the average angle is therefore expected to be greater than the normal tetrahedral value. In the similar six membered ring of hunterburnine ( $\beta$ -methiodide (14) the average angle is  $114^\circ$ , the value also found in the corresponding ring of macusine - A(16). Ring B has an average angle of  $113^\circ$ . This is not significantly greater than  $109^\circ 28'$  and in any case would be expected to be slightly higher due to the presence of the  $sp^2$  hybridised carbon atom, C(5), in the ring. A

value showing closer agreement with the tetrahedral angle is the average of  $112^\circ$  in ring C which contains only  $sp^3$  hybridised carbon atoms. The average angle in ring D is made somewhat unreliable by the presence of an abnormally narrow angle of  $98^\circ$  at C(13). The average angle is  $103^\circ$  but excluding the angle at C(13), becomes  $105^\circ$ . This is below the value of  $108^\circ$  expected of a planar five membered ring but agrees better with values of  $106^\circ$  in clerodin bromolactone (17),  $105^\circ$  in isoclovene hydrochloride (18) and  $105^\circ$  in himbacine hydrobromide (19). These values indicate non-planarity with consequent enlargement of the angles as was demonstrated in the case of cyclopentane (20). It is difficult to explain the cause of the narrow angle at C(13) and it was at first thought that the atom, C(13), had been misplaced. However the other angles involving C(13) are quite credible and furthermore similar values for this angle have been found in related compounds. For example the corresponding angle of the 17- $\beta$ -iodoacetoxy derivative of 3-keto-4,4-dimethyl androstane (21) is  $98^\circ$ , the structure being at an advanced state of refinement. The small angle is presumably due to strain in the ring caused by the rigidity of the molecule which does not allow a conformation of minimum angular strain to be adopted.

In the crystal the molecules form a three-dimensional network held together by normal van der Waals forces and a system of hydrogen bonds involving the oxygen of the acetone molecule of crystallisation and O(2) and O(3) of symmetry related molecules. Figure 2.6 shows the crystal packing in projection along the C axis and the system of hydrogen bonds is shown by means of broken lines. The lengths of

these bonds are  $O(7) \cdots O(2)_1$ ,  $2.94\text{\AA}$  and  $O(7) \cdots O(3)_3$ ,  $2.91\text{\AA}$  where the subscripts 1 and 3 refer to the equivalent positions  $x, y, z$  and  $\frac{1}{2} + x, \frac{1}{2} - y, -z$ , respectively. The hydrogen bond angles,  $C(11) - O(2) \cdots O(7)$  and  $C(17)_3 - O(3)_3 \cdots O(7)$  and have respective values of  $119^\circ$  and  $103^\circ$ . These are somewhat distorted from the tetrahedral angle but are well within the tolerance suggested by Fuller (22) for hydrogen bond formation. The dimensions of these bonds are in favourable agreement with the values found for other such bonds. For example the hydrogen bond in  $\alpha$ -glucose is  $2.87\text{\AA}$  (23) and angles between  $83^\circ$  (24) and  $123^\circ$  (25) have been reported. The other intermolecular distances (table 2.10) correspond to normal van der Waals contacts.

REFERENCES

1. R.E. Marker, R.B. Wagner, P.R. Ulshafer, E.L. Wittbecker, D.P.J. Goldsmith, C.H. Ruof, J. Amer. Chem. Soc., 69, 2167, 1947.
2. R.B. Woodward, F. Sondheimer, D. Taub, K. Heusser, W.M. MacLamore, J. Amer. Chem. Soc., 73, 2403, 1951; 73, 4057, 1951.
3. R.B. Woodward et al., J. Amer. Chem. Soc., 74, 4223, 1952.
4. J. Fried, E.F. Sabo, J. Amer. Chem. Soc., 79, 1130, 1957.
5. J.M. Robertson, J. Sci. Instr., 20, 175, 1943.
6. G. Tunnel, Am. Min., 24, 448, 1939.
7. G.A. Sim. Acta Cryst., 10, 536, 1957.
8. V. Shomaker, J. Waser, R.E. Marsh, G. Bergman, Acta Cryst., 12, 600, 1959.
9. R.A. Fisher, F. Yates, Statistical Tables, Oliver and Boyd, Edinburgh, 1957.
10. I.J. Grant, Ph.D. Thesis, University of Glasgow, 1962.
11. Miss J.A. Hamilton, A.T. McPhail, G.A. Sim, J. Chem. Soc., 708, 1962.
12. L.E. Sutton et al., Tables of interatomic distances; The Chemical Society, London, 1958.
13. L. Pauling, The Nature of the Chemical Bond, 3rd ed., Cornell University Press, New York, 1960.



14. J.D.M. Asher, Ph.D. Thesis, University of Glasgow, 1963.
15. J. Trotter, *Acta Cryst.*, 13, 86, 1960.
16. A.T. McPhail, J.M. Robertson, G.A. Sim, *J. Chem. Soc.*, 1832, 1963.
17. I.C. Paul, G.A. Sim, T. Hamor, J.M. Robertson, *J. Chem. Soc.*, 4133, 1962.
18. J.S. Clunie, J.M. Robertson, *J. Chem. Soc.*, 4382, 1961.
19. J. Fridrickson, A.M. Mathieson, *Acta Cryst.*, 15, 119, 1962.
20. K.S. Pitzer, W.E. Donath, *J. Amer. Chem. Soc.*, 81, 3213, 1959.
21. E.W. Macauley, Glasgow University, Personal Communication.
22. C.S. Fuller, *J. Phys. Chem.*, 63, 1705, 1959.
23. T.R.R. McDonald, C.A. Beevers, *Acta Cryst.*, 5, 654, 1952.
24. R.A. Pasternak, *Acta Cryst.*, 9, 341, 1956.
25. D.W. Smits, E.M. Wiebenga, *Acta Cryst.*, 6, 531, 1953.

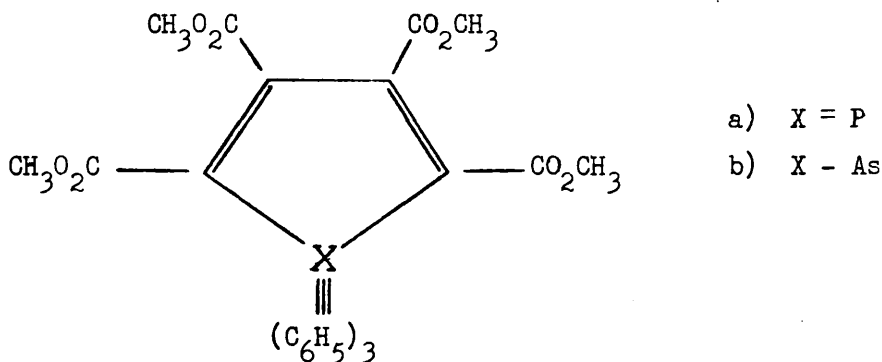
## C H A P T E R   3

THE CRYSTAL STRUCTURE ANALYSIS OF THE ADDUCT OF  
TRIPHENYLARSINE AND ACETYLENE DICARBOXYLIC METHYL ESTER

3.1

INTRODUCTION

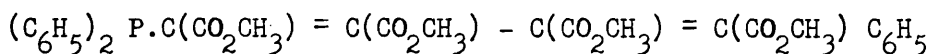
Addition of triphenylphosphine to acetylene dicarboxylic ester affords a pentacovalent phosphole shown in structure Ia.



I

Thus when a product from the corresponding reaction with  $(C_6H_5)_3As$  was gained it was accorded a similar structure, Ib. The shape of such a compound was of interest and hence this x-ray analysis was undertaken. From the outset it was obvious that this proposed structure was doubtful since determination of the density of the crystal afforded a calculated molecular weight well below that required for structure Ib. Chemical evidence was equally confusing (1, 2). The original analysis of the product had appeared reasonable with two carbonyl peaks at  $5.80\mu$  and  $6.00\mu$  present, similar to

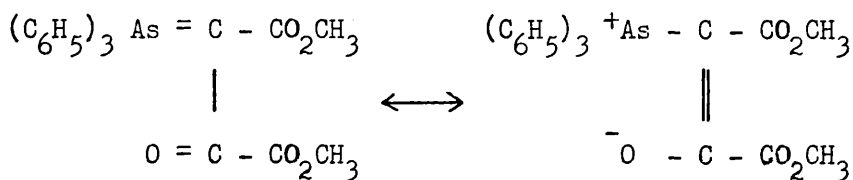
those in the phosphole, and although further corroborative evidence from proton magnetic resonance and U.V. spectra was gained, there were certain serious differences in the properties of the phosphorous and arsenic compounds. These differences could not be accounted for on the basis of such similar structures. Whereas the phosphole is yellow and undergoes a slow spontaneous transformation at room temperature, the major product being structure II, the arsenic compound is colourless and has very high thermal stability.



## II

Two further pieces of evidence led to Ib being discarded. Firstly, the strong indication that the adduct was in fact a higher oxidation state was provided by the later recovery of substantial yields of dimethyl fumarate from the original reaction. This implied reduction of some of the acetylene ester and hence oxidation of the product. Secondly, reaction of the adduct with hydrazine yielded the bishydrazide of oxalic acid.

These results suggested the structure to be III, the 1:1 adduct.

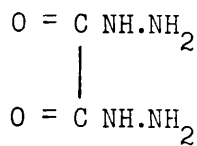
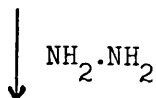
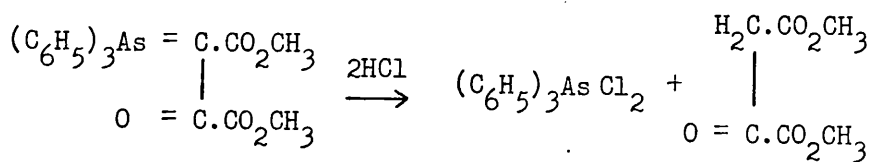
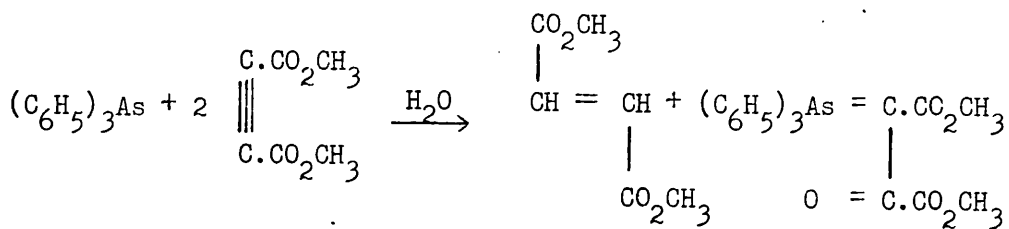


### III

The resonance involved is discussed later. This structure was verified by our analysis and by the following experiment (2).

Reaction of the adduct with dry hydrogen chloride in ethylene chloride at room temperature afforded a high yield of dimethyl oxalacetate as well as triphenylarsine dichloride. These products dictate III as the structure of the adduct on both mechanistic and stoichiometric grounds.

Hendrickson and co-workers (2) have postulated the following mechanism for the formation of structure III. The same sequential attack of the triphenylarsine on two molecules of the acetylene ester as occurs with triphenylphosphine is proposed with the following difference. The sequence is terminated by protonation instead of cyclisation. This reaction explains why the 1:2 adduct was expected. The reactions are summarised below.



### 3.2 EXPERIMENTAL DATA



CRYSTAL DATA

Initial formula  $C_{30}O_8As H_{27}$

Final formula  $C_{24}O_5As H_{21}$

Molecular weight 464.3

Assuming 4 molecules per cell D (calculated) = 1.45 gm/c.c.

D (observed) = 1.46 gm/c.c.

With initial formula D (calculated) = 1.84 gm/c.c.

The crystal is monoclinic with,

$$a = 10.91 \pm 0.03$$

$$b = 13.66 \pm 0.04$$

$$c = 14.61 \pm 0.04$$

$$\gamma = 101^\circ 37'$$

$$\text{Volume of the cell} = 2133 \text{ \AA}^3$$

Number of molecules per cell = 4

The space group is  $P2_1/a$

The space group was established from the systematic absences which are;

hk0 when h is odd

and 00l when l is odd

Linear absorption coefficient for x-rays ( $\lambda = 1.542\text{\AA}$ )

$$\mu = 26.5 \text{ cm}^{-1}$$

Total number of electrons per cell =  $F(000)$  = 952

$$\sum f^2(\text{light atoms}) = 1205 \quad (\sin \theta = 0)$$

$$\sum f^2(\text{heavy atoms}) = 1089 \quad (\sin \theta = 0)$$

$$r = \left[ \sum f^2(h) / \sum f^2(l) \right]^{\frac{1}{2}} = 0.95$$

## INTENSITY DATA

Rotation, oscillation, Weissenberg and precession photographs were taken with copper  $K_{\alpha}$  ( $\lambda = 1.5418\text{\AA}$ ) and molybdenum  $K_{\alpha}$  ( $\lambda = 0.7107\text{\AA}$ ) radiation. Rotation and precession photographs were used for the unit cell parameters. A small crystal completely bathed in a uniform x-ray beam was used for the recording of the intensities. No absorption corrections were made. The intensities were collected on zero layer and equi-inclination upper layer Weissenberg photographs obtained by rotating the crystal about the C axis, the reciprocal lattice nets,  $hk0$  ---  $hkl0$ , being recorded in this way. The estimation of intensities was carried out visually by comparison with a calibrated step-wedge. Correlation between planes of high and low intensity was achieved by the multiple film technique of Robertson (3) and the intensities on successive films of upper layer series were put on the same scale by means of a variable film factor due to Rossmann (4). This film factor was found to be consistent with observations.

### 3.3     THE STRUCTURE ANALYSIS

SOLUTION OF THE PATTERSON MAP FOR  
THE HEAVY ATOM POSITION

In the space group  $P2_1/a$  with unique axis  $\underline{c}$ , the expected heavy atom-heavy atom vectors will be at  $2x, 2y, 2z$ ;  $\frac{1}{2} + 2x, 2y, \frac{1}{2}$ ;  $\frac{1}{2}, 0, \frac{1}{2} + 2z$ ; because of the centre of symmetry, two-fold screw axis and the  $\underline{a}$  glide respectively.

