THE CRYSTAL STRUCTURES OF SOME BIOLOGICALLY SIGNIFICANT MOLECULES

A Thesis submitted for the degree of

Doctor of Philosophy in the

University of London

by

Edward Hough, B.Tech.

Chemical Crystallography Laboratory,

Imperial College,

London, SW7 2AY.

December, 1974.

Summary

This Thesis describes the determination of the structures of four organic compounds using single-crystal X-ray diffraction methods. Three of the compounds are biologically active natural products and the fourth is an intermediate product in the synthesis of pharmaceutically important isatin compounds. Two of the structures were solved by the heavy-atom method, the third by the symbolic-addition method and the fourth by the direct calculation of phases from the intensity differences which arise from the presence of a fluorescent atom in the molecule. An "extended FORTRAN" computer program which was written for this purpose is briefly described in the Appendix.

With the exception of the program mentioned above, all the calculations were carried out using the "X-ray 63" and "X-ray 67" crystallographic computer program systems and the data were measured on a Siemens AED four-circle diffractometer. Since there is no chemical or biological connection between the four compounds, they are presented separately.

Fusicoccin aglycone mercuribromide ($C_{21}H_{33}O_5HgBr$) is tetragonal with unit-cell dimensions $\underline{a}$ = 19.550(5), $\underline{c}$ = 13.375(4)Å, space group $\underline{I4}$ and Z = 8. Least-squares refinement, using 2552 independent reflections, gave a final R of 0.051. The crystal lattice was found to contain benzene molecules occupying disordered sites about the fourfold axes. The mechanism of the mercuration reaction and the decay of the compound under X-radiation are discussed.

Methyl 3,3,4-Trichloro-5-methoxy-indolenine-2-carboxylate ($C_{11}H_8NO_3C\ell$) is orthorhombic with unit-cell dimensions $\underline{a}$ = 22.450(5), $\underline{b}$ = 13.292(3), $\underline{c}$ = 8.558(2)Å, space group $\underline{Pna2_1}$ and Z = 8. The six chlorine atoms in the asymmetric unit were located by the symbolic-addition method and the complete structure was refined to an R factor of 0.036 for the 1373 observed reflections. The compound is compared with a series of substances which contain the indole ring system in different oxidation states.

The 2-bromo-4,6-dinitrophenolate of the erythrina alkaloid erythristemine ($C_{26}H_{28}N_3O_9Br$) is monoclinic with unit-cell dimensions $\underline{a}$ = 7.998(3), $\underline{b}$ = 16.159(5), $\underline{c}$ = 10.550(5)Å, β = 97.48(1)°, space group $\underline{P2_1}$ and $\underline{Z}$ = 2. Least-squares refinement, using 2036 observed reflections gave a final $\underline{R}$ of 0.067. The structure is compared with those of the three other determined erythrina alkaloids and its function as a curare agent is discussed. The nature of the bonding between the alkaloid moiety and the 2-bromo-4,6-dinitrophenolate moiety is considered. The crystal packing is partly attributable to charge-transfer effects.

Ferri-mycobactin P ($C_{47}H_{72}N_5O_{10}Fe$) is monoclinic with unit-cell dimensions $\underline{a}$ = 16.061(2), $\underline{b}$ = 12.193(1), $\underline{c}$ = 13.282(3)Å, β = 101.39(2)°, space group $\underline{P2_1}$ and $\underline{Z}$ = 2. The structure was solved using anomalous dispersion effects and has so far been refined to an R value of 0.121 for 2409 observed reflections. The role of the compound as an iron transport agent in the Mycobacteria is discussed in the light of the crystal structure and a mechanism for the insertion and reductive removal of the iron atom from the compound is proposed.

Acknowledgements

The author is most grateful to his supervisor Professor D. Rogers for his invaluable assistance and encouragement, to his colleagues past and present in the Chemical Crystallography Laboratory at Imperial College and to friends in the University of Oslo.

The financial support of Associated Lead Manufacturers Ltd., Greenford, Middlesex is aknowledged with thanks.

CONTENTS

	Page
CHAPTER 1.	
The Crystal Structure of Fusicoccin Aglycone Mercuri-bromide	6
CHAPTER 2	
The Crystal Structure of Methyl-3,3,4-Trichloro-5-methoxy--Indolenine-2-Carboxylate	69
CHAPTER 3	
The Crystal Structure of Erythristemine 2-Bromo-4,6-Dinitro-Phenolate	126
CHAPTER 4	
The Crystal Structure of Ferri-Mycobactin P	185
APPENDIX	
Phase Determination Using Anomalous Dispersion Effects	267

CHAPTER 1

The crystal structure of Fusicoccin aglycone mercuri-bromide.

INTRODUCTION

The fungus Fusicoccum Amygdali was first observed by Delacroix[1] in 1905 during an investigation of almond tree diseases in Provence. Four years later Arnaud[2] reported the fungus as the causative agent of "*la maladie ponctuee des rameaux des armandiers*". Possibly because it was known only in the lower Rhone valley no further study of the species was carried out until shortly before the second World War, when it was described[3] in connection with a new plant disease, "peach tree canker" in New Jersey U.S.A. Although, at that time, no absolute proof of the identity of the fungi could be established, it was considered that the diseases were caused by the same fungus. In the course of the next fifteen years peach canker spread widely through the New England States and caused severe damage to peach harvests. The identity of the American peach tree fungus with the French species was confirmed in 1953 by Guba[4], although peach tree canker had not been observed in France at that time.

In 1956, however, Grosclaude[5] published the results of an extensive study of peach canker in which he reported the widespread occurrence of Fusicoccum Amygdali on peach trees in France. He pointed out that the symptoms of peach canker disease are readily confused with those of two other fungal diseases and may also be masked by the effects of secondary infection by Cytospora Leucostoma, a fungus which he found to be commonly associated with Fusicoccum Amygdali. Grosclaude described the biology of the fungus in detail and succeeded in culturing it "*in vitro*". In a study of the toxin produced by the fungus, he also made the important discovery that culture filtrates exhibited strong phytotoxin activity, thus paving the way to the successful isolation and structure determination of the toxin, Fusicoccin A.

THE ISOLATION AND CHEMICAL CHARACTERISATION OF FUSICOCCIN A

Eight years after Grosclaude's work, Ballio, Chain *et al.*[1] published details of the isolation and preliminary characterisation of the most abundant phytotoxic component in culture filtrates of Fusicoccum Amygdali and named the compound "Fusicoccin A".

The compound was obtained by a series of solvent extractions and finally crystallised from ethyl acetate as a white solid, M. Pt. 150-152°C. Elemental analysis combined with molecular weight determinations by crystallography (M = 733.5 ± 2%) and isothermal distillation (M = 697.3 ± 4%) suggested a formula $C_{38}H_{58}O_{13}$ (M = 722.9). The compound was optically active and was found to contain two acetoxy groups by NMR spectroscopy. One methoxy group was found by direct analysis, and the presence of hydroxyl and ethylenic functions and the absence of aromaticity were deduced from infra-red spectra. Catalytic hydrogenation resulted in the uptake of one mole of hydrogen to give dihydro-fusicoccin-A, and acid hydrolysis yielded one mole of glucose.

The mass spectrum of the compound was described as "very complex" with no peak corresponding to a molecular ion. The highest mass peak was at m/e 704 and was considered to arise by loss of water from fusicoccin A.

A re-examination of the mass spectrum by Barton *et al.*[7] and, independently, by Ballio *et al.*[8] revealed the presence of a weak peak (relative abundance 0.025%) at m/e 722, which appeared to fragment into three segments, a prominent base peak at m/e 69 (isopentyl, C_5H_9), di-acetyl-glucose at m/e 247 which fragmented further by loss of CH_2CO and CH_3CO_2H, and a fragment at m/e 408 termed "mono-acetyl-aglycone". In addition, a peak was observed at m/e 680 with a relative abundance of 0.05%.

Hydrolysis of fusicoccin A at pH 9 yields desacetyl-fusicoccin (fusicoccin D) with a parent ion m/e 596. The discrepancy in mass between this ion and the Ballio[6] formula suggested the loss of three acetyl groups although NMR results indicated that only two such groups were present. Acetylation of fusicoccin D gave a penta-acetate, $C_{42}H_{62}O_{15}$, (penta-acetyl fusicoccin) which was identical with the acetylation product of fusicoccin A itself. Finally, acetylation of fusicoccin A with deutero-acetyl-chloride (CD_3COCl) yielded the same penta-acetate but with a parent ion 9 mass units heavier than that of the normal penta-acetate, thus confirming that fusicoccin A is di-acetylated and implying a formula $C_{36}H_{56}O_{12}$ (M = 680).

Both penta-acetyl fusicoccin and fusicoccin D give strong parent ions in mass spectra in contrast to the relatively weak peak at m/e 722 in the fusicoccin A spectrum. However, the latter shows a more intense peak at m/e 680 which corresponds to the formula $C_{36}H_{56}O_{12}$. It is probable that the minor ion at m/e 722 requiring three acetyl groups and the associated di-acetyl-glucose ion at m/e 247 arise by trans-acetylation in the ion source of the mass spectrometer.

A re-examination[8] of the crystals used in the original crystallographic molecular weight determination[6] showed that the crystallographic asymmetric unit contains two molecules of fusicoccin A and one molecule of the solvent, ethyl acetate, giving a "molecular weight" of 733 (later amended to 728, see below).

The Substituent Groups

The chemical and spectroscopic studies outlined above showed that the molecule of fusicoccin A is a di-acetylated glycoside consisting of three sub-units; first, an isoprene (C_5H_9) group, secondly, a substituted glucose residue and, thirdly, a substituted aglycone of formula $C_{23}H_{36}O_6$.

The first two of these sub-units and the substituents on the aglycone were characterised by chemical studies which are summarised below:-

a) The Isoprene Unit

An important characteristic of the mass spectrum of fusicoccin A is the large peak at m/e 69. This peak is readily removed by acid to yield 1:1 dimethyl-allyl-alcohol ($(CH_3)_2COH.CH:CH_2$), and from this, by loss of water, isoprene ($CH_3C(:CH_2)CH:CH_2$). Acetylation of fusicoccin after removal of this isoprene group gives a hexa-acetate, the extra acetate being on the glucose residue. The isoprene is, therefore, acid-labile ether linked to the glucose moiety. The characteristic ABX system of isoprene is present in the NMR spectrum of fusicoccin A and is replaced by the spectrum of the isopentyl group on reduction to dihydro-fusicoccin.

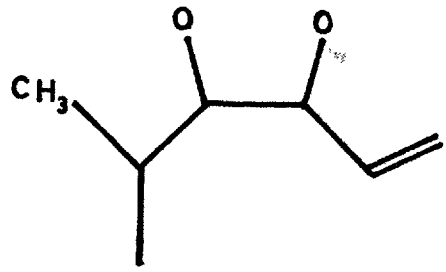
b) Substitution on the glucose residue

From the data above it is apparent that the glucose residue in Fusicoccin A carried one acetate group and the isopentyl group. Examination of the effect of periodate oxidation of both fusicoccin and its de-acetylation products showed that, while fusicoccin A is, itself, unchanged, derivatives with a de-acetylated sugar residue are degraded, taking up two moles of the reagent. The product was shown by mass spectroscopy to contain an intact aglycone moiety and to have lost one carbon atom from the sugar residue, requiring that the latter is a gluco-pyranoside unsubstituted at C2, C3 and C4 and implying that the acetate group is at C3 and the isoprenyl ether at C6.

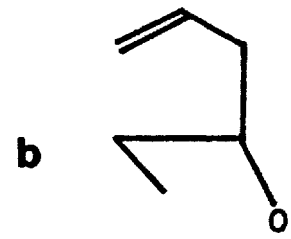
Attempts to determine the stereochemistry of the glucose residue by enzymic hydrolysis of fusicoccin to the aglycone were unusccessful, but the α configuration was established by NMR studies.

c) Substitution on the Aglycone Nucleus

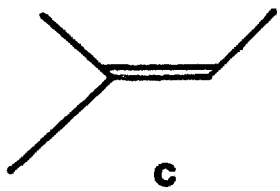
From the evidence outlined above the aglycone nucleus carries the



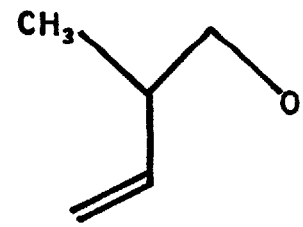
a



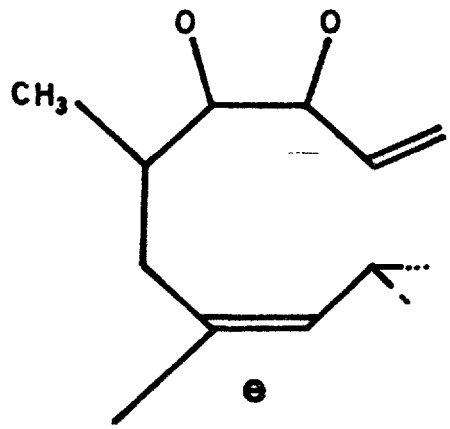
b



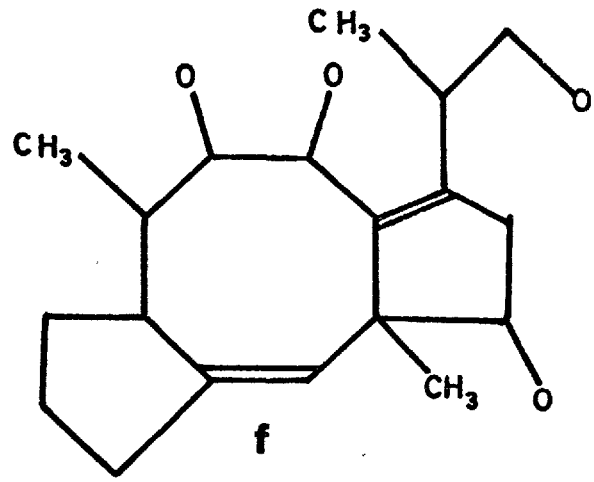
c



d



e

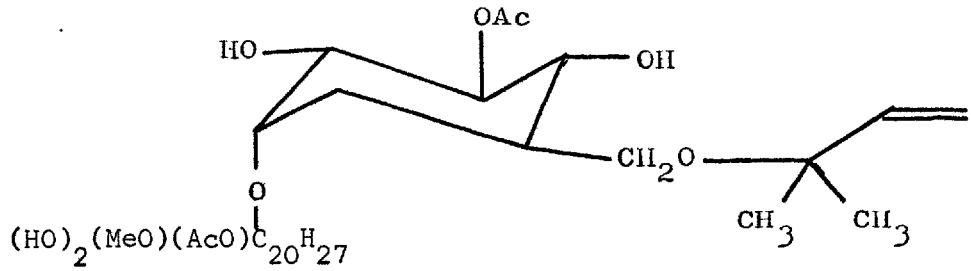


f

Figure 1. a - e are part structures derived from NMR decoupling experiments. The actual structure, f, is included for comparison.

following substituent groups:- acetate (mass spectra), methoxy (direct analysis), readily acetylated hydroxyl (mass spectra), hydroxyl which is only acetylated under vigorous treatment (IR). This completes the identification of all the oxygen atoms in the fusicoccin A molecule leaving a hydrocarbon nucleus of formula $C_{20}H_{27}$.

The substitution pattern of fusicoccin A may be summarised therefore as:-



d) The Aglycone

The aglycone of fusicoccin A is prepared by a three stage reaction sequence from the parent compound. It forms a tetra-acetate, a tetra-benzoate and an acetonide and is oxidised by sodium metaperiodate to an α - β unsaturated di-aldehyde, which implies the presence of an α-glycol system alongside a double bond. The NMR spectrum indicates two secondary and one ~~quarternary~~ tertiary methyl groups, a methoxy group and a single vinylic proton at τ = 4.51. Changes in the NMR spectra on acylation imply the presence of one primary and three secondary alcohol functions, and extensive spin decoupling experiments enabled the identification of four part structures (Figure 1, a - d). The vinylic proton signal arises from the proton on the double bond in part structure d and indicates that this bond is not the same as that in part structures a and b. This implies that there are two double bonds in the aglycone structure and that one of them is tetra-substituted. Part structure e was also obtained from decoupling experiments but, as the decoupling resulted in several instances in very vague spectral changes, it was regarded as very tentative, (the actual structure, represented by f in the figure, is included for comparison, being at that time, not known).

Since four of the oxygen atoms in the aglycone are acylable, the fifth is in a methoxy group and there are two double bonds in the molecule the structure must be tri-cyclic. It is apparent from f that all the part structures were correct, although it was not possible to assemble them into a complete structure until the crystal structure of the aglycone derivative, which is the subject of this work, was completed.

THE CRYSTALLOGRAPHY OF FUSICOCCIN A AND ITS DERIVATIVES

An X-ray crystallographic investigation of fusicoccin A was first reported in the original paper on the isolation of the compound[6]. This was done to determine the molecular weight but, due to incorporation of the solvent (ethyl acetate) in the crystal lattice, it gave an erroneous value. Later crystallographic details were[8]:-

Monoclinic, $a = 20.78$, $b = 14.46$, $c = 13.40$Å, $\beta = 95.50^{\circ}$,

$D_m = 1.207$ g.cc^{-1}, space group $P2_1$, solvent ethylacetate.

This gives an asymmetric unit of weight 1456, close to that of 2:1 ethyl acetate solvated fusicoccin A. (The formula weight of $2C_{36}H_{56}O_{12}.CH_3COOC_2H_5$ is 1450.)

Fusicoccin was found to crystallise[10] from a variety of solvents (methanol, benzene or acetone-light petroleum) with slightly differing habits, but all having essentially the same unit cell and space group (monoclinic, $a = 20.7$, $b = 14.4$, $c = 13.5$ Å, $\beta = 96.6$, space group $P2_1$) containing 4 molecules of fusicoccin ($C_{36}H_{56}O_{12}$) per cell (*ie.* two molecules per asymmetric unit) and varying levels of solvation which made them unreliable for molecular weight determination. It was found possible to incorporate bromobenzene, iodobenzene, m-dibromobenzene and bromothiophene into the lattice with only trivial changes of cell dimensions, but the heavier solvate molecules never exceeded 25% molar concentration compared to fusicoccin. The magnitude of the task of solving and refining such a structure and the absence of information

about rigid groups in the molecule dissuaded us from further work on solvated crystals.

While measuring the density of the various solvated crystals using aqueous solutions of potassium iodide it was observed that a chemical reaction took place. This was reported to the chemists working on fusicoccin and they found that the acetate groups in fusicoccin compounds are very sensitive to pH. The culture medium for the growth of Fusicoccum Amygdali normally gave a mixture of fusicoccins which were troublesome to separate but, as a result of the crystallographic observation, it was found that by controlling the pH of the culture medium at *ca.* 7 very high yields of fusicoccin A could be obtained. Evidently the B,C, and D forms (differing in degree of acetylation) are artifacts.

Work was then concentrated on attempts to form suitable heavy atom derivatives using the hydroxyl groups in fusicoccins A and D. It has been mentioned above that the fusicoccins yielded a number of crystalline esters and an acetonide, but in each case where a halogenated reagent such as p-bromobenzoylchloride, dibromo-acetone or p-bromobenzaldehyde was used in place of the normal reagent, it was not possible to crystallise the resulting compound.

Attention was then focused on the possibility of forming a metal complex with one or both of the olefinic groups known to be present in the molecule. In view of the many silver nitrate/olefin complexes which have been reported in the literature, a series of fusicoccin A/silver nitrate solutions in aqueous and absolute methanol and ethanol were prepared. About two months later crystals were found in one of the flasks, but X-ray photographs showed them to be silver nitrate.

However, in the meantime, Steven Neidle, a colleague, suggested that it might be possible to prepare a mercury derivative of fusicoccin A. This proved to be straightforward and as in a short time suitable crystals

of two mercury derivatives had been grown, the silver nitrate complex approach was abandoned.

In a 1951 review article Chatt[11] attributes the discovery and characterisation of olefin/mercury salt addition compounds to Hofmann and Sands who published a series of papers on the subject in the period 1900-1908. Their work stimulated further activity and the literature is now very extensive. The method used in the present case was based on that reported by Henbest and Nicholls[12]. In the initial experiment 0.19 g of fusicoccin A (0.28×10^{-3} moles, a 10% excess) were dissolved in methanol and 0.5 ml of a 0.5 molar solution of mercuric acetate in methanol were added, followed by 0.25 ml of water. A small portion of the reaction solution was then tested for mercuric ions by the addition of a few drops of 1 molar sodium hydroxide solution. A negative result was obtained, suggesting that an unusually rapid mercuration reaction had occurred. The reaction which normally occurs between olefins and mercuric acetate in a methanol solution is:-

$$Hg(OAc)_2 + MeOH + R_1CH{:}CHR_2 = MeO(R_1)CH\text{-}CH(R_2)HgOAc + HOAc$$

In view of the comparatively low stability of these alkoxy-mercuric acetate compounds it is normal to convert them to the equivalent halide by treatment with an aqueous solution of a sodium or potassium halide. In this instance 1 ml of a saturated solution of potassium bromide was added producing a white flocculent precipitate which was extracted by shaking with benzene. After separation, the benzene extract was boiled to dryness, leaving an oil which was induced to crystallise by scratching the wall of the container with a glass rod. The resulting white crystalline compound had a melting point of 105-110°C. A thin layer chromatographic examination showed that the solid was a mixture of five components, four of which were identified as fusicoccins A,B,C and D and a new spot which represented about 70% of the sample. This was demonstrated to be a mercury derivative and a larger scale preparation

was carried out. This time, to ensure as complete a mercuration as possible, a 20% excess of mercuric acetate was used and, in the second stage, the acetate derivative was converted to the crystallographically more suitable bromide derivative. The benzene extract from this preparation was purified by Dr. U. Ohnsorge by preparative thin layer chromatography. The purified compound was successfully crystallised from a benzene/light petroleum ether solution as elongated needles, with an irregular hexagonal cross section, M.Pt. 120°C, showing parallel extinction.

Rather surprisingly, the NMR spectrum of the compound showed that the isoprene system was still present but indicated the replacement of an olefinic proton by a new singlet at $\tau = 7.78$. This suggests the formation of an H-C-Hg-Br system in the aglycone nucleus. A more detailed examination of the reaction products[9], however, showed that three mercuration products were formed, the mercury addition being either to the isoprene double bond or to the protonated double bond in the aglycone moiety or, thirdly, to both of them.

X-ray photography of the crystals grown by Dr. Ohnsorge gave the following data:-

Monoclinic, $a = 11.3$, $b = 9.4$, $c = 42.7$Å, $\beta = 96°$, $V = 4511$Å^3, $D_m = 1.43$ g.cm^{-3}.

The absent spectra (hkℓ, h+k = 2n; hOℓ, h = 2n; OkO, k = 2n) are consistent with space groups C2, Cm, C2/m, but the optical activity of the compound eliminates all but the first of these. The cell parameters give a calculated molecular weight of 971 assuming a cell occupancy of 4. The normal net result of a reaction between an olefin and methanolic mercuric acetate followed by work up with bromide ions is the addition of CH_3OHgBr, thus producing a molecular weight increase of 311. In the case of fusicoccin A this would give a molecular weight of 991, rather than the observed value, 971. The latter value

is close to that which would be produced by the addition of HgBr (960) and suggests that one of the aglycone hydroxyl groups has participated in the attack on the olefinic double bond to produce a cyclic ether (cyclisations of this type are not uncommon[11,12]). NMR spectroscopy indicated further that the crystals were, like fusicoccin A itself, partly solvated.

A 0.3 x 0.1 x 0.1 mm crystal was selected and mounted on a quartz fibre to rotate about the needle (b) axis and after alignment on the diffractometer a series of θ and θ/ϕ scans was carried out on axial reflections in order to determine the unit-cell dimensions.

The reflections were found, in general, to be weak and diffuse, which caused considerable difficulties in the accurate determination of the angle β. In the h0ℓ zone ($\chi = 0$) θ/ϕ scans gave peaks with plateaux up to 0.3° wide in the ϕ direction and in addition, the h00 reflections showed a tendency to splitting which made direct measurement of the β angle impossible. Several other crystals were examined on the diffractometer, but the problems were found to apply to all of them.

During the course of the continued chemical investigation of the fusicoccin series, Ohnsorge found that mercuration could also be carried out on the desacetyl aglycone of fusicoccin A and he was able to crystallise the resulting derivative from benzene as extremely good square section prisms which showed sharp extinction parallel to the main direction of growth but no extinction in the square section. An NMR investigation of the compound indicated that the mercuration had occurred in a similar way to that in the mercuration of fusicoccin A.

Up to this time, the site of glucose substitution on the aglycone nucleus had been uncertain. Since the α-glycol hydroxyl groups (see above) are inert to periodate oxidation before removal of the glucose moiety, the latter must be linked to one of these groups. Further, since removal of the glucose moiety and the mercuribromide group from

the mercuration product of fusicoccin followed by a manganese dioxide oxidation and acetylation gave a product with the characteristic UV spectrum of an αβ-unsaturated ketone, the allylic hydroxyl group in the α-glycol could not have participated in the oxymercuration reaction and the glucose moiety must be linked to the aglycone at this oxygen atom.

X-ray photography of these crystals showed them to be far superior to those of the fusicoccin derivative and since the conformation and substitution of the glycoside moiety had already been firmly established from chemical studies, it was decided to continue the crystallographic investigation using this compound.

The crystallographic data were:-

$C_{21}H_{33}O_5HgBr.0.25(C_6H_6)$, M = 665.6, M.Pt. 153-155°C, Tetragonal, a = 19.550 + 0.005, c = 13.375 + 0.004Å, U = 5112Å^3, D_m = 1.73 g. cm^{-3}, D_c = 1.729 g. cm^{-3} for Z = 8, Space Group I4.

The absent spectra ($\underline{hk\ell}$, $\underline{h+k+\ell}$ = 2n) and Laue symmetry are consistent with space groups I4, $I\bar{4}$ and I4/m but the optical activity of the compound eliminates all but the non-centrosymmetric space group I4.

The calculated molecular weight (M = 666 for Z = 8) is slightly higher than expected (646) for a mercuration product of the same type as that from fusicoccin A, but the fact that crystals grown from chloroform, though similar in appearance to those from benzene, melt at 163-165°C suggests that the crystals were likely to be solvated.

In this case no difficulty was experienced in the determination of the cell parameters on the diffractometer and the collection of a set of intensity data out to θ = 70° was commenced. The diffractometer was operated in the normal[13] coupled mode. The 0 0 8 reflection was measured after each group of 20 reflections as an intensity reference and the "5-value" scan mode was used.

After approximately 30 hours, 280 reflections had been measured but it was apparent that the compound was suffering from severe radiation damage. The intensity of the reference reflection had dropped to 78% of its original value and the crystal was noticeably darker. The problem was aggravated by a tendency for the diffractometer to misread the steering tape with the result that in several instances, it stopped with the primary shutter open. (This fault was one of a number of "teething" troubles with the instrument and was a persistent difficulty during the collection of the intensity data for the subject compound.)

As several good crystals were available, it was decided to replace the first crystal after 300 reflections and to use a succession of crystals in order to maintain a high standard of data. The crystal morphology was extremely good, enabling accurate measurement of each crystal to be made for absorption correction, and it was hoped that, after scaling to a constant value of the reference reflection, "inter-crystal" scaling would be unnecessary.

In space groups of Laue symmetry 4/m the use of more than one crystal for the collection of intensity data introduces the problem of polarity. As a consequence of this the intensities of reflections of the type $\underline{h}\ \underline{k}\ \underline{\ell}$ are not the same as those of type $\underline{k}\ \underline{h}\ \underline{\ell}$ and it is necessary to determine the "sense" of the $\underline{c}$ axis and ensure that all the crystals used for data collection are mounted in the same way. Provided this is done, the same steering tape can be used for all crystals.

From a film of the hk0 zone, four pairs of reflections were selected where the intensity of hk0 was markedly different from that of kh0 and for each crystal considered for data collection 5 value scans were carried out for these pairs. If the intensity differences were not in the correct sense, the crystal was not used. The pairs used, together with a typical set of the number of counts obtained on the main

scans (point 5 in the 5 value sequence) were:-

h	k	ℓ	I (counts)
5	7	0	649
7	5	0	6923
6	10	0	6456
10	6	0	607
3	5	0	11189
5	3	0	1083
6	2	0	12887
2	6	0	7312

In all, the intensities of 2701 reflections were measured from five crystals. These were processed in the normal way[13], sixteen reflections being lost due to paper-tape errors, to give a final set of 2685 reflections of which 145 were classified as unobserved ($I < 2.58\sigma(I)$). A sharpened origin-removed Patterson synthesis was computed using these data and from peaks in the Harker section at $w = 0$, the x and y coordinates of the mercury and bromine atoms in the structure were calculated. The u and v components of the intramolecular mercury-bromine vector were calculated from these coordinates and the vector peak was readily located at $w = 0.108$. This gave an Hg-Br distance of 2.4Å - a reasonable value.[14]

The z coordinate of the mercury atom was then arbitrarily fixed at 0.0, enabling the completion of the coordinate set for the two heavy atoms. By extrapolation of the bromine-mercury bond and assuming an Hg-C bond length of 2.1Å,[14] the position of the first carbon atom in the structure was calculated and the resulting 3 atoms were then subjected to three cycles of full-matrix least-squares refinement, keeping the z coordinate of Hg fixed. This gave an R factor of 28%, and a difference Fourier synthesis was calculated. All three atoms retained reasonable temperature factors.

The Fourier map, not unexpectedly, contained pseudo symmetry which complicated the early stages of the determination of the structure. A re-examination of this map after completion of the structure showed that, with the exceptions of the atoms having z coordinates close to that of the mercury atom, all the peaks were mirrored but, in the majority of cases, the correct position was that with the greater calculated electron density. The situation was aggravated by the fact that two of the rings in the final structure lay approximately parallel to the z axis and at heights close to the pseudo mirror plane. (The absence of perfect pseudo symmetry is, of course, due to the fortunate circumstance that the bromine atom does not have the same z coordinate as the mercury atom.)

On the basis of peak heights and from a three dimensional model, 18 new "atoms" were located which formed a chemically feasible model consisting of fused seven-and five-membered rings with 4 side chains. 16 of these "atoms", together with the two heavy atoms and C1 were then used in the calculation of a second difference map.

Two further fused rings and a possible spiro five-membered ring (on C11) were located from this map, giving a structure which was close to the final form. One of the new rings appeared to be six-membered, but due to its bad geometry and the low height of one of the members, only five of the atoms were included in the next Fourier synthesis. The spiro system was also poorly defined and was not included. These 25 "atoms" were used in the calculation of a further difference map (R = 20.0%) and from this three new atoms were identified, two of them being part of the "spiro" five-membered ring.

The 28 assigned atoms were then subjected to three cycles of block-diagonal least-squares refinement, giving an R factor of 16.9%. All the proposed atoms survived this refinement and the only peaks present on a new difference Fourier map were attributable to anisotropic

thermal movement. This was particularly marked around the mercury and bromine atoms so a further three cycles of refinement were therefore carried out with anisotropic thermal parameters for these atoms; this gave an R factor of 11.8%.

On the basis of the chemical knowledge of the structure, the five oxygen atoms were assigned and the structure subjected to three further cycles of refinement: R fell to 11.2%. The data were then corrected for absorption effects using the ICABS[15] program, which resulted in a further improvement in the R factor ro 10.1%.

The most noticeable feature of a model constructed at this stage to illustrate the molecular packing was a series of continuous cylindrical holes of radius *ca.* 2Å sited about the fourfold axes. The observed density of the crystals and the variation of the melting point of the derivative when crystallised from different solvents (see above), suggested the presence of solvent in the crystal lattice but no evidence of any was found in a difference Fourier synthesis.

An examination of the list of calculated and observed structure factors showed bad agreement for large groups of reflections in the data from the fourth crystal (10 14 10 to 18 10 6), and the R factor calculated for this block of data was found to be 6-7% higher than that for the other crystals. The absorption correction and processing of crystal four were checked, but found to be correct and it was therefore decided to remeasure the intensity data for the bad block. Three crystals were used for this remeasurement, bringing the number of crystals used to eight, and giving a final data set of 2696 reflections of which 135 were classified as unobserved ($I_{net} < 2.58\sigma(I)$).

Further refinement with the new data set converged to an R factor of 9.0% but the general quality of the structure was still rather poor. The temperature factors for the carbon and oxygen atoms ranged from 0.5 to 8.5 and the bond lengths were considerably different from "normal"

values and had standard deviations of *circa* 0.08Å. The introduction of anisotropic temperature factors to the side chain atoms gave no further improvement in either the structure or agreement factor.

It was decided, therefore, to attempt to refine the structure further with separate scale factors for the individual crystals. This produced an immediate improvement to R = 8.0% and after three cycles of full-matrix refinement using an extended version of the ORFLS program, R fell to 7.3%. Two atoms (C5 and C19) wandered away from their positions at the start of the refinement, but C5 returned during successive cycles. The temperature factor of C19 in its new, unbonded position, became very high, but a large peak was found on a difference map in the chemically correct position for this atom.

Further inspection of the difference map showed clearly the presence of a ring of positive electron density around the central fourfold axis at $\underline{z} = 0.665$ and lying parallel to the 001 plane. The radius of the ring was 1.4Å and it was therefore assumed to represent a disordered benzene molecule.

Since the electron-density ring was centred on a fourfold axis, the model used to describe it must have fourfold symmetry and this is most readily achieved assuming two concentric and coplanar molecules displaced by a 30° rotation about the common centre and each given a multiplicity of 0.5. This model results in a total of two benzene molecules per unit cell and a consequent benzene/aglycone ratio of 1:4, which is in excellent agreement with the measured density. An integrated NMR spectrum of a solution of some of the crystals in deuteroacetone confirmed that this ratio was approximately correct.

The ring of electron density was observed to have maxima displaced slightly from the planes $x = 0.5$ and $y = 0.5$ and the positions of the necessary three half carbon atoms were calculated assuming the benzene to be slightly rotated with respect to these planes.

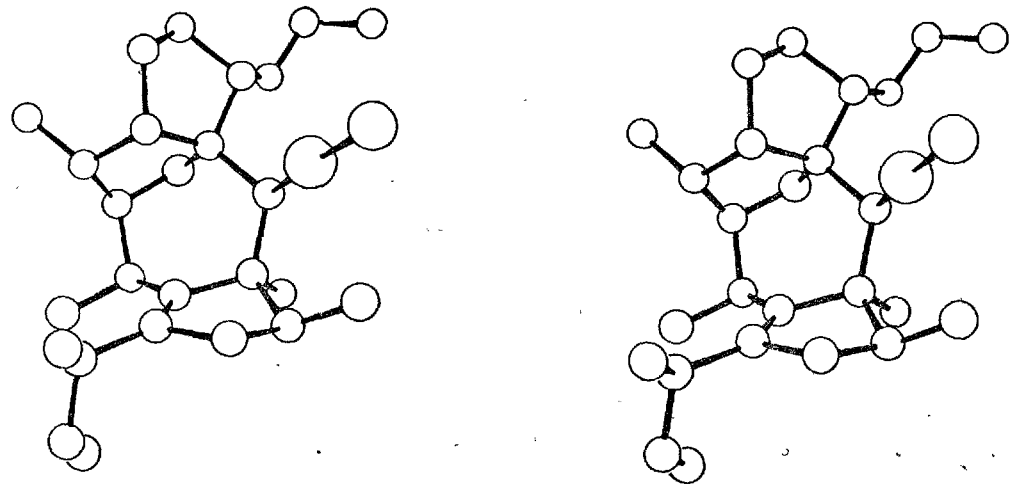


Figure 2. Stereoscopic drawing of the fusicoccin aglycone mercuri-bromide molecule.

The inclusion of the benzene together with the correction of the positions of C5 and C19 resulted in an improvement in the R factor to 6.7% and the positions of the 18 locatable hydrogen atoms in the structure were then calculated, assuming a C-H bond length of 1.0Å. Refinement of all the non-hydrogen atoms, with anisotropic thermal parameters assigned to all the side chain atoms and including the hydrogen atoms as fixed atoms, gave R = 6.3% and this was improved to 6.0% by the removal of nine reflections for suspected extinction.

Refinement for a further two cycles with this data, using the program "CRYLSQ" in the X-RAY 70 system reduced R to 5.5%. A correction for anomalous scattering by mercury and bromine was then applied using the same program and was found to give an increase of 0.2% to R with the structural absolute configuration which had been used up to that time. Reversal of the sign of the imaginary component of the dispersion factors for the two fluorescent atoms gave an R factor of 5.3%, *ie.* 0.4% lower than in the first case, implying that the absolute configuration was the opposite of that in use (see below).

Finally the whole structure was refined with anisotropic temperatures factors and converged to an $\underline{R}$ factor of 5.1% using 2552 observed reflections. The final statistics were:-

$\underline{R}$ = 0.051 over 2552 observed reflections

$\underline{R}_w$ = 0.063 over 2552 observed reflections

$\underline{R}$ = 0.111 for the 135 unobserved reflections

and the final refinement was of 259 variables.

A stereoscopic "ORTEP" drawing of the structure is shown in figure 2. Figures 3 and 4 show the bond lengths and angles in the structure and Table 1 gives a list of the observed and calculated structure factors and details of the ten reflections removed for suspected extinction. Tables 2 and 3 contains the positional and thermal parameters for the non-hydrogen atom and the positional parameters for the hydrogen atom are given in table 4.

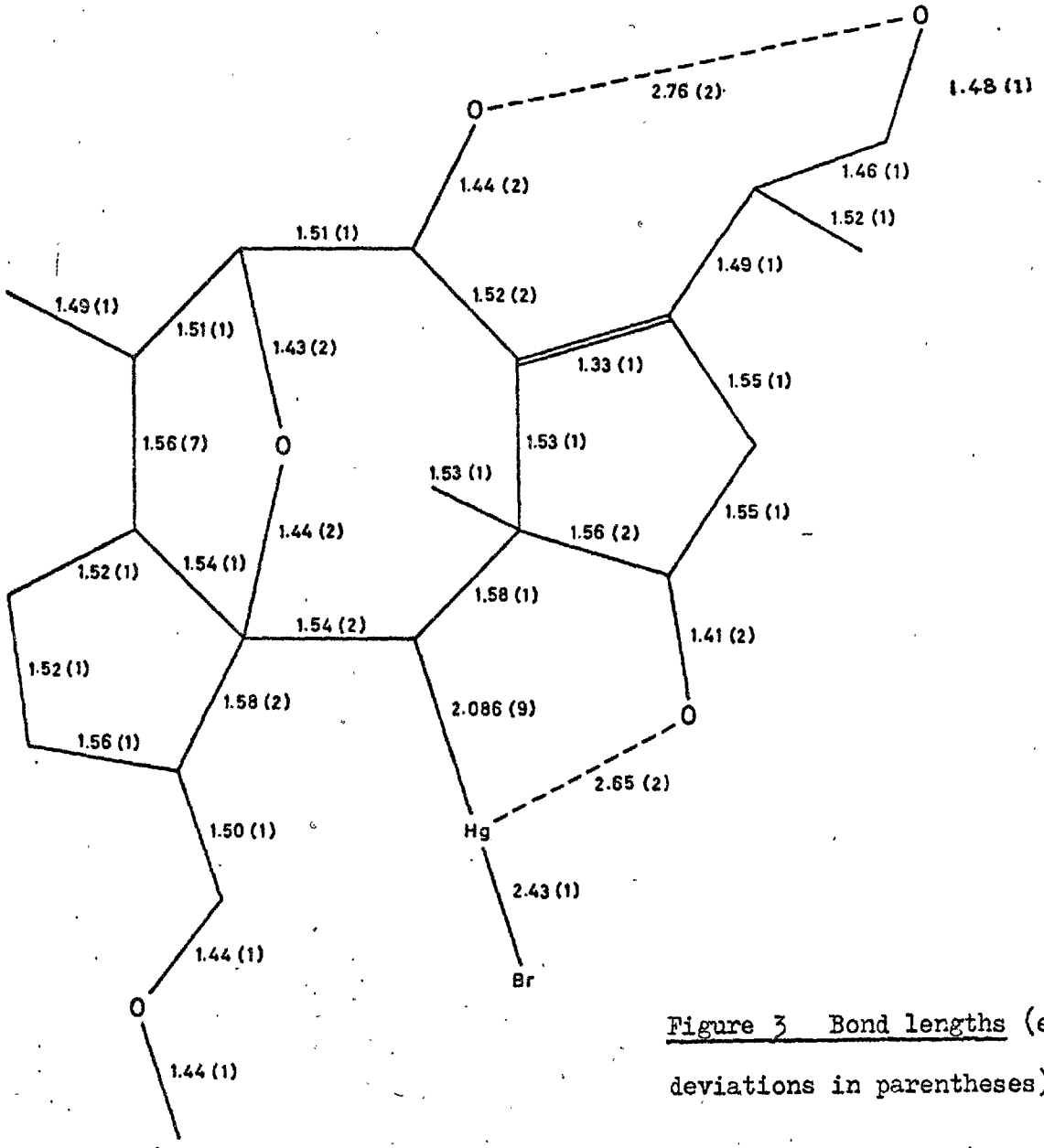
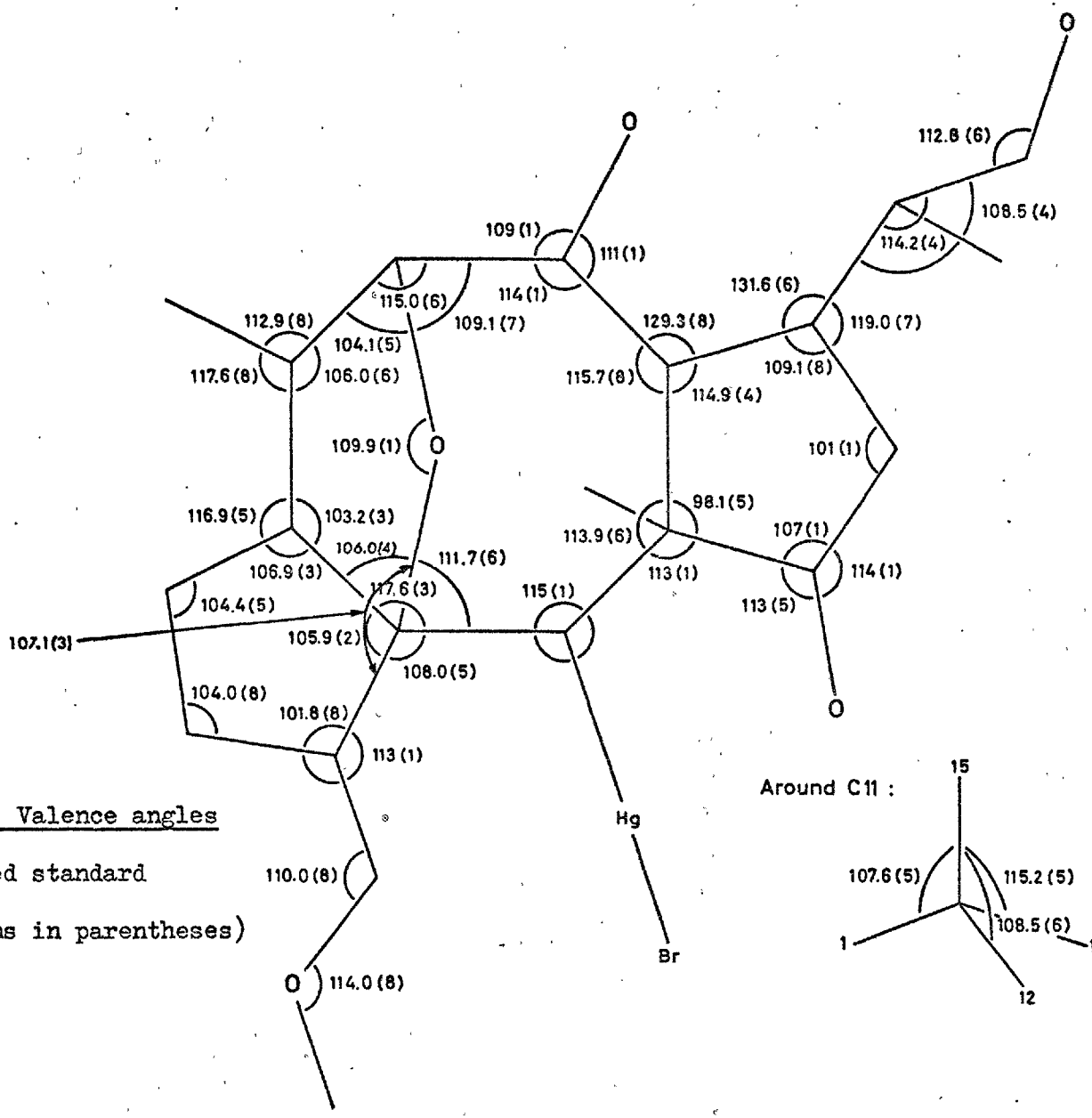


Figure 3 Bond lengths (estimated standard deviations in parentheses)



Around C11 :

Figure 4 Valence angles

(estimated standard deviations in parentheses)

Results and Discussion

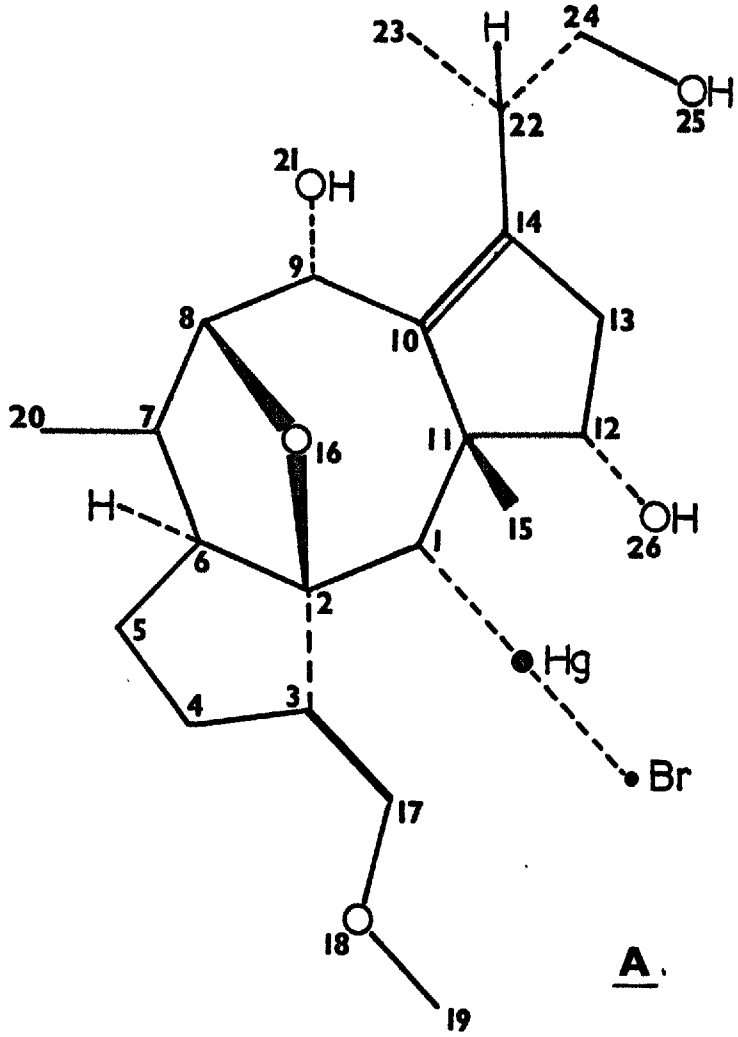
1) Fusicoccin and the Sesterterpenes.

From Figures 2 and 5a it is apparent that the structure studied consists of a tetra-cyclic nucleus containing an eight membered ring bridged by an ether oxygen which is attached to the carbon atom adjacent to the site of mercury addition. From the changes in the NMR spectrum produced by mercuration (*vide infra*) Barton *et al.* were able to deduce structure 5b for the aglycone. Then, knowing the absolute configuration of the aglycone (determined from anomalous dispersion effects; see below), and that the α-D-glucose moiety was attached at O_{21}, they were able to write down the full absolute stereochemistry of fusicoccin A.[17]

Some 24 hours before the submission of a preliminary communication on this work, the structure of the p-iodobenzene sulphonate of the complete molecule of fusicoccin A was published by Vaciago *et al.*[8] This work, of which we were unaware, was in complete agreement with Barton's deductions. In Vaciago's structure the absolute configuration was deduced by using the α-D-glucose unit as an internal reference standard. Our assignment, based on anomalous scattering of Hg and Br provided an independent crystallographic determination of the absolute configuration.

The aglycone consists of fused 5,8 and 5-membered rings forming a diterpenoid (C29) skeleton which is only known otherwise in the series of mildly phytotoxic fungal metabolites known as the ophiobolins[17], and in a wax secreted by the parasitic insect Ceroplastes Albolineatus to prevent desiccation.[18] These compounds are illustrated in Figure 6 (B-F).

With the exception of fusicoccin, all are sesterterpenes (C25) and carry an isoprenyl group on the side chain at C14 in the diterpenoid system. Fusicoccin was initially classified as a sesterterpene in view



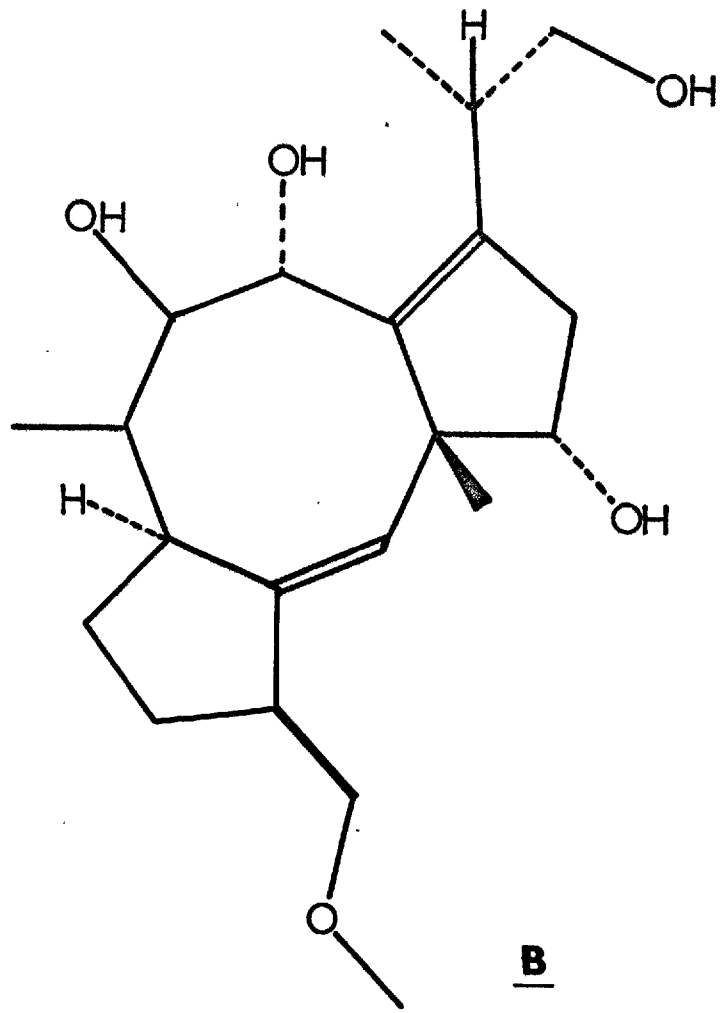
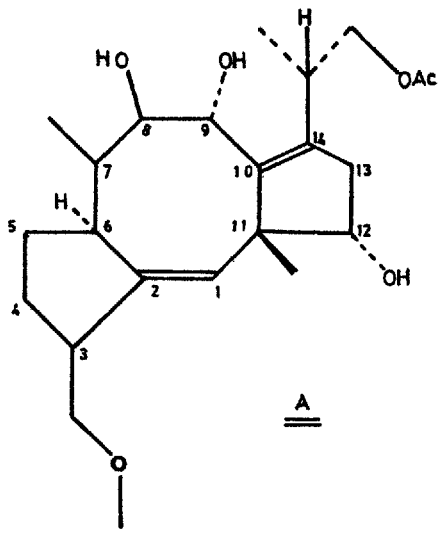
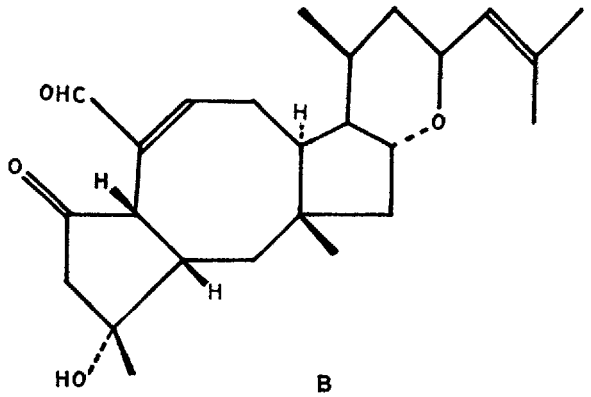


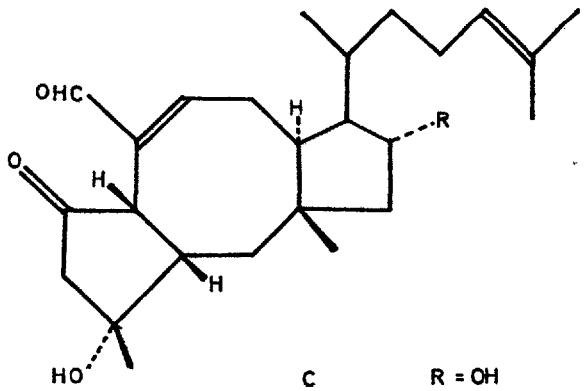
Figure 5. a) Fusicoccin aglycone mercieri-bromide and b) fusicoccin aglycone.



A

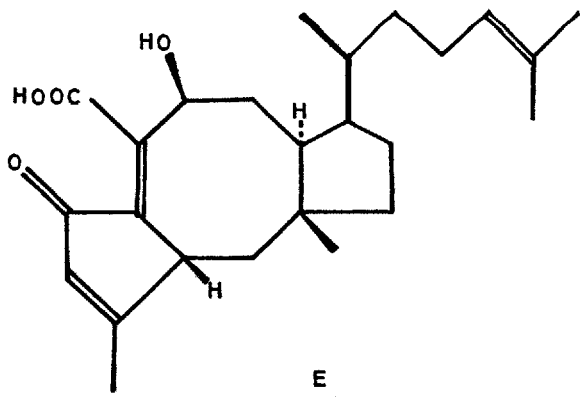


B

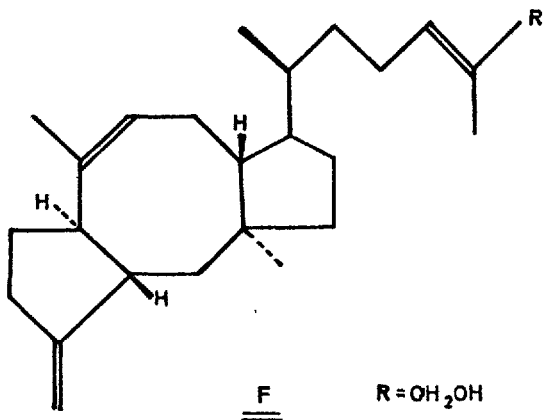


C R = OH

D R = H



E



F $R = OH_2OH$

G R = COOH

Figure 6. The sesterterpenes.

A Fusicoccin A
B Ophiobolin A
C Ophiobolin B
D Ophiobolin C
E Ophiobolin D
F Ceroplastol
G Ceroplasteric acid

of the presence of an isoprenyl system on the glycoside moiety. It was considered possible that this had been attached to the aglycone during the biosynthesis of the compound. However, later work[19] on fusicoccin H, a biosynthetic precursor of the main fusicoccins, has shown that the isoprenyl group is appended after the synthesis of the aglycone nucleus. The latter is, therefore, correctly termed a diterpenoid.

From Figure 6 it is clear that the diterpenoid nuclei in the fusicoccins, ophiobolins and ceroplastols differ in their substitution pattern, degree of unsaturation, the location of the double bonds, and the configuration of the asymmetric centres at C6 and C11. At C6 the fusicoccins resemble the ceroplastols, whereas at C11 they resemble the ophiobolins. Crystal-structure determinations have been reported for derivatives of ophiobilin A[17] (Fig. 6B), ophiobolin D[20] (Fig. 6E), ceroplastol[18] (Fig. 6F) and fusicoccin A (Fig. 6A + glycoside).

In fusicoccin A the α-D-glucose moiety is attached at O21, and its removal yields the "α-glycol system alongside a double bond" which is attacked by sodium periodate to yield an unsaturated di-aldehyde (*vide infra*).

The structure and conformation of the mercury derivative

The bond lengths in the structure can justifiably be described as normal, the mean lengths of the various bond types being:-

Csp^3-Csp^3	1.53Å
Csp^3-Csp^2	1.52Å
Csp^2-Csp^2	1.33Å
Csp^3-Osp^3	1.44Å

The only Csp^3-Csp^3 bonds which deviate significantly from the mean value are C7-C20 (1.49(1)Å) and C22-C24 (1.46(1)Å), both of which are exocyclic and are not subject to strain. The oxygen atom O25 is

hydrogen bonded both intra-molecularly to O21 (2.76Å) and inter-molecularly to O26 (2.68Å), but it is difficult to imagine that this could affect the length of the C22-C24 bond.

The structure contains three five-membered rings (A, B and D in Figure 5a), all of which are in the "envelope" conformation. The mean bond angle in ring A is 104.6°, close to that for a cyclopentane ring when all the bond lengths are equal (104.1°).[21] The standard deviation of the atoms in this ring from the best plane through all five atoms is 0.20Å but atoms 2,3,5 and 6 are coplanar with C4 0.61Å out of this plane. The torsion angles in the ring are given in Table 6, and Table 5 contains the least squares planes.

Ring B is distorted from the pure envelope form and the standard deviation of the four atoms forming the most planar part of the ring (C2, C6, C7, C8) is 0.05Å with O16 0.43Å out of the plane. On the other hand, the standard deviation of the atoms from the best plane through the complete ring is smaller than in ring A, 0.15Å. The more nearly planar character of the ring is also reflected in the mean bond angle, 105.8°, which is closer to the 108° in a plane regular sided pentagon.

The third five-membered ring (ring D, a cyclopentene ring) is again in the envelope conformation, the four atoms associated with the C10-C14 double bond (C10,C11,C13 and C14) being planar, and the fifth atom, C12, lying 0.48Å out of this plane (see Table 4). The mean bond angle within this ring is 104.1°, the same as that for cyclopentane[21] but significantly different from that in cyclopentene[22] (106.4°). The angle at C11 (98.1(5)°) is significantly smaller and that at C10 (114.9(4)°) almost 4° larger than their equivalents in cyclopentene. These deviations are probably attributable to the involvement of C10 and C11 in the eight-membered ring system and to repulsive forces between the mercury atom and O26. (The distance between these atoms

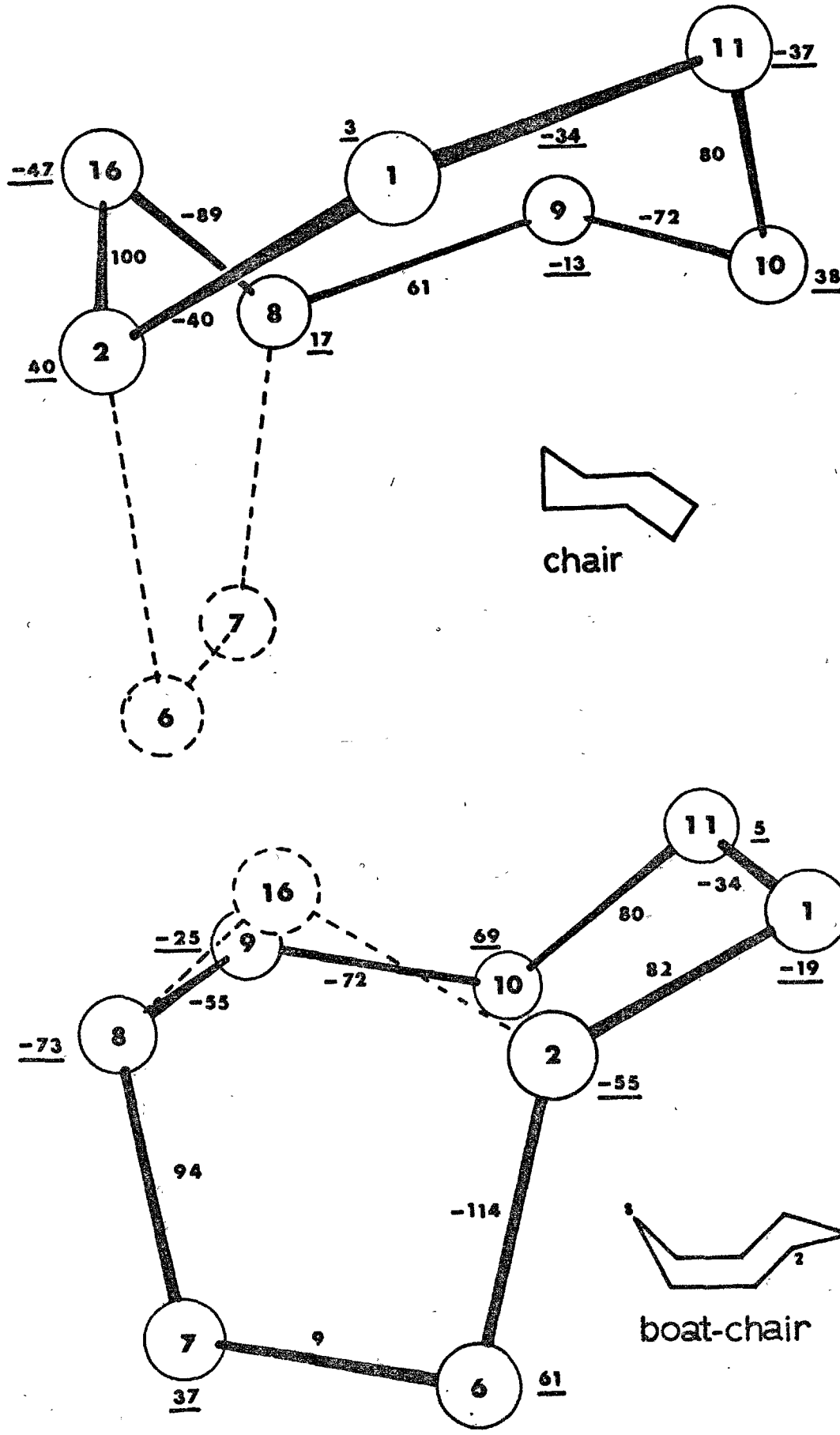


Figure 7. Conformations of the seven- and eight-membered rings. The values inside the rings are the torsion angles (in degrees) and those outside the rings the deviations of the atoms from the best planes through the rings (in Å x 10^2).

is only 2.65Å, considerably less than the sum of the van der Waals radii of mercury and oxygen (2.9Å)).

In addition to these rings, the structure contains seven- and eight-membered rings with six common atoms. The conformations, torsion angles and deviations from the best planes through these rings are shown in Figure 7.

Cycloheptane can exist in two conformations which are only interconvertible *via* major deformations of the bonding angles.[21] These are termed the "chair" and the "boat" form, and are analogous to the similarly named conformations in cyclohexane. From theoretical considerations, Hendrickson[21] showed that the chair form was the more energetically favourable, but pointed out that, in the pure form, both conformers are subject to severe H-H repulsion forces, which can, however, be relieved by a torsional twist to give the "twist chair" and "twist boat" conformations. In a more sophisticated calculation Bixon and Lifson[23] came to a similar conclusion but applied the names "chair" and "boat" to conformations which, though close to Hendrickson's, were twisted from the pure chair and boat. These have non-zero torsion angles about the "foot" of the chair (20°) and the "stern" of the boat (40°). In both series of calculations, the most energetically favourable conformation was found to be the "twist" or "skew" chair, followed by the "chair". In spite of energy considerations, however, almost all the seven-membered rings reported in the literature are closer to the chair conformation than to its skew form. This applied to both simple rings such as those in dimeric cycloheptanone peroxide,[24] in dextrorotatory 4-bromo-6,10-dimethyl-bycyclo (5,3,0)decane-3-one[25], in ferric mycobactin P[26], and to the complex fused rings encountered in the sesquiterpenes (*eg.* 11,13-dibromo-pulchellin[27] and refs. therein). It appears that no seven-membered carbocycle has yet been described as a "twist chair", but a small number of sesquiterpenes have been

reported[28,29,30], where the seven-membered ring is in the boat formation.

Figure 7a, which depicts the 7-membered ring in the present structure, shows that it can be regarded as a distorted chair with C_1, C_2, C_8 and C_9 nearly coplanar (see Table 7) and with C_{11} badly displaced from its ideal position - almost certainly by the Hg - O_{26} repulsion forces mentioned in connection with the discussion of ring D. C_{10} is an sp^2 carbon atom but according to Hendrickson this is unlikely to introduce any serious distortion. However, the C_{10} - C_{14} double bond in ring D, and the consequent increased rigidity of that ring allows the transfer of the Hg - O_{26} repulsion directly to C_{11}, forcing it some 0.5 Å out of the "ideal" position for the chair conformation.

Inspection of the torsion angles in this ring system however, shows that there is an approximate diad axis through C_1 and the midpoint of the C_8 - C_9 bond. Atoms C_1, C_2, C_8, C_9, and C_{11} are nearly coplanar and C_{10} and O_{16} are approximately symmetrically displaced from this plane. Since Hendrickson's "twist boat" is twofold symmetric in this way the 7-membered ring in the present structure is best termed a "twist boat".

The mean angle in the 7-membered ring is 112.6°, similar to that in Hendrickson's[21] twist-chair conformation (112°), but lower than that for both the twist chair (114.1°) and the chair obtained by Bixon and Lifson[23]. The three smallest angles in the ring are those involving the atoms which are common to ring B and indicate the dominance of the far less flexible five-membered ring. It must be noted that, in his treatment of ring conformations, Hendrickson[21] calculated the energy contribution from bond-angle strain as a function of the departure of an angle from the tetrahedral angle, 109.7°, and two of the angles in the present seven-membered ring (at C8 and O16) are closely tetrahedral. Bixon and Lifson stated that the assumption that zero bond-angle strain occurs at the tetrahedral angle is unjustified

since a) it is not valid unless all four substituents are the same and b) non-bonded interactions cannot expand the tetrahedral angle to the values observed in normal alkanes. They, therefore, calculated bond-angle strain in a ring with "n" members as a function of departure from the experimental bond angle in a normal alkane with "n" carbon atoms.

In strain-energy calculations of the same type as those used for cycloheptane, Bixon and Lifson[23], Wiberg[31] and Hendrickson[32] have shown that cyclo-octane can exist in four favourable conformations, but physico-chemical studies by a variety of methods (see references in Ferguson *et al.*[33]) have not indicated a preference for any one of these conformations either in solution or in the gaseous phase.

In this case, however, unlike that in seven-membered rings, the results of a number of crystallographic investigations have shown that the theoretically most favourable conformation is also that which occurs most commonly. This again applied to simple systems such as trans-1,4-dichloro-octane[34], dimeric cyclo-octane-peroxide[35] and cis- and trans-cyclo-octane-dicarboxylic acid[36,37] (trans-syn-trans-1,2,5,6-tetrabromo cyclo-octane[33], however, occurs in the twisted crown conformation). It also applies to complex ring systems such as those found in the sesterterpenes and in Vaciago's structure of fusicoccin A. In the latter, the ring is distorted by the presence of double bonds and the fused 5 membered rings.

In the present structure the eight-membered ring occurs in a distorted form of the boat chair conformation. The distortion plainly arises from the closure of ring B by addition of O16 to the C1-C2 double bond in fusicoccin itself. The mean angle in the ring is 112.5°, significantly lower than that calculated by Bixon and Lifson[23]. This is again due to the predominance of ring B, since the angles at C6 and C7 are both very small for an eight-membered ring but normal for a five-membered ring.

Although the eight-membered rings in both fusicoccin and our mercury derivative can be regarded as distorted forms of the boat chair conformation, the two rings are not directly superposable and the oxy-mercuration reaction involves a conformational change. This will be considered in the discussion of the mechanism of mercuration.

From Table 7, which contains a survey of reported values for mercury - carbon and mercury - bromine bond lengths, it is clear that both bonds in the present structure lie well within the normal range of values. The C-Hg-Br bond angle is 177.5(5)°, deviating slightly from linear but this is fairly common when the mercury atom is not sited in a special position. (The mercury atoms in di-p-tolyl mercury[55], potassium-iodo-dicyanato-mercurate[48] and diphenyl mercury are located in special position positions which force the C - Hg - C bond to be linear.)

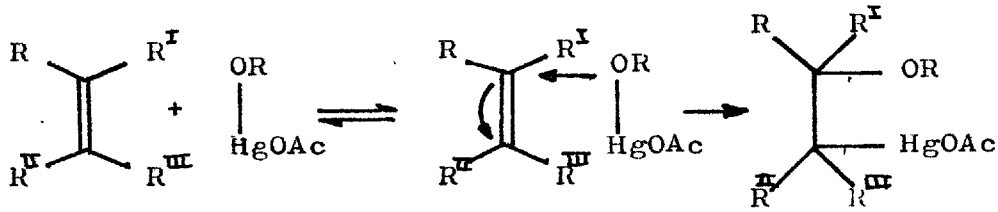
The most unexpected aspect of the C1-Hg-Br system, however, is the close contact between the mercury atom and the hydroxylic oxygen atom O26 (2.65 Å), a distance significantly shorter than the sum of the van der Waals radii of mercury and oxygen (1.50 and 1.40 Å respectively[66]). Short non-bonded contact distances between mercury and oxygen have been reported in several papers and seem to fall into two classes dependent on the length of the oxygen to mercury contact. These results are summarised in Table 8, but it is difficult to see any definite relationship between the type of oxygen atom and the metal to oxygen distance. The two short contacts in the Millon complex (ref. 57) are termed "carbonyl" on the basis of the predominance of the quinonoid (or keto) form of the 2-nitroso-4-methyl-phenol moiety inferred from the bond lengths. The oxygen atoms are not purely carbonylic in nature and comparison with the present results suggests that they may have considerable hydroxylic character.

Mercury (II) is able to form complexes in the normal way, but since it contains fully filled "d" orbitals, the coordination is more random than in first-row transition metal complexes. The characteristic coordination numbers and stereochemical arrangements are two-coordinate linear and four-coordinate tetrahedral, but this will be considered further in the discussion of the molecular packing.

The Mercuration Reaction

Studies on the mechanism of oxymercuration have been considered in detail in review articles by Chatt[11] (1951) and later by Kitching[63]. In 1951 two possible mechanisms were under examination, that of Wright[64]:-

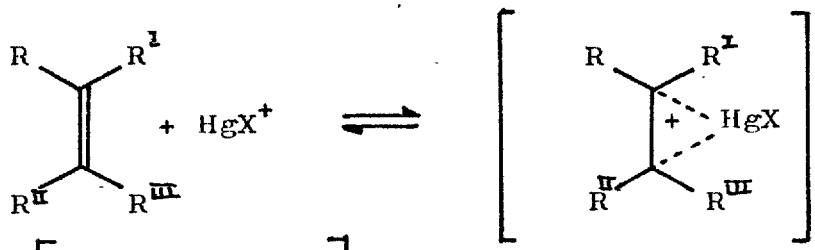
(I)



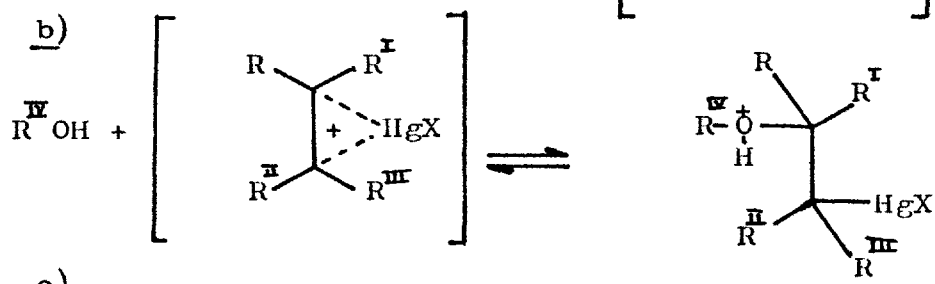
and that of Lucas, Heppner and Winstein[65]:-

(II)

a)



b)



c)

The first of these mechanisms required the existence of an alkoxy or hydroxy mercuric salt in the reaction medium and leads naturally to a cis or mixed cis and trans addition, whereas in practice the addition is found almost always to be trans. For these reasons and in view of the overwhelming evidence for the second mechanism, Chatt proposed the latter as the most feasible.[11] This is now generally accepted.[63]

The involvement of a "mercurinium ion" in the mechanism is analogous to that of a "bromonium ion" in bromine addition to olefins, and their structures are probably similar to that of the well-known silver/olefin complex, involving "forward" bonding by overlap of the olefinic orbitals and "back" bonding by interaction of the filled metal "d" orbitals with the olefin π^* orbitals. Kitching[63] applied this to the mercurinium ion by envisaging "a 6sp hybrid orbital of a mercury interacting equally with the carbon 2p orbitals of the olefin to form a triangular three-orbital, two-electron system in which each orbital has considerable overlap with the other two."

In the mercuration of fusicoccin A, three products were identified[9] where addition of one or two mercury atoms had occurred (see introduction). In the reaction of the aglycone with methanolic acetate, however, only one oxy-mercurated product was obtained, although the molecule contains two olefinic systems. In view of the known structures of the aglycone nucleus and its oxy-mercuration product, it is interesting to consider reasons for the non-reaction of the C10-C14 double bond and the mechanism of attack on the C1-C2 double bond.

From Vaciago's structure of the iodo-tosyl derivative of fusicoccin A it is plain that attack on the C10-C14 olefinic system by the $(HgOAc)^+$ ion to form a mercurinium ion is severely hindered by adjacent substituent groups, particularly the branched side chain on C14. Although some degree of rotation about the C14-C22 bond is possible, there is no position which leaves the double bond exposed, and the latter is also

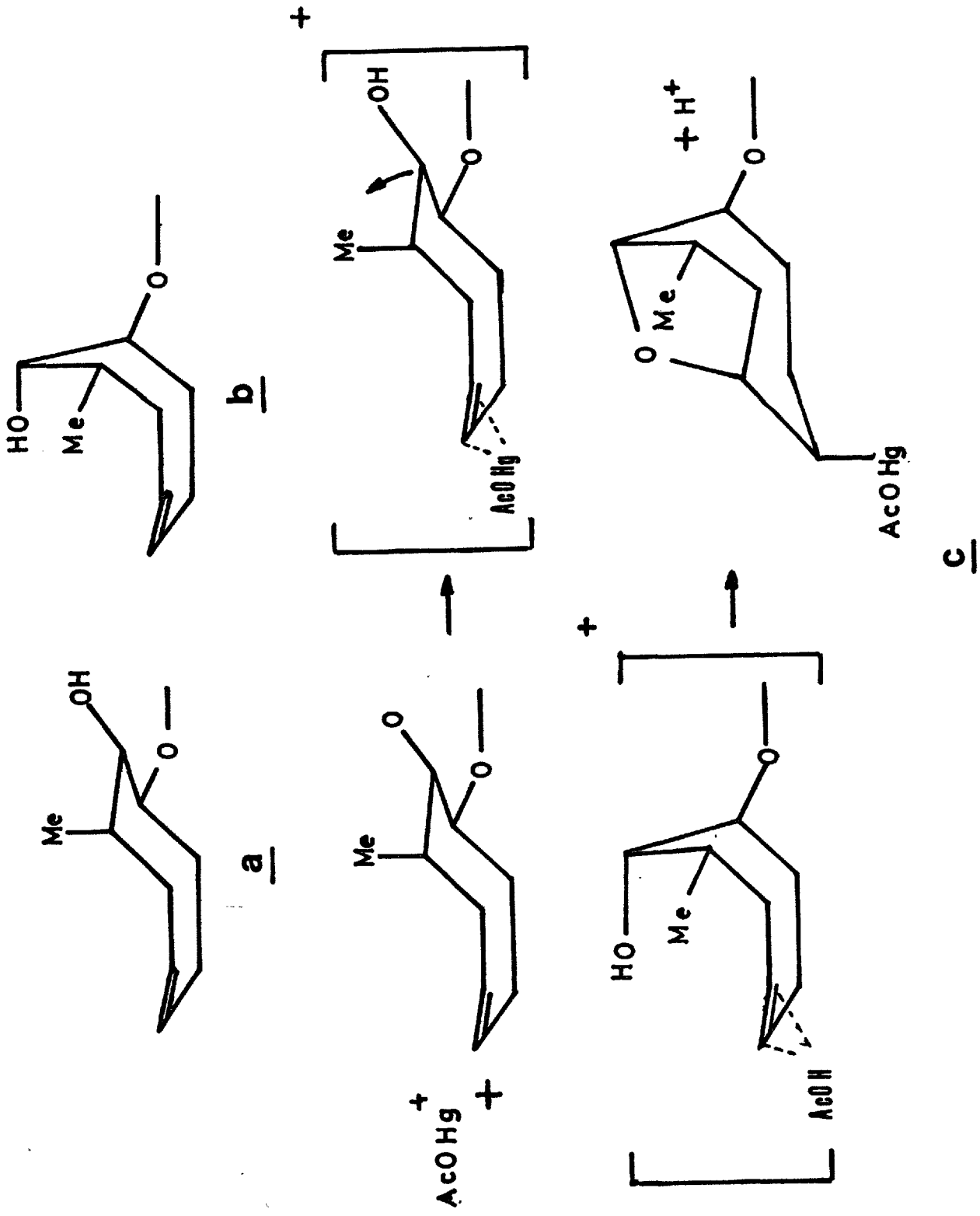


Figure 8. A possible mechanism for the mercuration reaction.

covered by O21 and the hydrogen atom on C9.

The C1-C2 olefinic system, on the other hand, is almost completely exposed to attack with only the side chain and hydrogen atom on C3 anywhere near it. The side chain is again free to rotate and can be positioned well away from the double bond. It is not surprising, therefore, that attack by the $HgOAc^+$ ion occurs at the C1-C2 double bond.

Formation of the mercurinium ion is normally followed by attack from "behind" by the solvent (see equation IIb above) to give a trans disubstituted product, but the formation of cyclic products by internal ether formation is well known (see Chatt[11], Kitching[63] and Ref. 12). In fusicoccin A the aglycone nucleus was found to be in a distorted form of the boat-chair conformation with a torsion angle of -9° about the C1-C2 double bond. C1 lies slightly above the C2, C6, C10, C11 plane and the oxygen atoms on C8 and C9 are "gauche" to each other. This is illustrated in Figure 8a. If C8 is flipped into its other equilibrium position, the two oxygens are in the "anti" configuration and that on C8 is brought into a very convenient position for attack on the C1-C2 double bond (see Figure 8b). Since C2 is already trisubstituted, such an attack must force C1 "down", changing the conformation of the "intermediate" from a boat into the new, but distorted boat chair, which is found in the mercury derivative. The mechanism of the oxymercuration therefore, appears to be that shown in Figure 8c. The sequence of these events is, of course, open to question.

The conformation change from the boat-chair to the boat (8a and 8b) is the only one which can be carried out without a major deformation of the structure and the breakdown of the π bonding in the C1-C2 bond. In a model of the aglycone moiety, the operation is easy to carry out and introduces no serious strain. It is therefore, possible that in solution an equilibrium exists between the two conformers and that the

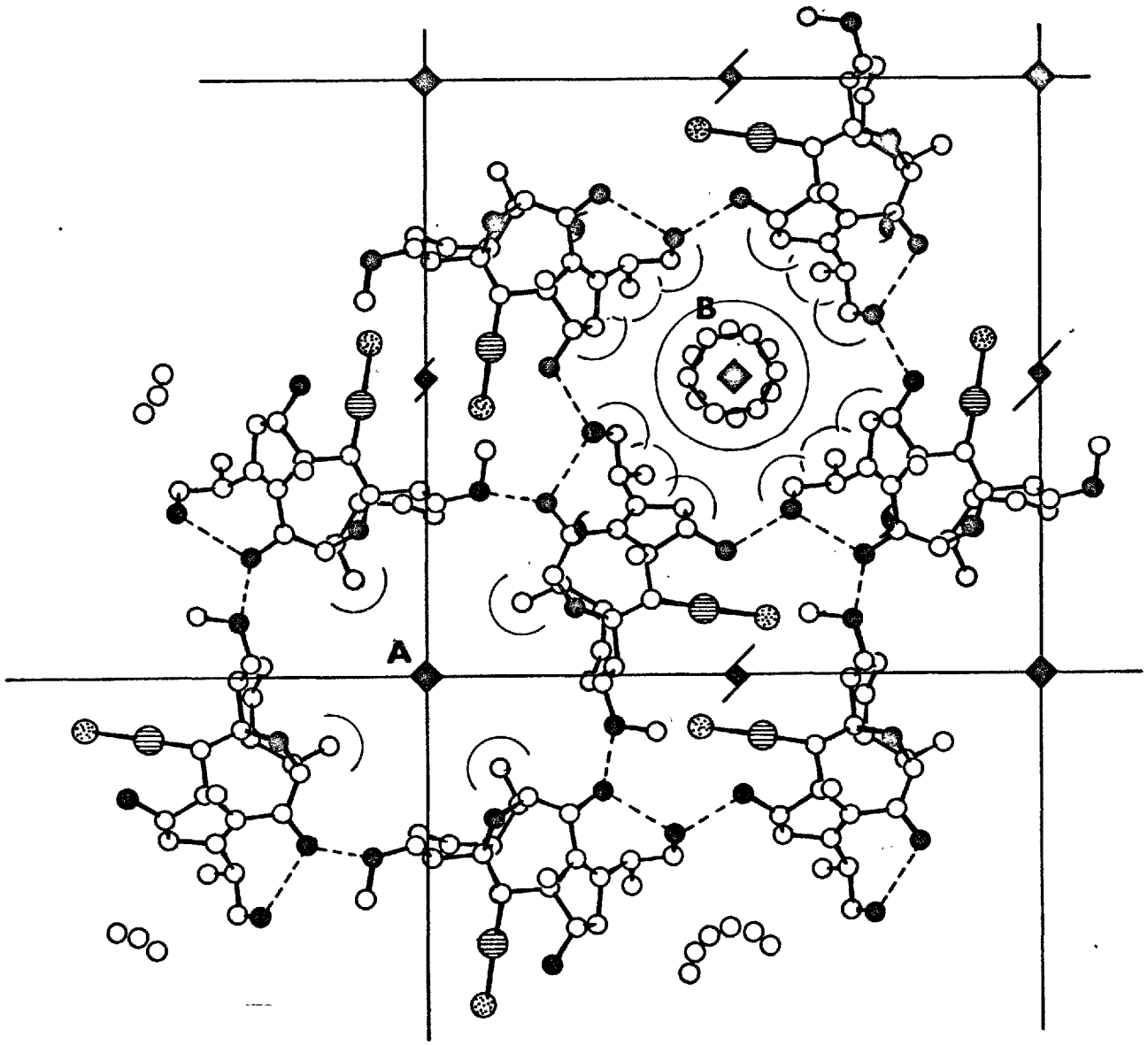


Figure 9. [001] projection of one layer of the crystal structure.