The two-dimensional Patterson function in (001) was computed with 165  $hk0$  data and is shown in figure 3.1. The peaks in this projection corresponding to the above vectors are designated B, C and A respectively. Peaks B and C were arbitrarily assigned since in projection, (UV), it is impossible to distinguish between them. However by calculating line sections through the three-dimensional Patterson distribution at points B and C it was possible to interpret unambiguously the vector set and arrive at coordinates for the arsenic at

$$x/a = 0.35481$$

$$y/b = 0.22151$$

$$z/c = 0.09224$$

## SOLUTION OF THE STRUCTURE

The value of  $r$  ( $= [\sum f^2(h) / \sum f^2(1)]^{\frac{1}{2}}$ ) was calculated to be 0.95, indicating that the first set of phasing calculations based on the arsenic atom should give a reasonable approximation to the correct phases.

The first set of structure factors was calculated using an isotropic temperature factor of  $U(\text{iso}) = 0.056$ , a value estimated from a knowledge of temperature factors of similar organic structures. The atomic scattering factors for arsenic were those of International table volume III. The residual  $R$ , where  $R = 100 \cdot (\sum ||F_o| - |F_c||) / \sum |F_o|$ , was 47.6%. The various layers,  $hk0 - hk10$ , of the structure factors were scaled by comparing  $\sum |F_o|$  and  $\sum |F_c|$  and a first electron density map was computed using all the structure factors except those with a very high value of  $\Delta$ , where  $\Delta = |F_o| - |F_c|$ .

Since the structure was known to contain three benzene rings bonded to the arsenic atom, a search of the electron density map was made for these features. Two of the rings were immediately evident. The third ring was found with a little more difficulty due to a larger amount of distortion and a tendency for this ring to consist of rather small peaks. With the position of the three benzene rings fixed, the other part of the molecule was found fairly easily. This proved to be an open chain as in structure III and not the ring structure, I, first proposed. The maxima of the 30 peaks assigned

to atoms were calculated by the method due to Booth (5) and included in the second set of phasing calculations. Apart from the arsenic atom, all were included as carbon atoms. The residual R dropped by 12.6% to a new value of 35.0%.

From the subsequent electron density map, new coordinates for the atoms were calculated. From chemical knowledge, geometry and peak height it was possible to differentiate between oxygen and carbon atoms and these were included as such in a further structure factor calculation. The discrepancy was now 32.5%. A third electron density distribution was calculated, both  $F_o$  and  $F_c$  maps being computed. This afforded new coordinates and as a result of comparison of peak heights on both maps a new temperature factor  $U(\text{iso}) = 0.049$  was employed. Somewhat surprisingly this produced a drop in R of only 1.5%, the value now being 31.0%. It was noticed that while the peak heights were in general increasing, that of C(21) was not. This was a carbon atom belonging to the ring which in the original electron density map had been distorted. In order to check the position of this atom bond lengths and angles were calculated. These were quite reasonable within the distortion expected at this stage, apart from those involving C(21). The previous structure factor calculation was repeated excluding C(21) and this produced a further drop in R of 1% showing that the position attributed to C(21) was incorrect. The misplacement of C(21) can be attributed to the fact that its position in the cell was very close to a diffraction

ripple directly above the arsenic atom and it may well have been this diffraction ripple which was taken for the atom in the first Fourier distribution. Small sections of an electron density map were calculated in the vicinity of the benzene ring which included C(21). A peak was established which was shown by considerations of geometry to be satisfactory for the position of C(21).



## STRUCTURE REFINEMENT

A further round of structure factor calculations with improved atomic positions yielded an R factor of 29.0%. The structure factors were scaled and  $F_o$  and  $F_c$  maps computed. New atomic coordinates were calculated and on the basis of peak height discrepancies between the  $F_o$  and  $F_c$  maps, individual isotropic temperature factors were allotted to the atoms. This had the effect of lowering R by a more encouraging amount viz. 3.7%, to a value of 25.3%. By repeating this process and assigning more accurate temperature factors to the atoms, a further drop in R, to 24.0%, was achieved. A final R factor of 23.1% was gained in this way and refinement was then continued by the method of least squares.

Block diagonal isotropic least squares on K.D.F.9 produced coordinates which when used in a further round of calculations reduced R to 19.2%. In all three cycles of isotropic structure factor least squares refinement were computed. Even after this small number of cycles, the drop in R fell off very rapidly. After the third cycle the discrepancy was 18.6% and it was obvious that further cycles of isotropic refinement would prove fruitless. Therefore in the fourth cycle it was decided to switch some of the atoms from isotropic to anisotropic temperature factors. At the time of this refinement, computing facilities were such that conversion of all the atoms to anisotropic values produced too large a number of parameters for the computer to handle. Since

the side chain is likely to be less rigid than the rest of the molecule, further refinement of the atoms As, C(1) - C(6) and O(1) - O(5) was carried out employing anisotropic temperature factors while isotropic refinement was continued for the rest of the atoms. This course of action had an immediate effect on the R factor, the value after least squares four being 16.7% and the values of the anisotropic temperature factors,  $U_{11}$ ,  $U_{22}$ ,  $U_{33}$ , showed marked anisotropic temperature vibration in this part of the molecule. A further two cycles of mixed isotropic and anisotropic refinement furnished discrepancies of 15.3% and 15.0% respectively. Refinement was continued by refining the atoms of the benzene rings anisotropically and suspending refinement of the other atoms. One such cycle and one final cycle refining all the atoms with anisotropic temperature factors afforded a final R factor of 14.3%. Throughout the refinement a check was kept on progress by the R factor, the value of  $\sum w\Delta^2$  and by frequent bond length and angle calculations. The course of the analysis is shown in table 3.1.

#### 3.4 TABLES AND FIGURES

TABLE 3.1

Course of the Analysis

<u>S.F.</u>	<u>Atoms Included</u>	<u>R%</u>	<u>Temperature Factor</u>
1	1	47.6	Overall isotropic
2	30	35.0	"
3	30	32.5	"
4	30	31.0	"
4a	29	30.0	"
5	30	29.0	"
6	30	25.3	Individual isotropic
7	30	24.0	"
8	30	23.1	"
<u>S.F.L.S.</u>	<u><math>\sum w\Delta^2</math></u>	<u>R%</u>	<u>Temperature factor</u>
1		19.2	Individual isotropic
2	54156	18.7	"
3	32700	18.6	"
4	25743	16.7	12 atoms anisotropic, 18 isotropic
5	22314	15.3	"
6	19099	15.0	"
7	17289	14.5	18 atoms anisotropic, allowance for 12
8	16871	14.3	Fully anisotropic

TABLE 3.1 (Contd.)

Weighting scheme

$$\sqrt{w} = 1 \quad \text{if} \quad |F_o| \leq p$$

$$\sqrt{w} = p/|F_o| \quad \text{if} \quad |F_o| > p$$

where

$p$  = average structure amplitude

### The Measured and Calculated Values of the Structure Factors

[illegible]

TABLE 3.2 (Contd.)

[illegible]

TABLE 3.2 (Contd.)

[illegible]



TABLE 3.3

Triphenylarsine Adduct

Final Atomic Coordinates

<u>Atom</u>	<u>x/a</u>	<u>y/b</u>	<u>z/c</u>
C(1)	0.37612	0.55287	0.09076
C(2)	0.29548	0.40442	0.00964
C(3)	0.28852	0.29827	0.00146
C(4)	0.23457	0.24212	-0.07338
C(5)	0.16059	0.28846	-0.14289
C(6)	-0.03234	0.33277	-0.17078
C(7)	0.49337	0.17310	0.04014
C(8)	0.55695	0.11454	0.09878
C(9)	0.66461	0.08633	0.06142
C(10)	0.70825	0.11442	-0.02668
C(11)	0.64293	0.16965	-0.08324
C(12)	0.53347	0.19740	-0.04802
C(13)	0.22506	0.11331	0.13302
C(14)	0.25499	0.04602	0.19485
C(15)	0.15997	-0.03333	0.22607
C(16)	0.03816	-0.04338	0.19277
C(17)	0.01079	0.02580	0.13008
C(18)	0.10291	0.10387	0.09874
C(19)	0.42074	0.29115	0.19787

TABLE 3.3 (Contd.)

<u>Atom</u>	<u>x/a</u>	<u>y/b</u>	<u>z/c</u>
C(20)	0.35065	0.28233	0.27726
C(21)	0.40258	0.33016	0.35670
C(22)	0.51808	0.38656	0.35619
C(23)	0.58980	0.39737	0.27460
C(24)	0.53933	0.35031	0.19801
O(1)	0.35334	0.44568	0.08100
O(2)	0.25748	0.45461	-0.05111
O(3)	0.23879	0.15190	-0.08766
O(4)	0.19606	0.31197	-0.21994
O(5)	0.04931	0.29244	-0.10970
As	0.35407	0.22112	0.08914

TABLE 3.4

Triphenylarsine AdductStandard Deviations of the Atomic Coordinates ( $\text{\AA}$ )  $\times 10^3$ 

<u>Atom</u>	<u>x</u>	<u>y</u>	<u>z</u>	<u>Atom</u>	<u>x</u>	<u>y</u>	<u>z</u>
C(1)	18	11	17	C(16)	12	11	13
C(2)	9	7	11	C(17)	11	11	14
C(3)	9	8	11	C(18)	10	11	12
C(4)	9	9	10	C(19)	9	9	11
C(5)	10	10	14	C(20)	13	13	15
C(6)	14	17	19	C(21)	16	16	17
C(7)	8	9	11	C(22)	15	14	16
C(8)	12	13	14	C(23)	14	15	17
C(9)	12	13	15	C(24)	12	13	15
C(10)	11	12	13	O(1)	9	7	8
C(11)	12	12	13	O(2)	9	7	11
C(12)	10	10	13	O(3)	7	7	8
C(13)	9	9	11	O(4)	10	10	11
C(14)	10	10	12	O(5)	10	8	8
C(15)	11	10	13	As	1	1	1

TABLE 3.5

Triphenylarsine AdductAnisotropic Temperature Factors  $\times 10^4$ 

<u>Atom</u>	<u>U(11)</u>	<u>U(22)</u>	<u>U(33)</u>	<u>2U(23)</u>	<u>2U(31)</u>	<u>2U(12)</u>
C(1)	1111	188	1115	-248	311	154
C(2)	451	183	336	268	67	356
C(3)	405	258	293	-121	-144	302
C(4)	384	305	233	-28	-9	351
C(5)	467	345	335	-37	-39	440
C(6)	665	867	821	131	-582	803
C(7)	484	281	254	68	-188	686
C(8)	544	841	683	455	158	1457
C(9)	827	911	484	244	194	1762
C(10)	618	766	185	-15	85	947
C(11)	473	516	457	-49	1	272
C(12)	587	199	496	9	-82	373
C(13)	580	293	121	-160	35	574
C(14)	706	316	272	295	-30	435
C(15)	924	335	130	148	632	598
C(16)	741	320	422	-394	487	97
C(17)	409	691	321	-181	137	84
C(18)	351	619	434	195	67	611
C(19)	384	325	526	-264	-380	510

TABLE 3.5 (contd.)

<u>Atom</u>	<u>U(11)</u>	<u>U(22)</u>	<u>U(33)</u>	<u>2U(23)</u>	<u>2U(31)</u>	<u>2U(12)</u>
C(20)	633	601	539	-696	207	10
C(21)	1377	782	146	-521	48	828
C(22)	1233	568	462	-676	-1070	1100
C(23)	666	991	675	-464	-875	250
C(24)	523	588	895	-613	-875	595
O(1)	865	216	537	-174	-120	184
O(2)	915	309	647	174	-309	530
O(3)	467	338	502	-178	-125	359
O(4)	829	878	539	349	358	929
O(5)	414	575	483	146	-49	528
As	369	298	325	-34	-55	337

TABLE 3.6

Triphenylarsine AdductBonded Distances ( $\text{\AA}$ )

<u>Bond</u>			<u>Bond</u>		
<u>Distance</u>			<u>Distance</u>		
C(1)	O(2)	1.44	C(13)	C(14)	1.37
C(2)	O(1)	1.29	C(14)	C(15)	1.42
C(2)	O(2)	1.24	C(15)	C(16)	1.40
C(2)	C(3)	1.44	C(16)	C(17)	1.39
C(3)	C(4)	1.40	C(17)	C(18)	1.39
C(4)	O(3)	1.26	C(18)	C(13)	1.41
C(4)	C(5)	1.51	C(19)	C(20)	1.38
C(5)	O(4)	1.21	C(20)	C(21)	1.39
C(5)	O(5)	1.32	C(21)	C(22)	1.34
C(6)	O(5)	1.45	C(22)	C(23)	1.42
C(7)	C(8)	1.44	C(23)	C(24)	1.35
C(8)	C(9)	1.42	C(24)	C(19)	1.38
C(9)	C(10)	1.40	As	C(3)	1.89
C(10)	C(11)	1.40	As	C(7)	1.91
C(11)	C(12)	1.42	As	C(13)	1.93
C(12)	C(7)	1.38	As	C(19)	1.92

TABLE 3.7

Triphenylarsine AdductInterbond Angles

C(3)	C(2)	O(1)	115°	C(14)	C(13)	C(18)	121°
C(3)	C(2)	O(2)	123	C(13)	C(14)	C(15)	119
O(1)	C(2)	O(2)	122	C(14)	C(15)	C(16)	120
C(2)	C(3)	C(4)	123	C(15)	C(16)	C(17)	120
C(2)	C(3)	As	124	C(16)	C(17)	C(18)	121
C(4)	C(3)	As	113	C(17)	C(18)	C(13)	119
C(3)	C(4)	C(5)	120	C(20)	C(19)	As	120
C(3)	C(4)	O(3)	125	C(24)	C(19)	As	121
C(5)	C(4)	O(3)	115	C(20)	C(19)	C(24)	119
C(4)	C(5)	O(5)	110	C(19)	C(20)	C(21)	120
C(4)	C(5)	O(4)	124	C(20)	C(21)	C(22)	121
O(5)	C(5)	O(4)	125	C(21)	C(22)	C(23)	120
C(8)	C(7)	As	118	C(22)	C(23)	C(24)	119
C(12)	C(7)	As	120	C(23)	C(24)	C(19)	122
C(8)	C(7)	C(12)	122	C(1)	O(1)	C(2)	120
C(7)	C(8)	C(9)	116	C(5)	O(5)	C(6)	116
C(8)	C(9)	C(10)	122	C(13)	As	C(19)	104
C(9)	C(10)	C(11)	120	C(3)	As	C(7)	111
C(10)	C(11)	C(12)	118	C(3)	As	C(13)	111
C(11)	C(12)	C(7)	121	C(7)	As	C(13)	112
C(14)	C(13)	As	120	C(7)	As	C(19)	104
C(18)	C(13)	As	119	C(3)	As	C(19)	116

TABLE 3.8

Triphenylarsine AdductSome intramolecular non-bonded distances ( $\overset{\circ}{\text{\AA}}$ )

C(1)	O(2)	2.66	C(9)	C(12)	2.79
C(2)	C(5)	2.95	C(13)	C(16)	2.78
C(3)	O(4)	3.40	C(13)	O(3)	3.27
C(3)	O(5)	3.06	C(14)	C(17)	2.79
C(3)	C(7)	3.13	C(15)	C(18)	2.80
C(3)	C(13)	3.14	C(18)	O(3)	3.11
C(3)	C(19)	3.22	C(19)	C(22)	2.76
C(4)	C(13)	3.48	C(19)	O(1)	2.92
C(4)	C(18)	3.30	C(20)	C(23)	2.76
C(5)	O(2)	2.66	C(20)	O(1)	3.63
C(6)	O(2)	3.70	C(21)	C(24)	2.74
C(7)	C(10)	2.80	O(1)	As	3.07
C(7)	C(19)	3.01	O(2)	O(1)	2.21
C(7)	C(24)	3.31	O(2)	O(5)	2.96
C(7)	C(13)	3.18	O(2)	O(4)	3.13
C(7)	C(14)	3.61	O(3)	As	2.94
C(8)	C(11)	2.87			



TABLE 3.9

Triphenylarsine AdductIntermolecular Distances  $< 3.8\text{\AA}$ 

C(1)	C(12) <sub>3</sub>	3.41	C(11)	O(4) <sub>2</sub>	3.45
C(1)	O(4) <sub>4</sub>	3.50	C(11)	C(14) <sub>3</sub>	3.73
C(2)	C(22) <sub>2</sub>	3.57	C(11)	C(15) <sub>3</sub>	3.75
C(5)	C(16) <sub>3</sub>	3.68	C(15)	O(3) <sub>4</sub>	3.46
C(6)	O(4) <sub>2</sub>	3.33	C(16)	O(3) <sub>3</sub>	3.45
C(8)	C(15) <sub>2</sub>	3.58	C(16)	O(5) <sub>3</sub>	3.56
C(8)	C(16) <sub>2</sub>	3.71	C(17)	O(3) <sub>3</sub>	3.32
C(9)	C(15) <sub>2</sub>	3.50	C(20)	C(23) <sub>2</sub>	3.60
C(9)	O(3) <sub>3</sub>	3.64	C(20)	C(24) <sub>2</sub>	3.71
C(9)	C(14) <sub>2</sub>	3.76	C(22)	O(1) <sub>2</sub>	3.70
C(10)	C(14) <sub>3</sub>	3.38	C(23)	O(1) <sub>2</sub>	3.52
C(10)	C(15) <sub>3</sub>	3.52	C(24)	O(2) <sub>3</sub>	3.77
C(10)	C(13) <sub>3</sub>	3.68			

The subscripts refer to the following equivalent positions

1.  $x, y, z$
2.  $\frac{1}{2} + x, y, \frac{1}{2} - z$
3.  $-x, -y, -z$
4.  $\frac{1}{2} - x, -y, -\frac{1}{2} + z$

TABLE 3.10

Triphenylarsine Adduct

Deviations of the atoms from mean planes  
calculated through the benzene rings ( $\text{\AA}$ )

	<u>Atoms</u>	<u>Deviation</u>	
Ring A	C(7)	-0.0156	$\chi^2 = 1.89; p = 0.6$
	C(8)	0.0024	
	C(9)	0.0116	
	C(10)	-0.0129	
	C(11)	0.0000	
	C(12)	0.0144	
Ring B	C(13)	-0.0054	$\chi^2 = 0.28; p = 0.9$
	C(14)	0.0044	
	C(15)	-0.0032	
	C(16)	0.0031	
	C(17)	-0.0041	
	C(18)	0.0052	
Ring C	C(19)	-0.0106	$\chi^2 = 0.66; p = 0.8$
	C(20)	0.0101	
	C(21)	-0.0042	
	C(22)	-0.0012	
	C(23)	0.0007	
	C(24)	0.0053	

TABLE 3.10 (contd.)

Deviation of As from the plane through C(7), C(13), C(19) = 0.7252

<u>Dihedral angles</u>	Rings A - B	=	52.15°
	A - C	=	68.27°
	B - C	=	63.01°

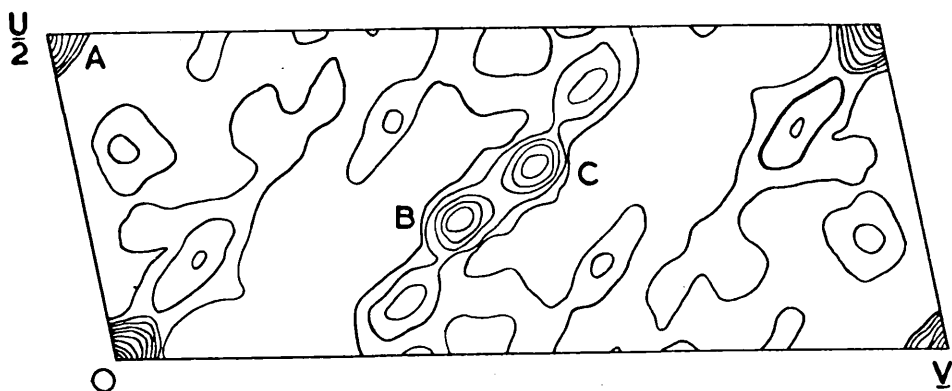
	<u>Atoms</u>	<u>Deviation</u>	
Ring A	C(7)	0.0311	
	C(8)	0.0298	
	C(9)	0.0000	
	C(10)	-0.0433	$\chi^2 = 10.20 ; p = 0.04$
	C(11)	-0.0109	
	C(12)	0.0426	
	As	-0.0433	
Ring B	C(13)	-0.0118	
	C(14)	0.0005	
	C(15)	-0.0016	$\chi^2 = 0.63 ; p = 0.94$
	C(16)	0.0074	
	C(17)	-0.0026	
	C(18)	0.0014	
	As	0.0068	

TABLE 3.10 (contd.)

	<u>Atoms</u>	<u>Deviation</u>
Ring C	C(19)	0.0124
	C(20)	0.0236
	C(21)	-0.0098
	C(22)	-0.0162
	C(23)	-0.0047
	C(24)	0.0185
	As	-0.0237

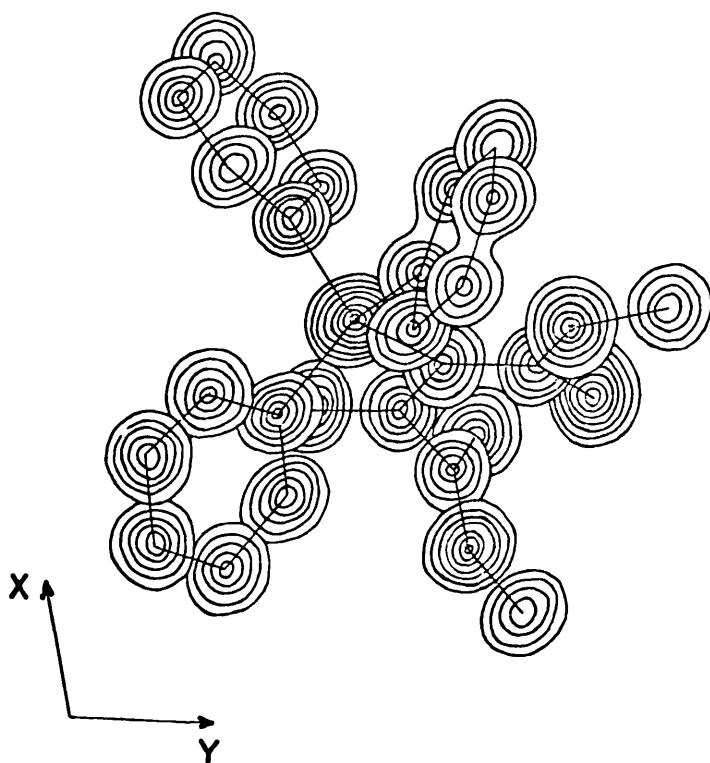
$$\chi^2 = 5.01; \quad p = 0.25$$

Planarity is probable when  $p > 0.01$



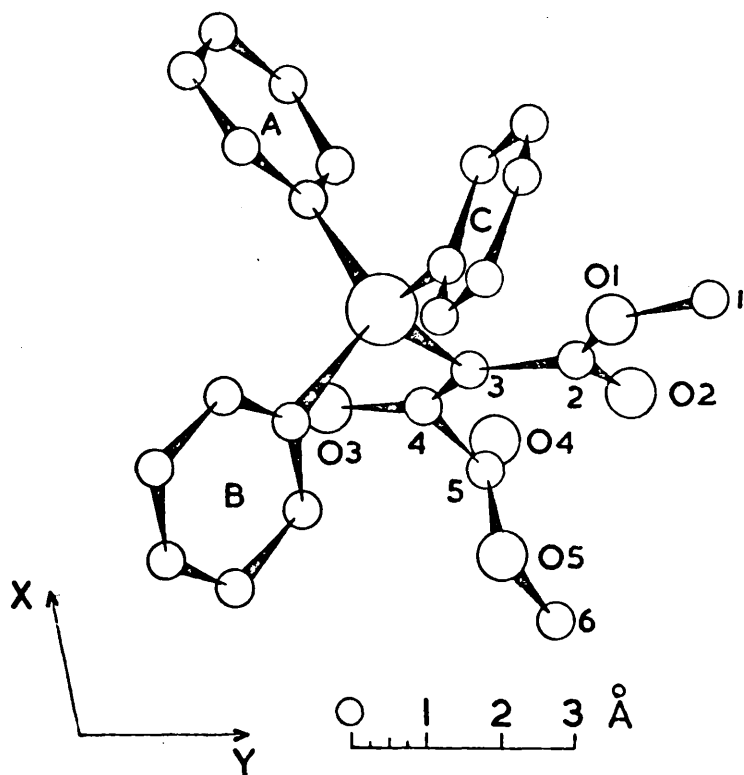
Triphenylarsine Adduct

Figure 3.1 The Patterson projection along the c axis. The contour scale is arbitrary.



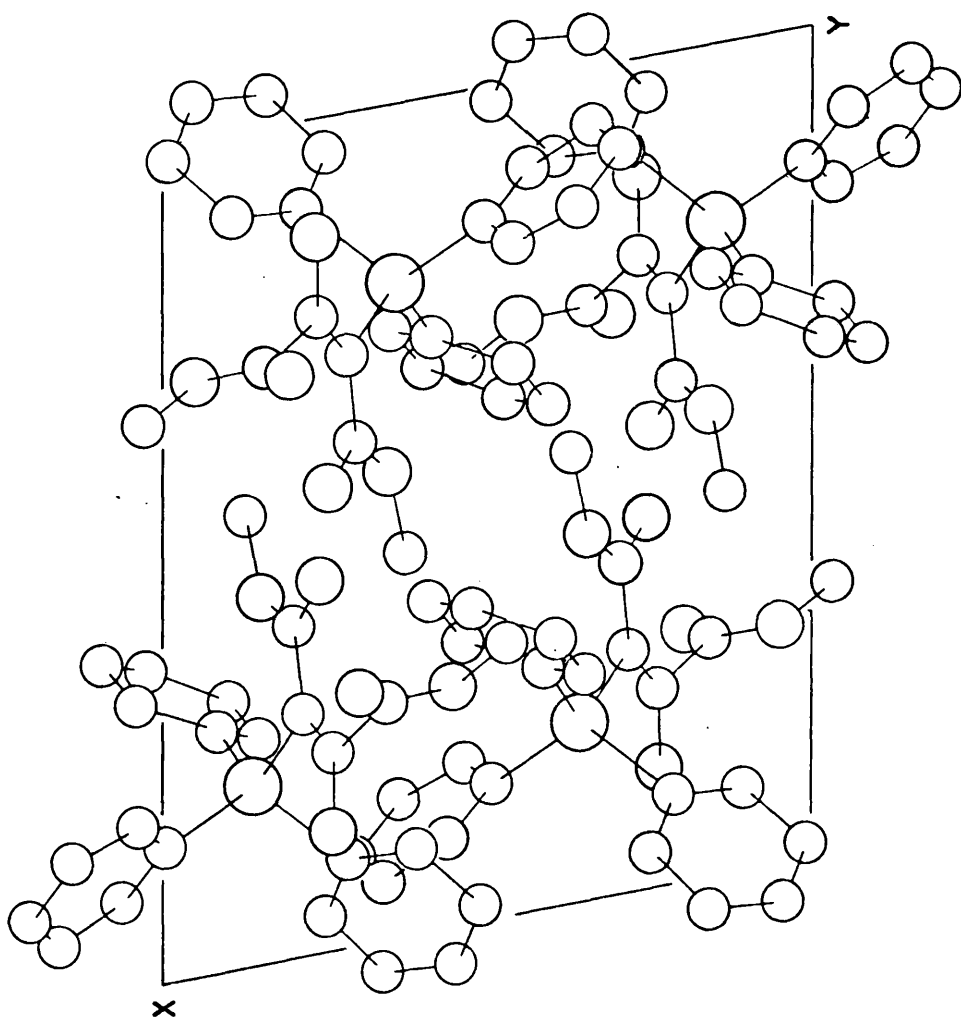
Triphenylarsine Adduct

Figure 3.2 The final three-dimensional electron density distribution shown by means of superimposed contour sections drawn parallel to the (001) plane. The contour intervals are  $1e/\text{\AA}^3$  except around the arsenic atom where they are  $4e/\text{\AA}^3$



Triphenylarsine Adduct

Figure 3.3 The atomic arrangement corresponding to figure 3.2



Triphenylarsine Adduct

Figure 3.4 The crystal packing viewed along the c axis



### 3.5      RESULTS OF THE ANALYSIS

The final electron density distribution is shown in figure 3.2 by means of superimposed contour sections parallel to the (001) plane and the atomic arrangement corresponding to this is shown in figure 3.3.