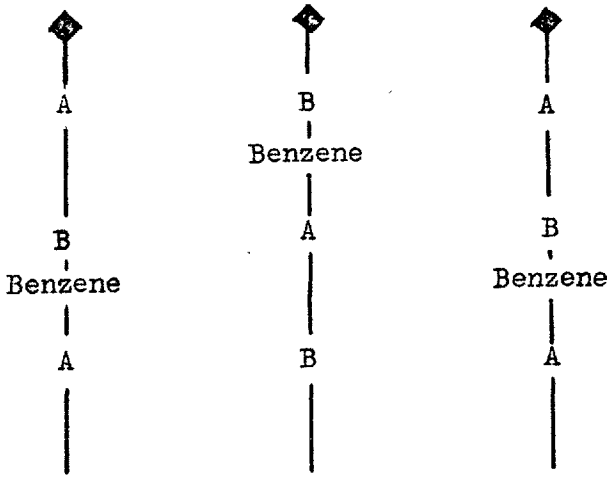
comparatively short oxygen to Csp^2 distance in the boat conformer causes an enhancement of the rate of mercuration (this was unusually rapid - see introduction).

The Crystal Structure

The crystal structure of the compound is built up from two layers of molecules which are related to each other by the body centering $\frac{111}{222}$ translation. Each layer contains four molecules per cell, and these are arranged in different patterns around the two independent four fold axes. This is shown in Figure 9, where the patterns are labelled A and B. The individual molecules in each layer are linked by inter-molecular hydrogen bonds to form continuous hydrogen bonded layers of molecules. In pattern A the hydrogen bonds occur between O_{18} and O_{21} (2.72Å) and in pattern B they occur between O_{25} and O_{26} (2.68Å). An intra-molecular hydrogen bond also occurs between O_{21} and O_{25} (2.76 Å) and participates in a string of hydrogen bonds having the sequence:-

$$O_{26} - H \cdots O_{25} - H \cdots O_{21} - H \cdots O_{18}$$

The body centering displaces the layers relative to each other such that an alternation of the A and B environments occurs in the c direction along any one tetrad. This produces "pill box" cavities of the B type, in which the benzene molecules lie. They are entrapped by the C_{20} methyl groups in the A pattern in the adjacent layers. The stacking sequence is



The benzene molecules lie closer to the plane through the molecules comprising pattern B than those in pattern A and are retained by van der Waals contacts with the various groups which protrude into or comprise the walls of the cavities. They are held from above by the C_{20} methyl groups and from below by contacts with C_{12}, C_{13}, C_{24} and O25. The interatomic distances for these contacts are:-

C20-C33	3.86Å
C12-C32	3.95Å
C13-C32	3.91Å
C24-C33	3.94Å
O25-C33	3.79Å

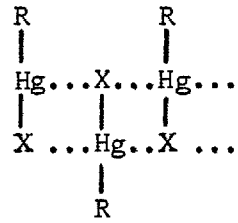
and are all longer than the sum of the van der Waals radii of the relevant atoms. Using the data summarised by Bondi[66] (R_w for C_{aliph} = 1.70Å, R_w for C_{arom} = 1.77Å and R_w for O = 1.52Å), the normal van der Waals contact distance for O - C_{arom} is 3.29Å and for C_{aliph} - C_{arom} is 3.47Å, so this suggests that the benzene molecules are only loosely entrapped in the structure, and rotational disorder is understandable.

It has been observed in the introduction that both fusicoccin and its derivatives show a tendency to include solvent molecules in the crystal structure. However, after the determination of the NMR spectrum of some crystals from the batch used for the structure determination (see above), it was observed that removal of the deutero-acetone solvent by evaporation left a transparent oil which crystallised after a short time to give good, but very small, single crystals. An optical examination of these crystals showed them to be tetragonal and very similar in appearance to those used in the structure determination. A redetermination of the NMR spectrum shown that they no longer contained benzene, implying that this is probably unimportant to the stability of the crystal structure.

Bonding between the layers seems to be purely of the van der Waals type with the close C-C contacts given in Table 9. This table includes short contacts within the layers.

The van der Waals volume of the aglycone mercuribromide molecule, calculated using the increments tabulated by Bondi[67] and a C-Hg volume (18.8 cm^3/mole) calculated according to Bondi[66] using the van der Waals radii given in that paper is 234 cm^3/mole giving a packing density[68] of 0.608. This is increased to 0.639 by the inclusion of the two benzene molecules. Both of these values are fairly low but are not surprising in view of the unoccupied space in the crystal structure.

In addition to cylindrical holes, the crystal structure contains a second feature which seems to be common in mercuric halide compounds and has been termed "anti-parallel" packing since it results in an alternation of Hg groups in the crystal structure.



Russian workers have investigated a considerable number of mercuric halide derivatives[44,45,55,64,65] which show this type of packing and similar arrangements have been reported for α and β-1-chloromercuric-2-methoxycyclohexanes[54]. In most of the above examples the arrangement is continuous, but in the present case continuity is prevented by the intervention of the bulky organic molecule and is restricted to pairs of molecules. The pairs lie symmetrically across the 4_2 screw axes on the unit cell edges at $\frac{1}{2}0\ \underline{z}$ and $0\frac{1}{2}\ \underline{z}$ and each screw generates a second pair rotated 90° with respect to the first pair and translated by one half cell edge in the $\underline{z}$ direction (*ie.* into the next layer). There is, however, no vertical bonding since the shortest contact interatomic distance in the $\underline{z}$ direction is the Hg-Br contact of 5.49Å.

A wide diversity of coordination around mercury has been reported. In trans-β-chlorovinyl-mercuric chloride[51] the three mercury atoms in the asymmetric unit have three differing environments, one a distorted trigonal bipyramid, the second an octahedron while the third atom has a coordination number of 8 (the Hg-Cl "coordinate" bond lengths range from 2.81 to 3.77Å). In the Millon complex[49] the mercury coordination is described as square planar with intramolecular Hg-O contacts of 3.05 and 2.96Å.

In methyl mercuric cyanide[46] the mercury atom is surrounded by four cyanide nitrogen atoms at distances of 2.70Å in addition to the two covalent Hg-CN bonds, but the authors concluded that there was no Hg-N coordinate bonding, and that the structure is held together solely by van der Waals forces.

Canty and Gatehouse[61] have determined the structures of diphenyl mercury complexes in which the coordination is best described as distorted octahedral with two of the four planar ligands absent. In mercuric cyanide[53] the coordination is distorted octahedral.

In a 1965 review on the structural chemistry of mercury Grdenic[69] proposed that all atoms surrounding mercury at a distance less than the sum of the van der Waals radii ($D_{Hg\text{-}Ligand} < R_w Hg + R_w$ Ligand) be considered to belong to the coordination sphere of mercury. In the present case this would mean that only the intra-molecular O26-Hg contact (2.65Å) should be termed a coordinate mercury atom. The next nearest contact is from the "anti-parallel" bromine atom at 4.07Å, some 0.75Å greater than the sum of the van der Waals radii of mercury (1.5Å) and bromine (1.8Å), followed by the Hg-Hg contact of 4.54Å. The former distance is significantly longer than the normal Hg ... Br non covalent contact in the anti parallel arrangement (3.45Å)[44] and the latter is near to that reported in methoxycarbonyl-mercuric chloride[43] (3.45Å) and longer than that in the Millon complex (3.74)[57].

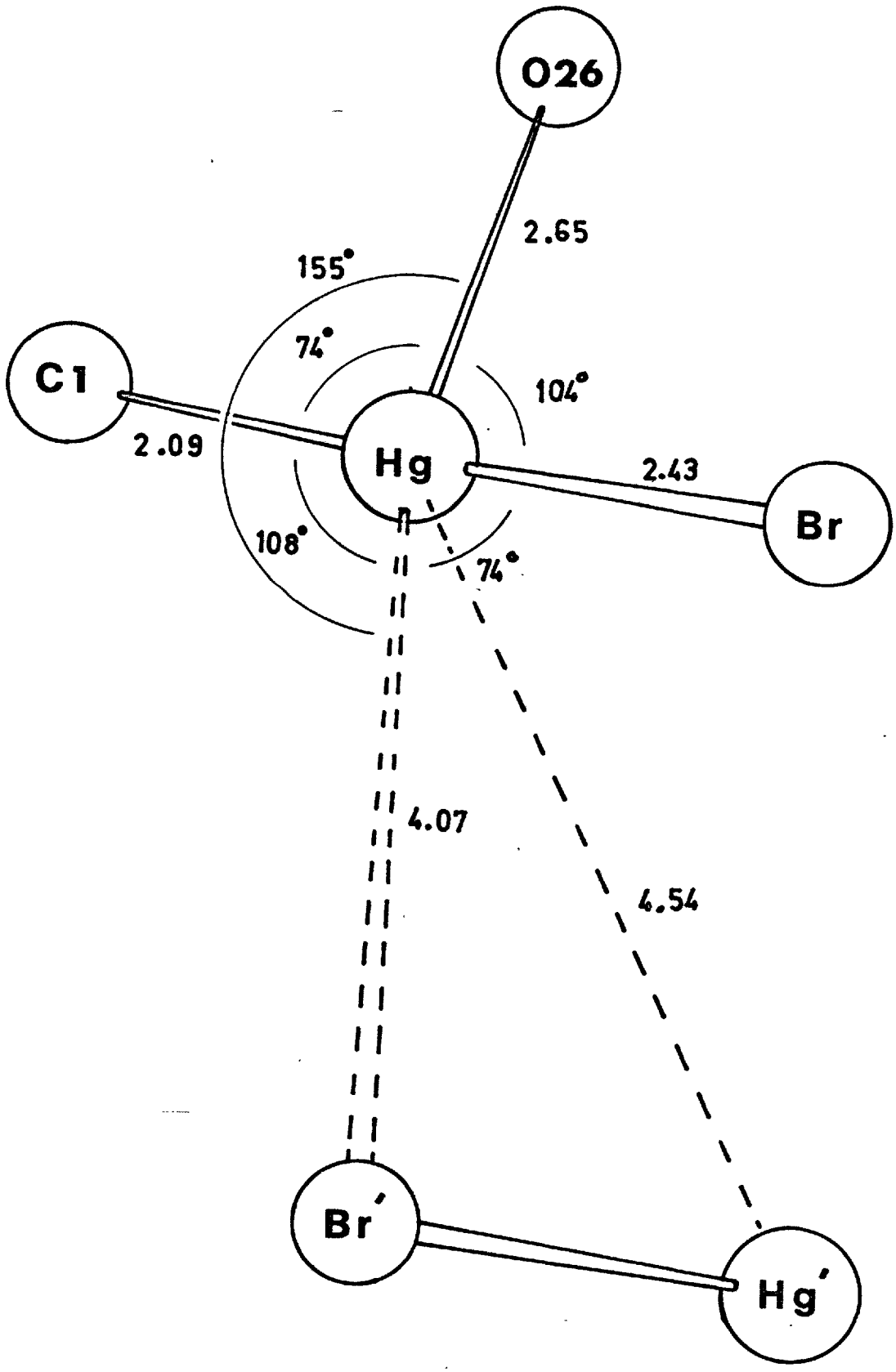


Figure 10. The mercury "coordination".

Figure 10 shows the "bond lengths" and angles associated with each mercury atom and suggests that the coordination is approximately square planar with one contact distance which is, according to Grdenic, not a coordinate bond.

The sum of the C_1-Hg-O_{26}, O_{26}-Hg - Br, C_1-Hg - Br' and Br' - Hg -Br angles is 360° which suggests that the system is planar. However, the O_{26}- Hg - Br' angle is 155° and the plane is therefore, folded along the C_1- Hg - Br bond.

The Absolute Configuration

It was mentioned in the section on the solution and refinement of the structure that the application of a correction for anomalous dispersion by mercury to the full set of F data and bromine resulted in an increase in R or 0.2% using the positive sign for Δf" whereas refinement with a negative Δf" reduced R by 0.2%. This was taken to imply that the refinement was based on a coordinate set in the wrong absolute configuration.

The statistics of the refinements were:-

		$\Delta f"_{-ve}$(normal)	$\Delta f"_{+ve}$(reversed)
1.	R	5.7	5.3
2.	R_w	6.8	6.4
3.	$\Sigma\omega\Delta^2$	0.1534×10^5	0.1358×10^5
4.	$\Sigma\omega\Delta^2/\lvert m-n \rvert$	2.542	2.392

All these quantities support the hypothesis that the absolute configuration is the reverse of that used for the structure determination (it had been selected arbitrarily at the Patterson stage).

The ratios of the R factor are:-

$$R = R_1/R_2 = 5.7/5.3 = 1.075$$

$$R_w = R_{w_1}/R_{w_2} = 6.8/6.4 = 1.062$$

and an application a Hamilton significance test[70] to the hypothesis that

the coordinates used for the structure determination are in the correct absolute configuration shows that the proposition can be rejected at a significance level very considerably lower than 0.005, the last entry in Hamilton's tables.

This absolute configuration for the compound was confirmed by visual comparison of a set of selected "Bijvoet pairs". This was the method first used by Bijvoet *et al.*[71] in 1951 for exploiting anomalous scattering, and which, with variants, has subsequently led to the determination of the absolute configuration of some 400 compounds.[72]

In space group I4, anomalous scattering introduces intensity differences between reflections of the type $\pm\underline{h}\ \pm\underline{k}\ \ell$ and $\pm h, \pm k, \bar{\ell}$. A crystal of the compound was mounted about the "b" axis in order to compare $hk\ell$ and $hk\bar{\ell}$ reflections. Using the X-ray 67 "CRYLSQ" program, the intensities of a set of Bijvoet pairs were calculated and the sign of the intensity difference $s\,(I_{\underline{h}1\underline{\ell}} - I_{\underline{h}1\bar{\underline{\ell}}})$ for pairs showing a large calculated difference was compared with the observed difference on an $\underline{h}\ 1\ \underline{\ell}$ Weissenberg film, with the following results:-

$\underline{h}\ \underline{k}\ \underline{\ell}$	calculated $s(I_{\underline{h}1\underline{\ell}}-I_{\underline{h}1\bar{\underline{\ell}}})$	observed $s(I_{\underline{h}1\underline{\ell}}-I_{\underline{h}1\bar{\underline{\ell}}})$
12 1 3	+	-
12 1 7	-	+
13 1 2	-	+
13 1 4	-	+
13 1 8	+	-
15 1 8	+	-
16 1 1	+	-
16 1 3	-	+
16 1 5	+	-
17 1 4	-	+
18 1 1	-	+
18 1 3	-	-
19 1 2	-	+

This indicates clearly that the absolute configuration is the opposite of that used in the refinement of the structure and is in agreement with that deduced from the full set of data and also that published for the fusicoccin A derivative by Vaciago *et al.* The latter was obtained by internal comparison since the glucose moiety was known from chemical studies to be in the D configuration.

The atomic coordinates (Table 2) and torsion angles (Table 5) conform to the absolute configuration.

X-radiation Damage to the Crystals

It has been mentioned in the section on the collection of the intensity data that crystals of desacetyl aglycone mercuribromide were severely damaged by X-radiation. The initial intensity data were collected from a set of five crystals and the data from one of these crystals were later re-measured using three further crystals, thus bringing the total number of crystals examined to eight. The decay was evident both in the decline of the intensity of the reference reflection and as a pronounced blackening of the crystals. In spite of the fairly large number of organo-mercury compounds which have been examined by X-ray methods, radiation damage has apparently only been reported once before[51] but, since the majority of workers have used film techniques for intensity collection, it is possible that radiation damage may have been present, but unnoticed in other cases.

In an attempt to form an impression of the chemical consequences of exposure of the compound to X-rays, a large crystal was irradiated for 250 hours with Cu-Kα radiation from a 1250 watt fine-focus X-ray tube. After this period the crystal was completely opaque but its external morphology was unchanged. The mass spectrum of the crystal was then determined by Dr. E.S. Waight and compared with that of an unirradiated crystal. In the latter case no peaks are observed higher than an m/e

of approximately 360 and this set of peaks shows the isotropic distribution of $HgBr_2$. The spectrum of the irradiated crystals, however, while similar to the normal spectrum up to m/e 360, shows a whole series of low intensity peaks above this, particularly two groups around m/e = 593 and 613. Both are probably mercury compounds but apart from observing that X-radiation seems to cause a "polymerisation", it is not possible to assign these peaks without extensive chemical studies.

Table10 contains details of the initial intensity for the test reflection, and its rate of decay together with the $\Sigma Fo/\Sigma Fc$ scale factor for each crystal and some comments about the type of decay curve. The rates of decay are given in the units "counts/reflection measured" since the rate of data collection is not constant but varies with the length of the scan, the "driving time" between reflections and, because of the way the Siemens diffractometer works, the intensity of the reflections. The time interval between measurement of successive reference reflections is in general shorter at low Sin θ because of the higher average intensity. But, as each crystal was used to scan a trapezoidal volume in reciprocal space and not a shell of an approximately constant Sin θ, the average number of reference measurements per hour is virtually independent of the crystal in use - except when the shutter stuck in the open position. The decay should be exponential, but, with one notable exception (crystal 10) was found to be linear or nearly so and in the worst of the latter cases (crystal 5) the rate given in the table is that of the best line through the data. Crystal 10, the largest crystal, showed a very rapid exponential decay and does not fit into the general sequence of decay statistics. The initial rate of decay for this crystal was 8×10^4 counts per reflection, and the overall rate (*ie.* calculated from the initial and final intensities of the reference reflection) over the 260 reflections measured was 3.1×10^4

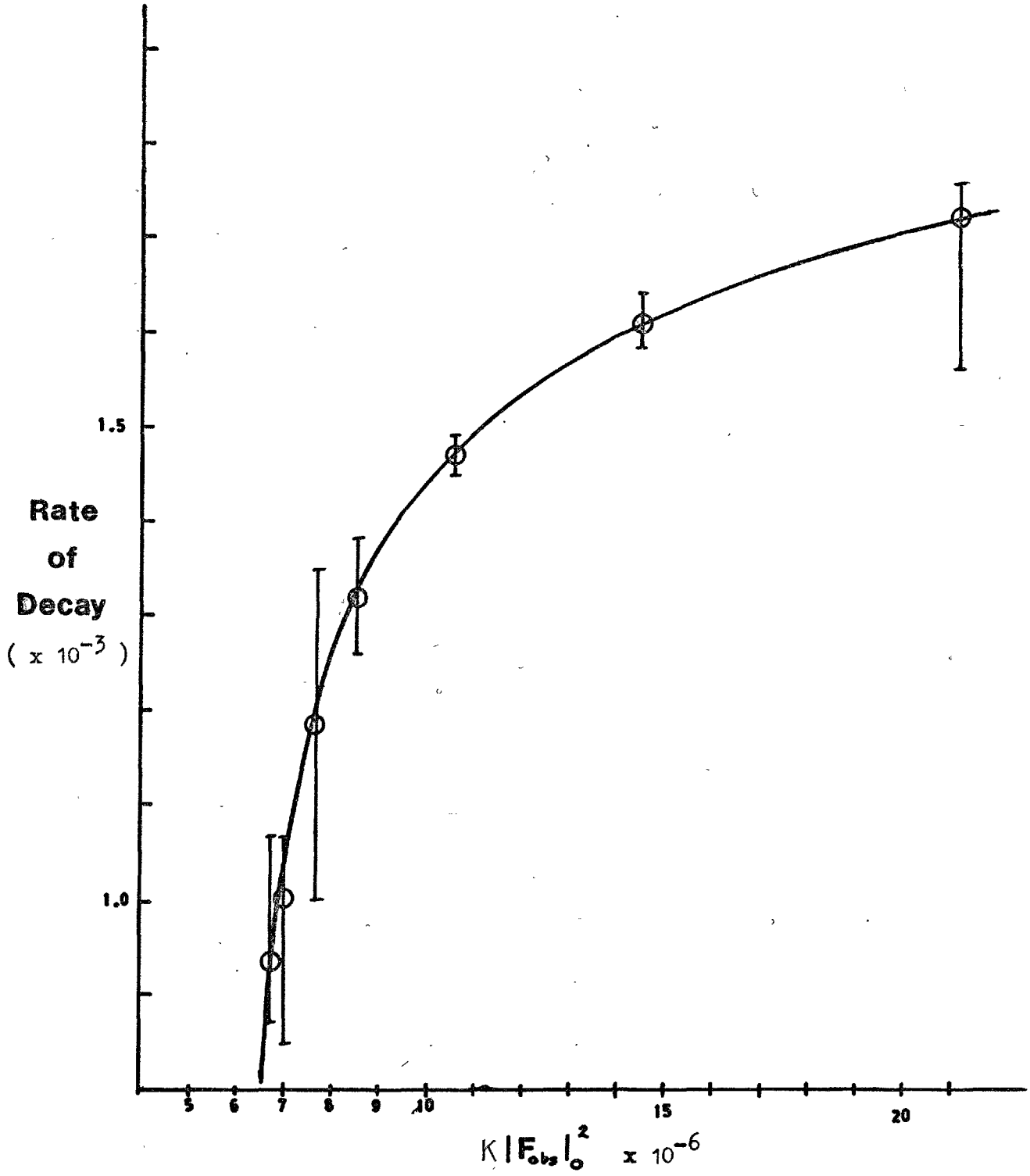


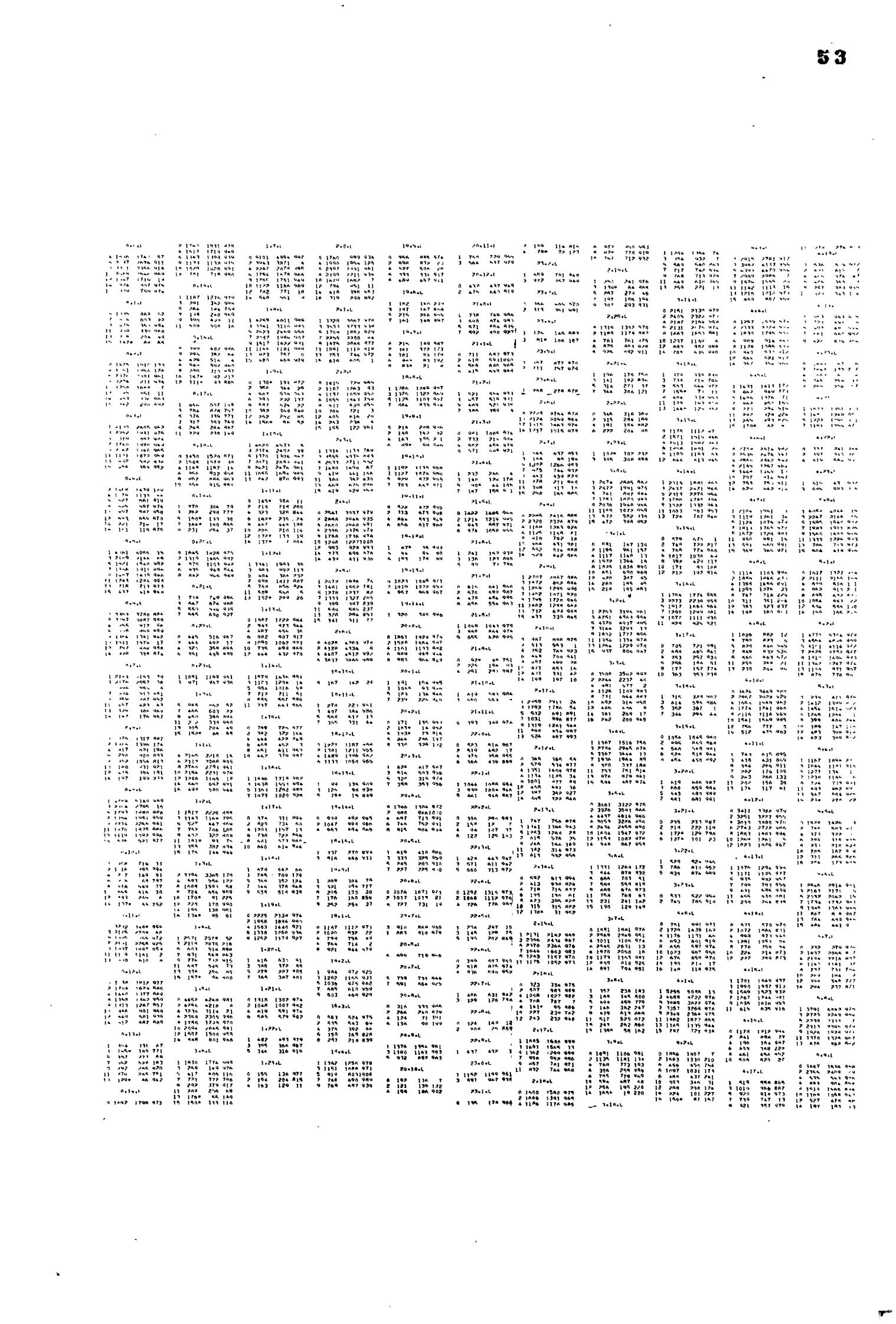
Figure 11. Relationship between initial intensity and decay rate.

counts/reflection. When the statistics of the decay were examined after the completion of the structure, it was found that crystal 10 had fallen in intensity to only 15% of its original value and that crystal 6 (used for 860 intensity measurements) had decayed to 42% of its original intensity. The R factors given in the table are the last which were calculated for individual crystals and give an overall R of 6.8% (the intensity data was treated as a single entity with seven internal scale factors after this stage). Although crystals 6 and 10 have the highest R factors they do not differ seriously from the others, and it seems that use of a crystal "too long" does not have any serious effect on the R factor. Since the original problem was one of stereochemistry, it was not considered necessary to remeasure these two sets of intensity data.

From Table 10 it is apparent that both the initial intensity and the rate of decay of the 0 0 8 reflections increase with the volume of the crystal in a roughly parallel but not wholly consistent fashion. The correlation between the initial intensity value of $K|F_{obs}|_o^2$ (or the initial intensity) for the test reflection and the rate of decay are shown in Figure 11. The possible deviations in the rate of decay given in this figure were measured directly from the decay curves and are least for the crystals which show linear decay. This is the only "smooth" relationship between the various parameters given in Table 9 and seems to imply a final situation where $K|F_{obs}|_0^2$ increases without further increase in the rate of decay. It is possible that the decay may be associated with the surface of the crystals, due to absorption of X-radiation and since decay products can only escape through the surface, whereas the intensity of scattered radiation is normally a linear function of the volume of scattering matter. Unfortunately, trial plots of the rate of decay against $(K|F_{obs}|_0)^{4/3}$ and against the surface area of the crystals do not support this, both curves being similar in form to Figure 11.

TABLE 1 Comparison of observed and calculated structure amplitudes for Fusicoccin Aglycone Mercuribromide

The data are listed in groups of constsnt h and k and consist of ℓ, $10|F_o|$, $10|F_c|$ and the phase in millicycles. Reflections marked "*" were classified as unobserved.



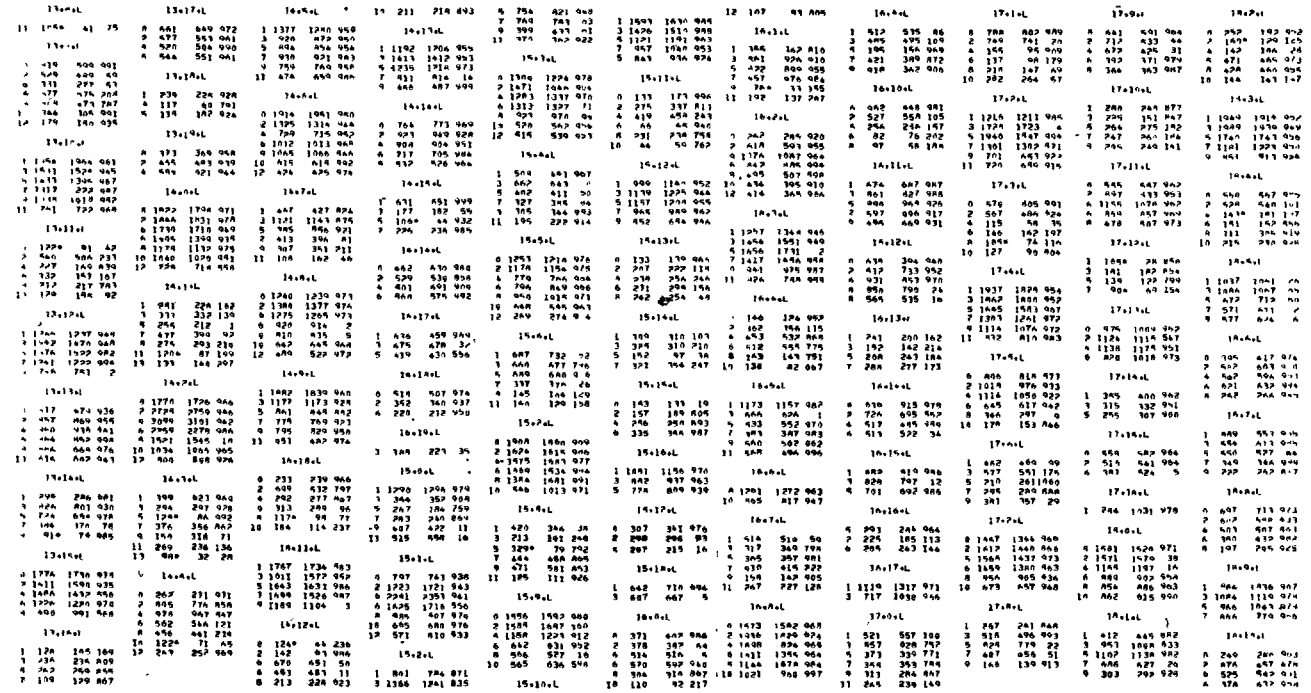


Table 2 Atom positions for non-hydrogen atoms in fractional coordinates x 10^4.

Standard deviations in parentheses.

Atom	x	y	z
HG	4523(0)	1059(0)	5000(0)
BR	5545(1)	904(1)	6012(2)
C1	3671(6)	1219(6)	4091(11)
C2	3022(7)	903(7)	4551(13)
C3	3084(8)	100(7)	4468(13)
C4	2662(11)	-139(8)	5386(15)
C5	2845(10)	369(8)	6198(14)
C6	2873(9)	1055(8)	5661(13)
C7	2196(8)	1479(9)	5621(15)
C8	2074(8)	1637(8)	4530(16)
C9	2354(6)	2314(7)	4181(13)
C10	3113(6)	2412(7)	4379(12)
C11	3594(6)	1985(6)	3732(19)
C12	4255(6)	2418(6)	3883(21)
C13	4205(7)	2745(7)	4922(13)
C14	3422(6)	2807(7)	5050(12)
C15	3422(9)	1973(9)	2620(14)
O16	2418(5)	1099(4)	4012(9)
C17	2865(9)	-173(8)	3471(15)
O18	3060(7)	-879(6)	3380(11)
C19	3757(14)	-979(12)	3077(21)
C20	1581(10)	1187(11)	6126(20)
O21	1960(5)	2855(5)	4629(10)
C22	3146(8)	3307(8)	5785(13)
C23	3563(11)	3320(10)	6742(15)
C24	3091(9)	4004(9)	5405(17)
O25	2710(7)	4044(6)	4453(12)
O26	4863(5)	2045(5)	3730(10)
C31	5049	4280	3350
C32	4689	4359	3350
C33	4410	4589	3350

Table 3 Anisotropic temperature factors x 10^3.

Standard deviations in parentheses.

The temperature factor has the form:

$$\exp\left|-2\pi^2(u_{11}a^{*2}\underline{h}^2+u_{22}b^{*2}\underline{k}^2+u_{33}c^{*2}\underline{\ell}^2+u_{12}a^*b^*\underline{hk}+u_{13}a^*c^*\underline{h\ell}+u_{23}b^*c^*\underline{k\ell})\right|$$

Atom	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
HG	28(1)	42(1)	50(1)	6(0)	-7(5)	-3(5)
BR	45(14)	109(22)	80(20)	24(13)	-26(13)	-18(17)
C1	19(62)	26(66)	32(82)	-7(51)	1(59)	-0(63)
C2	12(63)	26(73)	46(99)	-6(53)	0(63)	-3(69)
C3	42(88)	21(73)	47(88)	3(65)	0(79)	5(70)
C4	74(99)	23(81)	48(99)	-1(86)	6(98)	13(73)
C5	61(98)	38(90)	44(97)	8(84)	-9(94)	10(85)
C6	33(91)	34(89)	35(99)	-2(70)	-4(76)	2(75)
C7	25(83)	34(94)	53(98)	14(67)	16(82)	6(82)
C8	25(77)	27(84)	64(99)	3(63)	-0(80)	-7(79)
C9	17(64)	29(75)	57(99)	2(53)	0(69)	0(75)
C10	14(56)	30(71)	40(90)	6(51)	-1(61)	5(65)
C11	20(55)	24(57)	46(95)	5(46)	0(97)	-2(91)
C12	23(59)	41(70)	50(94)	-3(53)	1(99)	8(98)
C13	30(75)	35(76)	39(93)	-2(63)	3(71)	-1(75)
C14	22(65)	28(68)	33(83)	2(54)	-6(63)	0(64)
C15	51(98)	43(94)	41(99)	-0(80)	0(87)	4(83)
O16	26(50)	25(48)	48(70)	9(41)	-7(51)	0(50)
C17	53(99)	25(77)	56(98)	-2(72)	-5(93)	0(77)
O18	366(85)	28(60)	73(99)	0(59)	-13(81)	-6(66)
C19	99(99)	59(98)	81(99)	20(99)	32(99)	-4(99)
C20	46(99)	69(98)	93(99)	9(97)	27(99)	15(98)
O21	31(54)	27(50)	86(99)	12(42)	-3(61)	-3(59)
C22	40(87)	38(86)	40(97)	1(72)	7(79)	-5(77)
C23	66(99)	63(98)	42(99)	-1(97)	-1(99)	-15(99)
C24	46(98)	33(88)	73(99)	-12(93)	4(98)	-6(86)
O25	55(82)	32(62)	93(99)	1(63)	-15(83)	6(69)
O26	23(52)	45(64)	59(80)	3(47)	6(55)	-11(63)

Table 4 Positions of the hydrogen atoms in fractional coordinates x 10^3.

Atom	x	y	3
H1	375	94	346
H3	354	-13	449
H4A	329	23	660
H4B	248	36	682
H5A	323	130	614
H5B	223	193	608
H6	157	162	448
H7	230	232	343
H8	443	243	544
H9	444	319	493
H12	235	-12	336
H13A	429	278	336
H13B	308	10	291
H17A	280	-61	568
H17B	216	-13	535
H22	267	310	600
H24A	354	422	543
H24B	281	429	601

TABLE 5

Least-Squares Planes in the Structure

Planes are expressed in the form $Px + Qy + Rz + S = 0$, referred to the crystallographic axes (x, y and z in Å). Deviations of atoms from the planes are tabulated in Å $\times 10^3$.

1/ Ring A

P = 19.18	Q = 1.18	R = 2.46	S = -7.02
C(2)	-2	C(6)	2
C(3)	1	C(4)*	-611
C(5)	-1		

2/ Ring B

P = 11.44	Q = 15.68	R = -1.58	S = -4.04
C(2)	109	C(8)	180
C(6)	1	O(16)	-190
C(7)	-101		

P = 11.50	Q = 15.81	R = 0.17	S = -5.02
C(2)	-35	C(8)	37
C(6)	53	O(16)*	-430
C(7)	-55		

3/ Ring C

P = 3.11	Q = 1.16	R = 13.18	S = -6.64
C(1)	33	C(9)	-130
C(2)	401	C(10)	378
O(16)	-474	C(11)	-374
C(8)	166		

4/ Ring D

P = 0.02	Q = 14.40	R = -9.05	S = 0.50
C(10)	19	C(14)	-18
C(11)	-10	C(12)*	479
C(13)	10		

Table 5 continued

5/ The 8-membered ring

P = 8.55		Q = 8.86		R = 10.39		S = -8.66
C(1)	-191		C(8)	-728		
C(2)	-548		C(9)	-254		
C(6)	614		C(10)	687		
C(7)	370		C(11)	48		

6/ The four "planar" atoms in ring C and the 8-membered ring

P = 4.93		Q = 5.48		R = 12.39		S = -7.58
C(1)	-31		C(8)	-44		
C(2)	45		C(9)	31		

7/ The olefinic system

P = 0.34		Q = 14.33		R = -9.09		S = 0.42
C(9)	13		C(13)	22		
C(10)	0		C(14)	-33		
C(11)	-7		C(22)	5		

TABLE 6

The Torsion Angles

1/ Ring A

C6 - C2 - C3 - C4	-24.06
C2 - C3 - C4 - C5	39.34
C3 - C4 - C5 - C6	-40.06
C4 - C5 - C6 - C2	24.31
C5 - C6 - C2 - C3	0.40

2/ Ring B

O16 - C2 - C6 - C7	10.80
C2 - C6 - C7 - C8	9.07
C6 - C7 - C8 - O16	-25.76
C7 - C8 - O16 - C2	34.51
C8 - O16 - C2 - C6	-28.84

3/ Ring D

C14 - C10 - C11 - C12	20.89
C10 - C11 - C12 - C13	-29.75
C11 - C12 - C13 - C14	28.81
C12 - C13 - C14 - C10	-15.93
C13 - C14 - C10 - C11	-3.60

4/ Ring C

C11 - C1 - C2 - O16	-40.40
C1 - C2 - O16 - C8	100.41
C2 - O16 - C8 - C9	-88.69
O16 - C8 - C9 - C10	61.08
C8 - C9 - C10 - C11	-71.53
C9 - C10 - C11 - C1	79.55
C10 - C11 - C1 - C2	-34.08

5 / The 8-membered ring

C11 - C1 - C2 - C6	82.49
C1 - C2 - C6 - C7	-114.93
C2 - C6 - C7 - C8	9.07
C6 - C7 - C8 - C9	93.54
C7 - C8 - C9 - C10	-55.38
C8 - C9 - C10 - C11	-71.53
C9 - C10 - C11 - C1	79.55
C10 - C11 - C1 - C2	-34.08

TABLE 7

Hg - Br and Hg - C bond lengths and X - Hg - Y bond angle

Compound		C - Hg	Hg - Br	Angle	Method	Ref.
$HgBr_2$			2.48	180	XD	38
CH_3HgCl		2.06			MW	39
		2.06			XD	40
CH_3HgBr		2.07	2.40		MW	39
					XD	41
CCl_3HgBr			2.36		XD	42
$(CH_3)_2Hg$		2.2			ED	43
$(CN)_2Hg$		1.99			ND	45
CH_3HgCN	(Me-Hg)	2.15		180	XD	46
	(NC=Hg)	2.01		176		
	(Me-Hg)	2.08		180	ND	
	(NC-Hg)	2.05		180		
$CH_3O.OCHg$		1.96		180	XD	47
$(CN)_2Hg.KI$		2.08		180	XD	48
C_2H_5HgBr			2.50		XD	49
trans-ClC_2H_2HgBr		2.06	2.43	167	XD	50
$O(CH_2CH_2)Hg$		2.14		176	XD	51
C_6H_5HgBr		2.06	2.43	180	XD	52
$-CH_3OC_6H_{10}HgCl$		2.34		178	XD	53
$-CH_3OC_6H_{10}HgCl$		2.15		180	XD	
$(C_6H_5)_2Hg$		2.13			XD	54
$(C_6F_5)_2Hg$		2.09		176	XD	55
$(p\text{-}CH_3C_6H_4)_2Hg$		2.08		180	XD	56
$C_7H_7OHg.C_7H_6O_2N$		2.01			XD	57
		2.07				
$C_{18}H_{22}N_2Hg$		2.07		172	XD	58
		2.12				
$C_{16}H_{16}O_2Hg$		2.07		174	XD	58
$(C_6H_5)_2Hg(tmp)_2$		2.11		180	XD	59
$(C_6H_5)_2Hg(dmp)_2$		2.10			XD	
$(C_{15}H_{12}N_2)_2Hg(ClO_4)_2$		2.06			XD	60
$C_{49}H_{22}As_2F_{20}Hg_2$		2.15			XD	61
$C_{19}H_{28}O_2)_2HgCl_2$				165	XD	62

TABLE 8

Short "non-covalent" Hg - O Contacts

Oxygen Type	Hg - O Distance	Ref.
ether	2.3	52
"carbonyl"	2.57	57
	2.53	
hydroxyl	2.65	this work
ether	3.01	58
	3.19	
carbonyl	3.01	48
"hydroxyl"	3.05	57
	2.96	
methoxy	3.06	54(β)

Table 9 The shorter intermolecular contacts (Å)

In the notation used "C3 - Br 3/010" implies that the distance quoted is from C3 in equivalent position 1 to Br in equivalent position 3 and translated one unit-cell in the x direction.

C3 - Br	2/000	3.912
C4 - Br	2/000	3.902
C19 - O26	2/000	3.518
C19 - O21	3/000	3.332
O25 - C13	3/010	3.479
O25 - C12	3/010	3.456
C9 - O18	4/000	3.404
C23 - O16	6/000	3.767
C23 - O18	7/000	3.644
Br - C15	7/000	3.762
O26 - C5	$7/00\bar{1}$	3.535

Equivalent positions in $I\bar{4}$:-

1 x , y , z

2 $-x , -y , z$

3 $y , -x , z$

4 $-y , x , z$

symmetry operations 5, 6 , 7 and 8 are generated from 1 , 2 , 3 and 4 respectively by the 1/2 , 1/2, 1/2 body-centring operation.

TABLE 10

Statistics for the 8 Crystals used in data Collection and Decay of the Test Reflection

Crystal	Volume	I_0	$\|F_{obs}\|_o^2$	Scale	Decay Rate	% Decay	R	Comments
1	$.53 \times 10^{-2}$	113×10^4	676×10^4	3.316	$.94 \times 10^3$	22	7.0	Linear
2	$.66 \times 10^{-2}$	162×10^4	851×10^4	3.558	1.32×10^3	26	6.5	Linear
3	$.45 \times 10^{-2}$	156×10^4	762×10^4		1.19×10^3	25	6.5	Linear
4	$.54 \times 10^{-2}$	132×10^4	703×10^4	3.404	1.00×10^3	21	6.5	Slowly increasing
5	$.49 \times 10^{-2}$	$220^* \times 10^4$	1059×10^4	3.722	1.47×10^3	58	7.3	Lag then linear
6	$.79 \times 10^{-2}$	$245^* \times 10^4$	1446×10^4	3.376	1.61×10^3	30	5.8	Lag then linear
7	1.62×10^{-2}	$234^* \times 10^4$	2124×10^4	3.382	1.72×10^3	20	6.5	Slowly increasing
8	2.65×10^{-2}	1014×10^4	101×10^5	2.393	8×10^4 &	85	7.2	Rapid exp.

* - Values from intercept of best line or curve.

£ - Rate for best straight line.

& - Initial decay rate.

Units:- Volume - mm^3, I_0 - counts, Decay rate - counts per reflection.

REFERENCES

1. G. Delacroix, Bull. Soc. Myc.de France, 1905, 21, fasc. 3, 13.
2. G. Arnaud, Prog. agric.et. vitic., 1909, 1, 455.
3. C.M. Hoenseler and R.H. Daines, Phytopath, 1941, 31, 10.
4. E.F. Guba, Plant Dis. Rep., 1953, 37, 560.
5. C. Grosclaucle, Ann.des Epiphyties, 1956, 7, 397.
6. A. Ballio, E.B. Chain, P.de Leo, B.F. Erlanger, M. Mauri and A. Tonolo, Nature, 1964, 203, 297.
7. K.D. Barrow, D.H.R. Barton, E.B. Chain, C. Conlay, T.V. Smale, R. Thomas and E.S. Waight, Chem. Commun., 1968, 1195.
8. A. Ballio, M. Bvufani, C.G. Casinovi, S. Cerrini, W. Fedeli, R. Pellicciari, B. Santurbana and A. Vaciago, Experientia, 1968, 24, 631.
9. K.D. Barrow, D.H.R. Barton, Sir. E. Chain, U.F.W. Ohnsorge and R. Thomas, J. Chem. Soc., (C), 1971,1265.
10. M.B. Hursthouse and D. Rogers, private communication.
11. J. Chatt, Chem. Rev., 1951, 48, 1.
12. H.B. Henbest and B. Nicholls, J. Chem. Soc., 1959, 227.
13. F.H. Allen, D. Rogers and P.G.H. Troughton, Acta Cryst., 1971, B27, 1325.
14. L.E. Sutton, "Interatomic Distances (Supplement), 1964, The Chemical Society, London.
15. Based on P. Coppens, L. Lieserowitch and D. Robinowitch, Acta Cryst., 1965, 18, 1035.
16. K.D. Barrow, D.H.R. Barton, E.B. Chain, U.F.W. Ohnsorge and R. Thomas, Chem. Commun., 1968, 1198.
17. M. Morisaki, S. Nozoe, Y. Iitaka, Acta Cryst., 1968, B24, 1327.
18. Y. Iitaka, I. Watanabe, I.F. Harrison and S. Harrison, Acta Cryst., 1969, B25, 1299.
19. K.D. Barrow, D.H.R. Barton, Sir E. Chain, U.F.W. Ohnsorge and R.P. Sharma, J.C.S. Perkin I, 1973, 1590.

20. A. Itai, S. Nozoe, S. Okuda, Y. Iitaka, Acta Cryst., 1969, B25, 872.

21. J.B. Hendrickson, J. Amer. Chem. Soc., 1962, 84, 4537.

22. M.I. Davis and T.W. Muecke, J. Phys. Chem., 1970, 74, 1104.

23. M. Bixon and S. Lifson, Tetrahedron, 1967, 23, 769.

24. P. Groth, Acta Chem. Scand., 1967, 21, 2631.

25. T. Sato, H. Minato, M. Shiro and H. Koyama, Chem. Sommun., 1966, 363.

26. E. Hough and D. Rogers, Biochim. Biophys. Res. Commun., 1974, 57, 73. and Chapt. 4 of this thesis.

27. T. Sekita and S. Inayama, Acta Cryst., 1971, B27, 877.

28. C. Pascard-Billy, Acta Cryst., 1972, B28, 33.

29. V. Amirthalingam, D.F. Grant and A. Senol, Acta Cryst., 1972, B28, 1340.

30. M-ul-Haque, C.N. Caughlan, J. Chem. Soc. (B), 1967, 335.

31. K.B. Wiberg, J. Amer. Chem. Soc., 1965, 87, 1070.

32. J.B. Hendrickson, J. Amer. Chem. Soc., 1964, 86, 4854.

33. S. Ferguson, D.D. MacNicol, W. Oberhansli, R.A. Raphael, J.A. Zabkiewicz, Chem. Commun., 1968, 103.

34. J.V. Egmond and C. Romers, Tetrahedron, 1969, 25. 2693.

35. P. Groth, Acta Chem. Scand., 1967, 21, 2695.

36. M. Dobler, J.D. Dunitz and A. Mugnoli, Helv. Chim. Acta, 1966, 49, 2492.

37. J.D. Dunitz and A. Mugnoli, Chem. Commun., 1966, 166.

38. H.J. Verweel and J.M. Bijvoet, Z. Kryst., 1931, 77, 122.

39. W. Gordy and J. Sheridan, J. Chem. Phys., 1954, 22, 92.

40. D.R. Grdenic and A.I. Kitaigarodskii, Zh.Fiz. Khim., 1949, 23, 1161.

41. **See reference 68.**

42. T.A. Babushkina, E.V. Brijukhova, F.K. Velichko, V.I. Pakhamov, and G.K. Semin, Zh. Strukt. Khim., 1968, 9, 207.

43. L.O. Brockway, H.O. Jenkins, J. Amer. Chem. Soc., 1936, 58, 2036.

44. A.H. Gregg, B.C. Hampson, G.I. Jenkins, P.L.F. Jones, and L.E. Sutton, Trans. Farad. Soc., 1937, 33, 852.

45. J. Hvoslef, Acta Chem. Scand., 1958, 12, 1568.

46. J.C. Mills, H.S. Preston and C.H.L. Kennard, J. Organometallic Chem., 1968, 14, 33.

47. T.C.W. Mak and J. Trotter,

48. F.H. Kruse, Acta Cryst., 1963, 16, 105.

49. D.R. Grdenic and A.I. Kitaigorodskii, Zh.Fiz. Khim., 1949, 23, 1161.

50. V.I. Pakhamov and A.I. Kitaigorodskii, Zh. Strukt. Khim., 1966, 7, 860.

51. D. Grdenic, Acta Cryst., 1952, 5, 367.

52. V.I. Pakhamov, Zh. Strukt. Khim., 1963, 4, 594.

53. A.G. Brook and G.F. Wright, Acta Cryst., 1951, 4, 50.

54. B. Ziolkowska, Zh. Strukt. Khim., 1964, 5, 678.

55. N.R. Kunchur and M. Matthew, Chem. Commun, 1966, 71.

56. M. Matthew and N.R. Kunchur, Canad. J. Chem., 1970, 48, 429.

57. Y. Kobayashi, Y. Iitaka and Y. Kido, Bull. Chem. Soc. Japan, 1970, 43, 3070.

58. F.W. Kupper and H.J. Lindner, Z. Anorg. Chem., 1968, 359, 41.

59. A.J. Canty and B.M. Gatehouse, Acta Cryst., 1972, B28, 1872.

60. P. Luger and G. Ruban, Acta Cryst., 1971, B27, 2276.

61. A.J. Canty and B.M. Gatehouse, J.C.S. Dalton, 1972, 511.

62. A. Cooper, E.M. Gopalkrishna and D.A. Norton, Acta Cryst., 1968, B24, 935.

63. W. Kitching, Organometallic Chem. Rev., 1968, 3, 61.

64. G.F. Wright, J. Amer. Chem. Soc., 1935, 57, 1993.

65. H.J. Lucas, F.R. Heppner and S. Winstein, J. Amer. Chem. Soc., 1941, 63, 136.

66. A. Bondi, J. Phys. Chem., 1964, 68, 441.

67. A. Bondi, Physical Properties of Molecular Crystals and Glasses, Wiley and Son, New York.

68. A.J. Kitaigorodski, 1957, Organic Chemical Crystallography, Consultants Bureau.

69. D. Grdenic, Quart. Rev., 18, 303.

70. W.C. Hamilton, Acta Cryst., 1965, 18, 502.

71. J.M. Bijvoet, A.F. Peerdeman and A.J. van Bommel, Nature, 1951, 271.

72. F.H. Allen, S. Neidle and D. Rogers, Chem. Commun., 1969, 452.

CHAPTER 2

The crystal structure of Methyl-3,3,4-Trichloro-5-methoxy-indolenine-2-carboxylate.

INTRODUCTION

Recent investigations of the products and mechanism of attack by the positive chlorine ion (Cl^+) on Indole-2-carboxylic acid[1,2,3] have provided a novel route for the synthesis of substituted isatins. R.J. Bass of Pfizer (U.K.) Ltd., who aroused our interest in this matter, proposed in discussions a reaction sequence (Fig. 1) involving an unstable "intermediate compound" (Fig. 1,I). Later Muchowski[3] suggested that this intermediate consists rather of a sequence of reactions (Fig. 2) leading to the formation of a trichloro species (Fig. 2.7) which subsequently loses the elements of hydrogen chloride and carbon dioxide before tautomerising to an oxindole (Fig. 2.8). Attack by a further Cl^+ ion then yields the 3,3,4-trichlorooxindole shown in Figure 2.2.

Bass[4] found that treatment of the intermediate (Fig. 1,I) with diazomethane yielded a more stable and isolable compound (Fig. 1,II) of formula $C_{11}H_8NO_3Cl_3$ which could also be obtained from indole-2-carboxylic acid *via* its methyl ester. Spectroscopic data were compatible with two possible structures for this compound (Figs. 3,A and B). The present crystal structure determination was undertaken to resolve the ambiguity.

EXPERIMENTAL

The title compound crystallises from isopropanol, diethyl ether or benzene as fibrous yellow needles exhibiting parallel extinction.

Crystallographic Data*

($C_{11}H_8NO_3Cl_3$); M = 308.5; M.Pt. 185 - 190°C; Orthorhombic; a = 22.450(5); b = 13.292(3); c = 8.558(2)Å; U = 2554 Å^3; D_m = 1.60(3); D_c = 1.605 g.cm^{-3} for Z = 8; Space Group $Pna2_1$; $F(000)$ = 1248; μ(Cu-K_α radiation) = 41.3cm^{-1}.

*In practice the data were collected assuming the space group $Pnma$ and the refinement was carried out in the non-standard space group $Pn2_1a$. All cell position and thermal parameters and the symbolic addition are reported in space group $Pna2_1$. The final structure amplitudes, however, retain their original Miller indices.

The absent spectra (0$k\ell$, $k + \ell = 2 + 1$; $h0\ell$, $h = 2n + 1$) are consistent with the space groups Pnam and $Pna2_1$. Difficulty was experienced in measuring the density of the compound due to its high solubility in non-polar liquids and its apparent reactivity with polar liquids. The crystals were observed to have approximately neutral buoyancy in carbon tetrachloride, suggesting a density of about 1.6 g. $cm.^{-3}$. This implies a unit-cell occupancy of 8 molecules and the presence of two molecules in the asymmetric unit if the space group proves to be $Pna2_1$.

Intensity data were collected on a Siemens automatic four-circle diffractometer operated in the coupled $\omega/2\theta$ scan mode and using filtered Cu-$K\alpha$ radiation. The "5-value" measurement procedure was utilised with a maximum step time for 0.01° of 0.6 seconds. The crystal was mounted on a quartz fibre to rotate about the needle axis (b). Details of the standard procedure used on the instrument in the laboratory have been published[5].

The intensity data were collected in two parts, the first to $\theta = 35^\circ$, containing 627 reflections, and the second from 34.9 to 50°, containing 765 reflections of which 9 lay in the overlap shell. The data were processed using the "SODI" program[6] with the 8 0 0 reflection as reference, and Lorentz and polarization corrections were applied. Four reflections were discarded due to miscellaneous paper-tape errors, and, of the 1383 unique reflections processed, 1297 were found to be significantly above the backgound radiation level, i.e. their net count was greater than 2.58 (I).

SOLUTION AND REFINEMENT OF THE STRUCTURE

The general intensity distribution within the data set suggested that the structure was non-centrosymmetric and a "DATFIX"[7] calculation was therefore carried out in an attempt to resolve the space-group ambiguity. The statistics calculated by the program are compared with

the theoretical[8] values in the following table:-

	$\langle\|E\|\rangle$	$\langle E^2\rangle$	$\langle\|E^2-1\|\rangle$
Observed	0.833	0.929	0.773
Acentric	0.866	1.000	0.736
Centric	0.798	1.000	0.968

This gave a fairly decisive indication in favour of the non-centrosymmetric space group $\underline{Pna2}_1$.

The 24 chlorine atoms in the unit cell occur in sets of either 3 or 6 independent atoms in the asymmetric unit, and this inevitably produces a highly complex Patterson map. It proved impossible to deduce a set of usable chlorine positions from the calculated Patterson. There was extensive overlap of the vectors on the $U,\frac{1}{2},\frac{1}{2}$ and $\frac{1}{2},V,0$ Harker lines and the correlation between the peaks on these lines and those on the $U,V,\frac{1}{2}$ Harker section was very poor. Consideration of molecular packing did not clarify the situation.

In view of the availability within the laboratory of the Drew[9]/Neidle "PHASEM" program for direct phase determination in some non-centrosymmetric space groups by the "Symbolic-Addition Procedure"[10], an attempt was made to solve the structure by this method.

The version of the program available in the laboratory had been written for space groups in class 222. From a unique list of reflections with phases $\phi_{hk\ell}$ it appended a list of reflections with phases of the form:-

$$\phi_{hk\bar{\ell}} = m_1\pi - \phi_{hk\ell} \qquad \ldots \quad 1$$

where m_1 is a space-group and index dependent integer.

This double list was then further extended to cover a hemisphere in reciprocal space by reversing the k indices and deriving the new phases from the relationship:-

$$\phi_{(h\bar{k}\ell)} = m_2\pi - \phi_{(hk\ell)} \qquad \ldots \quad 2$$

where m_2 is a second space-group and index dependent integer.

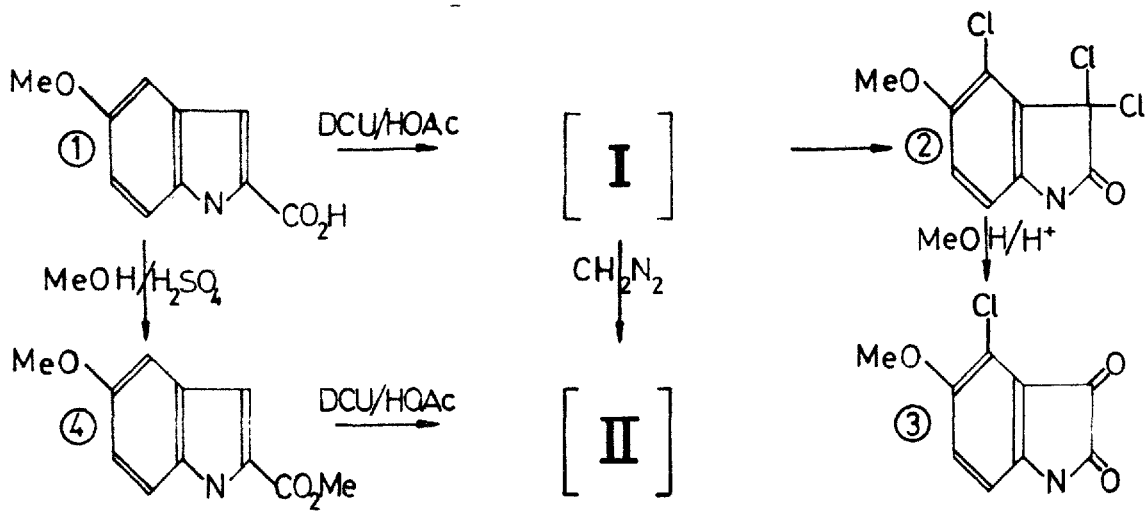


FIGURE 1

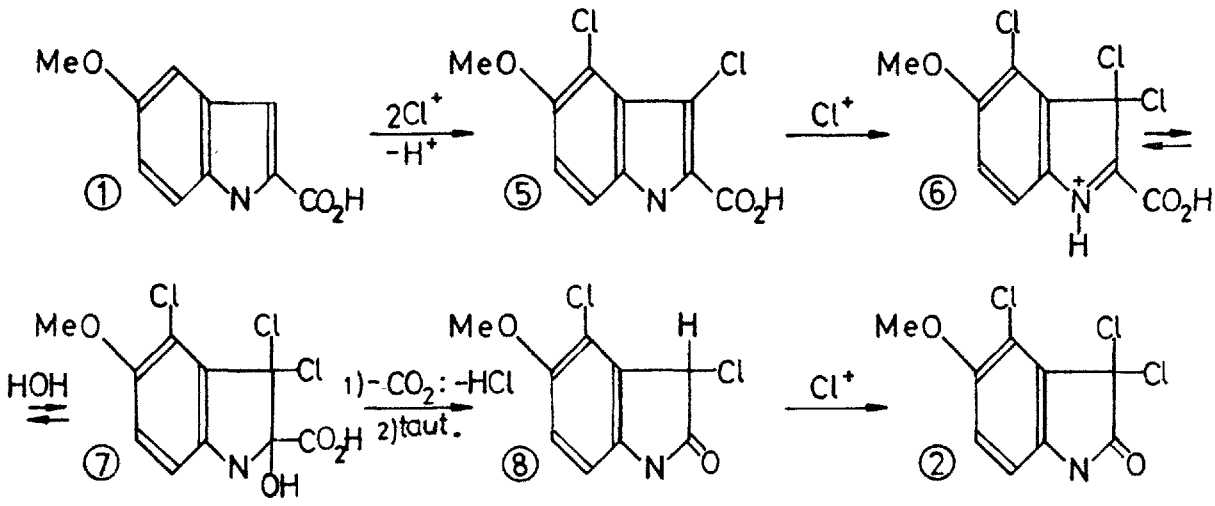


FIGURE 2

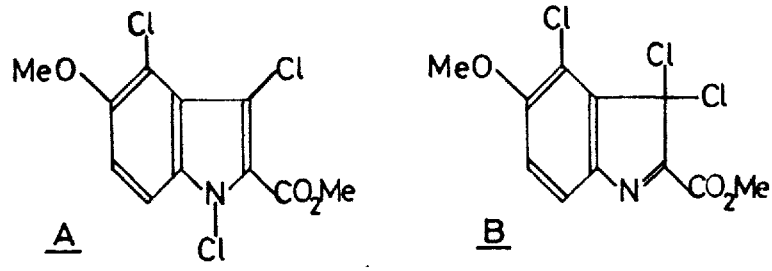


FIGURE 3

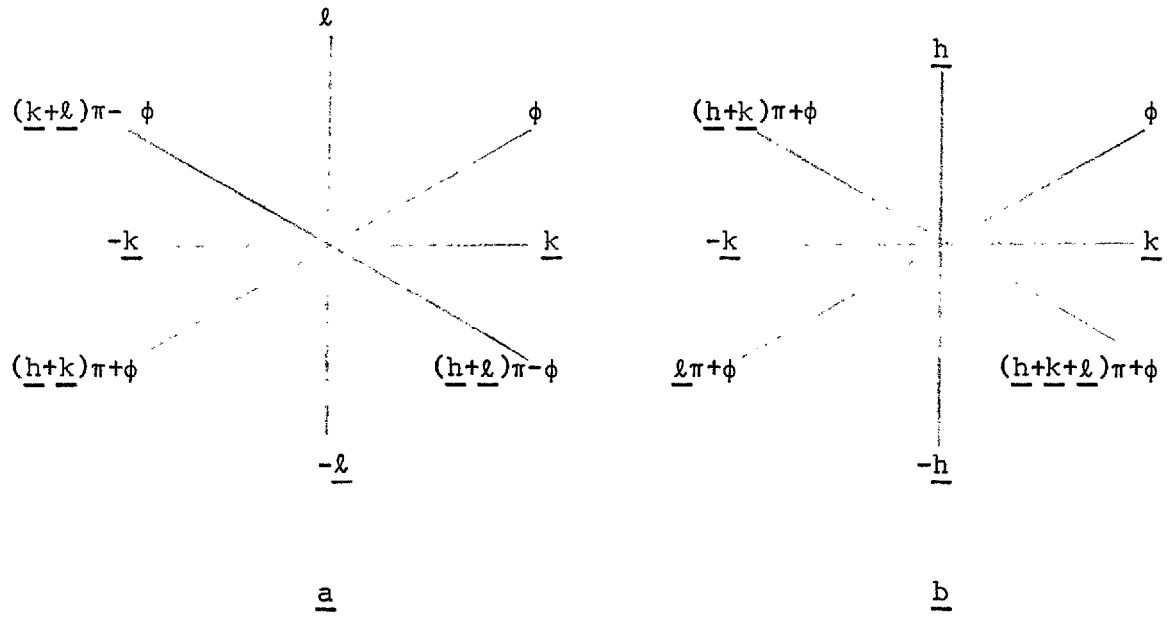


FIGURE 4

Relations between the phases for the reflections:-

a) $\underline{h} \pm \underline{k} \pm \underline{l}$ in space group $\underline{P2_12_12_1}$.

b) $\pm \underline{h} \pm \underline{k} + \underline{l}$ in space group $\underline{Pna2_1}$.

The effect of these transpositions for the four Laue equivalent reflections h ± k ± ℓ in the space group $\underline{P2_12_12_1}$ is illustrated in Figure 4a. In this the multiples of π are space-group dependent but the signs of ϕ are class dependent, conforming to the requirements of point-group symmetry 222.

Figure 4b illustrates the situation in the space group $\underline{Pna2_1}$ (class mm2), the signs of ϕ being positive if ℓ is positive. In order to accommodate these phase changes the program was altered by the present author as follows:-

Original form

$$\phi_{(hk\bar{\ell})} = \pi/2[-1 + (-1)^{(rh+sk+t\ell+1)}] - \phi_{(hk\ell)} \quad \ldots \quad 3.$$

Modified form

$$\phi_{(hk\bar{\ell})} = \pi/2[-1 + (-1)^{(rh+sk+t\ell+1)}] - (-1)^{u}\phi_{(hk\ell)} \quad \ldots \quad 4.$$

and similarly for the $\phi_{(h\bar{k}\ell)}$ transposition.

In the space group $\underline{P2_12_12_1}$ the values of the integers r,s,t,u for the generation of phases for $hk\bar{\ell}$ and $h\bar{k}\ell$ are:- 1,0,1,0; 0,1,1,0 and in the space group $\underline{Pna2_1}$ they are:- 0,0,1,0; 1,1,0,1.

The Symbolic-Addition Procedure in Outline[10]

In essence, the determination of phases for non-centrosymmetric structures by this method requires the application of two basic equations[10]:

$$\phi_h \approx \langle(\phi_k + \phi_{h-k})\rangle_{kr} \quad \ldots \quad 5.$$

and

$$\tan\phi \approx \frac{\sum_k |E_k.E_{h-k}| \sin(\phi_k+\phi_{h-k})}{\sum_k |E_k.E_{h-k}| \cos(\phi_k+\phi_{h-k})} \quad \ldots \quad 6.$$

where the subscripts h, k, h-k denote three reflections whose Miller indices are:-

$h_ik_i\ell_i$, $h_jk_j\ell_j$ and $h_{i-j}k_{i-j}\ell_{i-j}$ and $|E_h|$, $|E_k|$, $|E_{h-k}|$ are their normalised structure factors.[11]

At the start of the procedure a set of normalised structure factors[11] is calculated and a list of the reflections with high $|E|$ is examined to identify all the sets of triples h, k and h-k which occur between them. The list of triples is generally termed a Σ_2 ("Sigma Two") list.[10] The sorting may be carried out by hand but becomes very tedious when the number of reflections is large. In this case the sorting was carried out using "SIG 2", the second subroutine in the "PHASEM" program.

The probability distribution of a phase ϕ_h is given by:-

$$P_{(\phi_h)} = [2\pi I_0(\alpha)]^{-1}\exp[\alpha\cos(\phi_h - \beta)] \qquad \ldots \quad 7.$$

where

$$\alpha = [\Sigma_{k_r} 2\sigma_3\sigma_2^{-3/2}|E_h.E_k.E_{h-k}|]\ \{\Sigma_{k_r}\cos(\phi_k + \phi_{h-k}) + \Sigma_{k_r}\sin(\phi_k + \phi_{h-k})\} \qquad \ldots \quad 8.$$

$$\text{and} \quad \beta = \tan^{-1}\frac{\Sigma_{k_r}|E_k.E_{h-k}|\sin(\phi_k + \phi_{h-k})}{\Sigma_{k_r}|E_k.E_{h-k}|\cos(\phi_k + \phi_{h-k})} \qquad \ldots \quad 9.$$

I_o is a Bessel function, the suffix k_r implies that the summation is over values of k where $|E_k|$ and $|E_{h-k}|$ are large, and

$$\sigma_n = \sum_{j=1}^{N} Z^n_j \qquad \ldots \quad 10.$$

Z_j being the atomic number of the j^{th} atom and N the total number of atoms in the unit-cell.

Function 7 has a maximum when ϕ_h is equal to β, and the larger the value of α, the narrower and higher will this maximum be. This means that the larger the values of $|E_h|$, $|E_k|$ and $|E_{h-k}|$ the smaller will be the likely error, $(\phi_h - \beta)$. Accordingly, in practice, phase determination is carried out only on those reflections which have high $|E|$ values.

At the start of the phase determination the Σ_2 list is examined for reflections of high $|E|$ which enter into a large number of triples in

order to find a "starting set". This contains the terms necessary to define the origin of the cell and, if the space group is non-centrosymmetric, one other to define one enantiomorph (or the polarity); and also any reflections whose phases are directly obtainable from inequality considerations.[12] In the symbolic-addition procedure the starting set can be extended by the inclusion of a small number of extra reflections which are given symbolic phases, a, b, c, etc.

By means of these phases, equation 5 is applied to determine approximate phases for all reflections, h, for which ϕ_k and ϕ_{h-k} are known. The process is a cyclic one, newly determined phases being used in the starting set for the next cycle provided that their probability is sufficiently high. The process of propagation is usually complete after about 5-7 cycles and, if the starting set has been wisely selected, the phases of a high percentage of the reflections within the data list will then have been determined.

In the "PHASEM" program equation 5 is applied by the subroutine "PHISUM" and a function, α'_h, analogous to the Karle and Karle α function (equation 8) is evaluated. If the new phase is known in symbolic terms, an alternative function, α''_h, is evaluated, α'_h and α''_h are given by:-

$$\alpha'_h = |E_h|\{[\Sigma_{k_r} |E_k.E_{h-k}|\sin(\phi_k + \phi_{h-k})]^2 + [\Sigma_{k_r} |E_k.E_{h-k}|\cos(\phi_k + \phi_{h-k})]^2\}^{\frac{1}{2}} \quad \ldots \ 11.$$

$$\alpha''_h = \Sigma_{k_r} |E_h.E_k.E_{h-k}| \quad \ldots \ 12.$$

The calculated value for the appropriate function, α_h or α'_h, is compared with a value specified by the user, and, if greater, the phase is accepted for use in later cycles. Drew[9] recommends that the input value should have an upper limit of $1.25N^{\frac{1}{2}}$, where N is the number of atoms in the unit cell. Experience in this laboratory has suggested,

however, that values of α_h as low as $0.5N^{\frac{1}{2}}$ will give acceptable results. The new phase, itself, is evaluated in two parts, the first, a purely numeric component from equation 5, and the second, a symbolic component, representing the contributions from each symbol.

The final stage in the phase determination is the application of the "tangent formula" (equation 6). Numerical values of phase angles are given to the symbols and consequently to all the reflections which were determined symbolically and these are used as ϕ_k and ϕ_{h-k} in the tangent formula. The new values, ϕ_h, obtained in this way are then used as input into the next cycle of application of the formula if they are considered sufficiently reliable. A consistency index, t, defined in the program by:-

$$t_h = \alpha'_h / |E_h| \Sigma |E_k . E_{h-k}| \qquad \ldots \quad 13.$$

is evaluated for each reflection, h, over all contributing values of k (*ie.* over all the usable Σ_2 relationships for h). If its value is found to be less than a specifiable consistency parameter the phase ϕ_h is not accepted for the next cycle. Drew[9] recommends empirically that the limiting value of t should be 0.45.

The cyclic application of the tangent formula results in a refinement of the phases to a set with the maximum internal consistency and is generally termed "tangent refinement". The list of reflections used in the process is initially restricted to those uniquely determined by the symbolic addition and is then extended stepwise until phases have been determined for all the reflections which are involved in Σ_2 relationships and are within the list of reflections of high |E| used from the outset. The tangent formula may also be used to extend the phase determination to reflections of lower |E| or to refine phases obtained from a partial structure.[13,14]

After each cycle the program calculates the average consistency, $\bar{t}_h$, the average value of α_h and a function, Q, which is a measure of the overall

quality of the phases, Q is given by:-

$$Q = \Sigma_h |E_h - t_h E_h| / \Sigma_h |E_h| \quad \ldots \quad 14.$$

and compares with the Karle and Karle "R" factor[10]:-

$$R_K = \frac{\Sigma_{k_r} ||E_h|_{obs} - S|E_h|_{calc}|}{\Sigma_{k_r} |E_h|_{obs}} \quad \ldots \quad 15.$$

where

$$E_{h\ calc} = (A^2 + B^2)^{\frac{1}{2}} \quad \ldots \quad 16.$$

when

$$A = \langle |E_k . E_{h-k}| \sin(\phi_k + \phi_{h-k}) \rangle_k \quad \ldots \quad 17.$$

and

$$B = \langle |E_k . E_{h-k}| \cos(\phi_k + \phi_{h-k}) \rangle_k \quad \ldots \quad 18.$$

and

$$S \text{ (in 15)} = \Sigma_h |E_h|_{calc} / \Sigma_h |E_h|_{obs} \quad \ldots \quad 19.$$

The program was modified by Neidle to output the value of function 15 after the final cycle of tangent refinement.

Unless the alternative ϕ sums resulting from the symbolic addition have given indications of the numerical values of the symbols or of relationships between them, a series of tangent refinements is necessary. The symbols are assigned starting values in the range 0 to 2π at intervals of about $\pi/2$, and a refinement is carried out for all combinations of starting values. The refinement is extremely time consuming and if three symbols have been used and no information about their values has been deduced from the symbolic addition, 4x4x4 starting sets must be used. The results of each of these refinements must be examined to find the set with the best values for $\bar{t}_h$, $\bar{\alpha}_h$, Q and R. If several sets have equally good statistics an E-map must be calculated for each of them and examined for indications of the structure.

Fortunately, it is only ever necessary to examine a few of the sets. One of the reflections to which a symbolic phase has been assigned is sometimes used to define the enantiomorph and is consequently only allowed

to vary over a limited range, either $+\pi/2 \pm \pi/2$, or sometimes $0 \pm \pi/2$. In addition there are frequently indications of the values of some of the symbols from alternatives found in the symbolic addition. The combination of these two restrictions results in a very considerable reduction in the number of tangent refinements required.

The Application of the Symbolic-Addition Procedure in Space Group $\underline{Pna2}_1$

Using the "DATFIX"[7] program in the "X-RAY 63" system the data were placed on an absolute scale[15] ($B = 4.40$ Å^2, scale factor = 1.020) and a set of normalised structure factors was calculated. A Σ_2 listing over the 167 reflections having $|E|>1.5$ was prepared, which contained 6531 triple relationships. Reduction of the reflection list to the 150 strongest reflections yielded 4474 triples, a decrease which saved a good deal of computer time during the later stages of the phase determinations.

In the space group $\underline{Pna2}_1$ the standard origin lies on one of the following twofold screw axes in the unit cell:-

$0, 0, z$

$0, \frac{1}{2}, z$

$\frac{1}{2}, 0, z$

$\frac{1}{2}, \frac{1}{2}, z$

To restrict the origin to one of these axes the phases of two independent reflections in the hk0 class must be defined. Reflections in the gg0 parity class are invariant,[16] their phases being 0 or π, and for any one reflection the same for all four screw axes. Only ug0, gu0, or uu0 reflections may, therefore, be used to define the origin screw. For a unique location of the origin-carrying screw the two reflections chosen must not be in the same parity class, *ie.* must not give a term $h_1 + h_2$, $k_1 + k_2$, 0 which is in the gg0 parity class. In determinant notation:-

$$-\begin{vmatrix} P_{h_1} & P_{k_1} \\ P_{h_2} & P_{k_2} \end{vmatrix} \quad \text{must} = + \text{ or } - 1$$

where P_{h_1}, for example, represents the parity of the first h index, or more exactly is h_1 modulo 2, *ie.* 1 if h_1 is odd (u) and 0 if h_1 is even (g).

The two reflections selected were:-

8 1 0 $\quad |E| = 2.67 \quad \phi = 0$

7 8 0 $\quad |E| = 2.50 \quad \phi = 0$

In order to locate the origin in the third (c) direction the phase must be defined (anywhere in the range 0 to 2π) for an hkℓ reflection.

In determinant notation:-

$$\begin{vmatrix} P_{h_k} & P_{k_1} & 0 \\ P_{h_2} & P_{k_2} & 0 \\ h_3 & k_3 & 1 \end{vmatrix} \quad \text{must} = + \text{ or } - 1$$

The reflection selected was:-

2 4 1 $\quad |E| = 2.43 \quad \phi = 0$

In primitive space groups of the class mm2; the [001] projection is centrosymmetric and the phases of reflections in the gg0 parity class are given by the Σ_1 formula[17]:-

$$E_{2h2k0} = \frac{\sigma_2^{3/2}}{\sigma_3} \langle(-1)^{\ell}(|E_{hk\ell}|^2 - 1)\rangle_{\ell} \qquad \ldots \quad 20.$$

and the probability of the sign being positive is given by[18]:-

$$(P+) = \tfrac{1}{2} + \tfrac{1}{2}\tanh\left[\frac{\sigma_3}{\sigma_2^{3/2}} |E_{2h2k0}| \Sigma_1 (-1)^{\ell}(|E_{hk\ell}|^2 - 1)\right] \qquad \ldots \quad 21.$$

By use of these expressions the phases of five gg0 reflections were found

h	k	ℓ	|E|	φ	P(+)
14	4	0	2.42	0	0.98
10	8	0	2.21	π	0.04
12	6	0	1.92	π	0.05
6	6	0	1.84	0	0.94
14	2	0	1.59	π	0.06

In addition to these phases, equation 5 may be applied to determine the phases of reflection 2h00 and 02k0 by combining hkℓ (if present) with the equivalent reflections $h\bar{k}\bar{\ell}$ and $\bar{h}k\bar{\ell}$ respectively. The phase of 2h00 is given by:-

$$\phi_{2h00} = \phi_{hk\ell} + \phi_{h\bar{k}\bar{\ell}}$$

$$= (h+k+\ell)\pi \qquad \ldots \quad 22.$$

and for 02k0:-

$$\phi_{02k0} = \phi_{hk\ell} + \phi_{\bar{h}k\bar{\ell}}$$

$$= (h+k)\pi \qquad \ldots \quad 23.$$

Karle and Karle[10] recommend that a phase obtained in this way should be accepted if its variance, V, which they plot as a function of α(equation 8), is less than 0.5.

Using these relationships the phases of two further reflections were obtained:-

h	k	ℓ	|E|	φ	V
0	4	0	2.38	π	0.08
16	0	0	1.61	π	0.21

The starting set was then extended by the inclusion of three reflections which were given symbolic phases:-

h	k	ℓ	|E|	φ
7	1	7	2.93	a
5	4	5	2.70	b
2	11	0	2.37	c

but a preliminary application of equation 5 gave numerous indications that the phase of the 2 11 0 reflection was 0. It was therefore, given this value and a new reflection 6 3 7 ($|E|$ = 2.32) assigned the phase c.

The final starting set was therefore:-

h	k	ℓ	$\|E\|$	φ	
8	1	0	2.67	0	DEFINE
7	8	0	2.50	0	ORIGIN
2	4	1	2.43	0	
14	4	0	2.42	0	
10	8	0	2.21	π	
12	6	0	1.92	π	Σ_1
6	6	0	1.84	0	
14	2	0	1.59	π	
0	4	0	2.38	π	
16	0	0	1.61	π	Σ_2
2	11	0	2.37	0	
7	1	7	2.93	a	
5	4	5	2.70	b	
6	3	7	2.32	c	

Equation 5 was then applied by the subroutine "PHISUM" to the 150 reflections with $|E|$ greater than 1.5 and using an α'_h (see above) of 10.0, *ie.* $0.9N^{\frac{1}{2}}$ for the non-hydrogen atoms. After 5 cycles the phases of 84 reflections were uniquely determined. Alternative φ sums gave indications that the symbol "a" should have a value of approximately 0 and that b ≈ c ≈ π. Use of these values increased the number of uniquely determined phases to 116.

Karle and Karle[10] have shown that in this space group the polarity is best fixed using a reflection whose phase is not near 0 or π when

referred to the selected origin. Although none of the above listed reflections seems to fulfil this condition, it was decided to attempt the next stage in the phase determination, tangent refinement, using the 7 1 7 reflection to define the polarity. The starting values given to the symbols were:-

$a = \pi/4$

$b = 3/4\pi, \quad \pi, \quad 5/4\pi$

$c = 3/4\pi, \quad \pi, \quad 5/4\pi$

For each of the 9 combinations of these values the 84 uniquely determined reflections were subjected to 12 cycles of tangent refinement. The number was then increased to 120 for 13 more cycles and finally to 150 for the remaining 25 cycles. It was found that the average consistency, $\bar{t}_h$ (Equation 13), average α'_h (Equations 11 and 12) and the "Q" factor (Equation 14) generally converged in less than 5 cycles and, in retrospect, it appears to be unnecessary to use more than 10 cycles of tangent refinement after each increase in the number of reflections.

Examination of the results of tangent refinement revealed a very close similarity between them despite the different starting values of the symbols. This illustrates the power of tangent refinement to "pull" a set of phases into a general pattern despite wide differences in the starting values - provided one starts with a longish list of well connected $|E|$'s. The statistics obtained for almost every starting set were:-

$\bar{t}_h = 0.826$

$\bar{\alpha}_h = 156.15$

$Q = 0.17$

$R_K = 20.5$

The final phases, which are given in Table 6, refined to values which were within ± 50 millicycles of 0 or π, a strong indication that the structure was centrosymmetric or nearly so. Only two phases lay far

outside this range and they had very low consistencies:-

h	k	ℓ	|E|	φ(millicycles)	t_h
10	7	0	1.52	262	0.28
10	4	6	1.76	811	0.28

In view of this, a set of cards with centrosymmetric phases was prepared, the two reflections with low consistencies being rejected and a 3-dimensional Fourier map was computed using the 148 |E|'s whose phases were known. Examination of this map revealed the presence of 8 large peaks:-

Peak	x	y	z	Height (Arbitrary Units)
1	0.133	0.413	0.250	23
2	0.177	0.608	0.250	23
3	0.410	0.545	0.250	20
4	0.423	0.758	0.250	22
5	0.037	0.615	0.040	16
5'	0.037	0.615	0.460	16
6	0.297	0.668	0.046	13
6'	0.297	0.668	0.454	13

The first four peaks lay in crystallographically independent positions on the mirror plane at z = 1/4, and peaks 5 and 6 occurred as half-weight peaks, reproduced by the space group in the mirrored positions 5' and 6'.

Calculation showed that this arrangement could be regarded as four identical near-equilateral triangles occurring as two crystallographically independent pairs with one member of each pair of triangles arising either from disorder, or more likely, from the use of the space group Pnam for the calculation of the E-map. To comply with the space group $Pna2_1$ one must transfer 5 (or 5') and 6 (or 6') across the mirror plane. There are clearly two possibilities for this, such that the full-weight

"atoms" produced could be either cis or trans relative to the pseudo-mirror plane. Both possibilities must be considered.

It was already possible to reject the N-chloro-indole structure (Fig. 3,A), first because the shape and size of the triangles found in the E map did not match that expected from this molecule, and secondly because, if the triangles found did represent the molecular plane, no acceptable mode of packing could be envisaged - especially in view of the shortness of the b axis (8.6Å). Secondly, the observed triangles agreed well with that expected from the gem-dichloro-indole structure (Fig. 3,B) and an acceptable packing could be visualised with the plane of the molecule lying approximately perpendicular to the c axis. In addition to this, it was observed that the "intra-molecular" vector set linking the 6 peaks agreed well with the set observed around the origin in the Patterson map provided that peaks 5 and 6 were taken cis to the pseudo-mirror plane.

It was therefore, decided to reject peaks 5' and 6' and to attempt least-squares refinement of the other peaks treating them as full weight chlorine atoms and using the non-centrosymmetric space group $Pna2_1$.