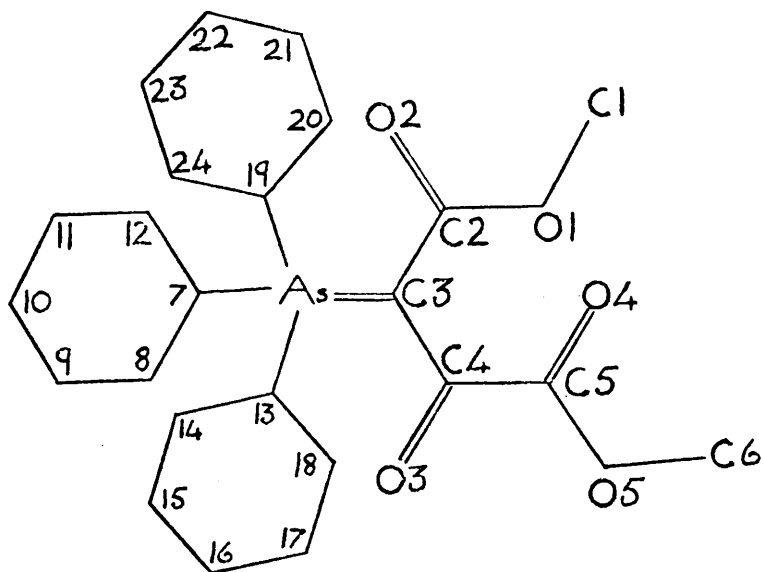
The final atomic coordinates are listed in table 3.3 and the standard deviations of these coordinates are listed in table 3.4. The anisotropic temperature factors for the atoms are shown in table 3.5.

Table 3.2 lists the observed and calculated structure amplitudes and the final R factor is 14.3%.

The interatomic bond lengths are compiled in table 3.6 while the interbond angles are shown in table 3.7. The average standard deviation of a bond is of the order of  $0.015\overset{\circ}{\text{\AA}}$ , those involving the heavy atom being  $0.01\overset{\circ}{\text{\AA}}$ .

Tables 3.8 and 3.9 show some intramolecular contacts and some intermolecular contacts respectively while the results of some mean plane calculations are compiled in table 3.10.

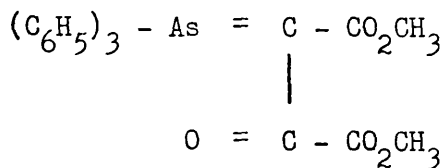
The packing of the molecules in the crystal is shown in figure 3.4 as viewed along the C axis and the numbering system employed is portrayed in structure IV.



IV

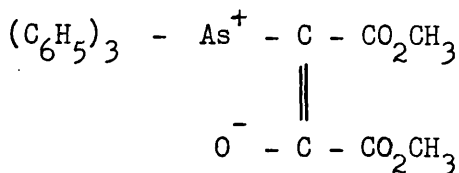
### 3.6 DISCUSSION

This analysis establishes that the adduct of acetylene dicarboxylic methyl ester and triphenylarsine has the structure V.



V

Chemical evidence suggests that the structure is a resonance hybrid of V and structure VI.



VI

Resonance between the  $\alpha$ -arsenyldene ketone, V, and the arsonium enolate zwitterion, VI, is indicated by analogy with

corresponding triphenyl phosphorylidene ketones which have high dipole moments and may be alkylated to give enol ethers. Furthermore the colour and stability of these compounds is attributed to the distribution of a negative charge on the molecule (6). The results of this x-ray analysis tend to support the postulate of resonance. The three bonds concerned in the resonance are C(3) - As, C(3) - C(4) and C(4) - O(3). The C(3) - As bond length is  $1.89\overset{\circ}{\text{\AA}}$  which is longer than the pure double bond expected for structure V. A single bond of the type in structure VI is an arsenic to  $sp^2$  carbon atom. Similar bonds are found in arsenobenzene,  $(C_6H_5)_6As_6$ , which has an average C - As bond of  $1.96\overset{\circ}{\text{\AA}}$  (7) and is considerably longer than the C(3) - As bond. Also the three bonds As - C(7), As - C(13) and As - C(19) are  $1.91\overset{\circ}{\text{\AA}}$ ,  $1.93\overset{\circ}{\text{\AA}}$  and  $1.92\overset{\circ}{\text{\AA}}$  respectively, all being longer than the As - C(3) similar bond of structure VI. The C(3) - C(4) bond length is  $1.40\overset{\circ}{\text{\AA}}$  and is intermediate between the values for a single bond in a similar environment and a pure double bond. Finally, the C(4) - O(3) double bond is significantly longer than a normal bond of this type. A comparison of expected bond lengths for structures V and VI and the observed values is shown below in tabular form.

<u>Bond</u>	<u>Expected Lengths (<math>\overset{\circ}{\text{\AA}}</math>)</u>		<u>Observed Lengths (<math>\overset{\circ}{\text{\AA}}</math>)</u>
	V	VI	
C(3) - As	1.78*	1.96	1.89
C(3) - C(4)	1.46	1.33	1.40
C(4) - O(3)	1.21	1.35	1.26

\* Calculated from Pauling's covalent radii (8)

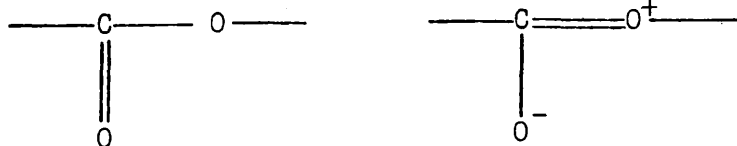
As shown in the table the observed lengths are all intermediate between the expected lengths for pure V and pure VI and thus support a resonance structure. In addition to the above considerations the close approach of the arsenic atom to the O(3) atom, ( $2.94\overset{\circ}{\text{\AA}}$ ), suggests electrostatic attraction between the two and hence the presence of a partial charge on both atoms.

Inspection of the anisotropic temperature factors, table 3.5, show a number of atoms of the side chain exhibiting extensive anisotropic motion. This is particularly marked in the group of atoms comprising C(1), C(2), O(1) and O(2). This is probably due to the fact that this group is relatively unhindered and is free to oscillate as can be seen from figure 3.3. Although a number of atoms of the side chain show anisotropic motion, this is not the case with O(3) which in addition has a lower value for the average of  $U(11)$ ,  $U(22)$  and  $U(33)$  than the other oxygen atoms. A possible explanation is that the proposed electrostatic attraction between O(3) and As restricts motion of this type.

The bonded distances agree well with the expected values. The average of the carbon-carbon bonds in the three benzene rings is  $1.40\overset{\circ}{\text{\AA}}$  compared with the expected value of  $1.397\overset{\circ}{\text{\AA}}$  reported by Sutton et al. for benzene (9). Of the four bonds round the arsenic, three are expected to be carbon-arsenic single bonds and the average length is  $1.92\overset{\circ}{\text{\AA}}$  which is considerably shorter than the carbon arsenic length in  $(\text{CH}_3)_3\text{As}$  listed by Sutton et al. as  $1.98\overset{\circ}{\text{\AA}}$ . The latter is however a pure  $\text{C}(\text{sp}^3)$  - As bond whereas the triphenylarsine adduct bonds are  $\text{C}(\text{sp}^2)$  - As and would therefore be expected to have

shorter lengths. In the structure of arsenobenzene,  $(\text{AsC}_6\text{H}_5)_6$ , the average carbon-arsenic bond length is  $1.96\text{\AA}$  (7) although this value is the mean of quite a large range, the individual values being  $1.92\text{\AA}$ ,  $1.97\text{\AA}$  and  $2.01\text{\AA}$ .

There are three types of carbon-oxygen bond. The average  $\text{sp}^3$  carbon-oxygen bond is  $1.45\text{\AA}$  in favourable agreement with the value of  $1.43\text{\AA}$  gained by summing Pauling's covalent radii (8) for carbon and oxygen and the value of  $1.46\text{\AA}$  found in other methyl esters such as methyl acetate (9). The average carbon-oxygen double bond is  $1.23\text{\AA}$  excluding the double bond at C(4). This value is consistent with the predicted value of  $1.21\text{\AA}$  due to Pauling (8) and the values of  $1.23\text{\AA}$  reported by Sutton et al. (9) for the similar bonds of carboxylic acids and esters. The  $\text{sp}^2$  carbon-oxygen single bonds are C(2) - O(1) and C(5) - O(5) and the average value of these is  $1.31\text{\AA}$ . This is somewhat shorter than the value of  $1.36\text{\AA}$  quoted by Sutton et al. (9) for the similar bond of methyl acetate and the value of  $1.35\text{\AA}$  in bromohydroisophoto- $\alpha$ -santonin lactone acetate (10). A possible explanation of this is found in the resonance structures which have been proposed for carboxylic acids, lactones and esters, as illustrated below.





This effect has been found in bromogeigerin acetate (11) and clerodin bromolactone (12) where the carbon-oxygen single bonds adjacent to a carbon-oxygen double bond are shorter than those not adjacent to a double bond. A short bond is also found in O-chlorobenzoic acid (13) and in oxalic acid dihydrate (14), the values being  $1.295\overset{\circ}{\text{\AA}}$  and  $1.285\overset{\circ}{\text{\AA}}$  respectively. Moreover the angles at C(2) and C(5) are not all close to  $120^\circ$ . The C - C - OMe angle is appreciably lower than  $120^\circ$  while the two remaining angles at each of the atoms C(2) and C(5) exceed  $120^\circ$ . This conforms to the results for bromogeigerin acetate (11) and clerodin bromolactone (12) and to the pattern observed in carboxylic acids where the C - C - OH angles approaches the tetrahedral value and the angles C - C - O and O - C - OH exceed  $120^\circ$  (13, 14, 15).

The remaining bonds are the  $\text{sp}^2$  carbon- $\text{sp}^2$  carbon single bonds, C(2) - C(3) and C(4) - C(5). These bonds vary in length somewhat but are in reasonable agreement with the value of  $1.46\overset{\circ}{\text{\AA}}$  -  $1.47\overset{\circ}{\text{\AA}}$  for similar bonds e.g. the central bond of butadiene (9).

The average angle in the benzene rings is  $120^\circ$  in full agreement with the anticipated values. Around the  $\text{sp}^2$  hybridised carbon atoms other than those of the benzene rings, the average angle is the expected  $120^\circ$ . There are six angles round the arsenic atom and the average of these is  $110^\circ$  which is consistent with the tetrahedral angle of  $109^\circ 28'$  which should be approximately true of an arsenic atom surrounded by four bonds.

In table 3.10 the results of mean plane calculations are shown.

The equation of the best plane through the atoms of ring A is;

$$0.5632 x + 0.7475 y + 0.3522 z - 4.1173 = 0$$

The value of  $\chi^2$ , calculated as in section 2.5 of this thesis, is 1.89 corresponding to a p value of 0.6. It is therefore probable that the ring A is planar. The values for rings B and C are also indicative of planar rings and the equations of the mean planes through them are;

$$0.2122 x - 0.6278 y - 0.7489 z + 1.6010 = 0$$

$$\text{and } -0.3732 x + 0.8937 y - 0.2490 z - 0.3408 = 0$$

respectively. The non planarity of substituent groups with the benzene ring has been found to be not an unusual occurrence. For example in 2-chloro-4-nitroaniline (16) the nitro group is out of the plane of the benzene ring and in some substituted benzoic acids (17) the non-planarity of substituent groups with the benzene ring has been found. Consequently the mean plane calculations were repeated including the arsenic atom in each of the calculations. The results are summarised in table 3.10 and show that the arsenic atom is in the plane of each of the benzene rings. The final point of interest as far as planarity is concerned is whether the three bonds to the arsenic atom from the benzene rings are in the one plane. These are in fact non-planar, the equation of the plane through C(7), C(13) and C(19) being;

$$0.3576x - 0.6782y + 0.6420z - 1.3933 = 0$$

and the distance of the arsenic atom from this plane is  $0.73\text{\AA}$ . Table 3.10 also shows the dihedral angles between the benzene rings.

Intramolecular non-bonded distances are normal. The average 1-4 distance in the benzene rings is  $2.79\text{\AA}$  as compared to the value of  $2.80\text{\AA}$  expected from a perfectly regular benzene ring. The closest approaches the benzene rings make with one another are C(17) ---- C(19),  $3.01\text{\AA}$  and C(7) ---- C(13),  $3.18\text{\AA}$ . All the intermolecular distances (table 3.9) are normal corresponding to van der Waals interactions and the packing of the molecules in the crystal is shown in figure 3.4 as viewed along the C axis.

REFERENCES

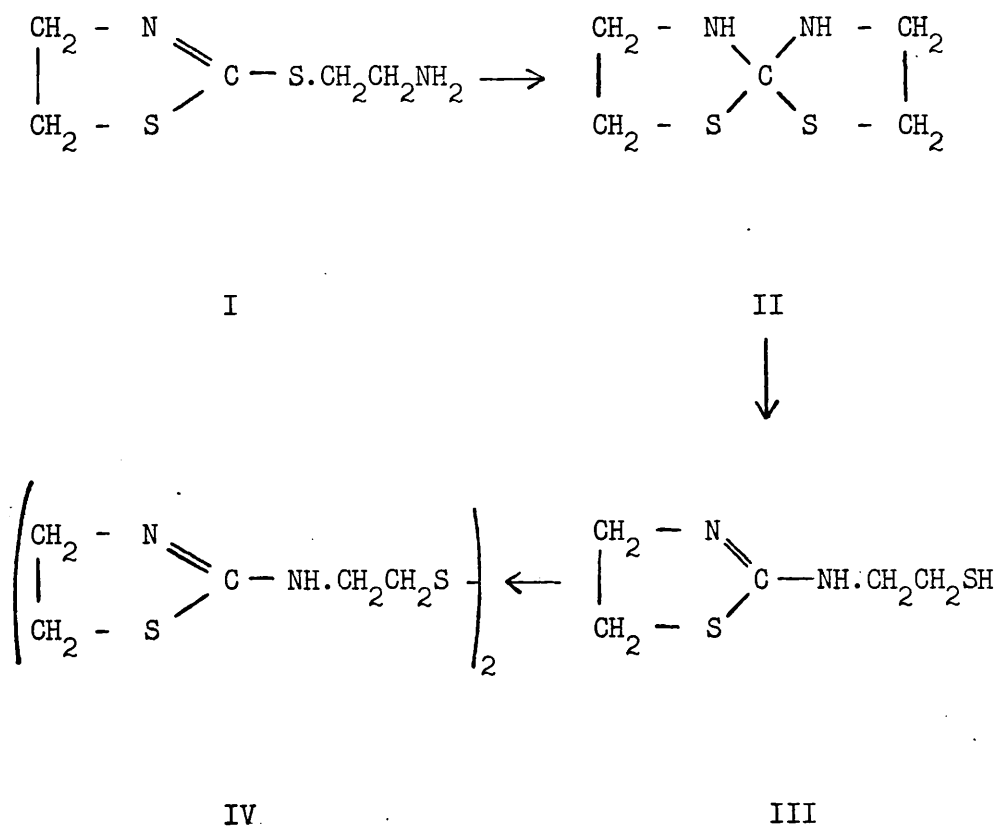
1. J.B. Hendrickson, R.E. Spenger, J.J. Sims, Tetrahedron Letters, 14, 477, 1961.
2. J.B. Hendrickson, R.E. Spenger, J.J. Sims, Tetrahedron, 19, 707, 1963.
3. J.M. Robertson, J. Sci. Instr., 20, 175, 1943.
4. M.G. Rossmann, Acta Cryst., 2, 819, 1956.
5. A.D. Booth, Fourier Technique in x-ray Organic Structure Analysis, p.63, Cambridge University Press, 1948.
6. F. Ramirez, S. Dershowitz, J. Org. Chem., 22, 41, 1957.
7. K. Hedberg, E.W. Hughes, J. Waser, Acta Cryst., 14, 369, 1961.
8. L. Pauling, The Nature of the Chemical Bond, 3rd ed., Cornell University Press, New York, 1960.
9. L.E. Sutton et al., Tables of Interatomic Distances, The Chemical Society, London, 1958.
10. J.D.M. Asher, Ph.D. Thesis, University of Glasgow, 1963.
11. Miss J.A. Hamilton, A.T. McPhail, G.A. Sim, J. Chem. Soc. 708, 1962.
12. I.C. Paul, G.A. Sim, T. Hamor, J.M. Robertson, J. Chem. Soc., 4133, 1962.
13. G. Ferguson, G.A. Sim, Acta Cryst., 14, 1262, 1961.
14. F.R. Ahmed, D.W.J. Cruickshank, Acta Cryst., 6, 385, 1953.
15. W. Cochran, Acta Cryst., 6, 260, 1953.
16. A.T. McPhail, Ph.D. Thesis, University of Glasgow, 1963.
17. G. Ferguson, Ph.D. Thesis, University of Glasgow, 1961.