The result of this refinement was encouraging, the overall agreement factor falling to 35.5% after four cycles of refinement and the temperature factors of all the atoms becoming less than 4.9Å^2. A difference map was then computed and found to contain no evidence of the mirror-image peaks 5' and 6'. The map contained 26 new peaks of height greater than 1eÅ^{-3}, 12 of which were greater than 2eÅ^{-3}. The latter were stereochemically acceptable although it was not possible to relate them to a structure. They were included as carbon atoms together with the 6 chlorine atoms in the next stage of the calculation and, after 3 cycles of least-squares refinement of the positional and isotropic thermal parameters and the overall scale factor, only one of them had an unacceptable temperature factor. The agreement had fallen to 27.8% and from the new difference

Fourier map a further 10 peaks were located. All survived refinement, and from the difference Fourier map at this stage, the remaining non-hydrogen atoms were located. The asymmetric unit found consisted of two molecules of the 3,3,4-trichloro-indolenine structure shown in Figure 3,B.

With all the atoms correctly identified and given individual isotropic temperature factors, the R factor fell to 9.0%. Examination of the difference map revealed the presence of considerable anisotropic thermal movement of the chlorine atoms and they were therefore, allowed to refine with anisotropic temperature factors, thereby almost completely removing the difference Fourier peaks in their neighbourhood and reducing R to 7.0%.

In view of the unexpectedly good agreement obtained from a crystal of apparently poor quality, it was decided to apply a correction for absorption effects using Troughton's "ICABS" program.[19] The positions of the external faces of the crystal were, therefore, measured as accurately as possible:-

h	k	ℓ	Distance (mm) from arbitrary origin
1	0	0	0.017
$\bar{1}$	1	0	0.017
$\bar{3}$	$\bar{2}$	0	0.022
0	0	1	0.402
0	0	$\bar{1}$	0.402

$\mu = 41.26\ cm^{-1}$ for Cu-K_α radiation.

A difference map was again computed and found to contain peaks of the order of 0.7 to $1.5 e\text{Å}^{-3}$ in the region of all the side-chain atoms except C9 and C29 (see below). The side-chains were therefore, refined with anisotropic temperature factors and R fell to 4.8% after 5 cycles of block-diagonal refinement.

As there are six chlorine atoms in the asymmetric unit the correction for anomalous dispersion by chlorine is considerable for reflections having a large contribution from these atoms. Unfortunately the X-ray-63 full-matrix and block-diagonal refinement programs do not apply the correction correctly in polar space groups. However, several workers[20] have reported the successful use of atomic form factors corrected for the real part of the anomalous scattering, so chlorine form factors treated in this way were used for subsequent calculations. This produced a small further improvement in the agreement between the observed and calculated structure factors, R falling to 4.6%, but the increase in the chlorine form factors seems to have been largely compensated by a 5.8% increase in the overall scale factor and a small increase in the chlorine temperature factors.

From a further difference map the positions of all 16 hydrogen atoms in the structure were identified as peaks of height 0.3 to 0.7e$\mathring{A}^{-3}$, but an attempt to refine the positional parameters and an isotropic temperature factor for each hydrogen atom was unsuccessful, the latter becoming very high or negative. Similar effects have been reported by Hvoslef[21] and Cooper and Norton[22] and have been attributed to errors in the atomic scattering factors for hydrogen atoms.

The procedure reported by Cooper and Norton[22] was then attempted, the hydrogen atoms being assigned temperature factors of 4.0$\mathring{A}^2$ while the non-hydrogen atoms were subjected to three cycles of least-squares refinement; the agreement factor fell to 3.8%. The hydrogen positional parameters were then refined holding their temperature factors and the rest of the structure fixed. Finally, six reflections were removed for suspected extinction and all the structural parameters except the hydrogen temperature factors were refined to convergence.

After 9 cycles of block-diagonal refinement, R had fallen to 3.6%, the maximum parameter shift being 0.1 times the standard deviation for

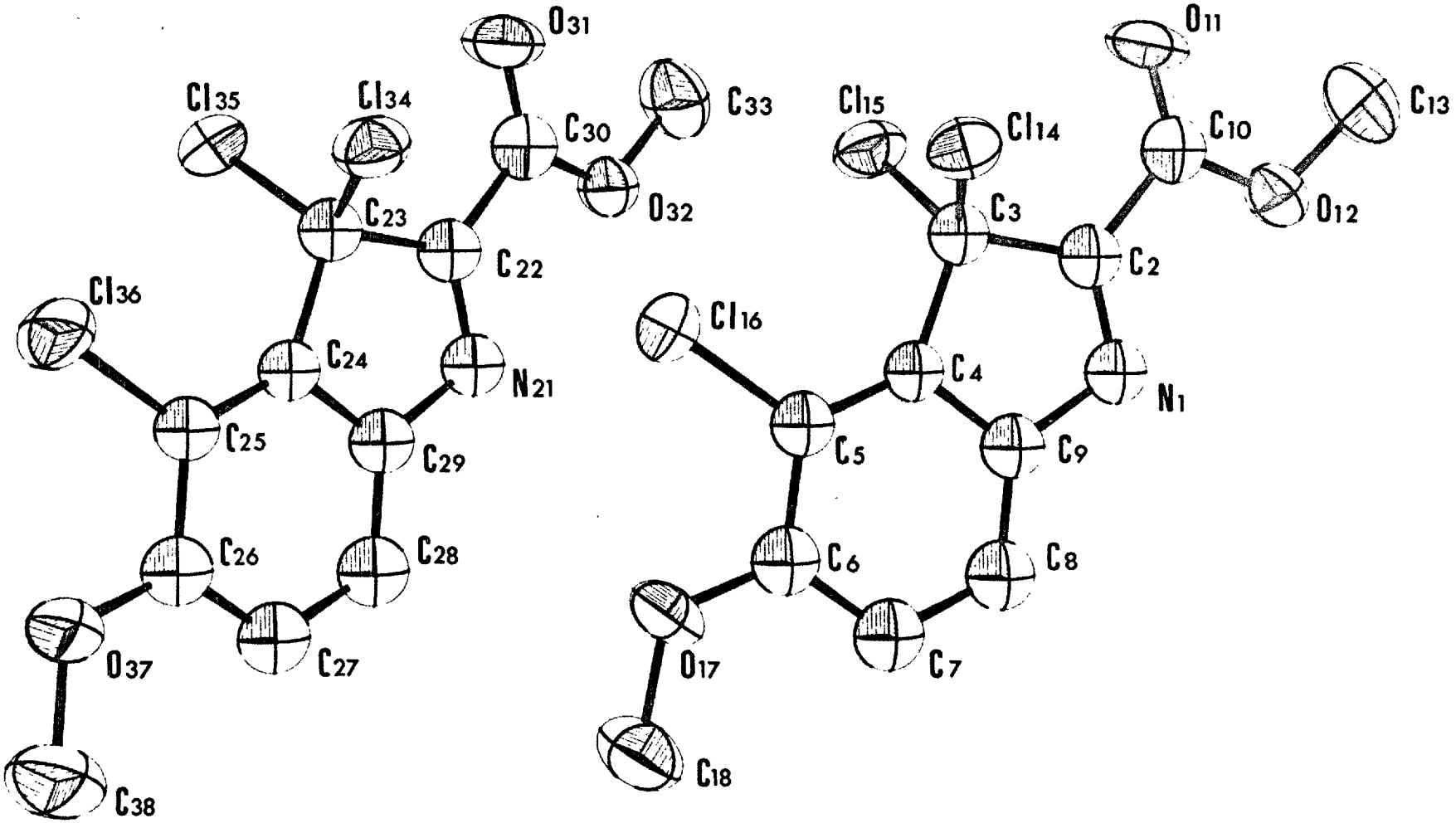


FIGURE 5 THE STRUCTURAL ASYMMETRIC UNIT

the non-hydrogen atoms and 0.1 times the standard deviation for all but two of the hydrogen parameters. The overall statistics were:-

Reflections measured	1383
Reflections lost in processing	4
Reflections removed for extinction	6
Total used in least-squares refinement	1373
R over 1373 reflections	0.0360
R over the 86 unobserved reflections	0.0540

Average Parameter Shift/Standard Deviation = 0.0549

Refinement was terminated at this stage.

A difference Fourier map computed with the final parameters was very flat, mostly in the range $\pm 0.1 e\text{Å}^{-3}$ but it also contained two peaks of height $0.2 e\text{Å}^{-3}$ neither of which could be related to any aspect of the structure.

Tables 1, 2 and 3 contain the positional and thermal parameters of the non-hydrogen atoms, Figure 5 illustrates the structural asymmetric unit and the atom numbering scheme used in the Tables. Table 4 contains the hydrogen positional parameters. The final structure-factor data are given in Table 5.

DISCUSSION

The Phase Determination

The successful refinement of the structure in the non-centrosymmetric space group $Pna2_1$ resolved the space group ambiguity and confirmed the indications for this space group from intensity statistics in spite of the preponderance of near real phases among the 150 strongest $|E|$'s (see Table 6). The distribution of normalised intensities is shown in Figure 6, an $N(z)$ plot[76] for $z \leq 1.0$. The distribution of the lower intensities is about half way between the theoretical centro- and non-centrosymmetric distributions but reflections with normalised intensities

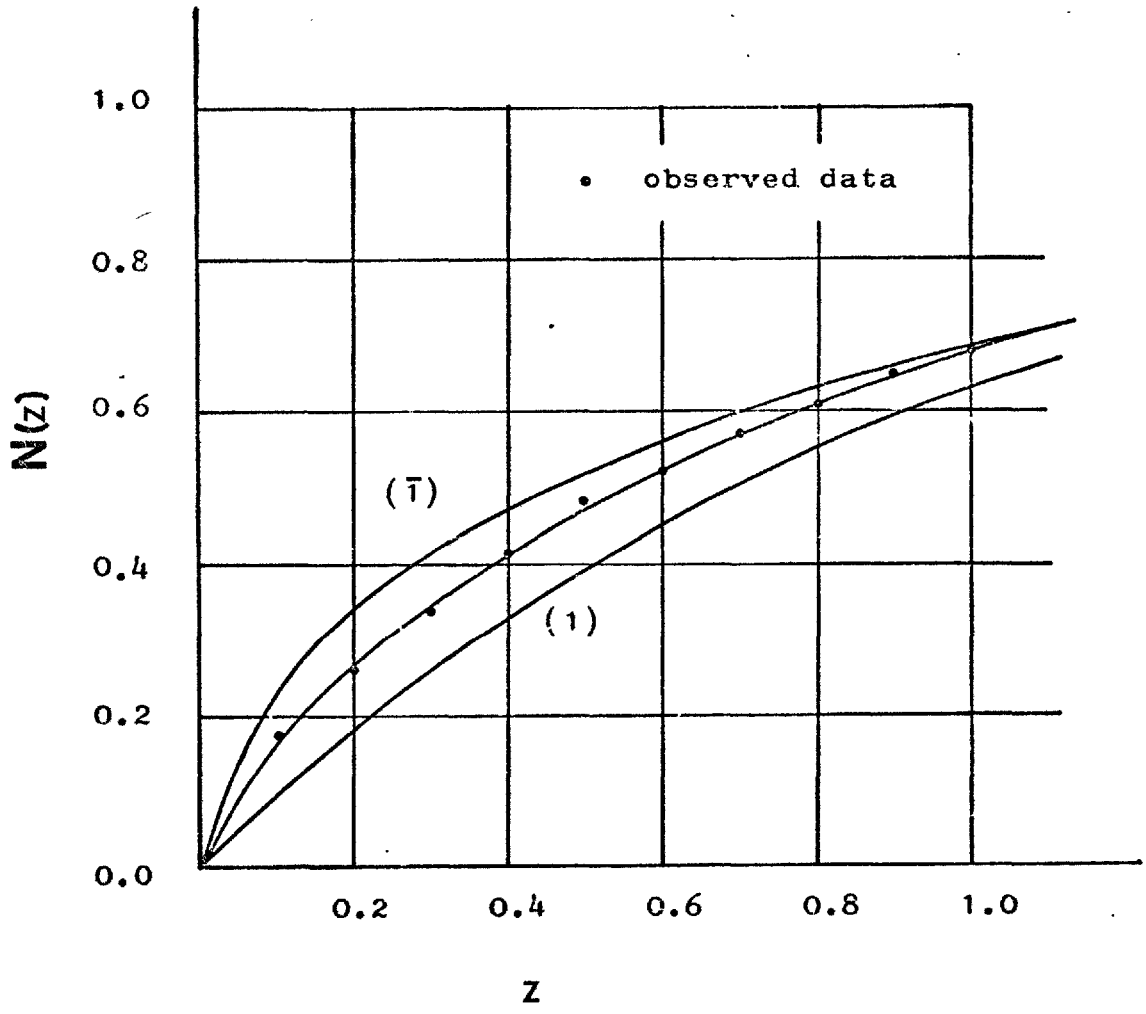


Figure 6. The distribution of observed normalised intensities compared with the theoretical curves for centro- and non-centrosymmetric data.

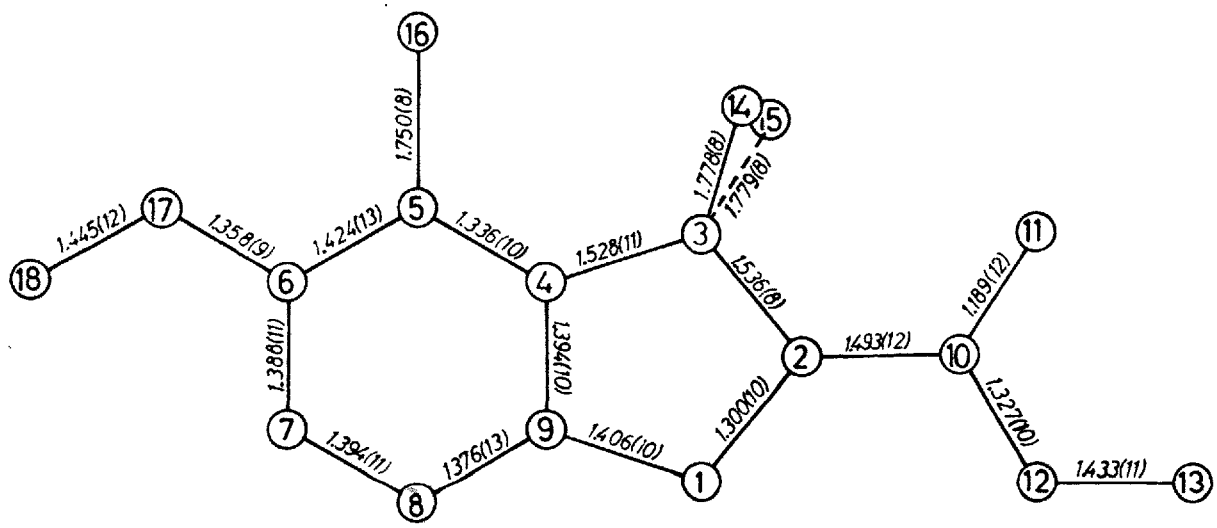
greater than 0.9 <I> show a distribution coincident with the theoretical values for a centrosymmetric structure. Since the phase determination utilises the strongest |E|'s it is not surprising that the phases from tangent refinement were close to 0 or π.

After the completion of the refinement, an E map was computed using the non-centrosymmetric phases obtained from the tangent refinement based on a = $\pi/4$, b = π, c = $3\pi/4$ rather than the centrosymmetric approximation to them. (A comparison between these phases and those calculated from the fully refined structure is given in Table 6.) The positions and heights of the major peaks in this map are given below:-

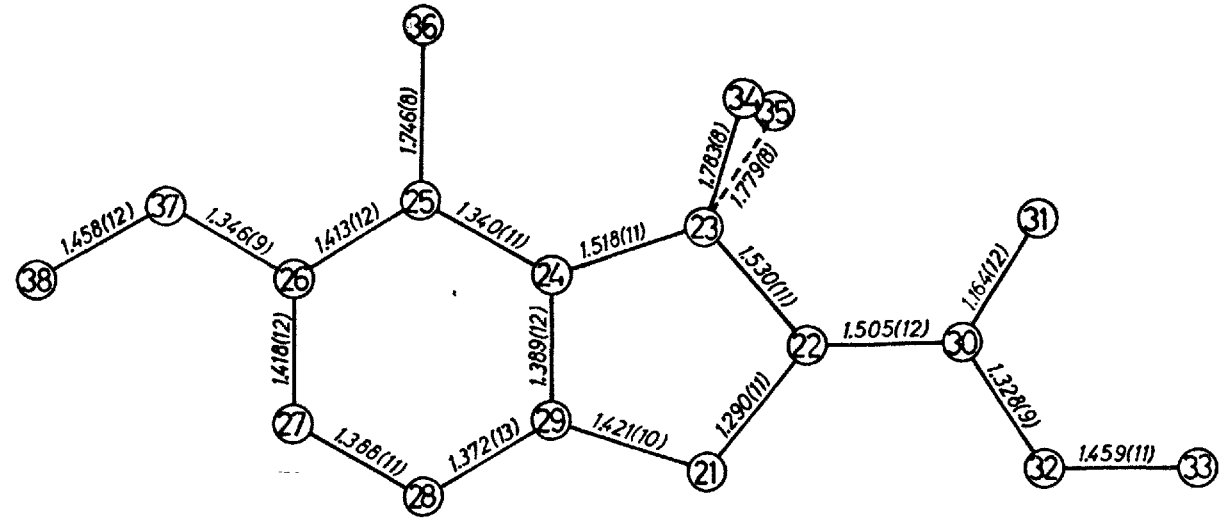
Peak	x	y	z	Height (arbitrary units)
1	0.133	0.413	0.246	25
2	0.171	0.608	0.242	27
3	0.413	0.540	0.250	20
4	0.421	0.760	0.242	23
5	0.037	0.615	0.035	15
5'	0.037	0.615	0.457	19
6	0.297	0.675	0.046	13
6'	0.297	0.675	0.457	14

Evidence of the pseudo mirror plane observed in the centrosymmetrically phased map is still present. Peaks 5' and 6' in the table above are pseudo-mirror images of peaks 5 and 6. Had the non-centrosymmetric map been computed initially, peaks 5' and 6' would have been selected as chlorine atoms as they are appreciably more electron dense, but the effect of the "wrong" selection is, in this space group, irrelevant.

It must be noted that the presence of a pseudo mirror plane in the non-centrosymmetrically phased E map could be attributed to the use of a reflection whose phase was near zero to define the polarity. However, the phases calculated from the final refined structure and listed



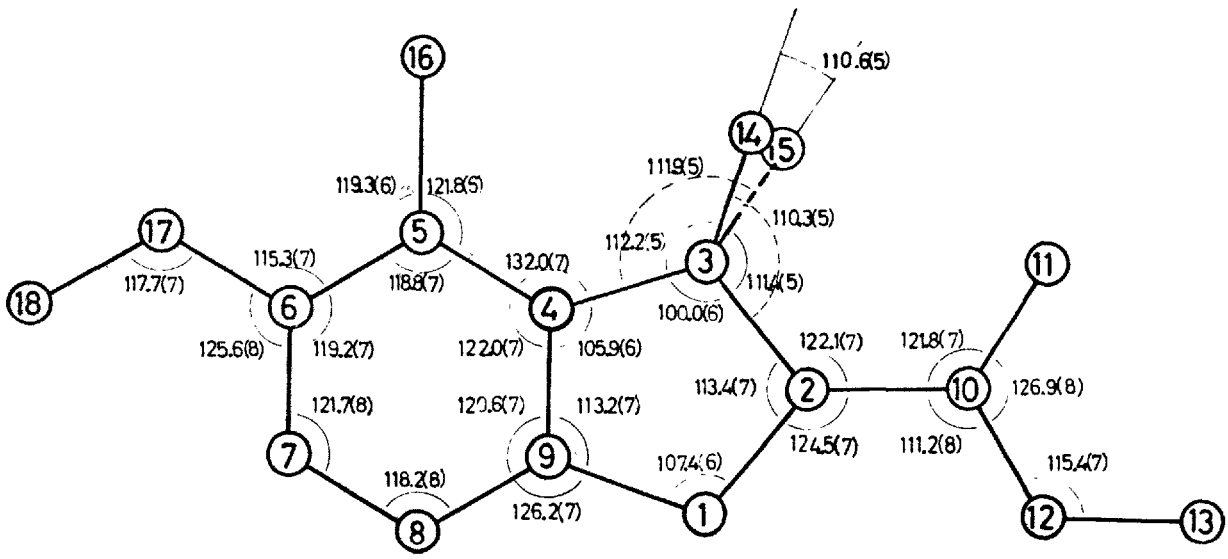
MOLECULE 1



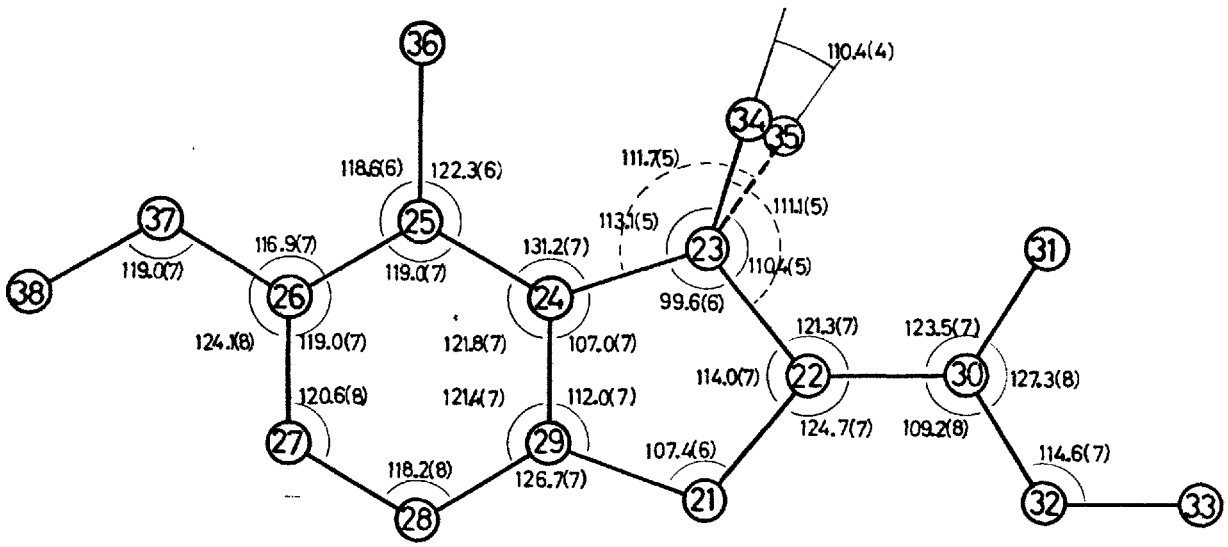
MOLECULE 2

FIGURE 7

BOND LENGTHS (e.s.d.'s in parenthesis)



MOLECULE 1



MOLECULE 2

FIGURE 8

BOND ANGLES (e.s.d.'s in parenthesis)

in Table 6, on the whole, retain the approximately centrosymmetric character found after tangent refinement. If, in the space group $Pna2_1$ all the atoms in a structure lie on the planes z = 0.25 and 0.75, centres of symmetry will be induced on the screw axes at z = 0 and 0.5. In the present structure 4 of the 6 chlorine atoms lie close to z = 0.25 and 0.75 and several of the light atoms also lie near these planes. This means that many of the vectors in the phase-amplitude diagram (including those of most of the heavier scatterers) will conform approximately to $\bar{1}$ symmetry and that this will be apparent in the final phase set.

A remarkably similar situation has recently been reported[23]. A partial structure lying on the mirror planes in the space group Pnma determined by iterative application of the Sayre equation,[24] but the remaining parts of the structure were only located after changing to $Pn2_1a$. Again the final calculated phases were near 0 or π.

The Structure

Figures 7 and 8 contain the bond lengths and angles in the two molecules comprising the asymmetric unit. It is readily apparent that to within the accuracy of the determination, the two molecules are virtually identical and that, with one exception the small differences between them are confined to the side chains and are probably attributable to packing forces. The exception is also the largest difference, the C6-C7 bond being 0.03Å shorter than that between C26 and C27. The close similarity between the two molecules also extends to the temperature factors (see tables 2 and 3).

Table 7 contains details of the least-squares planes through the indole nuclei, the aromatic and heterocyclic rings, and the diene system, N1-C2-C10-O11.

The indolenine nucleus can be accurately described as two planes intersecting at the ring bridgehead (C4-C9 and C24-C29) at a mean angle of 1.85°. The only non-hydrogen atoms, apart from the chlorine atoms

in the 3 positions, which deviate to a significant extent from the overall molecular planes are the methyl carbon atoms C18 and C38. The structure is considered in detail below:-

a) The Indolenine Nucleus

A large number of structures containing the indole nucleus in various oxidation states have been reported in the crystallographic literature, but none contains the indole nucleus in the indolenine form. This work, therefore, contributes to the completion of the series shown in Figure 9:-

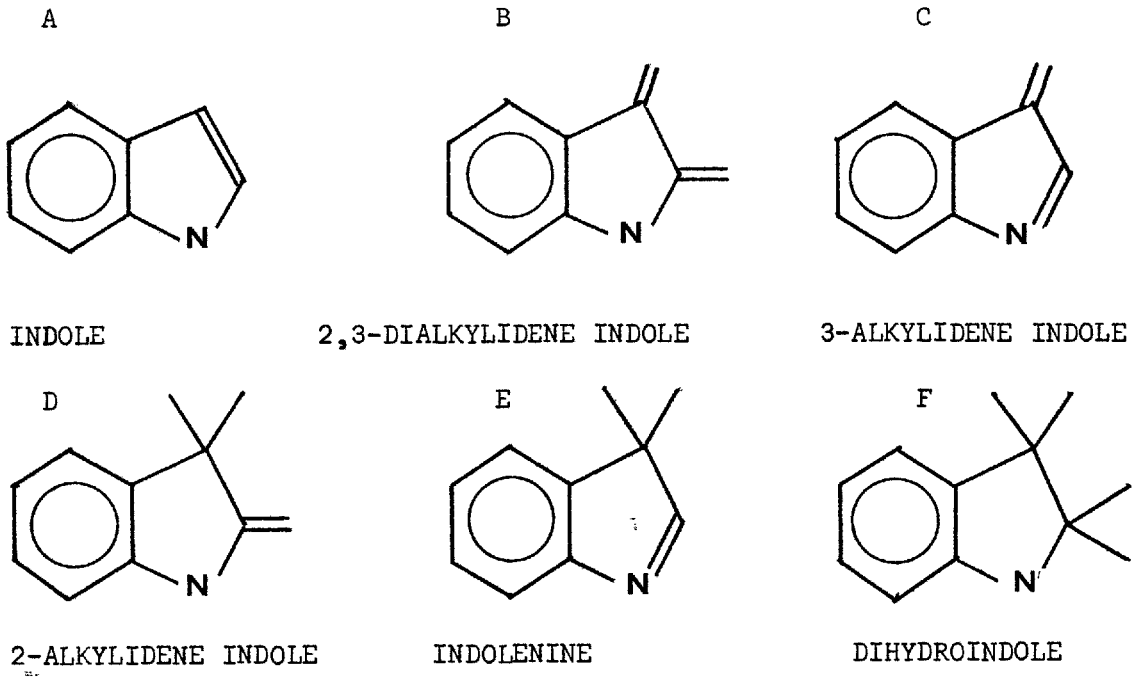


FIGURE 9

Tables 8-12 contain a summary of the reported bond lengths in the indole nucleus for nuclei A, B, D, E, and F, and Table 13 shows the estimated bond orders[25] determined from bond length versus bond-order plots for C-C and C-N bonds. Table 14 contains the mean bond lengths in the aromatic ring (ring A) throughout the series and shows them to be in good agreement with the accepted value for this bond length (1.394Å).

The bond lengths found in the indole nucleus (Figure 9,A) show good agreement with those calculated by Dewar and Gleicher[60] by self-consistent

field molecular orbital methods. It is interesting to note that, although the nucleus is a 10-π-electron aromatic system, the C2-C3 bond has predominantly double character and the N1-C2 bond rather little double-bond character.. This is consistent with the high electronegativity of the nitrogen atom and its consequent tendency to retain its electrons.

In the 2,3-dialkylidene-indole system the extension of the unsaturation results in a reversal of this situation, the N1-C2 bond order increasing considerably, and the C2-C3 bond order decreasing. This has been attributed[46] to the predominance of canonical structures having a double bond in the N1-C2 position, but it must be noted that, in all three reported structures[44,45,46] containing this ring system, the 2-substituent position is occupied either by an oxygen atom or by a carbon atom α to a carbonyl group.

No structure determinations have been reported on 3-alkylidene-indoles. Though synthetic methods are available[61,62], it is apparently only possible to isolate the products when they are stabilised by an electron donating side chain and usually as a salt after protonation of the indole nitrogen atom. Schellenberg[61] has proposed the involvement of 3-alkylidene-indoles in hydrogen transfer catalysed by yeast alcohol dehydrogenase and rabbit muscle lactate dehydrogenase.

Only two structures containing the 2-alkylidene-indole nucleus appear to have been reported[47,48], both being oxindole alkaloids. Unfortunately, the estimated standard deviations on the bond lengths are very high in both cases (0.05 and 0.07Å) and this is reflected in the marked differences between the indole nuclei in the two structures.

The structure of the indole nucleus found in the present determination is entirely consistent with the breakdown of the aromaticity of ring B. The nucleus is somewhat larger than might be expected but this is partly attributable to repulsion forces between the side chains and partly to coordination with the carboxy-methyl side chain.

The length of the N1-C2 bond (mean 1.295(7)Å) is significantly larger than that of the simple C - N double bond (1.26Å [63]) but is well within the range of values reported for C-N double bonds in coordinated systems.[64,65] The C9-N1 bond (mean 1.414(6)Å) is significantly longer than the accepted Csp^2 arom.- Nsp^2 bond, 1.39Å [63,66]. This may arise from the participation of the nitrogen atom in the conjugated trans-diene system N1-C2-C10-O11 (see below).

The C2-C3 bond (mean length 1.533(6)Å) and the C3-C4 bond (mean length 1.523(6)Å) are both longer than the accepted values for bonds of these types (1.51 and 1.505Å respectively), an effect which is probably due to repulsion forces between the substituent groups (see below).

The dihydro-indole nucleus is also very much as expected, the bond orders agreeing well with those predicted from the chemical formula. In dihydro-indole compounds the hetero ring (ring B) ahows the expected departure from planarity, adopting an envelope configuration with C2 raised from the plane through the remaining part of the molecule.

The mean value of the C2-C3-C4 bond angle in dihydro-indoles is 101° but values as widely spaced as 110° and 96° have been reported.[56,49] The angle obtained in the present determination, 99.8°, is therefore, quite acceptable for an sp^3 carbon atom at the C3 site.

From these considerations it seems legitimate to consider the present structure as one of the possible structures containing the indole nucleus and, as such, the compound falls correctly into the sequence.

b) The Substituent Groups.

1. <u>The 2-Carboxy-Methyl Group</u>

The system N1-C2-C10-O11 may best be considered as a trans 1,4-diene of the type $A{=}B{-}C{=}D$ and as such the geometry may be explained in terms of π-electron delocalisation.

A number of workers[67,68,69] have reported determinations of the structures of trans 1,4-dienes by electron diffraction and spectroscopic methods and the present results fit into the series as A and D are changed.

In butadiene where ($A=D=CH_2$) the single-bond length is 1.463(3)Å [67] and the bond has therefore considerable π character. In acrolein ($A=CH_2$, D=O) the bond length is 1.478(5)Å [68]. In glyoxal (A=D=O) the bond is further lengthened to 1.525(3)Å [69]. Qualitatively, the lengthening is related to the electronegativity of the groups A and D, a highly electronegative group tending to suppress the involvement of the Csp^2π-electrons in the Csp^2-Csp^2 single bond. On this basis, the C2-C10 bond length in the present structure should be intermediate between that in acrolein and that in glyoxal. This is, in fact, the case, the mean bond length being 1.499(8)Å.

The overall planarity of the diene system in this structure (the torsion angle about the C2-C10 bond is less than 0.1°) is also consistent with the presence of π-π interaction, but the influence of torsion on π-π orbital overlap is not large for angles less than *ca.* 20°, as the overlap integral is a cosine function.

The mean carbonyl bond length (1.177(8)Å) is considerably shorter than that found in either esters or ketones (1.233Å and 1.215Å respectively). It is also shorter than the value used by Hahn[70] in his paper on the structural characteristics of amino acids and carboxyl groups in the calculation of bond orders. A number of structures[70] have been reported with C=O bonds of approximately 1.19Å and in L-Tryptophane Hydrochloride[31] a value of 1.147Å was found. The geometry of the carboxyl group found in the latter structure closely resembles that found in the present work.

The shortness of the carbonyl bond may be partly attributed to the inductive effect of the terminal methyl group (C13), the C10-O12 bond also being significantly shorter (1.328(6)Å) than the normal value, 1.358Å.

2. The Chlorine Substituent Atoms and the 5-Methoxy Side Chain

In the gem-dichloro system Cℓ14,C3,Cℓ15 the carbon to chlorine bond lengths are close to the average found in chloroform by Soest and Peerdeman[71], (1.74Å). The Cℓ-C-Cℓ bond angle is also normal. The aromatic C-Cℓ bond, C5-Cℓ16, is rather longer (1.748Å) than the mean value for the bond calculated by Palenik, Donohue and Trueblood[72] (1.709Å) and termed an "isolated C_{arom}-Cℓ bond length" by Rudman[73].

The carbon to carbon bond length C4-C5 is exceptionally short (mean 1.338(7)Å) and is virtually a double bond, whereas the C5-C6 bond (mean length 1.418(9)Å) is rather longer than normal. Together with the elongation of the C5-Cℓ16 bond this may arise from the closeness of the non-bonded contact between O17 and Cℓ16 (mean distance 2.877(4)Å), a distance 0.31Å shorter than the sum of the van der Waals radii of oxygen (1.40Å) and chlorine (1.80Å [25]). However, the bond angles in the system tend to suggest that the opposite may be the case, Cℓ16 leaning towards O17 and away from Cℓ14 and Cℓ15 although the latter contacts (mean 3.543(2)Å) are considerably larger than any of the non-bonded Cℓ-Cℓ contacts tabulated by Rudman[73] and are only slightly less than twice the van der Waals radius of chlorine. As there are no exceptionally short non-bonded contacts to Cℓ16 or to the methoxy side chain (see Table 19) the effect cannot be attributed to packing forces. A similar effect is apparent in the bond angles around C2 and C10 although, in this case, the oxygen to chlorine distance (mean 3.239(3)Å) are not remarkable. The C18-O17-C6 bond angle (mean 118.4(3)Å) is large for an sp^3 oxygen atom and this, combined with the slightly short O17-C6 bond may, imply some π-π interaction between these atoms.

3. The Hydrogen Atoms

Table 15 contains the C-H bond lengths and 16 the corresponding bond angles. Most of the results lie within the normal range of error found in the X-ray crystallographic determination of hydrogen positions.

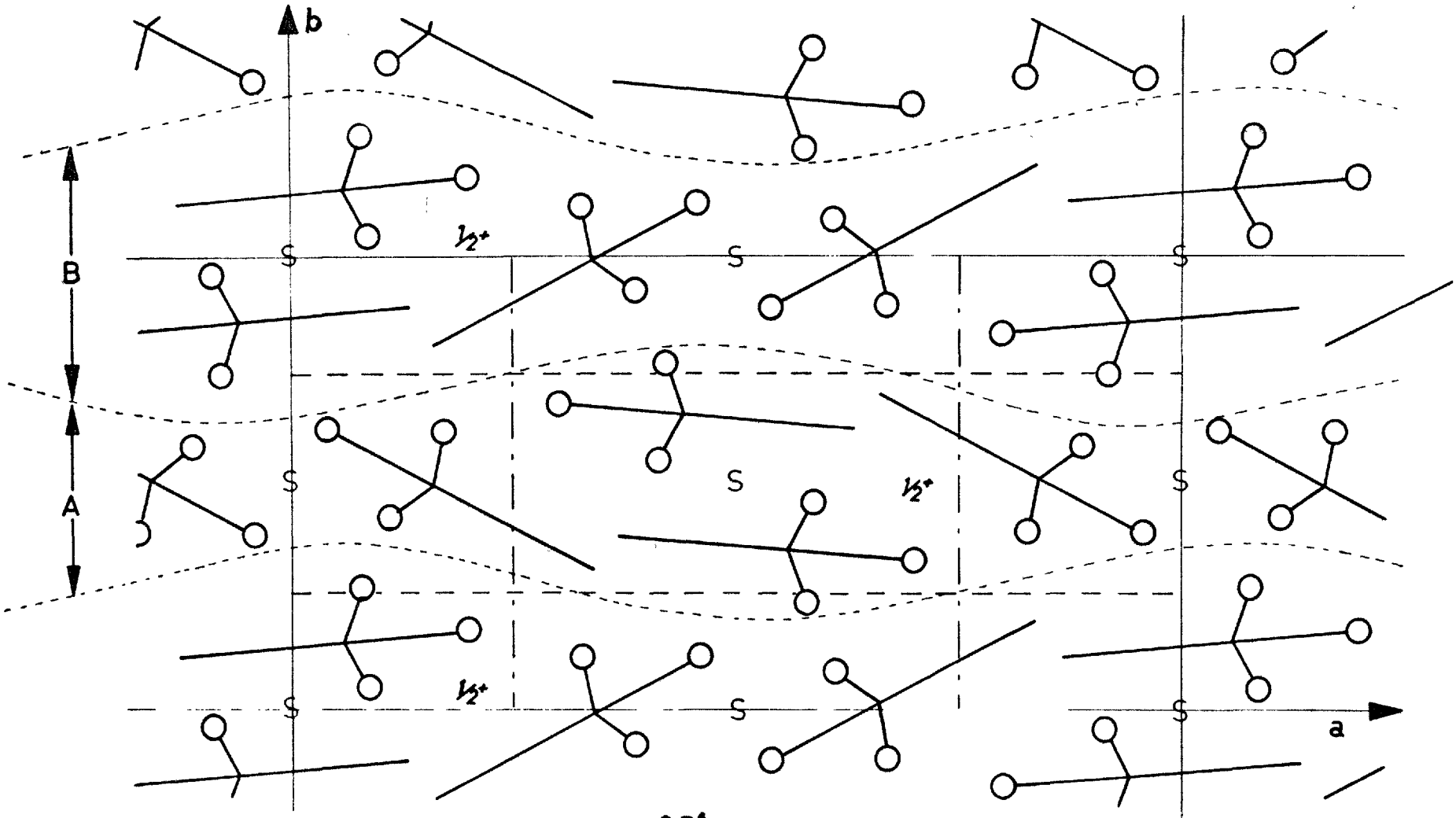


FIGURE 10 STYLISED 001 PROJECTION SHOWING LAYERING OF THE STRUCTURE

The C8-H2 and C27-H21 bond lengths are long but there is no evidence on the final difference Fourier map to suggest new positions for these hydrogen atoms. The geometry of the methyl groups is also quite acceptable, only one angle, H28-C33-H26 being exceptionally large.

Molecular Packing

From Figure 10 it can be seen that the structure consists of molecular layers of varying thickness centred on the xy0 and $xy\frac{1}{2}$ planes and related to each other by the n and a glide planes. Figure 11 shows the diad screw relationship of the molecules within a single layer and Table 17 contains the shorter intramolecular contact distances.

From these distances it is evident that the intramolecular binding forces within the crystal are solely of the van der Waals type. The van der Wasls volume of the molecule determined by the addition of the appropriate increments from the table given by Bondi[74] is 218Å^3. The "coefficient of molecular packing" or "packing density", defined by Kitaigorodskii[75] as:-

$$k = \frac{ZV_w}{U_o}$$

where Z is the number of molecules in the unit cell, V_w is the molecular van der Waals volume, and U_o is the unit-cell volume, is thus 0.679. This value is similar to that for benzene (0.697)[75] and is typical of aromatic structures.