C H A P T E R    4

THE CRYSTAL STRUCTURE ANALYSIS OF  
2-(2-HYDROXYETHYLAMINO)-2-THIAZOLINE

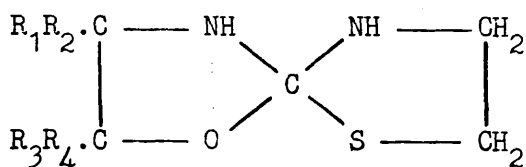
## 4.1 INTRODUCTION

Research by Clapp, Long and Hasselstrom (1) into the rearrangement of 2-(2-aminoethylthio)-2-thiazoline, I, to 2-(2-mercaptoethylamino)-2-thiazoline, III, by transthiazolination suggested that the reaction proceeded through a bicyclic intermediate II.



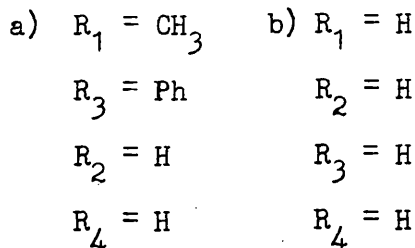
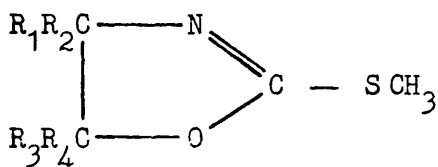
The product, III, was then readily oxidised in air to the disulphide IV.

In this hypothetical intermediate, the similarity of the rings permitted the formation of only a single product. Clapp et al. extended the rearrangement to 2-oxazoline derivatives in which unsymmetrical intermediates of type V might be involved.



V

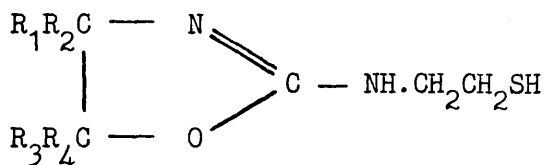
The preparation of the disulphides starting from 2-methylthio-4-methyl-5-phenyl-2-oxazoline, VIa, proceeded through a stable intermediate having unusual properties.



VI

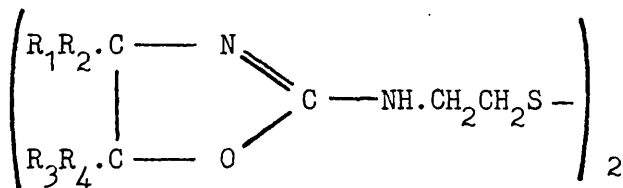


The molecular formula,  $C_{12}H_{16}N_2OS$ , corresponded to the mercapto compound, VIIa, rather than the expected disulphide. However chemical and spectroscopic evidence failed to show the presence of a mercapto group and the similar reaction of 2-methylthio-2-oxazoline, VIb, and 2-mercaptoethylamine hydrochloride produced a compound with properties analogous to the unusual intermediate of the original reaction.



VII

Both the former and the latter products afforded the disulphide, VIII, with aeration in alkaline solution.



VIII

The infra red spectrum of the  $C_{12}H_{16}N_2OS$  intermediate failed to show the presence of hydroxyl, primary amino or mercapto groups while its nuclear magnetic resonance spectrum presented evidence that two secondary amino groups were contained in the molecule. These observations suggested the possibility that a stable compound corresponding to the proposed bicyclic intermediate, V, in the trans-oxazolination had been obtained, in this case by ring closure of the 2-mercaptoethylamino side chain. Evidence to the contrary was a strong band in the infra red spectrum at  $6.16\mu$  which indicated that an unsaturated heterocyclic ring was present although this would require formation of a side chain having a terminal hydroxyl, amino or mercapto group. In order to establish the structure an x-ray crystal structure analysis was undertaken on the unsubstituted compound.

## 4.2 EXPERIMENTAL DATA

CRYSTAL DATA

Formula	=	$C_5H_{10}N_2OS$
Molecular weight	=	146
D (calc.)	=	1.340
D (obs.)	=	1.335 (by flotation method)

The crystal is monoclinic with

$$\begin{aligned}
 a &= 7.82 \pm 0.02 \text{ \AA} \\
 b &= 10.04 \pm 0.03 \text{ \AA} \\
 c &= 9.20 \pm 0.02 \text{ \AA} \\
 \beta &= 90^\circ
 \end{aligned}$$

$$\text{Volume of cell} = 722.3 \text{ \AA}^3$$

$$\text{Number of molecules per cell} = 4$$

Absent spectra - h0l when l is odd

0k0 when k is odd

Space group is  $P2_1/C$ .

$$\mu, \text{ absorption coefficient for x-rays } (\lambda = 1.5418) = 32.5 \text{ cm}^{-1}$$

$$\text{Total electrons per cell} = F(000) = 312$$

$$\sum f^2(\text{light atoms}) = 352 (\sin \theta = 0)$$

$$\sum f^2(\text{heavy atom}) = 256 (\sin \theta = 0)$$

$$r = \left[ \sum f^2(\text{heavy atom}) / \sum f^2(\text{light atoms}) \right]^{\frac{1}{2}} = 0.85$$

INTENSITY DATA

Rotation, oscillation, Weissenberg and precession photographs were taken with copper  $K\alpha$  ( $\lambda = 1.5418\text{\AA}$ ) and molybdenum  $K\alpha$  ( $\lambda = 0.7107\text{\AA}$ ) radiation. The cell dimensions were obtained from rotation and precession photographs. The monoclinic space group  $P2_1/C$  was determined uniquely from the systematic absences. In all 830 intensity data were obtained from Weissenberg photographs of a small crystal rotated about the  $C$  axis. Intensities were estimated visually by comparison with a calibrated intensity strip and correlation between reflections of very high intensity and low intensity was accomplished by using the multiple film technique (2). The intensities were converted to structure amplitudes by the usual corrections for Lorentz, polarisation and rotation factors (3). Since a small crystal was used, no correction for absorption was applied. The various layers were put on the same scale throughout the refinement by comparison with the calculated structure amplitudes so that  $\sum_k |F_o| = \sum |F_c|$  for each layer. The overall isotropic temperature factor was estimated from experience of values for similar organic structures.

#### 4.3 THE STRUCTURE ANALYSIS

## LOCATION OF THE HEAVY ATOM BY PATTERSON METHODS

In the space group  $P2_1/c$  with  $b$  unique, the vectors expected on a Patterson map due to the heavy atom are the general vectors  $(2x, 2y, 2z)$  and the Harker peaks  $(2x, \frac{1}{2}, \frac{1}{2} + 2z)$  and  $(0, \frac{1}{2} - 2y, \frac{1}{2})$  due to the two-fold screw axis and the  $C$  glide respectively.

A two dimensional Patterson projection,  $P(U,V)$ , was calculated. This could not be solved unambiguously and a sharpened Patterson function was calculated using the sharpening function  $M(s) = (\frac{1}{\hat{f}})^2 \exp. (-\frac{2.5 \sin 2\theta}{\lambda^2})$ . Figure 4.1 shows the two-dimensional distribution in which the expected vectors are  $(2x, 2y)$ ,  $(0, 2y - \frac{1}{2})$  and a double peak at  $(2x, -\frac{1}{2})$ . Peak A appears to be the most likely for assigning to the vector  $(0, 2y - \frac{1}{2})$  and peak B for  $(2x, \frac{1}{2})$  although in the latter case peak C is also possible. Assuming A and B to be the vectors in question, the general vector  $(2x, 2y)$  should appear between the crosses on the map corresponding to D which as can be seen is far from satisfactory. The combination of peaks A and C is equally inconclusive, corresponding to the point E.

In order to try and clarify matters a Harker section at  $V = \frac{1}{2}$  was computed and examined for the sulphur-sulphur and any other vectors present. The vector looked for was  $(2x, 2z - \frac{1}{2})$  and the section is shown in figure 4.2. The most prominent peak in a general position is labelled F and corresponds to the expected vector. The value of  $2x$  obtained from this agrees well with the

assignment of peak B of the two dimensional Patterson to the vector  $(2x, \frac{1}{2})$ . From these maps the likely position of the general sulphur-sulphur vector  $(2x, 2y, 2z)$  was calculated and hence the position of the sulphur atom in the unit cell. However due to the fact that this general peak could not be located in the Patterson projection a three dimensional Patterson map was calculated. Any remaining ambiguity was removed by the presence in this map of a large peak in a general position corresponding to the values already found. The position of the sulphur atom was therefore confirmed as

$$x = 0.39792 a$$

$$y = 0.04300 b$$

$$z = 0.21458 c$$



## SOLUTION AND REFINEMENT OF THE STRUCTURE

A first round of structure factor calculations with the sulphur atom alone produced a somewhat high discrepancy,  $R$ , of 57%. An electron density map computed using the observed structure amplitudes and the calculated phases. This was drawn out on glass sheets corresponding to sections of electron density at equal intervals of the  $z$  direction. Apart from the sulphur atom only two other peaks could be assigned to atomic positions and these were far from unambiguous. The value of  $r \left( = \left[ \frac{\sum f^2(h)}{\sum f^2(l)} \right]^{\frac{1}{2}} \right)$  is 0.85 and since the crystal is centrosymmetric and the scattering curves of the light atoms fall off more rapidly than the sulphur atom, a reasonable chance of solving the structure would be expected. The first Fourier map contained a very large number of obviously spurious peaks. Nevertheless a second round of structures factors were calculated including the sulphur atom and the only two plausible peaks mentioned above. Only a few planes were calculated and inspection of these showed further calculation to be pointless. Since the structure contains a very small number of atoms, it was decided to attempt a solution from the three dimensional Patterson map. The region around the origin of the Patterson map should contain peaks corresponding to vectors between the sulphur atom and the light atoms. A study of a volume bounded by a sphere of radius from the origin of about  $3\text{\AA}$  produced nothing conclusive and the Harker section at  $V = \frac{1}{2}$  was then studied. This contains a number of peaks and these were taken

in turn and positions in real space worked out assuming them to be vectors between the lighter atoms. These positions were checked against the general vector for these lighter atoms in Patterson space and the sulphur to light atom vectors of the Patterson. Unfortunately no useful information could be gleaned from this approach.

The next line of attack was to utilise the three dimensional Patterson distribution in a Patterson superposition. As discussed in section 1.9 there are a number of variations of the superposition method. The mechanics of the Beevers and Robertson method used in the present work may be summarised as follows. If  $x, y, z$  are the coordinates of light atoms and  $x(S_1), y(S_1), z(S_1)$  the coordinates of the sulphur atom, then vectors between light atoms and the sulphur atoms will be represented by peaks in Patterson space with coordinates;

$$\begin{aligned} X &= x - x(S_1) \\ Y &= y - y(S_1) \quad \dots\dots\dots (A) \\ Z &= z - z(S_1) \end{aligned}$$

A grid was drawn out representing the  $(x, y)$  plane and on it were marked the positions of the sulphur atoms  $S_1$  and  $S_2$  corresponding to the equivalent positions  $(x, y, z)$  and  $(-x, -y, -z)$ . Now  $z(S_1) = 51.5/240$  and  $z(S_2) = -51.5/240$  and applying the relationship (A) to the two sulphur atoms  $S_1$  and  $S_2$ ;

$$z = Z + z(S_1) = Z + 51.5/240$$

and  $z = Z - z(S_2) = Z - 51.5/240$

Thus in preparing the superposition function for say  $z = 15/240$  the Patterson section  $Z = 66.5/240$  was placed with its origin at  $S_1$  and and section  $Z = -36.5/240$  placed with its origin at  $S_2$ . The superposition function was plotted for each point of the grid at which peaks occurred. In this way the superposition function was prepared in sections at  $z = -3/240$  to  $z = 69/240$  at intervals of  $z = 6/240$ . In like manner the superposition function was built up with respect to the sulphur atoms  $S_3$  and  $S_4$  corresponding to the equivalent positions  $(x, \frac{1}{2} - y, \frac{1}{2} + z)$  and  $(-x, \frac{1}{2} + y, \frac{1}{2} - z)$ . The two functions were then directly superposed to give the complete function. There was then a fairly high probability that the peaks in the distribution correspond to genuine atomic sites of the light atoms. In fact, the skeleton of the molecule could be seen quite unambiguously, a very few spurious peaks being present. No conclusion were drawn from Patterson peak heights and a structure of point atoms was built up. At this stage, therefore, the distribution of scattering matter was known although chemical types apart from sulphur were not.

Atomic coordinates were estimated for the various atoms and were included in the calculation of a second set of structure factors. All the atoms except sulphur were included as carbons and the discrepancy,  $R$ , was calculated to be 28.6%. A second electron density map, was calculated and new coordinates obtained. Chemical

types were now allocated from consideration of the peak heights backed up by chemical knowledge and calculation of bond lengths and angles. This additional information was included in the third structure factor calculations, the R factor being 20.7%. The structure factors were scaled such that  $\sum_k |F_o| = \sum |F_c|$  for each layer and  $F_o$  and  $F_c$  maps were computed. Coordinates were calculated from both maps and back shift corrections due to Booth (4) were applied. The overall isotropic temperature factor  $U = 0.056$  gave fairly good agreement between the peak heights of the two maps and was therefore not adjusted. The fourth round of structure factor calculations produced an R factor of 19.3% and it was decided to continue refinement using the method of least squares.