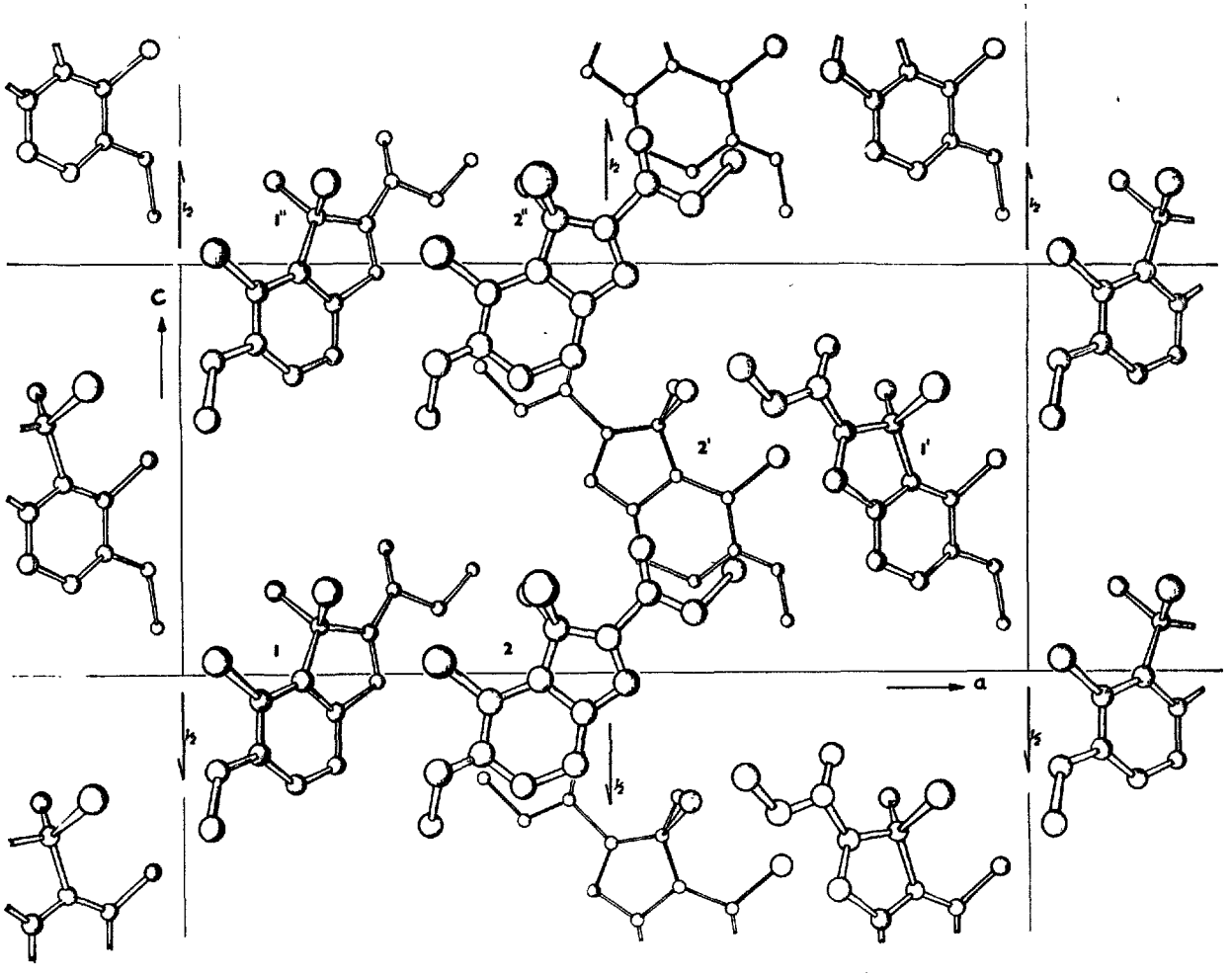


Figure 11. [010] projection of layer A of the structure.

TABLE 1 Atomic positions (fractional coordinates x 10^4)

Standard deviations in parenthesis.

Atom	x	y	z
N(1)	2271(3)	4201(5)	-307(8)
C(2)	2156(3)	4350(6)	1158(9)
C(3)	1576(3)	4937(6)	1438(9)
C(4)	1378(3)	5107(7)	-243(9)
C(5)	911(3)	5589(6)	-852(9)
C(6)	863(3)	5662(5)	-2501(11)
C(7)	1295(3)	5213(6)	-3426(10)
C(8)	1769(3)	4687(6)	-2769(11)
C(9)	1814(3)	4662(6)	-1172(10)
C910)	2543(3)	4004(5)	2467(12)
O(11)	2430(3)	4196(5)	3786(7)
O(12)	3008(2)	3470(4)	1948(6)
C(13)	3381(4)	3060(8)	3127(12)
Cℓ(14)	1061(1)	4187(2)	2500
Cℓ(15)	1715(1)	6090(1)	2424(3)
Cℓ(16)	364(1)	6142(2)	320(3)
O(17)	386(2)	6191(5)	-3028(6)
C(18)	343(4)	6367(7)	-4685(12)
N(21)	5192(3)	6300(5)	-4114(8)
C(22)	5052(3)	6351(6)	1043(10)
C(23)	4385(3)	6484(6)	1341(9)
C(24)	4161(3)	6498(5)	-327(9)
C(25)	3612(3)	6617(6)	-912(9)
C(26)	3532(3)	6588(5)	-2542(11)
C(27)	4037(4)	6459(6)	-3513(11)
C(28)	4602(3)	6353(6)	-2873(11)
C(29)	4656(3)	6388(6)	-1283(10)
C(30)	5490(3)	6298(5)	2366(12)
O(31)	5352(3)	6319(5)	3672(7)
O(32)	6038(2)	6234(5)	1794(6)
C(33)	6506(4)	6160(8)	2965(12)
Cℓ(34)	4097(1)	5440(2)	2396(3)
Cℓ(35)	4239(1)	7628(1)	2358(3)
Cℓ(36)	3985(1)	6737(2)	276(3)
O(37)	2972(2)	6702(5)	-3074(6)
O(38)	2856(4)	6572(8)	-4732(12)

TABLE 2 Anisotropic temperatures factors x 10^4

Standard deviation in parentheses.

(The anisotropic temperature factor is given by:-

$\exp - (\beta_{11}h^2 + \beta_{22}k^2 + \beta_{33}\ell^2 + \beta_{12}hk + \beta_{13}h\ell + \beta_{23}k\ell)$

Atom	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}
O(11)	30(2)	129(6)	9(1)	20(3)	-4(3)	-14(5)
O(12)	18(1)	86(4)	13(1)	9(2)	-2(3)	-2(5)
C(13)	24(2)	101(7)	23(2)	14(4)	-1(1)	1(1)
Cℓ(14)	20(1)	76(1)	115(3)	-3(1)	10(1)	8(2)
Cℓ(15)	24(1)	67(1)	111(3)	-2(1)	6(1)	-18(2)
Cℓ(16)	22(1)	84(2)	134(3)	16(1)	10(1)	-1(2)
O(17)	22(1)	95(5)	12(1)	8(2)	-3(3)	22(5)
C(18)	36(3)	10(1)	16(1)	8(4)	-13(6)	4(1)
O(21)	30(2)	144(6)	8(1)	7(3)	-10(3)	-18(6)
O(22)	18(1)	112(5)	13(1)	3(2)	-10(3)	6(6)
C(33)	27(2)	11(1)	22(2)	5(4)	-26(6)	6(6)
Cℓ(34)	27(1)	107(3)	69(3)	-3(1)	7(1)	16(2)
Cℓ(35)	23(1)	66(1)	115(3)	3(1)	8(1)	-23(2)
Cℓ(36)	16(1)	118(2)	134(3)	1(1)	6(1)	-5(3)
O(37)	30(2)	129(6)	9(1)	20(3)	-4(3)	-14(7)
C(38)	30(3)	11(1)	15(2)	6(4)	-22(6)	-2(1)

TABLE 3 Isotropic temperature factors

Atom	B(Å^2)	Atom	B(Å^2)
N(1)	3.71(13)	N(21)	3.82(14)
C(2)	3.71(16)	C(22)	3.83(17)
C(3)	3.53(16)	C(23)	3.55(16)
C(4)	3.17(15)	C(24)	3.19(15)
C(5)	3.35(16)	C(25)	3.61(17)
C(6)	4.00(16)	C(26)	4.11(16)
C(7)	4.32(18)	C(27)	4.54(18)
C(8)	4.51(17)	C(28)	4.50(18)
C(9)	3.73(16)	C(29)	3.77(16)
C(10)	4.23(16)	C(30)	4.21(16)

TABLE 4 Hydrogen positions (fractional coordinates x 10^3)

Atom	x	y	z
H(1)	118(3)	518(6)	-460(10)
H(2)	199(3)	398(6)	-346(10)
H(3)	68(3)	675(6)	-504(10)
H(4)	35(3)	585(6)	-512(10)
H(5)	-1(3)	678(6)	-480(10)
H(6)	370(3)	272(6)	284(10)
H(7)	308(3)	249(6)	353(11)
H(8)	351(3)	358(6)	366(11)
H(21)	384(3)	608(6)	-359(10)
H(22)	501(3)	623(6)	-462(10)
H(23)	313(4)	712(6)	-538(10)
H(24)	245(4)	665(6)	-480(11)
H(25)	295(3)	586(6)	-488(11)
H(26)	650(3)	548(6)	316(10)
H(27)	685(3)	624(5)	248(12)
H(28)	646(3)	675(6)	353(11)

TABLE 5 Comparison of observed and calculated structure amplitudes for Methyl-3,3,4-Trichloro-5-Methoxy-Indolenine-2-Carboxylate

The data are listed in groups of constant h and k† and consist of ℓ†, $10|F_o|$, $10|F_c|$. Reflections marked "*" were classified as unobserved.

† N.B. As explained in a footnote in the text, the data were collected and are also presented in this list only with b anc c (k and ℓ) interchanged, *ie.* as if for space group Pnma. The whole of the rest of this paper treats the cell and space group as $Pna2_1$

0, 0, L

6	46*	31
8	611	633
10	358	352
12	50*	12

0, 1, L

3	503	515
5	804	834
7	410	411
9	546	545
11	363	357
13	47*	41

0, 2, L

0	2308	2307
2	457	437
4	360	335
6	956	952
8	347	350
10	140	137
12	60	56

0, 3, L

1	601	580
3	385	381
5	852	864
7	142	143
9	491	502
11	125	123

0, 4, L

0	2179	2147
2	299	281
4	407	380
6	570	565
8	206	206
10	175	170

0, 5, L

1	432	456
3	177	177
5	501	487
7	96	105
9	203	187

0, 6, L

0	1099	1118
2	321	324
4	388	406
6	621	641
8	174	164

0, 7, L

1	130	129
3	190	226
5	280	285
7	57	47

0, 8, L

0	719	736
2	223	227
4	109	127

1, 0, L

1	92	90
2	520	546
3	1700	1754
4	1089	1141
5	360	382
6	224	232
7	944	978
8	578	608
9	599	593
10	48*	33
11	348	360
12	114	84
13	83	44

1, 1, L

2	792	806
3	174	203
4	1241	1290
5	616	639
6	164	164
7	381	395
8	336	366
9	151	171
10	156	154
11	225	225
12	115	107
13	92	87

1, 2, L

1	464	456
2	886	885
3	189	165
4	615	630
5	90	86
6	38	33
7	397	401
8	60	66
9	268	263
10	89	70
11	151	141
12	89	101

1, 3, L

1	736	695
2	489	498
3	383	394
4	778	801
5	459	461
6	83	77
7	130	140
8	79	75
9	65	52
10	178	164
11	125	120
12	162	150

1, 4, L

1	519	489
2	628	615
3	768	735
4	73	54
5	623	620
6	70	75
7	266	268
8	199	210
9	346	349
10	53	45
11	77	86

1, 5, L

1	336	345
2	207	202
3	202	205
4	360	365
5	317	320
6	447	459
7	54	71
8	146	145
9	84	89
10	275	274

1, 6, L

1	106	111
2	254	264
3	309	316
4	159	162
5	262	268
6	87	100
7	223	218
8	50*	37
9	182	191

1, 7, L

1	383	371
2	313	317
3	138	145
4	273	298
5	365	399
6	70	72
7	120	137

1, 8, L

1	198	194
2	155	162
3	53*	20
4	144	155

2, 0, L

0	514	501
2	279	275
3	533	525
4	733	756
5	615	625
6	277	281
7	731	749
8	802	817
9	98	104
10	285	309
11	530	529
12	305	304
13	70	63

2, 1, L

0	432	415
1	876	853
2	639	665
3	833	863
4	1868	1933
5	102	111
6	272	312
7	301	310
8	281	284
9	217	219
10	438	438
11	157	154
12	150	154
13	76	95

2, 2, L

0	259	246
1	285	279
2	253	256
3	165	158
4	729	736
5	1068	1096
6	313	325
7	201	192
8	350	359
9	169	177
10	273	262
11	183	176
12	141	144

2, 3, L

0	94	104
1	174	178
2	495	468
3	908	883
4	827	820
5	158	159
6	220	224
7	575	581
8	158	153
9	162	159
10	354	356
11	203	188
12	94	87

2, 4, L

0	146	137
1	405	392
2	354	327
3	227	219
4	443	433
5	592	595
6	409	407
7	255	255
8	49*	36
9	52*	54
10	132	131
11	259	253

2, 5, L

0	179	175
1	222	221
2	169	146
3	299	286
4	736	731
5	222	224
6	145	154
7	163	162
8	119	134
9	162	159
10	292	282

2, 6, L

0	193	196
1	240	259
2	168	173
3	231	228
4	293	302
5	433	441
6	279	289
7	227	234
8	212	223
9	105	115

2, 7, L

0	165	153
1	223	213
2	99	72
3	100	85
4	352	383
5	60	73
6	52	45
7	66	74

2, 8, L

0	110	123
1	218	224
2	185	190
3	91	102
4	82	93

3, 0, L

2	1213	1234
3	1734	1750
4	1167	1207
5	491	511
6	55	48
7	39*	26
8	524	544
9	363	372
10	78	73
11	51*	53
12	50*	63
13	47*	7

3, 1, L

1	527	525
2	964	974
3	1088	1093
4	494	519
5	294	285
6	405	436
8	174	177
9	338	316
10	136	154
11	154	157
12	173	150
13	128	123

3, 2, L

1	720	669
2	341	328
3	619	597
4	123	115
5	551	568
6	397	402
7	84	85
8	579	595
9	193	185
10	142	131
11	131	137
12	78	76

3, 3, L

1	660	620
2	829	806
3	536	520
4	228	224
5	140	140
6	384	393
7	324	307
8	93	89
9	511	520
10	133	145
11	129	128
12	183	173

3, 4, L

1	537	508
2	385	353
3	344	321
4	270	259
5	294	289
6	121	122
7	229	219
8	343	345
9	222	223
10	192	180
11	123	118

3, 5, L

1	632	595
2	217	217
3	476	446
4	144	147
5	417	410
6	72	74
7	135	145
8	93	114
9	94	86
10	75	68

3, 6, L

1	465	464
2	65	39
3	434	433
4	289	292
5	336	349
6	237	218
7	103	118
8	235	259
9	110	112

3, 7, L

1	115	132
2	402	404
3	217	222
4	62	58
5	141	126
6	206	229
7	81	59

3, 8, L

1	404	414
2	71	35
3	299	318
4	92	100

4, 0, L

0	537	522
1	157	158
2	732	762
3	742	741
4	1245	1288
5	577	579
6	40	32
7	381	399
8	197	186
9	185	193
10	82	88
11	58	79
12	124	126
13	50*	28

4, 1, L

0	788	811
1	306	299
2	253	251
3	227	227
4	567	580
5	164	187
6	511	513
7	291	306
8	78	77
9	250	231
10	114	116
11	206	228
12	50*	50

4, 2, L

0	2089	2052
1	1342	1785
2	451	442
3	588	567
4	530	534
5	232	225
6	511	506
7	352	348
8	170	168
9	137	132
10	59	65
11	79	66
12	48*	8

4, 3, L

0	1486	1443
1	602	584
2	365	341
3	129	127
4	60	46
5	273	256
6	272	261
7	155	158
8	243	247
9	315	302
10	92	83
11	170	165
12	104	96

4, 4, L

0	707	701
1	370	348
2	350	345
3	192	187
4	110	98
5	847	810
6	425	423
7	232	242
8	166	165
9	147	149
10	129	121
11	48*	35

4, 5, L

0	338	321
1	235	235
2	145	120
3	292	289
4	298	283
5	45*	8
6	288	293
7	65	62
8	151	145
9	296	297
10	192	176

4, 6, L

0	601	607
1	529	546
2	50*	61
3	237	258
4	259	255
5	387	398
6	296	287
7	252	265
8	209	207
9	52*	29

4, 7, L

0	656	655
1	200	191
2	186	172
3	101	85
4	129	131
5	195	207
6	140	142
7	68	64

4, 8, L

0	166	180
1	353	375
2	92	96
3	109	114
4	88	77

5, 0, L

1	86	83
2	385	380
3	1008	1008
4	144	136
5	564	525
6	128	135
7	106	106
8	321	333
9	159	145
10	93	92
11	152	135
12	48*	20

5, 1, L

1	122	118
2	1502	1522
3	714	726
4	1022	1032
5	421	435
6	302	318
7	84	105
8	141	141
9	356	373
10	137	114
11	64	71
12	127	134

5, 2, L

1	325	315
2	779	751
3	208	190
4	263	263
5	492	499
6	609	598
7	341	355
8	256	243
9	137	135
10	186	173
11	129	130
12	49*	39

5, 3, L

1	522	508
2	1276	1235
3	397	386
4	469	453
5	536	517
6	41*	32
7	111	116
8	239	225
9	341	354
10	197	197
11	49*	31
12	82	88

5, 4, L

1	151	141
2	403	384
3	285	303
4	181	191
5	176	174
6	113	107
7	215	205
8	132	157
9	122	125
10	133	123
11	152	145

5, 5, L

1	42*	46
2	643	593
3	205	183
4	870	854
5	141	137
6	69	69
7	255	249
8	251	257
9	124	115
10	95	96

5, 6, L

1	55	34
2	412	408
3	385	393
4	238	209
5	334	335
6	297	314
7	125	147
8	142	148

5, 7, L

1	135	137
2	243	221
3	357	342
4	229	255
5	194	203
6	175	189
7	123	121

5, 8, L

1	134	125
2	294	310
3	111	93

6, 0, L

0	377	379
1	37*	33
2	182	183
3	803	815
4	506	511
5	387	408
6	898	918
7	679	688
8	441	454
9	46*	21
10	456	459
11	137	120
12	139	136

6, 1, L

0	357	374
1	213	209
2	387	393
3	1325	1337
4	152	157
5	278	305
6	470	466
7	374	371
8	59	58
9	443	425
10	60	51
11	142	140
12	201	195

6, 2, L

0	734	730
1	326	317
2	370	353
3	374	357
4	524	519
5	115	118
6	453	440
7	400	395
8	106	100
9	249	255
10	251	245
11	201	209
12	85	68

6, 3, L

0	514	517
1	454	448
2	280	254
3	670	652
4	729	707
5	405	393
6	505	484
7	363	363
8	340	339
9	359	362
10	98	88
11	169	180

6, 4, L

0	640	617
1	144	134
2	161	153
3	784	751
4	585	565
5	228	207
6	245	250
7	291	298
8	117	121
9	47*	9
10	270	258
11	114	119

6, 5, L

0	284	260
1	103	99
2	163	161
3	491	467
4	144	128
5	324	328
6	324	328
7	140	138
8	174	184
9	227	228
10	325	314
11	55	54

6, 6, L

0	371	395
1	221	219
2	147	130
3	362	361
4	476	466
5	71	67
6	294	281
7	220	239
8	173	172

6, 7, L

0	177	186
1	86	94
2	97	77
3	461	460
4	196	192
5	99	135
6	129	142

6, 8, L

0	236	258
1	52*	48
2	55*	43
3	139	154

7, 0, L

1	441	449

21	421	1426
3	51	39
4	156	149
5	530	533
6	957	969
7	581	604
8	823	842
9	160	129
10	99	121
11	141	133
12	174	166

7,1,L

1	884	890
2	309	306
3	423	434
4	446	464
5	253	261
6	168	173
7	269	276
8	154	142
9	267	265
10	208	222
11	85	90
12	57	51

7,2,L

1	664	651
2	621	610
3	513	509
4	227	235
5	483	466
6	489	479
7	245	248
8	417	403
9	117	121
10	125	109
11	103	98
12	48*	41

7,3,L

1	813	742
2	370	345
3	55	57
4	534	525
5	111	118
6	200	190
7	196	192
8	187	173
9	172	163
10	90	93
11	85	90

7,4,L

1	477	444
2	778	755
3	146	132
4	133	123
5	80	76
6	397	390
7	247	249
8	379	396
9	172	176
10	50*	25

7,5,L

3	410	403
4	144	156
5	180	182
6	126	119
7	129	120
8	205	205
9	185	189

7,6,L

1	254	231
2	495	475
3	325	325
4	158	148
5	221	207
6	389	380
7	89	81
8	245	254

7,7,L

1	604	568
2	274	251
3	79	88
4	278	281
5	117	112
6	54	56

7,8,L

1	84	82
2	203	209

8,0,L

0	971	983
1	1944	1977
2	782	811
3	1220	1233
4	42*	27
5	40*	34
6	285	268
7	450	457
8	177	161
9	49*	11
10	224	238
11	116	114
12	105	129

8,1,L

0	954	947
1	712	710
2	119	121
3	476	476
4	73	81
5	494	505
6	210	208
7	270	256
8	322	331
9	63	89
10	257	265
11	50*	60
12	133	116

8,2,L

0	711	704
1	280	238
2	223	213
3	175	146
4	402	366
5	827	821
6	138	130
7	157	124
8	348	346
9	70	74
10	85	89
11	52*	58

8,3,L

0	768	748
1	447	434
2	82	75
3	218	228
4	681	655
5	465	467
6	128	120
7	412	425
8	232	230
9	73	76
10	242	231
11	80	92

8,4,L

0	594	588
1	829	799
2	317	304
3	560	536
4	60	64
5	270	265
6	54	52
7	168	177
8	187	179
9	84	94
10	97	112

8,5,L

0	772	760
1	619	601
2	94	82
3	146	139
4	94	82
5	103	108
6	234	238
7	232	235
8	51*	10
9	51*	26

8,6,L

0	497	492
1	205	235
2	104	105
3	144	151
4	263	252
5	341	327
6	199	199
7	180	200
8	211	212

8,7,L

0	322	305
1	257	264
2	111	95
3	186	186
4	70	40
5	197	197

9,0,L

1	331	337
2	403	385
3	412	407
4	73	75
5	122	123
6	275	273
7	499	501
8	201	178
9	271	281
10	297	297
11	166	166
12	47*	18

9,1,L

1	481	490
2	803	815
3	290	255
4	651	643
5	234	228
6	210	224
7	350	340
8	282	274
9	353	354
10	53	48
11	77	93
12	58	47

9,2,L

1	477	471
2	445	420
3	568	546
4	805	779
5	131	128
6	179	155
7	182	178
8	241	236
9	163	173
10	118	111
11	66	42

9,3,L

1	465	470
2	478	457
3	563	551
4	335	338
5	450	434
6	51	53
8	205	209
9	137	135
10	63	79
11	69	90

9,4,L

1	200	202
2	103	94
3	272	262
4	228	223
5	68	66
6	234	238
7	84	83
8	241	235
9	63	65
10	92	98

9,5,L

1	148	151
2	469	453
3	298	285
4	192	199
5	205	205
6	51*	46
7	211	214
8	110	113
9	249	256

9,6,L

1	312	290
2	276	258
3	121	139
4	404	412
5	59	52
6	266	262
7	51*	40

9,7,L

1	182	190
2	333	317
3	123	79
4	285	282
5	215	217

10,0,L

0	498	484
1	328	333
2	38*	2
3	285	289
4	207	199
5	197	205
6	41*	42
7	44*	23
8	609	622
9	556	581
10	272	289
11	134	147

10,1,L

0	939	928
1	140	158
2	148	158
3	190	189
4	467	467
5	148	149
6	201	212
7	249	234
8	164	178
9	184	200
10	354	361
11	116	128

10,2,L

0	644	648
1	280	262
2	534	595
3	652	637
4	429	403
5	587	577
6	391	383
7	314	315
8	152	140
9	177	159
10	105	98
11	196	198

10,3,L

0	648	625
1	322	304
2	277	254
3	195	187
4	732	717
5	264	255
6	53	57
7	318	310
8	151	164
9	224	223
10	245	247

10,4,L

0	498	507
1	159	142
2	130	133
3	346	333
4	205	212
5	294	282
6	284	282
7	73	70
8	324	336
9	206	224
10	98	106

10,5,L

0	398	371
1	125	118
2	48*	32
2	54	32
3	313	316
4	474	461
5	109	105
6	135	138
7	338	345
8	106	102

10,6,L

0	132	139
1	79	92
2	390	383
3	405	409
4	353	332
5	360	349
6	303	306
7	132	140

10,7,L

0	381	355
1	52*	10
2	73	67
3	90	93
4	252	258

11,0,L

1	252	240
2	74	92
3	119	128
4	451	442
5	58	52
6	289	288
7	173	173
8	47*	18
9	50*	24
10	69	90
11	61	43

11,1,L

1	187	199
2	200	216
3	218	228
4	269	257
5	99	82
6	342	329
7	392	402
8	132	130
9	140	153
10	274	287
11	105	108

11,2,L

1	846	796
2	575	588
3	515	497
4	197	193
5	423	426
6	189	188
7	194	196
8	48	37
9	306	306
10	212	219
11	131	140

11,3,L

1	241	232
2	271	266
3	208	205
4	349	337
5	46*	48
6	411	398
7	153	172
8	228	220
9	62	73
10	180	177

11,4,L

1	65	70
2	375	382
3	162	161
4	402	391
5	193	196
6	227	219
7	137	146
8	64	73
9	154	145

11,5,L

1	425	431
2	403	396
3	258	261
4	254	251
5	218	217
6	183	185
7	50*	28
8	136	132

11,6,L

1	362	326
2	477	460
3	173	196
4	119	118
5	171	179
6	50*	29

11,7,L

1	81	103
2	50*	22
3	196	185

12,0,L

0	72	90
1	335	330
2	338	336
3	43	29
4	732	712
5	309	302
6	601	578
7	408	424
8	66	67
9	133	104
10	85	75
11	47*	48

12,1,L

0	248	265
1	560	572
2	221	220
3	180	190
4	306	300
5	221	228
6	368	349
7	81	71
8	148	153
9	305	299
10	128	133
11	104	118

12,2,L

01	124	1123
1	56	44
2	109	93
3	655	653
4	176	148
5	221	218
6	282	263
7	75	71
8	180	181
9	165	173
10	61	57

12,3,L

0	452	449
1	831	832
2	103	117
3	302	309
4	45*	50
5	156	151
6	267	270
7	158	155
8	53	51
9	128	136
10	66	76

12,4,L

0	311	310
1	236	225
2	94	80
3	263	269

12,5,L

0	148	142
1	148	154
2	175	172
3	110	102
4	164	166
5	267	266
6	431	440
7	76	85
8	108	97

12,6,L

0	693	721
1	71	56
2	222	207
3	282	274
4	450	444
5	120	114
6	306	302

12,7,L

0	78	94
1	347	350
2	152	151

12,4,L

4	197	205
5	152	174
6	331	332
7	176	173
8	61	44
9	108	114

13,0,L

1	135	157
2	239	233
3	185	179
4	400	396
5	372	360
6	44*	11
7	50*	49
8	170	183
9	51*	29
10	105	66

13,1,L

1	535	529
2	240	233
3	331	316
4	174	162
5	241	226
6	237	227
7	207	210
8	287	291
9	147	147
10	52*	58

13,2,L

1	140	139
2	248	264
3	274	248
4	177	166
5	329	309
6	51*	55
7	222	234
8	184	178
9	262	253
10	141	136

13,3,L

1	549	545
2	233	240
3	187	191
4	316	306
5	316	316
6	205	213
7	130	131
8	344	334
9	48*	45

13,4,L

1	374	380
2	116	111
3	165	152
5	328	317
6	75	67
7	56	51
8	52*	40

13,5,L

1	421	427
2	187	178
3	121	109
4	186	189
5	345	335
6	52*	47
7	49*	22

13,6,L

1	189	169
2	134	156
3	274	276
4	59	66
5	100	96

14,0,L

0	488	490
1	197	193
2	579	585
3	41*	0
4	777	766
5	377	364
6	462	458
7	54*	49
8	182	186
9	248	251
10	266	260

14,1,L

0	490	476
1	159	166
2	42*	27
3	277	277
4	336	331
5	271	260
6	195	195
7	322	329
8	209	221
9	118	111
10	144	140

14,2,L

0	434	420
1	226	229
2	400	384
3	92	80
4	382	375
5	156	152
6	49*	25
7	242	251
8	157	161
9	168	173

14,3,L

0	331	325
1	188	189
2	118	116
3	56	42
4	244	233
5	305	293
6	137	143
7	263	269
8	151	164
9	50*	39

14,4,L

0	143	134
1	175	172
2	347	347
3	50*	24
4	409	418
5	202	196
6	232	231
7	137	148
8	50*	17

14,5,L

0	235	251
1	65	69
2	247	235
3	198	196
4	176	180
5	286	296
6	89	99

14,6,L

0	373	397
1	53*	68
2	339	346
3	107	87
4	271	275

15,0,L

1	609	605
2	191	187
3	543	536
4	249	258
5	423	396
6	260	251
7	48*	12
8	230	217
9	50*	38

15,1,L

1	155	154
2	368	377
3	326	320
4	171	182
5	249	240
6	211	223
7	197	200
8	49*	50
9	95	118

15,2,L

1	179	165
2	267	253
3	404	384
4	337	327
5	155	151
6	97	95
7	197	186
8	76	77
9	73	72

15,3,L

1	127	127
2	445	442
3	54	66
4	208	204
5	155	150
6	103	103
7	155	151
8	52*	49

15,4,L

1	301	315
2	173	181
3	310	326
4	131	130
5	304	303
6	89	93
7	54	57

15,5,L

1	203	208
2	211	219
3	111	117
4	84	89
5	109	107
6	211	227

15,6,L

1	187	191
2	170	144
3	166	178

16,0,L

0	688	706
1	608	603
2	246	242
3	247	255
4	386	389
5	339	346
6	129	135
7	81	84
8	50*	6
9	165	157

16,1,L

0	770	784
1	142	152
2	192	190
3	77	69
4	230	228
5	115	129
6	376	373
7	62	65
8	181	180
9	49*	26

16,2,L

0	90	75
1	190	181
2	191	179
3	181	175
4	140	145
5	177	183
6	125	134
7	282	289
8	154	150

16,3,L

0	408	420
1	93	111
2	71	83
3	179	177
4	111	134
5	98	105
6	335	343
7	219	224

16,4,L

0	289	307
1	339	360
2	158	157
3	114	122
4	150	137
5	138	138
6	49*	55

16,5,L

0	573	593
1	81	81
2	135	134
3	137	137
4	145	143
5	68	57

17,0,L

1	421	440
2	54	57
3	321	332
4	56	52
5	128	119
6	237	240
7	115	118
8	49	43

17,1,L

1	170	176
2	49*	38
3	219	224
4	442	448
5	204	216
6	186	198
7	268	269
8	99	105

17,2,L

1	380	373
2	84	82
3	163	152
4	181	192
5	272	270
6	156	161
7	61	62
8	48*	20

17,3,L

1	89	97
2	219	219
3	152	164
4	358	379
5	200	215
6	105	99
7	245	262

17,4,L

1	246	249
2	49*	30
3	114	114
4	120	121
5	76	75

17,5,L

1	133	152
2	150	156
3	52*	78

18,0,L

0	405	425
1	50*	3
2	135	147
3	75	84
4	98	95
5	72	77
6	183	187
7	239	226

18,1,L

0	351	356
1	219	232
2	136	125
3	240	236
4	50*	14
5	357	373
6	102	95
7	66	64

18,2,L

0	275	292
1	93	95
2	119	117
3	342	359
4	331	338
5	132	134
6	51*	53
7	143	148

18,3,L

0	141	172
1	285	309
2	187	189
3	189	193
4	91	98
5	297	306
6	102	105

18,4,L

0	172	167
1	62	63
2	70	92
3	196	205
4	160	168

18,5,L

0	47*	4
1	88	108

TABLE 6 Comparison of phases obtained from symbolic addition (ϕ_{dm}) and phases calculated from the final structure (ϕ_{str}). Phases given in millicycles.

h	k	ℓ	ϕ_{dm}	ϕ_{str}	h	k	ℓ	ϕ_{dm}	ϕ_{str}	h	k	ℓ	ϕ_{dm}	ϕ_{str}
7	1	7	955	908	14	10	0	500	500	6	4	6	470	566
5	4	5	462	458	0	0	2	495	419	12	6	4	470	445
6	3	7	433	475	4	0	2	981	15	10	2	6	470	519
8	1	0	0	0	18	4	2	982	956	5	2	5	461	412
7	8	0	0	0	1	10	5	461	459	9	4	7	445	483
2	4	1	993	26	2	10	0	500	500	2	4	7	451	400
14	4	0	0	0	17	7	3	473	488	12	3	2	985	33
0	4	0	500	500	2	10	0	495	500	11	1	6	906	958
2	11	0	0	0	7	8	4	965	920	15	2	3	34	888
10	8	0	500	500	14	5	5	970	932	18	3	2	479	452
12	6	0	500	500	8	0	5	467	406	4	4	0	500	500
6	6	0	0	0	6	7	0	500	500	3	8	2	985	973
16	0	0	500	500	10	4	7	940	904	13	1	3	475	484
14	2	0	500	500	4	9	5	465	455	8	5	6	463	522
12	0	6	422	572	15	5	0	500	500	10	8	4	475	502
10	9	0	0	0	11	1	5	961	40	9	9	5	975	914
12	1	3	487	384	7	8	6	456	435	4	5	6	5	818
11	2	6	434	615	16	7	2	485	467	1	1	7	444	492
12	1	7	451	406	1	6	5	967	982	8	4	3	0	794
4	1	2	981	20	6	9	1	491	446	15	5	4	461	500
2	4	5	964	929	1	3	4	949	57	16	1	4	468	490
7	6	0	0	0	1	5	4	949	46	2	11	4	978	981
4	1	8	444	348	3	1	5	965	65	7	2	6	961	759
12	6	6	953	817	9	2	5	959	942	18	5	1	516	489
2	8	0	0	0	1	11	0	500	500	6	9	5	472	428
4	0	7	446	491	17	1	2	980	61	3	3	8	445	326
16	1	0	500	500	2	7	0	500	500	13	8	3	482	390
10	4	3	983	939	14	0	6	959	792	16	6	3	980	983
1	7	0	0	0	4	0	3	492	402	6	3	4	960	51
8	3	0	0	0	18	7	0	0	0	18	5	3	0	870
8	1	5	463	451	5	4	1	499	511	13	1	5	960	999
6	10	0	0	0	1	9	4	462	467	5	2	8	934	878
7	2	0	500	500	1	4	1	990	52	9	2	7	444	471
12	4	0	0	0	2	10	3	975	4	14	4	6	476	303
17	4	3	980	835	13	3	6	491	251	10	6	6	952	112
16	0	1	995	912	2	7	3	492	341	2	10	5	461	421

TABLE 6 continued

h	k	ℓ	ϕ_{dm}	ϕ_{str}	h	k	ℓ	ϕ_{dm}	ϕ_{str}	h	k	ℓ	ϕ_{dm}	ϕ_{str}
6	3	1	993	25	17	1	0	500	500	3	2	7	433	308
13	5	5	455	454	10	5	2	481	529	2	5	6	977	900
11	1	2	959	124	17	7	1	973	926	4	1	6	933	104
14	4	4	961	961	7	4	7	944	905	10	4	6	811	352
2	5	2	990	975	3	3	6	906	84	9	4	2	480	582
1	9	0	500	500	11	2	5	968	958	5	3	7	445	270
15	1	0	0	0	6	10	4	976	937	12	6	5	961	949
10	7	5	966	990	10	0	7	262	953	3	1	8	937	901
10	10	1	991	24	6	4	3	990	773	16	0	5	965	931
7	2	4	453	499	3	1	6	417	594	4	5	4	470	556
16	6	1	490	556	6	9	3	983	993	5	2	3	975	937
5	2	1	495	503	12	9	1	980	983	14	2	6	958	875
15	3	0	500	500	12	3	6	958	993					
14	6	0	0	0	11	9	2	974	976					

Table 7 Least-Squares Planes in the Structure

Planes are expressed in the form $Px + Qy + Rz - S = 0$, referred to the fractional coordinates of Table 1. Deviations of atoms from the planes are tabulated in Å x 10^3.

The Indole Nuclei

P = 11.72	Q = 11.34	R = -0.07	S = -7.43
N(1)	-1	C(6)	20
C(2)	22	C(7)	23
C(3)	7	C(8)	-22
C(4)	-21	C(9)	-9
C(5)	-17		

P = 2.38	Q = 13.21	R = -0.17	S = -9.59
N(21)	47	C(26)	5
C(22)	-7	C(27)	-28
C(23)	5	C(28)	-44
C(24)	-2	C(29)	-12
C(25)	35		

The 6-Membered Rings

P = 11.88	Q = 11.28	R = 0.04	S = 7.39
C(4)	2	C(7)	9
C(5)	-11	C(8)	-19
C(6)	6	C(9)	14

P = 2.79	Q = 13.18	R = 0.31	S = -9.75
C(24)	-13	C(27)	-1
C(25)	9	C(28)	-3
C(26)	-2	C(29)	10

TABLE 7 continued

The Pyrrole Rings

P = 11.53	Q = 11.40	R = -0.18	S = 7.42
N(1)	-1	C(4)	0
C(2)	0	C(9)	1
C(3)	0	C(10)*	32

P = 2.29	Q = 13.22	R = -0.10	S = -9.55
N(21)	11	C(24)	0
C(22)	-4	C(29)	-23
C(23)	16	C(30)*	13

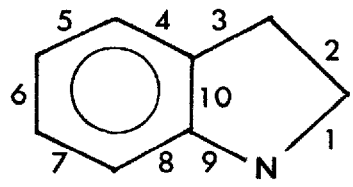
The 1,4-Diene Systems

P = 11.63	Q = 11.35	R = -0.44	S = 7.41
N(1)	14	O(12)*	-56
C(2)	-15	C(13)*	-139
C(10)	-14	C(3)*	-36
O(11)	15		

P = 1.85	Q = 13.25	R = -0.07	S = -9.34
N(21)	-1	O(32)*	24
C(22)	2	C(33)*	4
C(30)	3	C(23)*	53
O(31)	-4		

Tables 8 - 12 A summary of reported bond lengths (in Å) in indole nuclei of types A, B, D, E. (this work), and F (see Figure 8).

The bond numbering scheme is :-



Values marked "*" were omitted when calculating the mean bond lengths.