Block diagonal anisotropic least squares refinement (5) was carried out on the D.E.U.C.E. computer. Five least squares cycles reduced the R factor from 19.3% to 16.3%. With the structure factors from the fifth cycle of least squares a difference map was computed and all the hydrogen atom positions determined. These were included in the sixth structure factor least squares calculation, producing an R factor of 15.4% which with one further cycle dropped only 0.1% to 15.3%. The weighting scheme used throughout this part of the refinement was as follows;

$$\begin{aligned} \text{if } |F_o| &\leq p, & \sqrt{w(hkl)} &= 1 \\ \text{if } |F_o| &> p, & \sqrt{w(hkl)} &= p/|F_o| \end{aligned}$$

where  $p$  = average structure amplitude.

With the advent of the K.D.F.9 computer and the facility for full matrix least squares refinement it was decided to attempt further refinement. The restrictions on the number of parameters which could be refined in this way necessitated the use of isotropic temperature factors. The average of  $U_{11}$ ,  $U_{22}$  and  $U_{33}$  from the last least squares cycle on D.E.U.C.E. was taken to be  $U(\text{iso})$ . With the previous weighting scheme, the first cycle produced an R factor of 15.4%. One of the facilities of the K.D.F.9 computer is that a weighting analysis is produced and this showed that the weighting scheme could be improved, the following being adopted.

$$\sqrt{w} = 1/(p_1 + F + p_2 F^2)^{\frac{1}{2}}$$

where  $p_1 = 2F (\text{min.})$  and  $p_2 = 2/F(\text{max.})$

The R factor was reduced to 13.5% in a second cycle and a third cycle using block diagonal anisotropic least squares gave R as 12.7%. The various layers were scaled such that  $\sum k |F_o| = \sum |F_c|$  and one final cycle of least squares calculated, the final R factor being 12.1%. The course of the structure analysis is shown in table 4.1.

#### 4.4 TABLES AND FIGURES

TABLE 4.1

Course of the analysis

		<u>R%</u>	<u>Atoms included</u>	
Structure factors	1	57.0	S	
	2	28.6	C <sub>5</sub> N <sub>2</sub> OS	
	3	20.7	"	
	4	19.3	"	
				<u><math>\sum w\Delta^2</math></u>
Anisotropic S.F.L.S.	1	19.3	"	548
	2	17.5	"	489
	3	16.8	"	456
	4	16.5	"	441
	5	16.3	"	432
	6	15.4	C <sub>5</sub> H <sub>10</sub> N <sub>2</sub> OS	387
	7	15.3	"	384
Isotropic full matrix S.F.L.S.	1	15.4	"	317
	2	13.5	"	215
Anisotropic S.F.L.S.	3	12.7	"	194
	4	12.1	"	191

### Measured and Calculated Values of the Structure Factors

[illegible]



TABLE 4.3

2-(2-Hydroxyethylamino)-2-thiazoline

Final Atomic Coordinates

<u>Atom</u>	<u>x/a</u>	<u>y/b</u>	<u>z/c</u>
C(1)	0.17967	0.09419	0.19805
C(2)	0.29998	0.21210	0.01502
C(3)	0.45528	0.18311	0.10513
C(4)	-0.11525	0.06741	0.27756
C(5)	-0.16199	0.18868	0.36439
N(1)	0.14775	0.18340	0.10132
N(2)	0.06551	0.03621	0.28676
O(1)	-0.13304	0.17352	0.51553
S(1)	0.39488	0.04193	0.21412

TABLE 4.4

2-(2-Hydroxyethylamino)-2-thiazoline

Standard Deviations of the Atomic Coordinates ( $\text{\AA}$ )

<u>Atom</u>	<u>x</u>	<u>y</u>	<u>z</u>
C(1)	.008	.008	.008
C(2)	.009	.010	.012
C(3)	.009	.010	.012
C(4)	.009	.008	.011
C(5)	.008	.008	.010
N(1)	.006	.007	.008
N(2)	.007	.007	.008
O(1)	.005	.006	.007
S(1)	.002	.002	.003

TABLE 4.5

2-(2-Hydroxyethylamino)2-thiazolineAnisotropic Temperature Factors  $\times 10^4$ 

<u>Atom</u>	<u>U(11)</u>	<u>U(22)</u>	<u>U(33)</u>	<u>2U(23)</u>	<u>2U(31)</u>	<u>2U(12)</u>
C(1)	471	521	425	-52	46	107
C(2)	556	897	604	507	244	121
C(3)	474	874	744	113	69	-105
C(4)	484	636	517	229	-111	-234
C(5)	417	676	560	265	20	90
N(1)	463	651	574	152	22	51
N(2)	476	533	708	167	67	-83
O(1)	550	654	453	7	-32	-120
S(1)	497	632	651	34	9	282

TABLE 4.6a

2-(2-Hydroxyethylamino)-2-thiazolineIntramolecular Bonded Distances (Å)

<u>Bond</u>	<u>Distance</u>	<u>Bond</u>	<u>Distance</u>
C(1) N(1)	1.29	C(3) S(1)	1.80
C(1) N(2)	1.34	C(4) N(2)	1.45
C(1) S(1)	1.77	C(4) C(5)	1.51
C(2) N(1)	1.46	C(5) O(1)	1.42
C(2) C(3)	1.50		

TABLE 4.6b

<u>Bond</u>	<u>Distance</u>	<u>Bond</u>	<u>Distance</u>
C(2) H(1)	1.00	C(4) H(6)	0.94
C(2) H(2)	1.13	C(5) H(7)	0.81
C(3) H(3)	0.97	C(5) H(8)	0.90
C(3) H(4)	1.21	N(2) H(9)	1.23
C(4) H(5)	1.01	O(1) H(10)	1.04

TABLE 4.7

2-(2-Hydroxyethylamino)-2-thiazoline

Interbond Angles

S(1)	C(1)	N(1)	117°
S(1)	C(1)	N(2)	117
N(1)	C(1)	N(2)	126
C(3)	C(2)	N(1)	109
C(2)	C(3)	S(1)	104
C(5)	C(4)	N(2)	112
C(4)	C(5)	O(1)	113
C(1)	N(1)	C(2)	111
C(1)	N(2)	C(4)	121
C(1)	S(1)	C(3)	88

TABLE 4.8

2-( 2 - Hydroxyethylamino)-2-thiazolineIntramolecular non-bonded contacts ( $\text{\AA}$ )

C(1)	C(2)	2.27	C(4)	O(1)	2.44
C(1)	C(4)	2.44	C(4)	N(1)	2.87
C(1)	C(3)	2.49	C(5)	N(2)	2.45
C(1)	C(5)	3.23	C(5)	N(1)	3.43
C(1)	O(1)	3.90	N(1)	N(2)	2.35
C(2)	S(1)	2.61	N(1)	S(1)	2.61
C(2)	N(2)	3.57	N(2)	S(1)	2.66
C(3)	N(1)	2.41	O(1)	N(2)	2.96
C(3)	N(2)	3.78			

TABLE 4.9

2-(2-Hydroxyethylamino)-2-thiazolineSome Intermolecular Distances ( $\overset{\circ}{\text{A}}$ )

C(3)	C(5) <sub>1</sub>	3.82	O(1)	C(5) <sub>2</sub>	3.50
C(3)	C(4) <sub>1</sub>	3.89	O(1)	C(4) <sub>2</sub>	3.55
S(1)	C(4) <sub>1</sub>	3.88	O(1)	C(2) <sub>2</sub>	3.58
C(1)	C(2) <sub>2</sub>	3.63	S(1)	C(2) <sub>2</sub>	3.78
C(1)	O(1) <sub>2</sub>	3.77	C(1)	O(1) <sub>3</sub>	3.78
C(3)	O(1) <sub>2</sub>	3.62	C(4)	O(1) <sub>3</sub>	3.64
C(3)	C(5) <sub>2</sub>	3.94	C(5)	N(2) <sub>3</sub>	3.99
C(5)	N(1) <sub>2</sub>	3.50	N(2)	O(1) <sub>3</sub>	2.83
C(5)	C(2) <sub>2</sub>	3.99	S(1)	O(1) <sub>3</sub>	3.88
N(2)	C(2) <sub>2</sub>	3.76	S(1)	C(3) <sub>3</sub>	3.89
O(1)	N(1) <sub>2</sub>	2.74	N(2)	C(5) <sub>4</sub>	3.83

The subscripts refer to the following equivalent positions

1.  $x, y, z$
2.  $x, \frac{1}{2} - y, \frac{1}{2} + z$
3.  $-x, -y, -z$
4.  $-x, -\frac{1}{2} + y, -\frac{1}{2} - z$

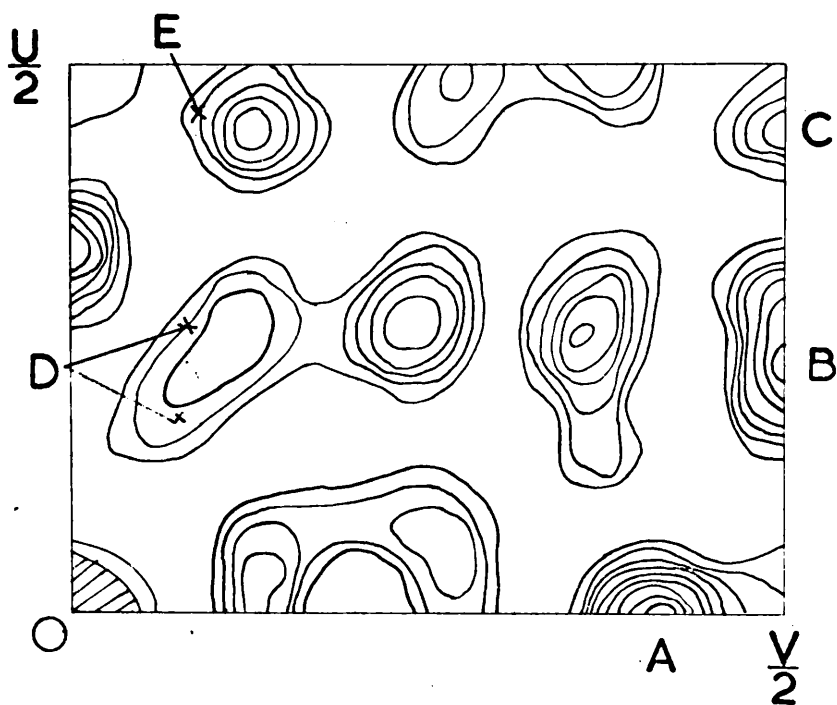
TABLE 4.10

2-(2-Hydroxyethylamino)-2-thiazolineMean Plane Calculations

	<u>Dist. from plane (<math>\text{\AA}</math>)</u>	<u><math>\sum \Delta^2 \times 10^8</math></u>	<u><math>\sigma^2 \times 10^6</math></u>	<u><math>\chi^2</math></u>	<u>p%</u>
C(1)	-0.0034				
N(1)	0.0013	1575	42	.375	60
N(2)	0.0009				
S(1)	0.0013				

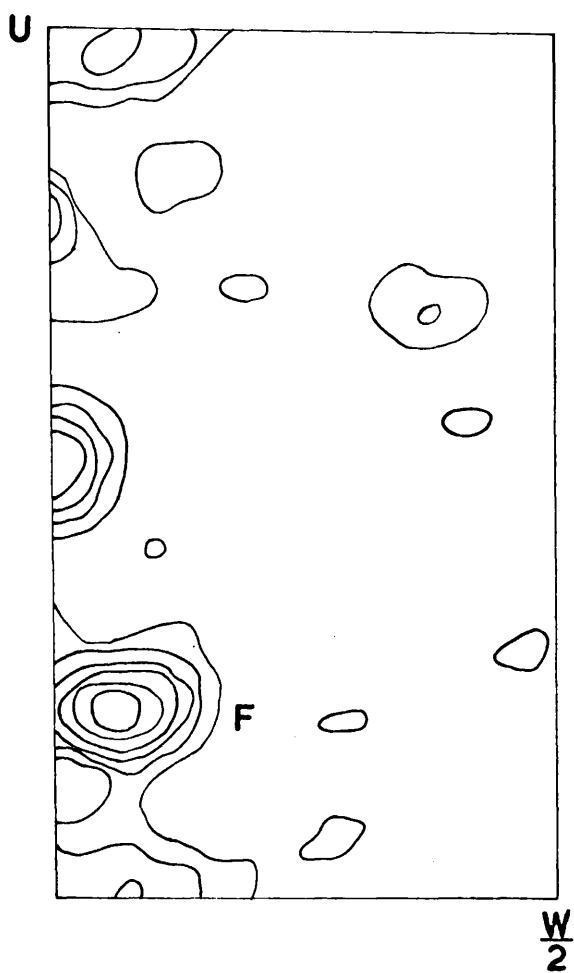
Planarity is probable when  $p > 1\%$





2-(2-Hydroxyethylamino)-2-thiazoline

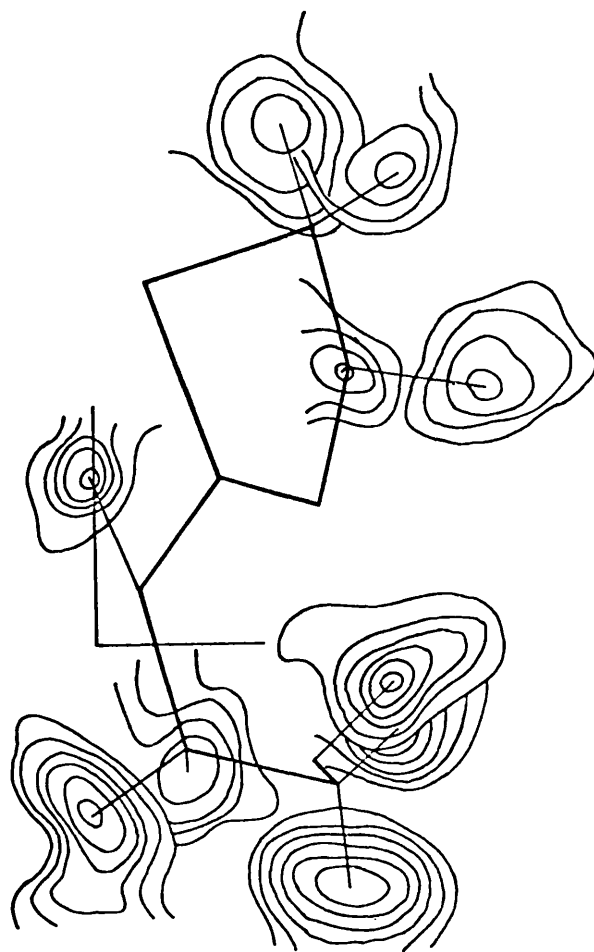
Figure 4.1    The sharpened Patterson projection along the c-axis.    The contour scale is arbitrary.