TABLE 8 Bond lengths in the indole nucleus (Å)

Ref.	Bond 1	2	3	4	5	6	7	8	9	10	e.s.d.	COMPOUND
26	1.40	1.34	1.49	1.37	1.37	1.42	1.39	1.40	1.38	1.39	0.007	Skatole
27	1.40	1.34	1.47	1.43	1.41	1.40	1.41	1.42	1.39	1.41	0.019	3-Indolyl-Acetic Acid
28	1.37	1.35	1.44	1.39	1.38	1.40	1.37	1.39	1.37	1.41	0.003	DL-Tryptophan Formate
29	1.43	1.34	1.42	1.40	1.41	1.35	1.39	1.40	1.31	1.39	0.014	Glycyl-L-Tryptophan. $2H_2O$
30	1.39	1.38	1.47	1.41	1.42	1.38	1.40	1.43	1.39	1.40	-	Seratonin-Creatin Sulphate. H_2O
31	1.38	1.34	1.45	1.41	1.40	1.39	1.40	1.40	1.39	1.38	0.016	L-Tryptophan, HCl
32	1.38	1.37	1.44	1.41	1.39	1.42	1.36	1.40	1.38	1.41	0.004	5-Methoxy-(N,N)-Dimethyl-Tryptamine,HCl
33	1.29*	1.28*	1.54*	1.29*	1.43*	1.35*	1.38*	1.35*	1.49*	1.50*	0.050	Hunterburnine- - Methiodide
34	1.32*	1.42*	1.44*	1.45*	1.28*	1.50*	1.38*	1.39*	1.35*	1.50*	0.070	Macusine A Iodide
35	1.36*	1.40*	1.47*	1.43*	1.30*	1.40*	1.36*	1.48*	1.39*	1.44*	0.050	Akuammidine-Methiodide
36	1.37	1.34	1.45	1.41	1.36	1.39	1.41	1.40	1.39	1.41	0.025	Mitragynine-Hydriodide
37	1.39	1.41	1.51	1.48	1.45	1.42	1.36	1.39	1.40	1.40	0.038	Ibogaine-Hydrobromide
38	1.31*	1.53*	1.52*	1.34*	1.48*	1.34*	1.41*	1.52*	1.45*	1.40*	0.050	Cleavamine-Methiodide
39	1.40	1.40	1.46	1.43	1.39	1.51*	1.38	1.39	1.39	1.49*	-	Reserpine
40	1.42*	1.50*	1.49*	1.53*	1.43*	1.47*	1.44*	1.36*	1.47*	1.46*	0.100	Leurocrystine-Methiodide. $2H_2O$
41	1.36	1.37	1.44	1.40	1.37	1.40	1.38	1.39	1.37	1.41	0.004	Serotonin Picrate. H_2O
42	1.37	1.44	1.48	1.37	1.41	1.39	1.36	1.40	1.49	1.40	0.015	Perlolyrine Hydrobromide
43	1.36	1.38	1.45	1.38	1.41	1.37	1.42	1.37	1.45	1.44	0.022	Yohimbane Hydrobromide
Mean	1.38	1.37	1.46	1.41	1.40	1.41	1.39	1.40	1.39	1.40		

TABLE 9 Bond lengths in the 2,3-Dialkylidene-Indole Nucleus (Å)

	Bond											
Ref.	1	2	3	4	5	6	7	8	9	10	e.s.d.	COMPOUND
44	1.35	1.49	1.47	1.36	1.37	1.39	1.37	1.36	1.38	1.37	-	Isatin
45	1.34	1.47	1.47	1.36	1.39	1.40	1.42	1.34	1.44	1.40	-	Indigo
46	1.33	1.50	1.46	1.43	1.44	1.34	1.42	1.39	1.38	1.40	-	Iso-Indigo
Mean	1.34	1.49	1.46	1.38	1.40	1.38	1.40	1.36	1.40	1.39		

TABLE 10 Bond lengths in the 2-Alkylidene-Indole Nucleus (Å)

	Bond											
Ref.	1	2	3	4	5	6	7	8	9	10	e.s.d.	COMPOUND
47	1.40	1.51	1.47	1.43	1.33	1.44	1.39	1.33	1.40	1.39	0.066	(-)-N-Methyl-Gelsemicine Hydroiodide
48	1.33	1.45	1.59	1.31	1.47	1.40	1.36	1.40	1.30	1.46	0.070	Rauvoxinine Methiodide
Mean	1.37	1.48	1.53	1.37	1.40	1.42	1.38	1.37	1.35	1.43		

TABLE 11 Bond lengths in the Indolenine Nucleus (Å) this work

	Bond											
Ref.	1	2	3	4	5	6	7	8	9	10	e.s.d.	COMPOUND
	1.30	1.54	1.53	1.34	1.42	1.39	1.39	1.38	1.41	1.39	0.011	Molecule 1
	1.29	1.52	1.52	1.34	1.41	1.42	1.39	1.37	1.42	1.39	0.011	Molecule 2
Mean	1.29	1.53	1.52	1.34	1.42	1.40	1.39	1.37	1.41	1.39		

TABLE 12 Bond lengths in the Dihydro-Indole Nucleus (Å)

Ref.	Bond 1	2	3	4	5	6	7	8	9	10	e.s.d.	COMPOUND
49	1.51*	1.54*	1.52*	1.42*	1.41*	1.44*	1.36*	1.35*	1.38*	1.35*	0.080	Chimonanthine Dihydrobromide
	1.36*	1.51*	1.52*	1.36*	1.38*	1.44*	1.34*	1.45*	1.49*	1.34*		
50	1.43	1.57	1.56	1.39	1.42	1.37	1.42	1.47	1.38	1.44	0.030	Echitamine Bromide. CH_3OH
51	1.46*	1.49*	1.59*	1.30*	1.30*	1.29*	1.34*	1.31*	1.37*	1.40*	±0.1	Strychnine Hydrobromide. $2H_2O$
52	1.51	1.54	1.49	1.39	1.44	1.36	1.40	1.40	1.43	1.42	0.030	N( )-Acetyl-7-Ethyl-Desethyl-Aspidospermine-n( )-Methiodide
53	1.34	1.55	1.52	1.41	1.40	1.35	1.45	1.38	1.45	1.36	0.030	(-)-Kopsanone-N( )-Methiodide
54	1.48	1.56	1.56	1.37	1.41	1.42	1.35	1.42	1.41	1.34	0.020	(+)-Haplophytine Dihydrobromide
	1.47	1.59	1.50	1.38	1.44	1.34	1.34	1.41	1.44	1.38		
55	1.48	1.64	1.48	1.39	1.39	1.38	1.43	1.39	1.43	1.38	0.026	Voaphylline Hydroxy-Indolenine Iodohydrate*
56	1.48	1.56	1.55	1.40	1.38	1.42	1.40	1.40	1.51	1.45	-	Echitamine Iodide
57	1.50*	1.48*	1.48*	1.52*	1.33*	1.33*	1.42*	1.43*	1.28*	1.35*	0.100	Caracurine-II-Dimethiodide
	1.44*	1.30*	1.58*	1.37*	1.44*	1.30*	1.21*	1.49*	1.36*	1.33*		
40	1.54	1.54	1.56	1.41	1.39	1.24	1.44	1.53	1.49	1.33	0.100	Leurocrystine Methiodide.$2H_2O$
58	1.36	1.52	1.52	1.37	1.41	1.39	1.39	1.39	1.41	1.40	0.006	1-Methyl-2-Picryl-imino-indoline
59	1.51	1.56	1.50	1.42	1.45	1.40	1.46	1.38	1.38	1.38	-	(±)-1-Acetyl-3-Methyl-Aspidospermine-9-Methiodide
Mean	1.45	1.57	1.52	1.39	1.42	1.38	1.40	1.40	1.43	1.37		

TABLE 13 Estimated Bond Orders in the Indole Series

Bond	Indole	Nucleus		Indolenine	Dihydro Indole	Ring
		Dialkyldene Indole	2-Alkylidene Indole			
1	1.3	1.5	1.35	1.8	1.05	
2	1.8	1.2	1.2	1.0	0.9	B
3	1.3	1.3	1.0	1.1	1.1	
9	1.25	1.2	1.45	1.2	1.1	
10	1.6	1.7	1.45	1.7	1.8	
4	1.55	1.7	1.8	2.0	1.7	
5	1.6	1.6	1.6	1.5	1.5	A
6	1.55	1.7	1.5	1.6	1.7	
7	1.7	1.6	1.7	1.7	1.6	
8	1.6	1.8	1.8	1.8	1.6	

The standard bond lengths used in the compilation of this table were:

C - N	1.47 Å	C - C	1.54 Å
C = N	1.26 Å	C = C	1.34 Å
C ≡ N	1.16 Å	C ≡ C	1.20 Å

TABLE 14 Mean Aromatic Ring Bond Lengths (Å) (Ring A)

Indole	Dialkylidene Indole	2-Alkylidene Indole	Indolenine	Dihydro Indole
1.40	1.39	1.39	1.39	1.39

TABLE 15 Carbon-Hydrogen Bond Lengths (Å)

C7 -H1	1.043(87)	C27-H21	1.154(81)
C8 -H2	1.214(82)	C28-H22	1.088(87)
C18-H3	0.968(80)	C38-H23	1.103(77)
C18-H4	0.784(83)	C38-H24	0.925(84)
C18-H5	0.974(80)	C38-H25	0.975(81)
C13-H6	0.883(75)	C33-H26	0.920(78)
C13-H7	1.077(78)	C33-H27	0.876(80)
C13-H8	0.877(83)	C33-H28	0.928(83)

Mean C - H bond length = 0.987 Å

TABLE 16 Bond Angles Involving Hydrogen Atoms (Å)

C6 -C7-H1	114(4)	C26-C27-H21	103(4)
C8 -C7-H1	124(4)	C28-C27-H21	134(4)
C7 -H8-H2	120(4)	C27-C28-H22	123(4)
C9 -C8-H2	116(4)	C29-C28-H22	136(2)
O17-C18-H3	110(5)	O37-C38-H23	108(5)
O17-C18-H4	109(6)	O37-C38-H24	103(6)
O17-C18-H5	102(5)	O37-C38-H25	102(5)
H4 -C18-H3	108(8)	H24-C38-H23	116(7)
H5-C18-H3	109(7)	H24-C38-H25	108(7)
H5-C18-H4	117(8)	H25-C38-H23	118(7)
O12-C13-H6	119(5)	O32-C33-H26	100(5)
O12-C13-H7	97(4)	O32-C33-H27	107(6)
O12-C13-H8	105(6)	O32-C33-H28	103(5)
H7-C13-H6	104(6)	H27-C33-H26	104(7)
H8-C13-H6	106(7)	H28-C33-H26	137(8)
H7-C13-H8	126(8)	H28-C33-H27	104(7)

TABLE 17 The Shorter Intermolecular Contacts (Å)

In the notation used "N1-C2 2/0$\bar{1}$0" implies that the distance quoted is from N1 in equivalent position 1 to C2 in equivalent position 2 and translated one unit cell in the -y direction. The figures in parentheses refer to molecules illustrated in figure 10.

Cℓ(34) -	C(13)	1/000	3.604	(1 - 2)
Cℓ(34) -	O(12)	1/000	3.614	(1 - 2)
Cℓ(36) -	Cℓ(15)	1/000	3.505	(1 - 2)
Cℓ(35) -	C(8)	3/000	3.552	(2 - 1")
O(11) -	C(8)	1/001	3.373	(1 - 1")
O(31) -	C(28)	1/001	3.413	(2 - 2")
Cℓ(15) -	C(38)	1/001	3.599	(1 - 2")
C(33) -	N(1)	2/110	3.157	(2 - 1')
Cℓ(34) -	N(21)	2/110	3.381	(2 - 2')
Cℓ(14) -	N(21)	2/110	3.544	(1 - 2')
Cℓ(14) -	O(32)	2/110	3.536	(1 - 2')
Cℓ(14) -	C(30)	3/0$\bar{1}$0	3.574	
O(12) -	C(38)	3/0$\bar{1}$0	3.485	
C(10) -	O(37)	3/0$\bar{1}$0	3.303	
O(12) -	O(37)	3/0$\bar{1}$0	3.209	
O(31) -	C(18)	4/001	3.385	
C(26) -	Cℓ(14)	3/00$\bar{1}$	3.574	
O(37) -	C(10)	3/00$\bar{1}$	3.303	
N(1) -	C(33)	2/1$\bar{1}\bar{1}$	3.167	
Cℓ(35) -	Cℓ(16)	4/010	3.481	
C(22) -	Cℓ(16)	4/010	3.460	
N(21) -	Cℓ(16)	4/010	3.478	
C(28) -	O(31)	1/00$\bar{1}$	3.413	
Cℓ(14) -	O(17)	2/010	3.318	
C(18) -	O(31)	4/$\bar{1}$1$\bar{1}$	3.385	

Equivalent positions in $Pna2_1$:-

1 x, y, z

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

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

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

REFERENCES

1. R.L. Hinman, O.P. Bauman, J. Org. Chem., 1964, 29, 1206.
2. T.A. Foglia, D. Swern, J.Org. Chem., 1968, 33, 4440.
3. J.M. Muchowski, Canad. J. Chem., 1970, 48, 442.
4. R.J. Bass, M.S. Tute, Private Communication.
5. F.H. Allen, D. Rogers, P.G.H. Troughton, Acta Cryst., 1971, B27, 1325.
6. A program written by P.G.H. Troughton and described in 5.
7. J.M. Stewart, Tech. Report 64-6, Univ. of Maryland (U.S.A.).
8. I.L. Karle, K.S. Dragonette, S.A. Brenner, Acta Cryst., 1965, 19, 713.
9. M.G.B. Drew, D.H. Templeton, A. Zalkin, Acta Cryst., 1969, B25, 261. and Private Communications.
10. J. Karle, I.L. Karle, Acta Cryst., 1966, 21, 849.
11. J. Karle and H. Hauptman, Acta Cryst., 1956, 9, 635.
12. D. Harker and J.S. Kasper, Acta Cryst., 1948, 1, 70.
13. I.L. Karle and J. Karle, Acta Cryst., 1968, B24, 81.
14. J. Karle, J.A. Estlin and I.L. Karle, J. Amer. Chem. Soc., 1967, 89, 6510.
15. A.J.C. Wilson, Nature, 1942, 150, 151.
16. H. Hauptman and J. Karle, Acta Cryst., 1956, 9, 45.
17. H. Hauptman and J. Karle, "The Solution of the Phase Problem", A.C.A. Monograph N. 3, 1953.
18. W. Cochran and M.M. Woolfson, Acta Cryst., 1955, 8, 1.
19. A. Gaussian Integration Program Based On:
    a) P. Coppens, L. Leiserowitch, D. Rabinowitch, Acta Cryst., 1965, 18, 1035.
    b) W.R. Busing, H.A. Levy, Acta Cryst., 1957, 10, 180.
20. eg. J.M. Ohrt, B.A. Haner, A. Cooper, D.A. Norton, Acta Cryst., 1968, 24B, 312.

21. J. Hvoslef, Acta Cryst., 1968, 24B, 23.

22. A. Cooper and D.A. Norton, Acta Cryst., 1968, 24B, 811.

23. A. Camerman and N. Camerman, Acta Cryst., 1971, B27, 2205.

24. D. Sayre, Acta Cryst., 1952, 5, 60.

25. L. Pauling, "The Nature of the Chemical Bond", Cornell Univ. Press, 1960.

26. A.W. Hanson, Acta Cryst., 1964, 17, 559.

27. I.L. Karle, K. Britts, P. Gum, Acta Cryst., 1967, 17, 496.

28. E. Bye, A. Mostad, Chr. Rømming, Acta Chem. Scand., 1971, 25, 364.

29. R.A. Pasternak, Acta Cryst., 1956, 9, 341.

30. I.L. Karle, K.S. Dragonette, S.A. Brenner, Acta Cryst., 1965, 19, 713.

31. T. Takigawa, T. Ashida, Y. Sasada, M. Kakudo, Bull. Chem. Soc. Japan, 1966, 39, 2369.

32. G. Falkenberg, D. Carlstrom, Acta Cryst., 1971, B27, 411.

33. J.D.M. Asher, J.M. Robertson, G.A. Sim, J. Chem. Soc., 1965, 6355.

34. A.T. McPhail, J.M. Robertson, G.A. Sim, J. Chem. Soc., 1963, 1832.

35. S.J. Silvers, A. Tulinsky, Acta Cryst., 1963, 16, 579.

36. D.E. Zacharias, R.D. Rosenstein, G.A. Jeffrey, Acta Cryst., 1965, 18, 1039.

37. G. Arai, J.L. Coppåla, G.A. Jeffrey, Acta Cryst., 1960, 13, 553.

38. N. Camerman, J. Trotter, Acta Cryst., 1964, 17, 384.

39. I.L. Karle, J. Karle, Acta Cryst., 1968, B24, 81.

40. J.W. Moncreif, W.N. Lipscomb, Acta Cryst., 1966, 21, 322.

41. U. Thewalt, C.E. Bugg, Acta Cryst., 1972, B28, 82.

42. G.A. Jeffrey, J. Chem. Soc. (C), 1970, 1091.

43. J.P. Fennessey, W. Nowacki, Z. Kryst., 1970, 131, 342.

44. G.H. Goldschmidt and F.J. Llewellyn, Acta Cryst., 1950, 3, 294.

45. H. Von Eller-Pandraud, Bull. Soc. Chim. Fr., 1955, 14B, 3.

46. H. Von Eller-Pandraud, Acta Cryst., 1960, 13, 936.

47. M. Przybylska, Acta Cryst., 1962, 15, 301.

48. C. Pascard-Billy, Bull. Soc. Chim. Fr., 1968, 3289.

49. I.J. Grant, T.A. Hamor, J.M. Robertson, G.A. Sim, J. Chem. Soc., 1965, 5678.

50. J.A. Hamilton, T.A. Hamor, J.M. Robertson, G.A. Sim, J. Chem. Soc., 1962, 5061.

51. J.M. Robertson and C.A. Beevers, Acta Cryst., 1951, 4, 270.

52. A. Camerman, N. Camerman, J. Trotter, Acta Cryst., 1965, 19, 314.

53. B.M. Craven, Acta Cryst., 1969, B25, 2131.

54. D.E. Zacharias, Acta Cryst., 1970, B26, 1455.

55. P.J. Guilhem, Acta Cryst., 1970, B26, 2029.

56. H. Manohar, S. Rameseshan, Z. Kryst., 1962, 117, 273.

57. A.T. McPhail, G.A. Sim, J. Chem. Soc., 1965, 1663.

58. A.S. Bailey, W.A. Warr, G.B. Allison, C.K. Prout, J. Chem. Soc. (C), 1970, 956.

59. O. Kennard, K.A. Kerr, D.G. Watson, J.K. Fawcett, L. Rivadi Sanseverino, J. Chem. Soc. (A), 1970, 1779.

60. M.J.S. Dewar, G.J. Gleicher, J. Chem. Phys., 1966, 44, 759.

61. K.A. Schellenberg, G.W. McLean, J. Amer. Chem. Soc., 1966, 88, 1077.

62. G.F. Smith, J. Chem. Soc., 1954, 3842.

63. E. Bayer, G. Hafeliger, Chem. Ber., 1966, 99, 1689.

64. R. Karlsson, Acta Cryst., 1972, B28, 2358.

65. T.S. Rao, M. Sundaralingam, J. Amer. Chem. Soc., 1970, 92, 49.

66. a) L. Dupont, O. Dideberg, Acta Cryst., 1972, B28, 2340.

b) P. Groth, Acta Chem. Scand., in press.

67. W. Haugen, M. Tratteberg, Acta Chem. Scand., 1966, 20, 1726.

68. E.A. Cherniak, C.C. Costain, J. Chem. Phys., 1966, 45, 104.

69. E. Kuchitsu, T. Fukuyama, Y. Morino, J. Mol. Struct., 1968, 1, 463.

70. T. Hahn, Z. Krist., 1957, 109, 438.

71. T.C. van Soest, A.F. Peerdeman, Acta Cryst., 1970, B26, 1940.

72. G.J. Palenik, J. Donohue, K.N. Trueblood, Acta Cryst., 1968, B24, 1139.

73. R. Rudman, Acta Cryst., 1971, B27, 262.

74. A. Bondi, Physical Properties of Mol. Crystals, Liquids and Glasses, Wiley and Son., New York, 1968.

75. A.J. Kitaigorodskii, "Organic Chemical Crystallography" Consultants Bureau, 1957.

76. E.R. Howells, D.C. Phillips and D. Rogers, Acta Cryst., 1950, 3, 210.

CHAPTER 3

The crystal structure of Erythristemine 2-Bromo-4,6-dinitro phenolate.

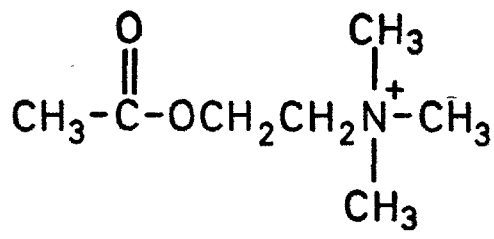
INTRODUCTION

The Transmission of Nerve Impulses[1]

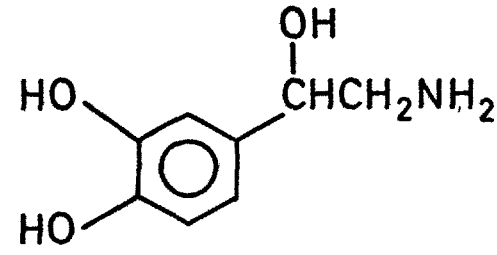
The transmission of impulses within a nerve is a complex process which, at any one point on the nerve, involves the rapid but brief exchange of sodium and potassium ions across the membrane of the nerve axon followed by an ATP dependent re-exchange which restores the nerve to its resting condition. The latter process is generally known as the "Sodium Pump". Transmission between nerves (*i.e.* across a synaptic junction) and across nerve to muscle interfaces in the parasympathetic system (the regulator of processes such as heart beat and respiration when the organism is at rest) is mediated by the chemical transmitter substance acetyl-choline (Ach - see figure 1a). In the sympathetic system (the system which controls the body processes in times of stress or emergency) the mediator substance is norepinephrine (adrenalin) - see figure 1b.

Synthesis of the transmitter substances occurs in vesicles located in the terminal portion of the transmitting nerve and is initiated by the receipt of a nerve impulse. The compound is then released into the synaptic or neuro-muscular junction in close proximity to receptor sites in the new nerve or in the motor endplate of a muscle. In the first of these cases the arrival of the transmitter substance initiates a new nerve impulse which is then transmitted and in the second case the muscle is stimulated to contract. In both types of junction the receptor site contains enzymes which are able to destroy the transmitter substance after its arrival, clearing the receptor for the receipt of further impulses.

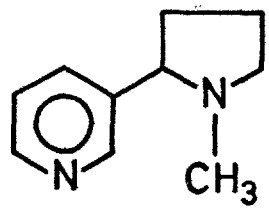
The arrival of a "quantum" of ACh molecules at a cholinergic receptor site is believed to cause a depolarisation of the muscle endplate giving rise to an endplate potential which excites the muscle fibre. It is also believed that the concentration of ACh to which a muscle will respond is confined within strict limits since the injection of ACh or of substances



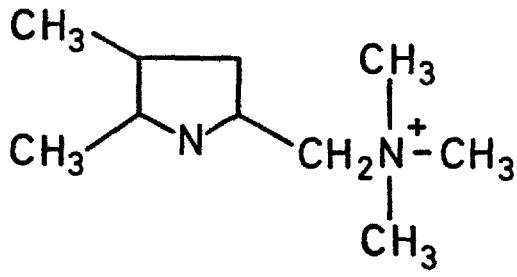
a) ACETYL CHOLINE



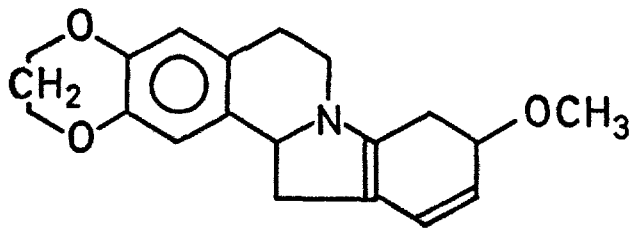
b) NOREPINEPHRINE



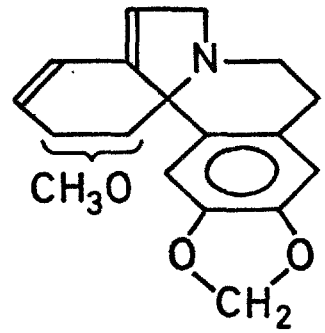
c) NICOTINE



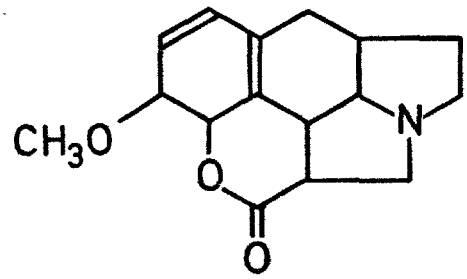
d) MUSCARINE



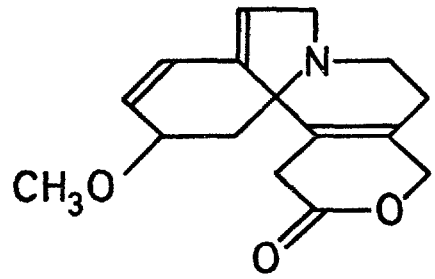
e)



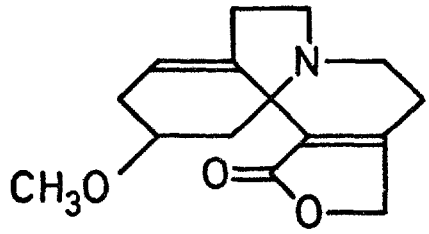
f)



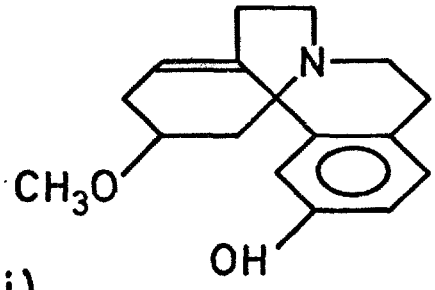
g)



h)



i)



j)

Figure 1

which inhibit the action of the hydrolysing enzyme Acetyl Choline Esterase (AChE), results in paralysis.

The classical experiments of Dale[2] have shown that ACh receptor sites are of two types which are termed "nicotinic" and "muscarinic" since they are inhibited by antagonists which resemble nicotine and muscarine respectively (see figure 1, c and d). Since these molecules have considerably different shapes it is considered that the receptor sites differ in form. ACh itself is able to adopt conformations which permit its acceptance at both types of site[3].

The blocking of transmission of nerve impulses across the neuromuscular junction produces paralysis of the muscles, an effect which is generally termed the "curare effect" since it was originally observed in connection with the South American arrow poison "curare". Transmission may be blocked by chemical substances at several points in the sequence[4]. The bacterial "botulinus" toxin, for example, prevents the synthesis of ACh, whereas the local anaesthetic procaine prevents the release of ACh. Various "anti-choline esterases" prevent the hydrolysis of ACh and thus cause paralysis by a build up of ACh beyond the acceptable limits. The curare alkaloids block the ACh receptor sites. All these agents produce symptoms characteristic of the curare effect.

The Curare Effect[4]

Crude plant extracts containing the paralysing alkaloid tubocurarine have been used as arrow poisons in South America for centuries. As long ago as 1811, Brodie[5] observed that death was due to respiratory failure, and shortly afterwards Bernard[6] stated that the physiological action occurred at the neuromuscular junction without apparent effects on either the nerve trunk or the muscle fibres themselves. The term "curare effect" is applied only to those compounds which prevent the transmission of impulses from the nerve to the muscle at the neuromuscular junction.

Curare poisoning produces a muscular paralysis which generally occurs in a definite order, the first signs being a drooping of the eyelids, drowsiness, loss of speech and paralysis of the neck muscles. The extremities are then affected followed by the muscles of the diaphragm and finally death occurs from respiratory failure. Death, however, can be averted by artificial respiration.

In 1868 Crum Brown and Fraser[7] attributed the curare effect to the "onium" ion and suggested that synthetic quarternary ammonium compounds could be used in place of crude extracts from plants. Boehm[8] showed that the active components of curare arrow poisons were all quarternary salts of alkaloids, and subsequent research has shown that even the simplest quarternary ammonium salts, as well as sulphonium, phosphonium, arsonium and stilbonium salts exhibit curare activity.[9] The alkaloids which exhibit curare effect are almost entirely derived from three plant genera, Chondodendron, Strychnos and Erythrina. The first of these genera yields a series of isoquinoline alkaloids which includes d-tubocurarine, a substance which has found widespread medical use as a muscle relaxant. The strychnos species produce a series of potent indole alkaloids known as the toxiferins, and the genus Erythrina yields a series of related compounds known as the erythrina alkaloids. Erythristemine, which is the subject of the present crystal structure investigation, is a member of this series.

Curare agents are used in abdominal, orthopaedic and ocular surgery and in many other situations requiring the relaxation of muscles, *e.g.* the softening of convulsions in shock treatment.

The first report of the medicinal use of an extract containing d-tubocurarine was published by Hoffmann[10], who used it to treat the convulsive seizures of tetanus. The pure compound was isolated in 1935 and its structure was elucidated by King[11]. Subsequently a number of synthetic compounds based on d-tubocurarine have been used but the natural product is still medically important.

The Erythrina Alkaloids

The erythrina alkaloids all exhibit the curare effect[4], and two of them, β-erythroidine and its reduced form dihydro-β-erythroidine, have found wide clinical use. Dihydro-β-erythroidine is some 5 times more effective than β-erythroidine itself and is *circa* one fifth as active as d-tubocurarine. The erythrina alkaloids are unusual in that they are effective when taken orally and, in addition to blockage of impulse transmission at the neuromuscular junction, they produce a mild effect on the central nervous system which causes a general relaxation. They cause a drop in blood pressure and stop respiration when administered in paralysing doses. Though the paralysing action of the β-erythroidines is less intense and of shorter duration than that of d-tubocurarine they have a greater margin of safety and are therefore widely used.

The erythrina alkaloids are unique among curare agents in that they are tertiary amines and that their activity is greatly diminished by quarternisation. At physiological pH's, however, the erythrina alkaloids are protonated and are thus not strictly tertiary.

The poisonous nature of extracts from an Erythrina species was first reported in 1890 by Greshoff[12] and early reports of other pharmacological effects ranging from narcotic to vermicidal have been reviewed by Folkers and Unna[13]. The curare-like action of extracts from the South American species E. crysta-galli was recognised in 1937[14] and shortly afterwards Folkers and Major[15] succeeded in isolating a pure alkaloid which they named Erythroidine from the South American plant E. americana. This compound was shown to produce the curare effect and in a survey of 50 out of the 105 Erythrina species known in 1944 Folkers and Unna[16] and Pichard and Luco[17] showed that all contained curare effective alkaloids. By 1952[18] erythroidine had been shown to exist in two isomeric forms and 10 further compounds had been isolated. In a recent book on the chemistry

of the alkaloids Mondon[19] listed 13 erythrina alkaloids which he described as naturally occurring and the compound which is the subject of this crystallographic study raises this number to 14.

The erythrina alkaloids fall into two groups, the "aromatic" series and the "erythroidines" in which the aromatic nucleus is replaced by a six-membered lactone ring. The aromatic erythrina alkaloids are subdivided into two groups which have been given the generic name stems "erythra-" and "eryso-". The elucidation of the structures of the erythrina alkaloids and extensive information on their general chemistry has been summarised in a series of articles in the publication "The Alkaloids"[4,18,20,21]. For the present purpose a brief outline of the elucidation of their structures and the correlation of the aromatic erythrina alkaloids with the erythroidines will be given.

a) The Aromatic Erythrina Alkaloids[4,18]

Work on the structures of the aromatic erythrina alkaloids has been concentrated on erythraline, $C_{18}H_{19}O_3N$, a crystalline base which was first isolated by Folkers and Koniuszy[22] as a hydriodide. It contains an aliphatic methoxyl and a methylene dioxyl group which account for all its oxygen atoms. The nitrogen atom is tertiary and the compound absorbs one or two moles of hydrogen, depending on the technique of hydrogenation, indicating the presence of two double bonds. Fusion with potassium hydroxide yields indole. On the basis of this evidence, combined with oxidation and spectroscopic studies, Folkers, Koniuszy and Shavel[23] proposed the structure shown in figure 1e.

This formula was generally accepted up to 1951 when, in the light of experimental evidence which was inexplicable from the Folkers structure, Carmack, McCusick and Prelog[24] proposed an isoquinoline structure (fig. 1f). This was confirmed in 1958 by the determination of the crystal structure of erythraline hydrobromide[25].

b) The Erythroidines[19]

The first of the erythrina alkaloids to be isolated in the pure form was given the name erythroidine[14] but a short while later was found to exist in two isomeric forms which were termed α and β since they were believed to be diastereoisomers. The α form is extremely sensitive to air and has a markedly different chemistry from that of the β form. In view of the greater stability and availability of the β form, research into the structures of the erythroidines was concentrated on this compound.

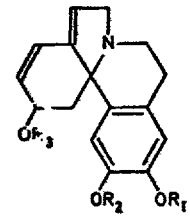
β-erythroidine is a tertiary base, of formula $C_{16}H_{19}O_3N$, with two double bonds capable of hydrogenation, one readily removable methoxy group and a lactone ring. It gives a characteristic red colour with sulphuric acid and ferric chloride. From the results of extensive chemical studies, Koniuszy and Folkers[26] proposed the structure shown in figure 1g but this was superceded by the structure proposed in 1953 by Boekelheide *et al.*[27] (fig. 1h). The latter structure was confirmed in 1962 by the determination of the crystal structure and absolute configuration of dihydro-β- erythroidine hydrobromide[28].

Hofmann degradation of dihydro-α-erythroidine was found to yield a base which is the same as that obtained from a similar treatment of dihydro-β-erythroidine, indicating a close similarity between the compounds. After this had been realised a conversion of the α form to the β form by migration of one of the double bonds was found to occur on treatment of α-erythroidine with 2N sodium hydroxide.[29]

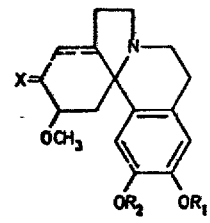
A third erythroidine, cocculolidine, was isolated from the leaves of Cocculus trilobus (a species allied to Erythrina spp.) in 1966[30] and was found to have insecticidal properties. The structure of this compound was elucidated by degradative methods and is shown in figure 1i.*

*Footnote: A recent determination of the crystal structure of cocculine hydrobromide[34] has shown that this compound, like its congener cocculolidine, is yet another aromatic erythrina alkaloid. Its structure is shown in figure 1j.

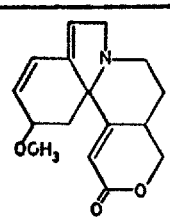
Figure 2 The Naturally Occurring Erythrina Alkaloids and the Erythrinane Skeleton



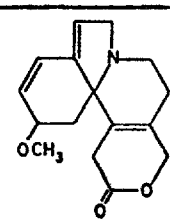
Name	Composition	R_1	R_2	R_3
Erythraline	$C_{18}H_{19}NO_3$	$-CH_2-$		CH_3
Erysodine	$C_{18}H_{21}NO_3$	H	CH_3	CH_3
Erysovine	$C_{18}H_{21}NO_3$	CH_3	H	CH_3
Erysopine	$C_{17}H_{19}NO_3$	H	H	CH_3
Erysonine	$C_{17}H_{19}NO_3$	H	CH_3	H



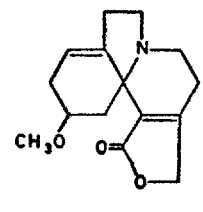
Name	Composition	R_1	R_2	X
Erythramine	$C_{18}H_{21}NO_3$	$-CH_2-$		H_2
Dihydroerysodine	$C_{18}H_{23}NO_3$	H	CH_3	H_2
Erythratine	$C_{18}H_{21}NO_4$	$-CH_2-$		H,OH
2-Ketoerythramine	$C_{18}H_{19}NO_4$	$-CH_2-$		=O
Erythratidine	$C_{19}H_{25}NO_4$	CH_3	CH_3	H,OH



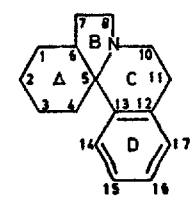
α-Erythroidine



β-Erythroidine



Cocculolidine



Erythrinane

The Relationship between the Aromatic Erythrina Alkaloids and the Erythroidines

The structures of the naturally occurring erythrina alkaloids (with the exception of cocculine - figure 1j) are shown in figure 2. They are divided into the three groups mentioned above. The "parent" hydrocarbon, erythrinane, is included and carries the standard numbering scheme and ring nomenclature. It is clear that all these alkaloids contain a tetracyclic skeleton with three rings (A, B, and C) and a spiro linkage in common. They differ in the nature of ring D, and their substituent groups. In the aromatic alkaloids ring D is benzenoid, with oxygenated groups at C15 and C16, while in the erythroidines this ring is an unsaturated δ-lactone. The close similarity in structure, which extends to stereochemical identity (but see below), is a strong indication of a common origin. The "eryso-" alkaloids differ from the "erythra-" group in having one or two phenolic groups in place of the methoxy group(s) at C15 and C16. The eryso- alkaloids are obtained from plant extracts as sulphoacetic esters, sulphonated at one of the phenolic groups. Hydrolysis of these "combined" forms yields the eryso- alkaloids and sulphoacetic acid. The latter can be readily removed as an insoluble barium or lead salt. The curare activity of the alkaloids in the combined form is more pronounced than that of the eryso- alkaloids themselves.

All the erythrina alkaloids possess two asymmetric centres - at the spiro atom, C5, and at C3. Erythratine, erythratidine and α-erythroidine contain a third asymmetric centre (C2, C2 and C12 respectively). Chemical, spectroscopic and crystallographic methods have been used to assign relative and absolute configurations at these centres and the results imply a common biosynthetic origin for the aromatic erythrina alkaloids and the erythroidines. Until the determination of the present structure, however, the absolute configuration of the aromatic alkaloids has depended on somewhat circumstantial evidence.

By degradative methods, Boekelheide and Wenzinger[31] established that the configuration at C3 in dihydro-β-erythroidine was R and this was shown to be correct by the determination of the absolute crystal structure of dihydro-β-erythroidine[28]. The configuration of the two asymmetric centres in β-erythroidine is thus, 3R, 5S. The configuration at C3 and C5 in the α form must be the same as that in β-erythro idine since the α form is converted to the β form on treatment with sodium hydroxide. The configuration at the additional centre (C12) was established by degradation.

The crystal structure work on erythraline hydrobromide did not include a determination of the absolute configuration of the compound, but it did show that the relative configurations at C3 and C5 are the same as in the erythroidines. This result was corroborated by chemical studies on erysonine (see figure 2) but could not be extended to the other aromatic alkaloids. The only evidence for the assignment of the 3R, 5S configuration to the aromatic series depends on the interpretation of optical rotations and ORD curves from various erythroidine and aromatic erythrina alkaloid derivatives. A theoretical interpretation of these results in terms of the chirality of skewed diene systems gave[32] the opposite chirality to that known to be present in β-erythroidine but the validity of this work has been questioned by Boekelheide and Wenzinger[31]. The determination of the structure and absolute configuration of the present compound, erythristemine, would provide important evidence for (or against) the assumption of a common chirality at C(3) and C(5) in all the erythrina alkaloids.

In addition to the evidence summarised above, Barton and his coworkers[33] have established a biological pathway which leads to the synthesis of both the aromatic erythrina alkaloids and the erythroidines by phenol-oxidative coupling from N-norprotosinomenine.

Erythristemine

During the course of the studies on erythrina alkaloid biogenesis which established the biosynthetic pathway mentioned above, Barton *et al.* attempted to extract erythraline (see figure 2) from a new erythrina species, Erythrina lysistemon. They obtained a non-polar alkaloid which, after purification, gave a crystalline solid, M.Pt. 127-129°C, which they named erythristemine. Micro-analysis, combined with mass spectroscopy, gave the formula $C_{20}H_{25}NO_4$. The spectroscopic properties of the compound were almost identical with those of erythraline and the mass spectroscopic breakdown pattern was consistent with an aromatic erythrina alkaloid having a 1,6 diene system and an additional oxygen function on either ring C or ring D. Extensive n.m.r. studies including the technique of internuclear double resonance (INDOR) showed that the oxygen function was located on ring C but were unable to identify the exact position or the configuration of this ring.

In view of the biological importance of curare effective compounds, the lack of substantial evidence for the assignment of the absolute configuration of the aromatic erythrina alkaloids and their relationship to the erythroidines, and the incomplete determination of the structure by spectroscopic methods, it was decided to attempt to determine the structure of the compound by X-ray crystallography.

The purified alkaloid ($C_{20}H_{25}NO_4$) crystallises from organic solvents in the orthorhombic space group $P2_12_12_1$. A preliminary photographic examination gave the following dimensions:-

$a = 9.7$, $b = 10.8$, $c = 17.4$Å,

but no further work was carried out on the alkaloid itself since the quality of the crystals was rather poor. Attempts to form a methiodide or hydrobromide were unsuccessful, the latter resulting in loss of one of the methoxy groups, but the compound was found to give a crystalline picrate and 2-bromo-4,6-dinitrophenolate salts.

Excellent crystals of the latter derivative were obtained from ethyl alcohol as thin hexagonal plates elongated along a and with (001) prominent. They showed sharp extinction parallel to the a axis. A photographic examination showed that the crystals had 2/m Laue symmetry and the absent spectra, 0 k 0, k = 2n, combined with the known optical activity, uniquely determined the space group as $P2_1$. The density of the compound was measured by flotation in an aqueous solution of potassium iodide.

Crystallographic Data

($C_{26}H_{28}N_3O_9Br$): M = 606.45: M.Pt. 144-148°C; Monoclinic; a = 7.998(3); b = 16.159(5); c = 10.550(5)Å; β = 97.48(1)°; U = 1352Å^3; D_m = 1.50(1); D_c = 1.49 g. cm^{-3}; for Z = 2; Space Group $P2_1$; F(000) = 624; μ(Cu-K_α radiation)= 29.5 cm^{-1}.

Intensity data for 2086 independent reflections were collected on a Siemens automatic single-crystal diffractometer operated in the coupled θ-2θ mode, using a maximum step time of 0.6 seconds per 0.01° in θ and out to θ = 60°. The reference reflection (0 10 0) showed no significant decline in intensity during the five days of the data collection. The raw data were processed in the normal manner and, of the 2086 reflections examined, 2044 were classified as significantly above the background radiation level (*i.e.* I > 2.58σ(I)).

The Solution and Refinement of the Structure

Examination of the Harker section of a three dimensional Patterson map revealed a single prominent peak which was assumed to be the bromine - bromine vector between the two molecules in the unit cell. The x and z coordinates of the bromine atom were thus deduced to be 0.036 and 0.173 respectively. The undefined coordinate, y, was then fixed at 0.5, thus locating the structure-factor origin in the b direction.

After three cycles of full-matrix least-squares refinement of the proposed bromine atom, the agreement factor, R, was 45.5% and no serious movement of the atom had occurred. A three dimensional bromine-phased Fourier map, using F_o terms for which $|F_c| \geq 0.3|F_o|$, was then computed and from this 10 of the 13 atoms comprising the 2-bromo-4,6-dinitro-phenol moiety were readily identified. However, the pseudo-mirror plane which occurs in space group $P2_1$ when a Fourier synthesis is based on only one atom, produced two mirror-image versions of this moiety sharing a common bromine atom. Either image can be accepted as the basis for further work and refinement; one was chosen arbitrarily and each atom was given a temperature factor of 3.5Å^2. Three cycles of isotropic refinement brought R down to 41.2%.

From a difference Fourier map, the remaining three atoms of the phenolate system were located and a fused five-and six-membered ring system was found. They were tentatively identified as rings B and C in the alkaloid skeleton. After refinement of the positions and isotropic temperature factors of all the atoms identified so far, the agreement factor was 35.9% and from anew difference map the remaining atoms in the structure were located. It was then clear that the preliminary assignment of the alkaloidal nitrogen atom had been incorrect, and that the bicyclic system postulated as rings B and C was in fact rings A and B, not B and C.

After three cycles of block-diagonal refinement, it was noticed that the temperature factors for atoms C16 , C21, C23 and O24 had become very large, so it was concluded that they were wrongly sited. From a difference Fourier map, new positions for C16, C21 and O24 were identified. Two of them survived refinement, but C21 moved to a chemically unlikely position. Another difference map suggested new positions for C21 and C23. It also revealed the presence of considerable residual electron density around the bromine atom. So, in the next refinement, C21 and C23 were placed in their newest positions and the bromine atom was allowed to

refine anisotropically. The newly shifted atoms survived refinement and the R factor, after 6 cycles of block-diagonal refinement of the whole structure, was 12.5%.

In view of the excellent morphology of the crystal used for the data collection, a correction for absorption effects was applied at this stage using Troughton's "ICABS" program[35]. After this correction had been applied, three further cycles of refinement reduced R to 11.0%.

The refinement of the structure was then transferred from the XRAY-63 system block-diagonal refinement program to the larger capacity full-matrix program "NUCLS"[36] and anisotropic temperature factors were applied to the oxygen and side chain carbon atoms. After three cycles of refinement in this way the R factor was 8.6%, but since the number of variables was up to the maximum permitted in "NUCLS", final refinement was carried out using the X-RAY-63 block-diagonal program and the whole structure was allowed to refine with anisotropic temperature factors. R converged to 8.2%.