2-(2-Hydroxyethylamino)-2-thiazoline

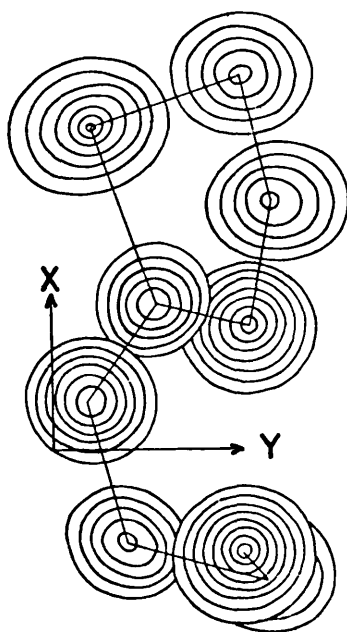
Figure 4.2

The Harker section at  $V = \frac{1}{2}$ .  
The contour scale is arbitrary.



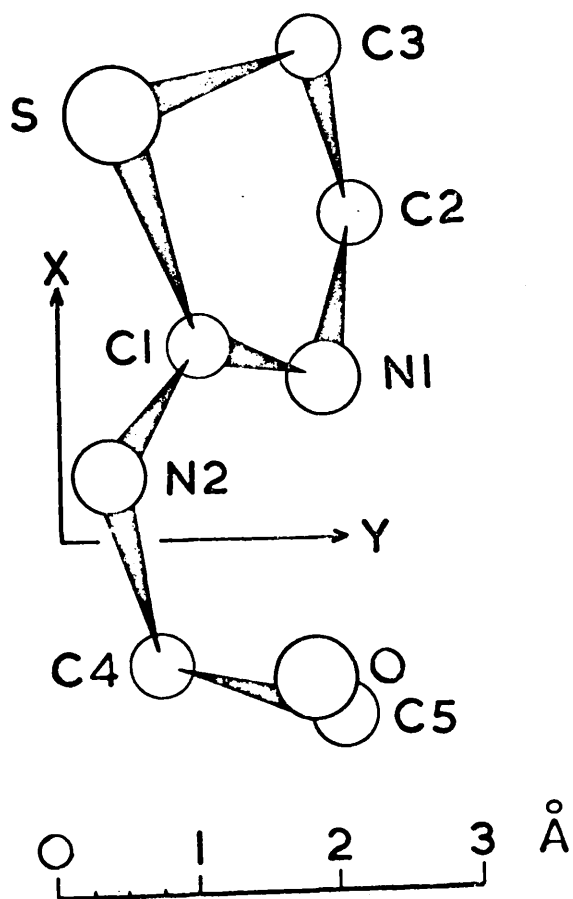
2-(2-Hydroxyethylamino)-2-thiazoline

Figure 4.3 The superimposed contour map of the difference Fourier showing the hydrogen atom positions



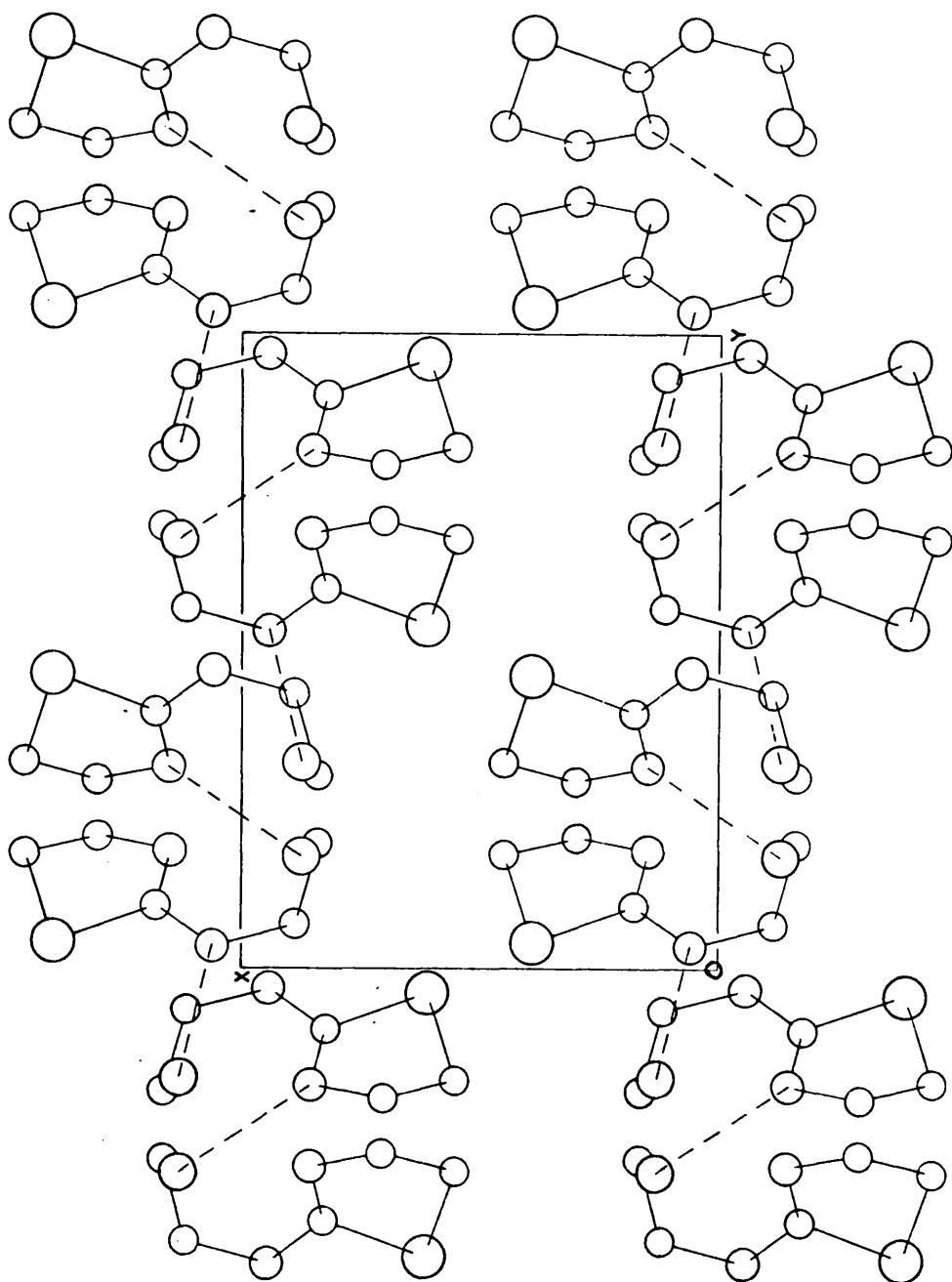
2-(2-Hydroxyethylamino)-2-thiazoline

Figure 4.4 The final three-dimensional electron density distribution shown by means of superimposed contour sections drawn parallel to the (001) plane. The contour intervals are  $1e/A^3$  except around the sulphur atom where the contour intervals are  $2e/A^3$



2-(2-Hydroxyethylamino)-2-thiazoline

Figure 4.5 The atomic arrangement corresponding to figure 4.4



2-(2-Hydroxyethylamino)-2-thiazoline

Figure 4.6 The packing of the crystal viewed along the c axis

#### 4.5 RESULTS OF THE ANALYSIS

The final electron density distribution is shown in figure 4.4 by means of superimposed contour sections parallel to the (001) plane. The atomic arrangement corresponding to this is shown in figure 4.5. As mentioned in section 4.3, the hydrogen atom positions were located by a difference Fourier map which afforded fairly well resolved peaks in sensible positions. In figure 4.3 a superimposed contour map parallel to the (001) plane, shows the peaks due to hydrogen atoms with the skeleton of the heterocyclic molecule superposed on this.

The final atomic coordinates are listed in table 4.3 and the final anisotropic temperature factors in table 4.5. The standard deviations of the coordinates were calculated as in section 2.5 and are shown in table 4.4.

Table 4.2 lists the observed and calculated structure amplitudes, the final R factor over the 830 observed structure amplitudes is 12.1%. The intramolecular bonded distances are listed in table 4.6a. Those concerning the hydrogen atoms are listed in 4.6b, while the interbond angles are shown in table 4.7. The standard deviations of bonds and angles were calculated as in section 2.5. The average standard deviation of a bond is of the order of  $0.010 - 0.015^{\circ}\text{\AA}$  and the standard deviation of a valency angle is about  $1^{\circ}$ .

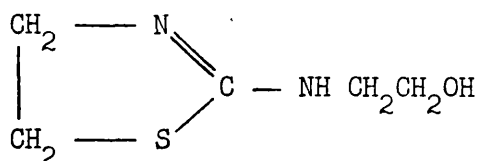
Table 4.8 lists some intramolecular non bonded contacts while table 4.9 shows some intermolecular distances. In table 4.10 the results of some mean plane calculations are shown.

The packing of the molecules, viewed along the  $\underline{c}$  axis, is shown in figure 4.6.

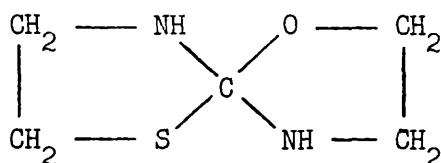


## 4.6 DISCUSSION

The results of this analysis unambiguously establish that the molecule is a substituted thiazoline, 2-(2-hydroxyethylamino)-2-thiazoline with constitution VIII and not the spiro compound, V, originally proposed.



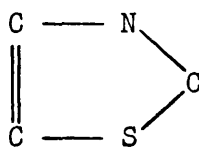
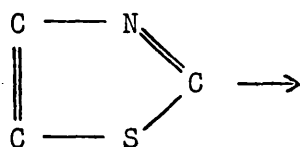
VIII



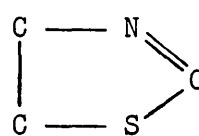
V

The position of the double bond was determined by consideration of the bond lengths and valency angles (tables 4.6 and 4.7) and was confirmed by the location of the hydrogen atoms.

Thiazolines are reduced thiazoles, such reduction producing either the 2- or 4- thiazolines as shown below.

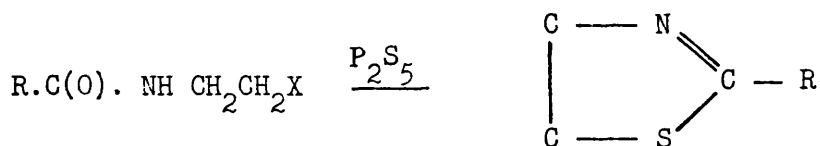


4-thiazole



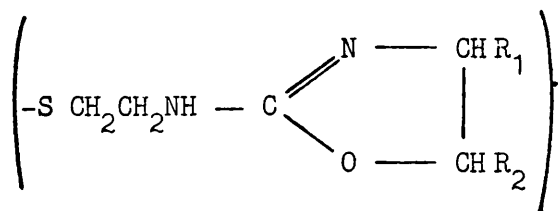
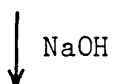
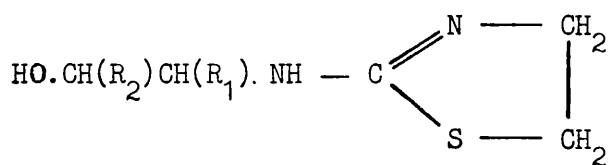
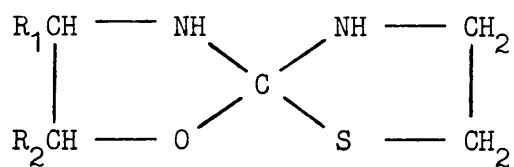
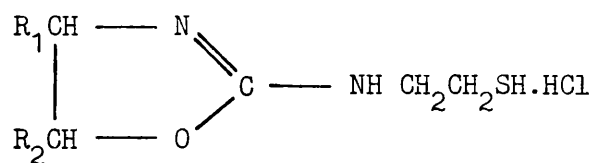
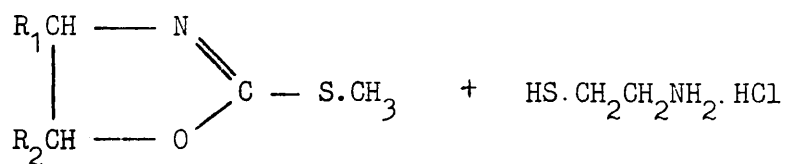
2-thiazoline

The fully reduced tetrahydro-thiazoles are well-known and have been extensively studied since the presence of one in penicillin was demonstrated (6). The 2-thiazolines are more common than the 4-thiazolines and have been fairly well studied, the known examples representing relatively few types. In the 4- or 5- position, substituent groups are usually confined to alkyl groups or phenyl groups. A number of 2-alkyl or 2-aryl derivatives are known but of the compounds with functional groups in the 2-position, the amino and mercapto compounds are the most common. A number of methods of preparing 2-thiazolines are available, the most general and widely applicable synthesis being represented by the equation shown below (7).



X = halogen, hydroxyl, acetoxy

The preparation by Clapp et al. (8) is a new route to the 2-thiazolines, and the spiro compound first proposed as the structure is expected to be an intermediate in the reaction shown overleaf.



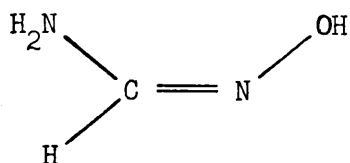
The molecular dimensions are summarised in tables 4.6 and 4.7. The valency angle of  $88^\circ$  at the sulphur atom is similar to the values found in other sulphur containing heterocyclic molecules, for example  $91^\circ$  in thiophene (9) and in thiofthen (10) but is distinctly smaller than the valency angles exhibited by oxygen in heterocyclic molecules, for example  $106.5^\circ$  in furan (11). It is not clear whether the smaller valency angles associated with sulphur are to be attributed to the use of pure p orbitals or to hybrid orbitals involving some d orbital contribution (12). The use of two 3p orbitals for bonding with unshared electron pairs in the 3s and 3p orbitals is favoured by Price and Oae (13) for the  $\text{H}_2\text{S}$  molecule which has an angle of  $92^\circ$ , and for simple alkyl sulphides which contain angles close to  $90^\circ$ . However, Burg (14) points out that if pure p bonding is the case, the angle in the  $\text{H}_2\text{S}$  molecule would be noticeably less than  $90^\circ$  due to repulsion by the other pair of p electrons. The fact that it is greater than  $90^\circ$  suggests some amount of hybridisation which would have the effect of widening the angle. This is supported by the fact that  $\text{CH}_3\text{SH}$  has a wider angle and  $(\text{CH}_3)_2\text{S}$  a still wider angle. The explanation of these facts being that the  $\text{sp}^3$  orbital of the carbon atom is wider than a hydrogen 1s electron cloud and therefore for maximum overlap the sulphur hybrid must have a wider angle. This can be achieved by greater use of the 3s orbital and in the series  $\text{H}_2\text{S}$  ( $92.5^\circ$ ),  $\text{CH}_3\text{SH}$  ( $100^\circ$ ) and  $(\text{CH}_3)_2\text{S}$  ( $105^\circ$ ) the full  $\text{sp}^3$  hybridisation angle is approached. It would appear that steric repulsion might account for the angle widening in the series but

Gibbs (15) has argued that this alone would not be sufficient. In the thiazoline derivative the angle has been narrowed from  $90^\circ$ . However, it is not significantly smaller than  $90^\circ$  and the strained ring in which the sulphur finds itself will most probably cause a decrease in the valency angle. The average valency angle in the thiazoline ring is  $106^\circ$  in good agreement with values reported for other non planar five-membered rings (16). The average angle round the  $sp^2$  hybridised carbon atom is the expected  $120^\circ$ .

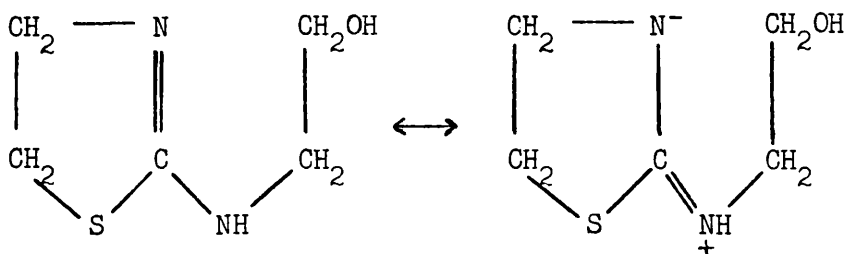
The C(3) - S bond is  $1.80\overset{\circ}{\text{\AA}}$ , in good agreement with the normal single bond distance of  $1.81\overset{\circ}{\text{\AA}}$  reported by Sutton et al. (17), while the C(1) - S bond is  $0.03\overset{\circ}{\text{\AA}}$  shorter, consistent with the accepted difference of  $0.02 - 0.03\overset{\circ}{\text{\AA}}$  between the radii of  $sp^2$  and  $sp^3$  hybridised carbon atoms. The carbon-carbon single bonds are slightly short although not significantly so. The  $sp^3$  carbon-oxygen distance of  $1.42\overset{\circ}{\text{\AA}}$  and the  $sp^3$  carbon-nitrogen distances of  $1.45\overset{\circ}{\text{\AA}}$  and  $1.46\overset{\circ}{\text{\AA}}$  do not differ significantly from the accepted values (17). The average hydrogen to light atom distance is  $1.02\overset{\circ}{\text{\AA}}$  as compared to the accepted value of  $1.0 - 1.1\overset{\circ}{\text{\AA}}$ .

The two bonds which as yet have not been considered, are the C(1) - N(1) and C(1) - N(2) bonds. Pauling's standard value for a carbon-nitrogen double bond length is  $1.27\overset{\circ}{\text{\AA}}$  (18) and the length of the similar bond in pyrazoline hydrochloride is  $1.255\overset{\circ}{\text{\AA}}$  (19). These are both somewhat shorter than the C(1) - N(1) bond length of  $1.29\overset{\circ}{\text{\AA}}$ . The C(1) - N(2) bond distance of  $1.34\overset{\circ}{\text{\AA}}$  is significantly lower than expected and the effect of the  $sp^2$  hybridisation of C(1) is not

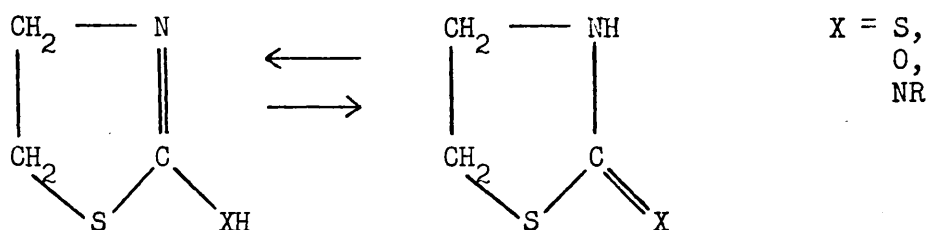
sufficient in itself to explain this. The two bonds are in more favourable agreement with the bonds of  $\text{H}_2\text{H}.\text{CH}.\text{NOH}$  the structure of which is shown below.



The C -  $\text{NH}_2$  bond of  $1.33\text{\AA}$  and the C -  $\text{NOH}$  bond of  $1.30\text{\AA}$  are explained on the basis of resonance involving these bonds (17). A possible explanation therefore of the C(1) - N(1) and C(1) - N(2) bonds may be found in the formulation of the structure as the resonance hybrid shown below.

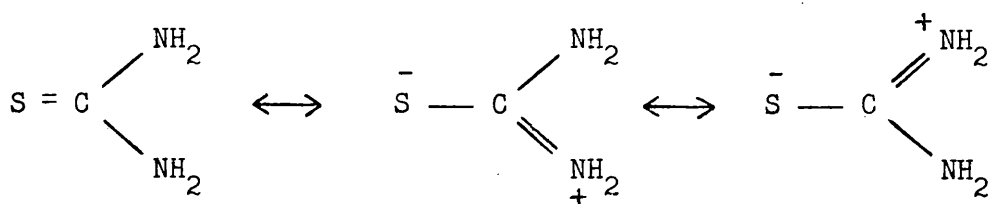


In solution the 2-thiazolines having a functional group such as mercapto, hydroxy or amino can undergo tautomerism of the type illustrated below (20).



This is known from the reactions of the 2-thiazolines which occasionally behave as the thiazolidine. Many of the reactions of 2-amino-2-thiazolines reflect their tautomerism and frequently two series of derivatives are found. In crystals containing molecules with the grouping  $\text{N} - \text{C}(\text{X}) - \text{N}$  resonance has been proposed where  $\text{X} = \text{S}$  or  $\text{O}$ . In the crystal of thiourea (21, 22, 23) the arrangement of the molecules in the crystal is such that the sulphur atom of one molecule is next to the amino end of another and is described by Bernal (24) as a typical polar molecular structure. The length of the  $\text{C} - \text{N}$  bond is too short for a pure single bond and the  $\text{C} - \text{S}$  distance is too long for a pure double bond. The structure can be explained as a resonance hybrid thus;





In the thiazoline derivative it is therefore possible that electron shift in the  $\text{N} - \text{C}(3) = \text{N}$  grouping takes place, the resultant resonance explaining the  $\text{C}(1) - \text{N}(1)$  and  $\text{C}(1) - \text{N}(2)$  bond lengths.

The molecules are held together in the crystal by van der Waals forces and a system of hydrogen bonds. The intramolecular non-bonded contact of  $2.96\text{\AA}$  between  $\text{O}(1)$  and  $\text{N}(2)$  is not attributed to hydrogen bonding since the angle  $\text{C}(5) - \text{O}(1) \cdots \text{N}(2)$  is only  $56^\circ$  and no hydrogen atom was located between the two atoms. The hydrogens bonds are shown by means of broken lines in figure 4.6. There are two distinct types of bond, an  $\text{NH} \cdots \text{O}$  and an  $\text{OH} \cdots \text{N}$ . The  $\text{NH} \cdots \text{O}$  distance is  $2.83\text{\AA}$  which is comparable to the similar hydrogen bond in, for example, L-serine phosphate (25) with values of 2.79, 2.83 and  $2.96\text{\AA}$ . The great importance of the  $\text{NH} \cdots \text{O}$  hydrogen bond is in fixing the molecular configurations of proteins and for this reason considerable attention has been focused on bonds of this type. Tables of lengths have been compiled and a histogram of the distribution of bond distances constructed by Pimentel and McClellan (26). By this means the average length of this type was

found to be  $2.93\text{\AA}$  for amides (average of 35 examples) and  $3.04\text{\AA}$  for amines (average of 30 examples). The length of the OH --- N hydrogen bond is  $2.74\text{\AA}$  which is comparable to the OH --- N distances in theophylline monohydrate (27) of  $2.87\text{\AA}$  and dimethylglyoxime (28) of  $2.83\text{\AA}$ . Pimentel and McClellan list various bond lengths of this type and show that the average length of such bonds is  $2.80\text{\AA}$  (26).

The properties which distinguish a hydrogen bond are not restricted to considerations of length alone although plausible distances are a fair indication of hydrogen bonding. That the distances involved are always less than the sum of the normal van der Waals distances has been observed by Donohue (29) who has stressed that molecules in crystals tend to arrange themselves so that the maximum number of hydrogen bonds are formed. In a more recent review (30), Fuller has supported this view and has laid down the further criterion that for hydrogen bond formation the angle between the hydrogen bond direction and the angle between the donor and the hydrogen should be within  $25^\circ$ . In the ideal case the donor atom, the hydrogen atom and the acceptor atom would form a straight line. Inspection of figures 4.3 and 4.6 show that the direction of the donor atom to the hydrogen atom is in the general direction of the acceptor atom. In the OH --- N hydrogen bond, the angle C(5) - O --- N is  $111^\circ$  while the angle C(5) - O - H(10) is  $118^\circ$  and in the NH --- O bond, the similar angles are  $95^\circ$  and  $112^\circ$ . Both cases are therefore within Fuller's criterion of  $25^\circ$ .

Although many hydrogen bonds are accepted as such on the basis of the positions of the donor and acceptor atoms, the final proof is furnished by the location of the hydrogen atom. This also serves to prove whether the hydrogen atom is symmetrically placed in the hydrogen bond. One of the proposed theories of hydrogen bond formation was that two resonance structures were involved in which the hydrogen atom alternates between the donor and acceptor atoms and is therefore symmetrically placed between the two. Many structure analyses have shown that the hydrogen atom is not centrally placed and the resonance theory has therefore been discounted. Occasionally hydrogen bonded oxygen atoms are related by a crystallographic centre of symmetry suggesting a central hydrogen atom. However, it is more probable that the hydrogen atoms are randomly distributed throughout the crystal at one or other of the two positions available to it. One such structure is potassium hydrogen bisphenylacetate (31), the infra-red spectrum of which does not support the idea of a symmetrical bond. In the structure of 2-(2-hydroxyethylamino)-2-thiazoline the bond lengths involving the hydrogen atoms are not too reliable but are accurate enough to suggest that the hydrogen atoms are as expected unsymmetrically placed in the hydrogen bonds. The involvement of the hydroxyl group in hydrogen bonding is probably the reason why the infra-red spectrum failed to indicate this terminal hydroxyl and gave initially a confusing picture of the structure.

REFERENCES

1. R.C. Clapp, L. Long Jr., T. Hasselstrom, J. Org. Chem., 26, 1666, 1961.
2. J.M. Robertson, J. Sci. Instr., 20, 175, 1943.
3. G. Tunnel, Am. Min., 24, 448, 1939.
4. A.D. Booth, Nature, 156, 51, 1945.
5. J.S. Rollet, Computing Methods and the Phase Problem in X-ray Crystal Analysis, paper 7, Oxford, Pergamon Press, 1961.
6. D. Crowfoot, C.W. Bunn, B.W. Rogers-Low, A. Turner-Jones, Chemistry of Penicillin, Princeton University Press, 1949.
7. R.C. Elderfield ed., Heterocyclic Compds., Vol. 1, p.76; Vol. 5, p.679, Wiley & Sons, New York.
8. R.C. Clapp, L. Long Jr., T. Hasselstrom, J. Org. Chem, 29, 2172, 1964.
9. V. Schomaker, L. Pauling, J. Amer. Chem. Soc., 61 1769. 1939.
10. E.G. Cox, R.J.J.M. Gillot, G.A. Jeffrey, Acta Cryst., 2, 356, 1949.
11. B. Bak. D. Christensen, W.B. Dixon, L. Hansen-Nygaard, J.R. Anderson, M. Schottlander, J. Mol. Spectr., 2, 124, 1962.
12. C.A. Burrus, W. Gordy, Phys. Rev., 91 313, 1953.
13. C.C. Price, S. Oae, Sulfur Bonding, The Ronald Press Co., New York, 1962.
14. A.B. Burg, Organic Sulfur Compounds, ed. by N. Kharasch, Ch.4, Pergamon Press, 1961.
15. J.H. Gibbs, J. Phys. Chem., 59, 644, 1955.

16. See inter alia; I.C. Paul, G.A. Sim, T.A. Hamor, J.M. Robertson, J. Chem. Soc., 4133, 1962.
17. L.E. Sutton et al., Tables of Interatomic Distances, the Chemical Society, London, 1958.
18. L. Pauling, The nature of the Chemical Bond, 3rd ed., Cornell University Press, New York, 1960.
19. M. Nardelli, G. Fava, Acta Cryst., 15, 214, 1962.
20. R.C. Elderfield ed., Heterocyclic Compds., Vol. 5, J. Wiley & Sons, New York.
21. S.B. Hendricks, J. Amer. Chem. Soc., 50, 2455, 1928.
22. R.W.G. Wyckoff, Z. Cryst., 81, 102, 1932.
23. R.W.G. Wyckoff, R.B. Corey, ibid, 89, 426, 1934.
24. J.D. Bernal, W.A. Wooster, Ann. Rep. of Chem. Soc., 303, 1929.
25. G.H. McCallum, J.M. Robertson, G.A. Sim, Nature, 184, 1863, 1959.
26. G.C. Pimentel, A.L. McClellan, The Hydrogen Bond, W.H. Freeman & Co., 1960.
27. D.J. Sutor, Acta Cryst., 11, 83, 1958.
28. L.L. Merrit Jr., E. Lanterman, Acta Cryst., 5, 811, 1952.
29. J. Donohue, J. Phys. Chem., 56, 502, 1952.
30. C.S. Fuller, J. Phys. Chem., 63, 1705, 1959.
31. J.C. Speakman, J. Chem. Soc., 3357, 1949.