The positions of the 15 "fixed" hydrogen atoms in the structure (*i.e.* the non-methyl and N-hydrogen atoms) were then calculated using the "BONDLA" program and an examination of a difference Fourier map at these positions enabled the tentative location of 14 of them together with 7 of the methyl hydrogen atoms. These atoms were assigned isotropic temperature factors equal to the spherical equivalents of those of the atom to which they were bonded and were included as fixed atoms in the refinement of the rest of the structure. Subsequently the hydrogen positional parameters were refined and finally the positional parameters of all the atoms and the temperature factors of the non-hydrogen atoms were refined for 3 cycles. The R factor fell to 6.8%, but examination of the carbon-hydrogen bond lengths and angles showed that several of the hydrogen atoms had moved to chemically unacceptable positions. A new difference map using all the non-hydrogen atoms was therefore computed. From this map

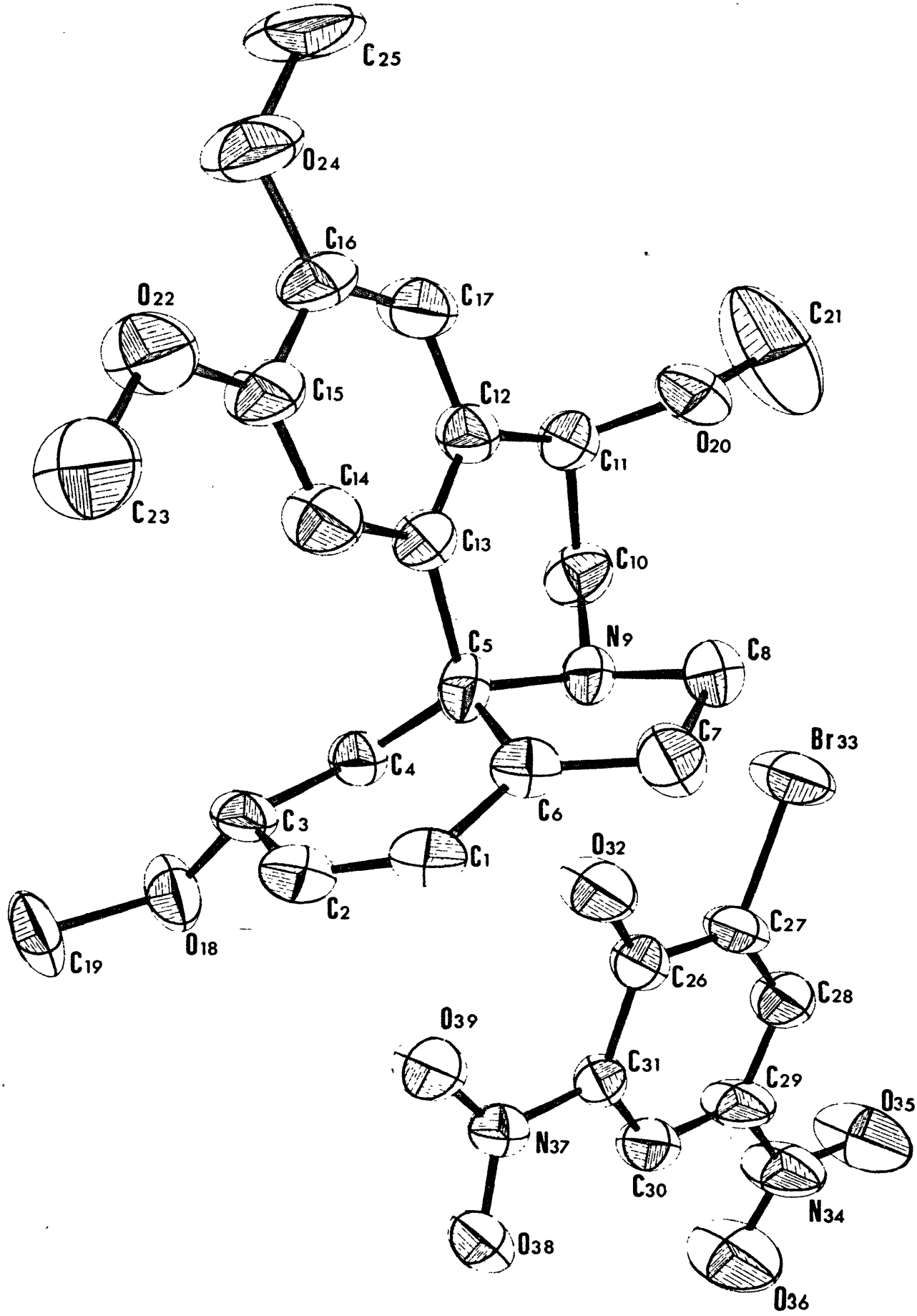


Figure 3 The structural asymmetric unit

24 possible hydrogen atoms were located, mostly in positions close to those found originally. These were included as fixed atoms in a refinement of the rest of the structure but a bond-length calculation again showed that several of the C-H bonds were rather long, particularly those in the methyl groups.

A final difference map showed that there was still residual electron density adjacent to the methyl hydrogen atoms but it was not possible to assign definite maxima to the unlocated hydrogens on C21 and C23. There was no sign of electron density for the hydrogen atoms on N9 and C28. The positions of the four unlocated hydrogen atoms were calculated using the XRAY-63 "BONDLA" program and are included in table 3.

Tables 1 and 2 contain the positional and thermal parameters of the non-hydrogen atoms and table 3 those of the hydrogen atoms. Table 4 contains the final structure-factor data based on the whole structure except for the methyl hydrogen atoms and the hydrogens on N9 and C28 (R = 7.1%) and figure 3 shows the structural asymmetric unit.

Results and Discussion

In accordance with the chemical and spectroscopic evidence, erythristemine was found to be a new aromatic erythrina alkaloid. It is the first known naturally occurring tetramethoxylated erythrina alkaloid, its nearest homologue being the trimethoxylated alkaloid derivative erysotrine (fig. 2a, R_1, R_2 and R_3 = Me), which is readily obtained by methylation of the three phenolic erythrina alkaloids erysodine, erysovine and erysopine (see fig. 2). Erythristemine carried its additional methoxy group in the 11 position and the presence of this group introduces a third asymmetric centre.

The absolute configuration of the compound was determined using the effects of the fluorescence of bromine under copper and molybdenum X-radiation (see below). The configurations of the three asymmetric

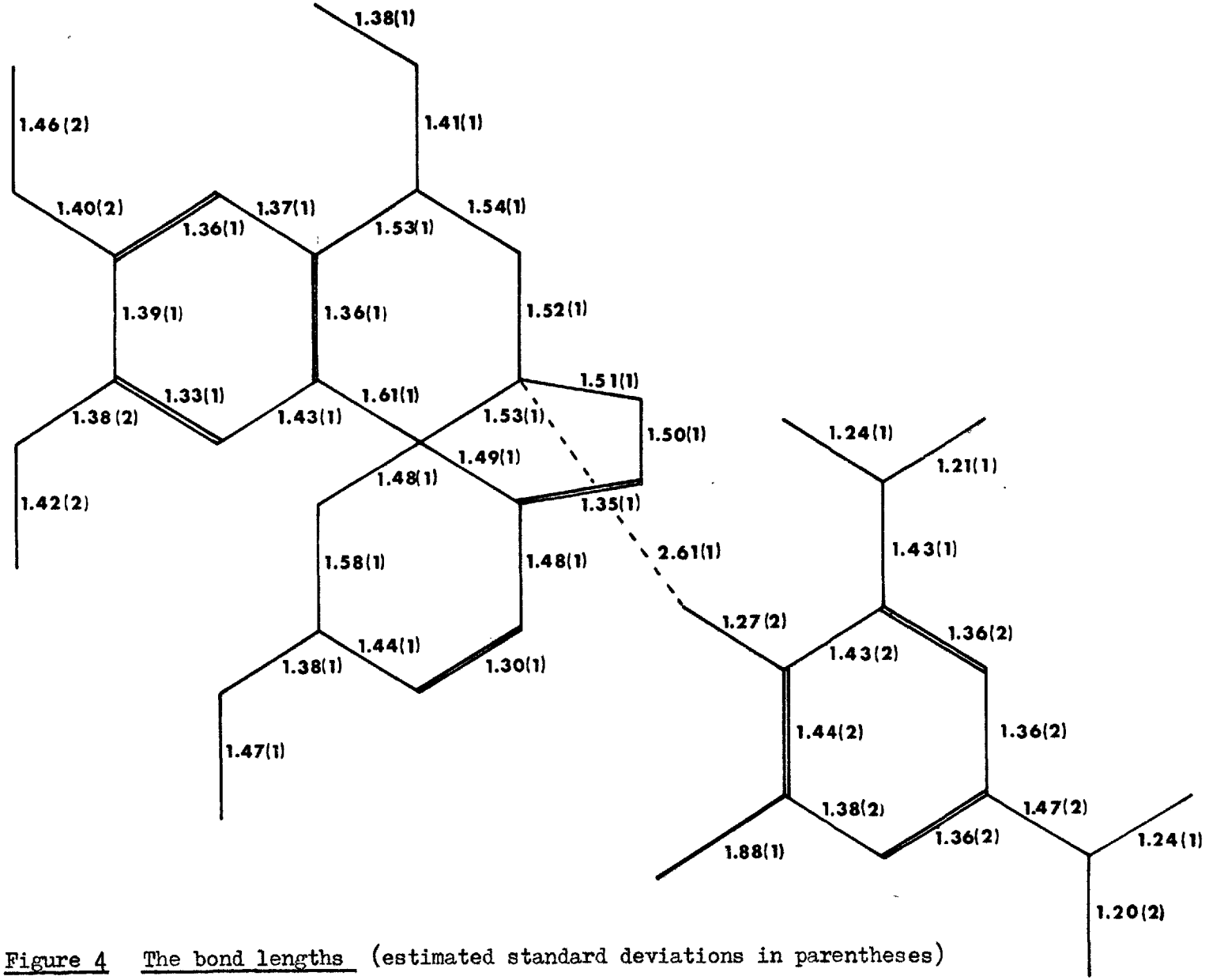


Figure 4 The bond lengths (estimated standard deviations in parentheses)

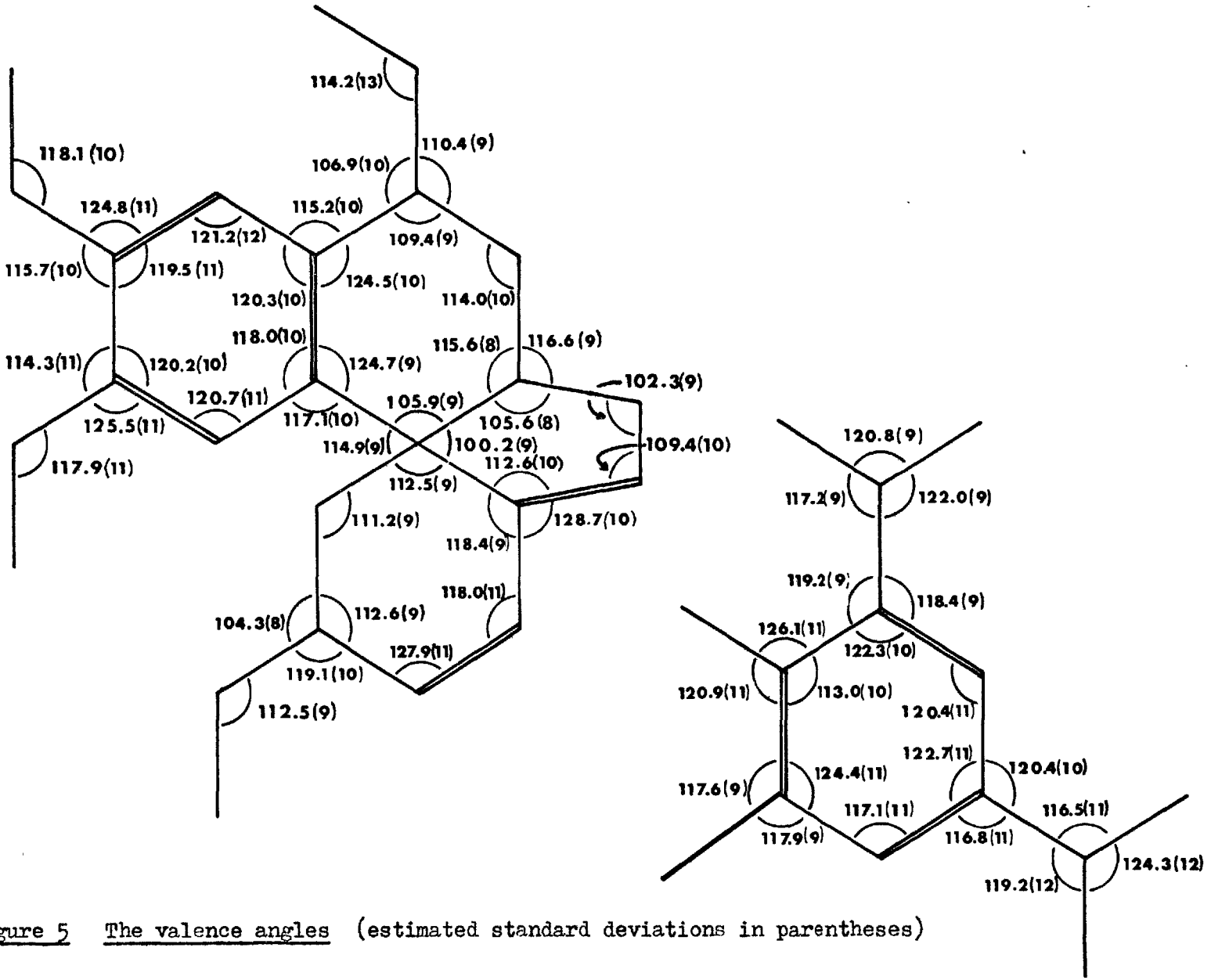


Figure 5 The valence angles (estimated standard deviations in parentheses)

centres are 3R, 5S, 11S, thus confirming the assumption that the 3R, 5S configuration of the erythroidines also occurs in the aromatic erythrina alkaloids. This result places the theory of a common biosynthetic origin for the two types of erythrina alkaloid on firm ground.

In the subsequent discussion of the molecular structure of the compound the following abbreviations will be adopted for the structures examined by X-rays:-

Erythristemine	ERTE
Erythraline[25]	ERTA
β-dihydro-erythroidine[28]	HERTO
Cocculine[34]	COCC,

and the 2-bromo-4,6-dinitro-phenolate moiety will be referred to as 2BDNP.

The molecule contains the expected tetra-cyclic erythrinan skeleton and is essentially similar in shape to ERTA[25] and COCC[34]. This similarity is enforced by the fusion of the A/B bicyclic ring system with the C/D ring system at the alkaloidal nitrogen atom (N9) and the spiro atom (C5). In ERTE rings C and D form a plane which lies at 84° to the best plane through rings A and B. In ERTA this angle is 88° but in HERTO the angle is 69°, reflecting the considerable difference between the latter molecule and the aromatic erythrina alkaloids.

The 2BDNP moiety lies beneath the plane of rings A and B, in a plane rotated some 20° out of that through rings C and D. The phenolic oxygen atom lies 2.61Å from the alkaloidal nitrogen atom, forming a short nitrogen to oxygen hydrogen bonded contact.

The bond-lengths and angles in the structure, together with their standard deviations are shown in figures 4 and 5. On the whole these values are similar to those reported for ERTA, COCC and, where relevant, HERTO. But, as no standard deviations were quoted for those structures, a detailed comparison of the individual bonds is difficult and the discussion will be restricted to common features.

a) The Erythristemine Moiety

The most marked deviation from normal bond lengths in this part of the molecule is that of the C5-C13 bond (1.61Å) which is some 10σ longer than the accepted value for a C_{arom}-C_{sp^3} bond (1.505Å). Table 5 contains the bond lengths and angles around C5 in the four crystallographically determined erythrina alkaloids and it is plain from these data that in three cases, HERTO, ERTE and COCC, either the C4-C5 bond or the C5-C13 bond is longer than normal. The sum of the two lengths is in each of these cases 3.09Å. In the fourth structure, ERTA, the effect seems to be shared between the two bonds and the sum of the two lengths is marginally higher (3.13Å).

In the three structures where bond angle data are available, the C4, C5, C13 angle is the largest of the six angles around C5 with a mean value of 114°. This is followed by the C4, C5, C6 angle (mean 112°). On the other hand, the C6, C5, N9 is considerably smaller than the normal tetrahedral angle with a mean value of 99° but this contraction is inevitable in a five-membered ring.

A second bond which deviates significantly from the normal value (1.51Å[87]) is the C_{sp^3}-C_{sp^2} single bond between C2 and C3 (1.44(1)Å) and the bonding system from C2 through C3, C4 and C5 to C13 is therefore generally irregular with alternate long and short bonds. This tendency is also apparent in COCC and ERTA and, together with opening of the associated valence angle around C5 is probably due to repulsion forces between the hydrogen atom on C14 and atoms in ring A.

Thus, it is clear from figure 3 that the fusion of the two "planar" moieties (rings A and B and rings C and D) at C5 and N9 produces a situation where C14 lies almost centrally over ring A and is close to all six atoms in that ring. In most cases the hydrogen atoms on ring A are directed away from that on C14 but in one case (H3) this is not so and a close hydrogen to hydrogen contact (2.14Å) occurs. This distance is

significantly shorter than twice the mean van der Waals radius of hydrogen calculated by Bondi[38] from a number of crystal structures (2.40Å) and therefore implies considerable steric repulsion. This repulsion is reinforced by H14 to C1 and H14 to C2 interactions since these contacts (2.65 and 2.54Å respectively) are also considerably shorter than the distance for an unstrained van der Waals contact between carbon and oxygen (2.90Å[38]). These forces are almost certainly responsible for the stretching of the C5-C13 and C13-C14 bonds, the shortening of the C14-C15 bond and the decrease in the bond angles at C13 and C16 from the 120° of a regular hexagon.

The two olefinic systems, C6, 1,2,3 and C5,6,7,8 are both planar (see table 6) but form a non-planar trans diene system with a torsion angle of 18° about the C1-C6 bond and with C1 lying 0.13Å out of the plane through C5,6,7,8. The torsion angle about the C1-C6 bond is similar to that about the same bond in ERTA (17.6°), the only other determined erythrina alkaloid containing the trans diene system. The $C_{sp}2$-$C_{sp}2$ single bond, C1-C6, with a length of 1.48(1)Å and the two double bonds, C1-C2 and C6-C7 with lengths of 1.30(1) and 1.35(1)Å respectively, are similar to those in the simplest trans diene, butadiene[39] (1.466(3) and 1.343(3)Å).

The three $C_{sp}3$-N bonds are significantly longer than the normal value for this type of bond[37] and are compared with C-N bonds in the other erythrina alkaloids in table 5a. In 1957 Hahn[40] concluded that protonation of a nitrogen atom resulted in an increase in $C_{sp}3$-N bond lengths and calculated an average value of 1.503Å. In 1962 Hamilton *et al*[41] suggested that this value should be 1.52Å, but five years later Birnbaum[42] calculated a weighted mean of 1.499(2)Å from 30 $C_{sp}3$-N^{+} bonds in a number of alkaloids. It is thus clear that the C-N bond lengths in all four determined erythrina alkaloids reflect the increase in length which arises on protonation of the nitrogen atom and are typical $C_{sp}3$-N^{+} bonds.

With the exception of the aromatic ring (D) and its substituents the rings in the structure are generally distorted from "ideal" conformations. This arises from either the presence of double bonds or the effect of fusion with neighbouring rings. The deviations of the atoms from the best or most prominent plane in each ring are shown in table 6 and table 7 contains the torsion angle around the bonds in each ring.

Ring A contains four accurately coplanar atoms, comprising the C6, 1, 2, 3 olefin system, and C4 and C5 are respectively above and below this plane. Ring B again contains a closely planar olefin system (C5, 6, 7, 8) and the departure of N9 from this plane brings the ring into the normal envelope conformation. Ring C is similar in shape to ring A, with four atoms (C11, 12, 13, 5) approximately planar and N9 and C10 above and below this plane. Ring D, the aromatic ring, is essentially planar and this planarity is extended to the majority of its substituent atoms, only C5 lying seriously out of the plane.

b) The 2-bromo-4,6-dinitro-phenolate moiety

Taken as a whole the 2BDNP moiety is approximately planar (the standard deviation of the component atoms from the plane through the whole is 0.1Å), but the benzene ring itself is closely planar (σ = 0.01Å). The details of these planes are given in table 6.

Of the "first" substituent atoms on the benzene ring, the bromine atom is the most seriously out of plane (0.1Å) but both nitro groups are rotated about their C - N bonds so that their oxygen atoms lie above and below the plane of the benzene ring.

The C27 - Br33 bond length lies well within the range of values for aromatic carbon to bromine bonds summarised by Christensen and Stromme[43] and is similar to that termed a "normal" length by Krigbaum and Wildman[44]: the benzene ring angle at the site of bromine substitution (C26, C31, C30) is 124.4°, significantly larger than in benzene itself and is also larger

than the normal value at a bromine substituted aromatic carbon atom (*ca.* 122°).

Table 8 contains a summary of the reported geometry of nitro groups in nitro-phenols and nitro-phenolates and shows that these groups in the present compound can justifiably be described as normal. Resonance interaction between the phenolic oxygen atom and the nitro group(s) in these compounds enhances the π character of the C-N bond. This results in bonds which are some 0.4Å shorter than those in aromatic nitro compounds where resonance interaction between the substituents does not occur. In p-nitro-toluene for example, where the nitro group does not interact with the methyl group, the C-N bond length is 1.482 ± 0.007Å [58] and in m-dinitro-benzene the C-N distances are 1.491(8) and 1.494(9)Å [59]. The ring angles at both sites of nitro substitution in 2BDNP are greater than 120° but this is again characteristic of aromatic nitro compounds and may be attributed to the electron withdrawing effect of nitro groups.

The mean N-O distance (1.228(6)Å) is again typical but one N-O bond in each nitro group is significantly shorter than the other. This seems to imply a predominance of one canonical form for these groups rather than an equal population of the two possibilities. Similar differences are apparent in several of the nitro groups in table 8 and have been observed in a number of non-phenolic aromatic nitro compounds (*e.g.* 1,3,5-trinitro-benzene)[45] but no systematic study of this discrepancy has been carried out.

The mean O - N - O angle (122.1(7)°) is close to that for the data in table 8 but the two angles themselves differ significantly. This difference fits the approximately linear correlation between the length of the C - N bond and the O - N - O angle in aromatic nitro compounds which was noted by Coppens in 1962 [60]. (Using his data one would predict O - N - O angles of approximately 121° at N37 and 124.3° at N34 from the observed bond lengths.) The correlation is generally explained in

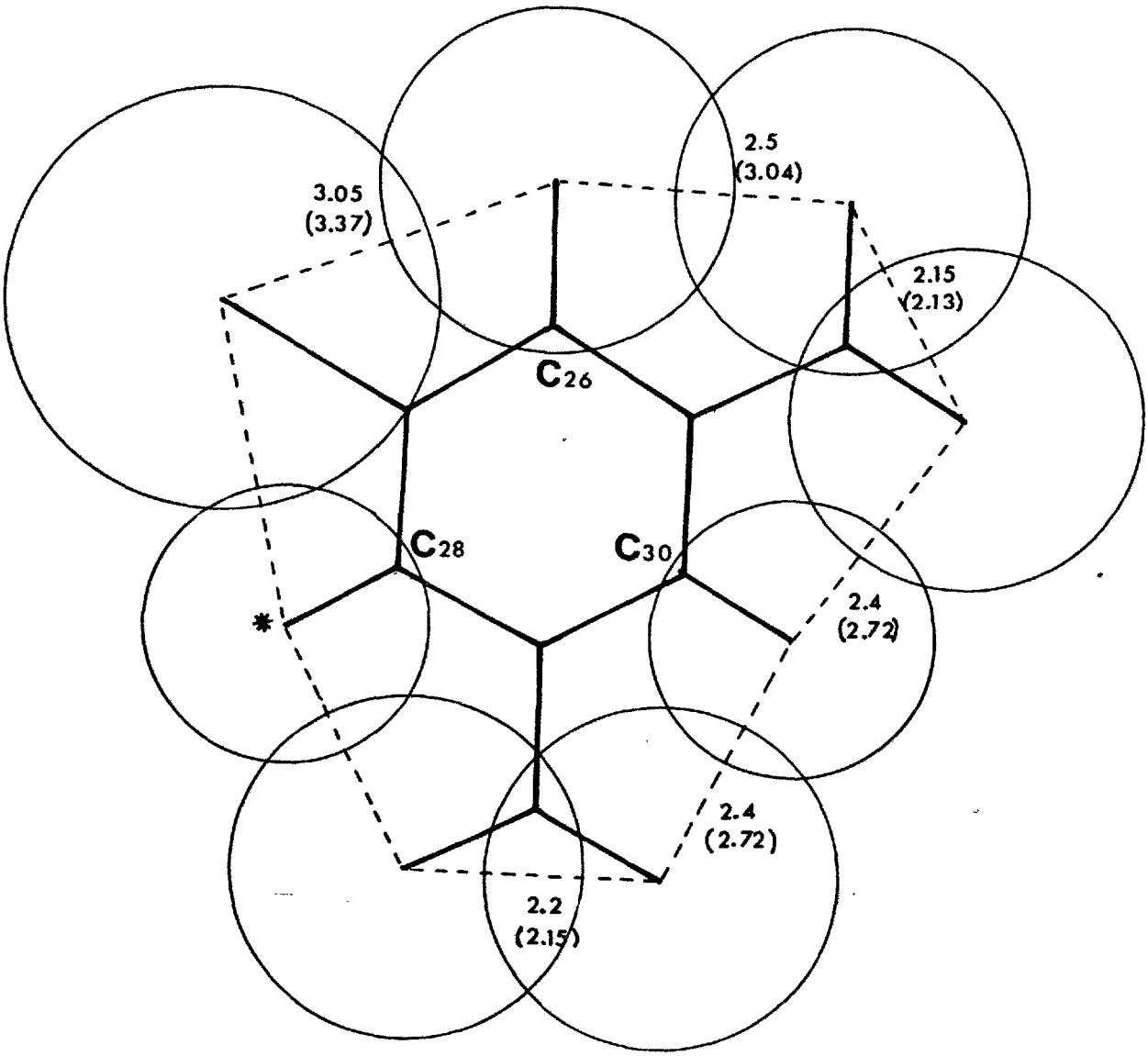


Figure 6 Van der Waals "surfaces" of the substituents in the 2BDNP moiety

terms of repulsive forces between the oxygen atoms in a nitro group (tending to increase the O - N - O angle) and similar forces between these oxygen atoms and the substituent atoms occupying the two positions ortho to the nitro group (increasing the C - N length to permit an opening of the O - N - O angle).

Figure 6 shows the van der Waals "surfaces" (using the radii of Bondi[38]) of the oxygen, hydrogen and bromine atoms in the 2BDNP moiety. The figure is drawn to scale using the observed bond lengths and angles but assuming the whole to be planar. The estimated interatomic distances between the non-bonded substituent atoms (measured from the drawing) are shown and are compared with the observed values (in parentheses). The differences in these lengths gives an approximate indication of the effect of rotation of the nitro groups about their C - N bonds and the areas of overlap of the van der Waals surfaces possibly indicates why the ortho nitro group is rotated more than the para group (approx. 13 and 7° respectively - see torsion angles in table 7). An approximation of this type was used by Trotter[59] in his discussion of the differences between nitrobenzene (where the nitro group is coplanar with the benzene ring since repulsion forces can be relieved by small, in plane, lateral movements of the ortho protons)and meta and para dinitro-benzene. Overlap forces of the same type are doubtless responsible for the displacement of the bromine atom from the plane of the benzene ring.

Tables 9, a and b contain a summary of some of the bonding data in nitro-phenols and nitro-phenolates and show the remarkably large change in geometry which occurs on loss of the phenolic proton. Comparison with the present structure shows very plainly that the 2BDNP moiety is anionic and that, as such, it is entirely typical.

The C - O bond is significantly shorter than in nitro-phenols but a comparison with ortho and para-benzoquinones (1.218(2)[61] and 1.222(8)Å [62] respectively) and aldehydes and ketones (1.215(5)Å [37]) shows that it has

only partial double bond character. It is generally accepted that the strength of an organic acid is directly related to the ability of the anion to delocalise the negative charge which results from dissociation. In nitro-phenols this should be related to the length of the C - O bond in the anion but, although the pK_a values of most nitro-phenols are published[63], the only crystal structures containing nitro-phenolate anions are those shown in table 9b and these are mostly picrates. Phenol itself has a pK_a value of 9.9 and progressive substitution by nitro groups in the 2,4 and 6 positions reduces this to the 0.38 of picric acid itself. In o-nitro-phenol (pK_a 7.17) the anionic C - O length is approximately 1.29Å (see table 9b) whereas in picric acid this length is 1.24Å. If one assumes a linear relationship, the length of the 2BDNP C - O bond suggests a pK_a value of approximately 4, similar to that of 2,4-dinitro-phenol (3.96 [63]). Since halogen substituents have little effect on the electron density in aromatic rings the bromine atom should not change the pK_a of 2,4-dinitro-phenol seriously and a value of 4 for 2BDNP is reasonable. Unfortunately no experimental value has been reported.

The most remarkable changes resulting from loss of the phenolic proton are those in the ring system itself. The two C - C bonds adjacent to the phenolic oxygen atom are lengthened from the 1.395(3)Å in the undissociated phenol (the normal C_{arom} - C_{arom} length) to approach that of the "single" bond between two sp^3 carbon atoms (1.466(3)Å in butadiene[39]).

The remaining four bonds in the ring are approximately equal (average 1.38Å) and the ring is planar so that the increased lengths of bonds 1 and 7 in the table results in a closing of the ring angle at the phenolic oxygen site from the 120° in benzene (and in nitrophenols - see table 9a) to a mean value of 112.4°.

c) The N9-O32 Hydrogen Bond

Hydrogen bonded contacts between nitrogen and oxygen have been reviewed by numerous authors[64-68]. Most workers have classified them into three basic types, N-H...O, N...H-O and $NH^+...O^-$, which differ in the position of the hydrogen atom and in the charge distribution within the N-H-O system. The mean N-O distances in the three types also differ significantly but the individual distances in each group are widely scattered due to variations in the substituents on the donor and acceptor atoms and crystal lattice effects (see histograms in refs. 64 and 65). In general, the nitrogen to oxygen distances in the three classes are: 2.9 - 3.05Å for N-H...O bonds, 2.8Å for N...H-O bonds and 2.83Å for $NH^+...O^-$ bonds.

The distance between the "bonded" atoms in the two parts of the ERTE-2BDNP molecule, 2.61(1)Å, is thus considerably shorter than the normal nitrogen to oxygen separation in hydrogen bonded systems of all three types. The contact is also very much shorter than the sum of the van der Waals radii of nitrogen and oxygen[39] (1.52 and 1.55Å respectively) and must imply a strong hydrogen bond.

The phenolic hydrogen atom was not located during the structure determination but, since the geometry of the 2BDNP moiety implies that it is anionic and the lengths of the C-N bonds in the ERTE moiety suggest $C-NH^+$ bonds, it seems reasonable to assume that the N9-O32 contact is a hydrogen bond of the $NH^+...O^-$ type. This assumption is supported by the pK_a values of 2BDNP (approx. 4 - see above) and of protonated erythrina alkaloids (erythraline hydrochloride has a pK_a of 5.97 and is the most acidic of the aromatic erythrina alkaloids[20])

The position of the "phenolic" proton H9 was calculated assuming a linear hydrogen bond of the $NH^+...O^-$ type and an N-H bond length of 1.06Å. This value was taken from the results of two neutron diffraction studies where $NH^+...O^-$ hydrogen bonds shorter than 2.7Å were encountered[69,70].

The bond angles and distances involving H9 are:-

N9 - H9	1.06Å	(assumed)
H9 - O32	1.55Å	
C5 - N9 - H9	102.7°	
C8 - N9 - H9	103.4°	
C10 - N9 - H9	111.4°	
N9 - H9 - O32	180.0°	(assumed)
H9 - O32- C26	155.4°	

From the bond angles at N9 it is clear that the assumption of a linear hydrogen bond gives a nitrogen with reasonably tetrahedral geometry. The H9-O32 distance is comparable with H-O distances reported by Jönsson and Hamilton[69] (1.646(6)Å) and Kvick *et al.*[70] (1.608(4)Å) in the neutron diffraction studies mentioned above. On the other hand, the H9-O32-C26 angle is larger than the 120° expected for an sp^2 phenolic oxygen atom with one lone pair involved in the hydrogen bond and the hydrogen atom also lies 0.08Å below the plane through O32,C26,C27 and C31. Calculation of a position for H9 assuming an O-H distance of 1.6Å [69,70], an H-O-C angle of 120° and that the hydrogen atom lies in the plane through O32,C26,C27 and C31 (*ie.* idealised involvement of the oxygen lone pair) gives an N-H distance of 1.61Å and angles at N9 ranging from 86° to 138°. These values are distinctly inferior to those assuming a linear hydrogen bond.

The N-O distance, although shorter than the mean value for $NH^+...O^-$ hydrogen bonds, is not exceptional. Two instances with N-O contacts shorter than 2.7Å have been mentioned above[69,70] and the tables given in references 65 and 66 contain other examples. Since the two picrates shown in table 9 which contain $NH^+...O^-$ hydrogen bonds have N - O distances of 2.70 (serotonin picrate) and 2.73Å (diazepin picrate), values which approach that found in the present case, it may be that this length is normal in picrates, although shorter than the overall average.

The Absolute Configuration

The absolute configuration of the compound was determined by visual comparison of a selected set of Bijvoet pairs[71]. The relevant reflections in space group $P2_1$ are $\underline{h}\ \underline{k}\ \underline{l}$ or $\underline{\bar{h}}\ \underline{k}\ \underline{\bar{l}}$ versus $\underline{h}\ \underline{\bar{k}}\ \underline{l}$ or $\underline{\bar{h}}\ \underline{\bar{k}}\ \underline{\bar{l}}$.

A crystal was therefore mounted about the unique axis (b) and a series of zero-layer (hk0) precession photographs were measured using Mo-K radiation (the fluorescence effect is more serious with Mo than with Cu radiation). Using the X-RAY-67 program "CRYLSQ", the intensities of a set of hk0, $h\bar{k}0$ Bijvoet pairs was calculated and the sign of calculated intensity differences is compared with the observed difference below:-

h	k	l	Calculated	Observed
6	3	0	+	-
2	4	0	-	+
5	4	0	-	+
1	5	0	-	+
2	5	0	-	+
3	5	0	-	+
5	5	0	-	+
7	5	0	-	-
1	6	0	-	+
2	6	0	-	+
1	7	0	-	+
2	7	0	-	+
4	7	0	+	-
5	7	0	+	-
7	7	0	-	+
6	5	0	+	-
4	5	0	+	-
5	6	0	+	-
6	6	0	+	-
8	1	0	-	+

This indicates clearly that the absolute configuration is the opposite of that used for the structure determination and supports the results of the structure-factor calculations outlined above.

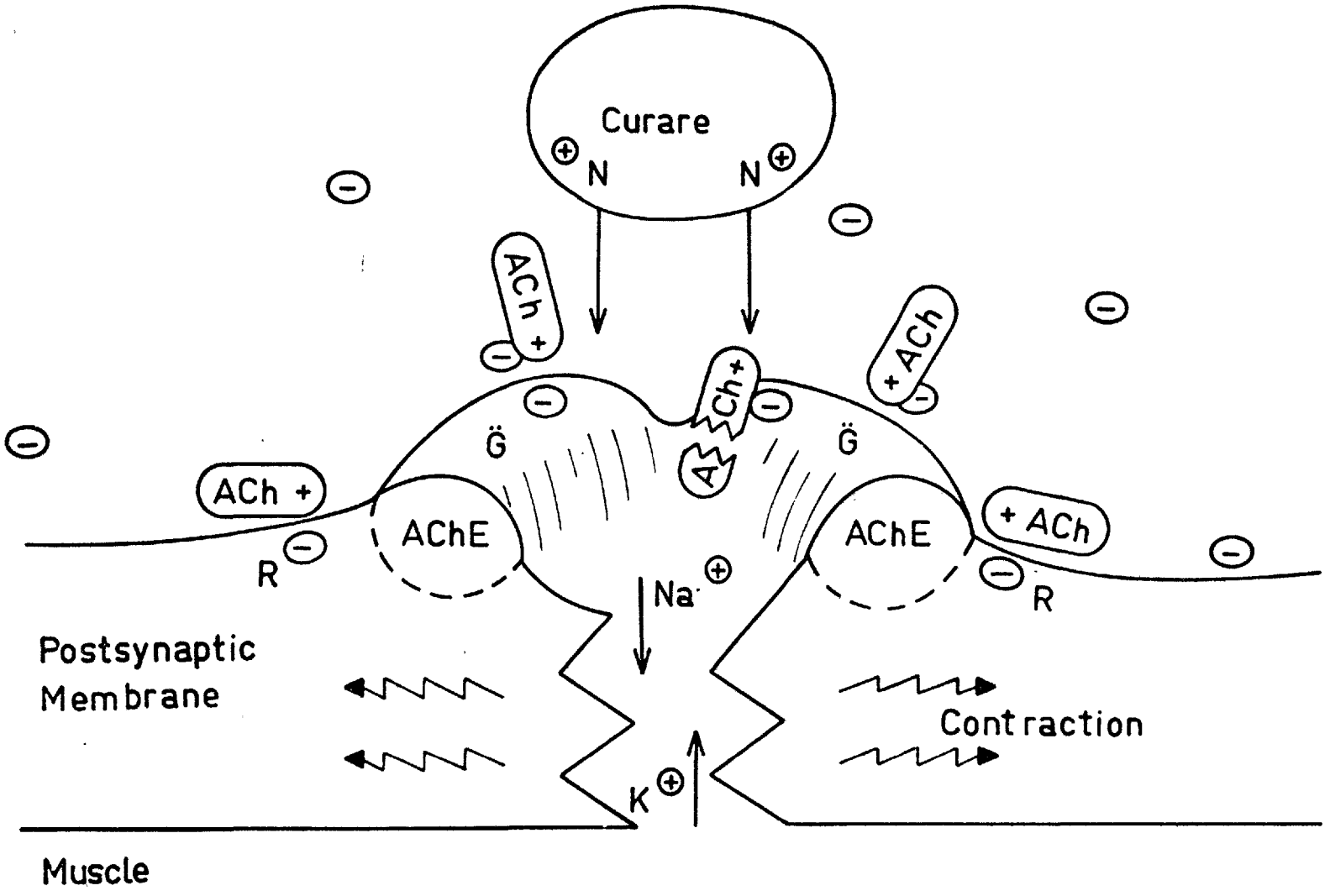


Figure 7

Schematic cross-section of a receptor area according to Waser[73]

The atomic coordinates and the diagrams in the chapter conform to the correct absolute configuration.

Pharmacology

A great deal of research has been carried out in recent years into the mode of action of neurotransmitters and their antagonists and into the nature and operation of receptor sites.

The arrival of a "quantum" of a neuro-transmitter at the post-synaptic membrane causes a change in the ionic permeability of the membrane which permits an ion flow. This results in a depolarisation which initiates muscle contraction or the transmission of a new nerve impulse. Waser[73] has proposed the model of the pores which permit Na^+/K^+ exchange across the post-synaptic membrane shown in figure 7. In this model two molecules of acetyl-choline-esterase (AChE) are arranged to form a circle which enclosed the mouth of the ion pore. In the unexcited state the pore is closed but the arrival of acetyl-choline (ACh) or ACh mimics at receptor sites behind the AChE molecules results in allosteric changes in the latter molecules which open the pore and thus permit an ion flux.

Tracer studies[72] have shown that the pore is surrounded by approximately 10 ACh receptor sites and that these are not the same as the active sites in AChE although they are closely connected to the AChE molecules. ACh receptor sites and the two active sites in AChE are ionic, so that a positive charge is a primary requirement for binding at either of these sites. Since many of the curare agents (*e.g.* d-tubocurarine and the toxiferines) are bulky and contain two positively charged nitrogen atoms, Waser[73] has suggested that they combine with the AChE molecules at the ion pore entrances, using two of the ionic sites, and thus prevent an ion flow. Smaller ACh agonists and antagonists, including the erythrina alkaloids are believed to bind to the ACh receptor sites.

An examination of the structures of a series of compounds which show a particular biological activity has long been considered to give an

indication of the basic structural requirements for that activity. Since the effect of neurotransmitters and their antagonists is readily observed, they have been widely studied from this viewpoint and the structures of many such compounds have been examined by crystallographic spectroscopic and theoretical methods. As yet, no single set of structural requirements has been universally accepted but most workers agree that a positively charged nitrogen atom and a hydrogen-bonding acceptor group are essential. Most workers also agree that the separation distance between these sites is centred around 4.7Å for nicotinic activity[3] and 3.3Å with a second acceptor site at 5.9Å for muscarinic activity[3]. Theoretical calculations have shown that the proposed hydrogen bond acceptor sites in many agonists and antagonists carry partial negative charges[74] and are therefore suitable for hydrogen bonding. Acetyl choline is able to satisfy both of these sets of conditions[74] although there has been some discussion as to how this is achieved[75,76,3].

Unfortunately many of the molecules under consideration have been non-rigid, necessitating the assumption of similar conformations in vivo to those found by crystallographic or theoretical methods in the solid or gaseous state. The erythrina alkaloids are, however, less subject to this difficulty since they are rigid compounds whose conformations are unable to change much on solution. In a study of "rigid" cholinergic agents which included β-erythroidine, Beers and Reich[3] concluded that the separation distance between the charged nitrogen atom and the hydrogen-bond acceptor group is about 4.7Å for nicotinic agents including the erythrina alkaloids and 3.2Å for muscarinic agents.

The crystal structures of the three determined erythrina alkaloids (insufficient data has been published for COCC) support this viewpoint. Beers and Reich suggested from models that the lactone "ether" oxygen atom at position 16 in the erythrinan skeleton (see figure 3c, III and 3d) was the acceptor atom, lying at approximately 4.7Å from the alkaloidal

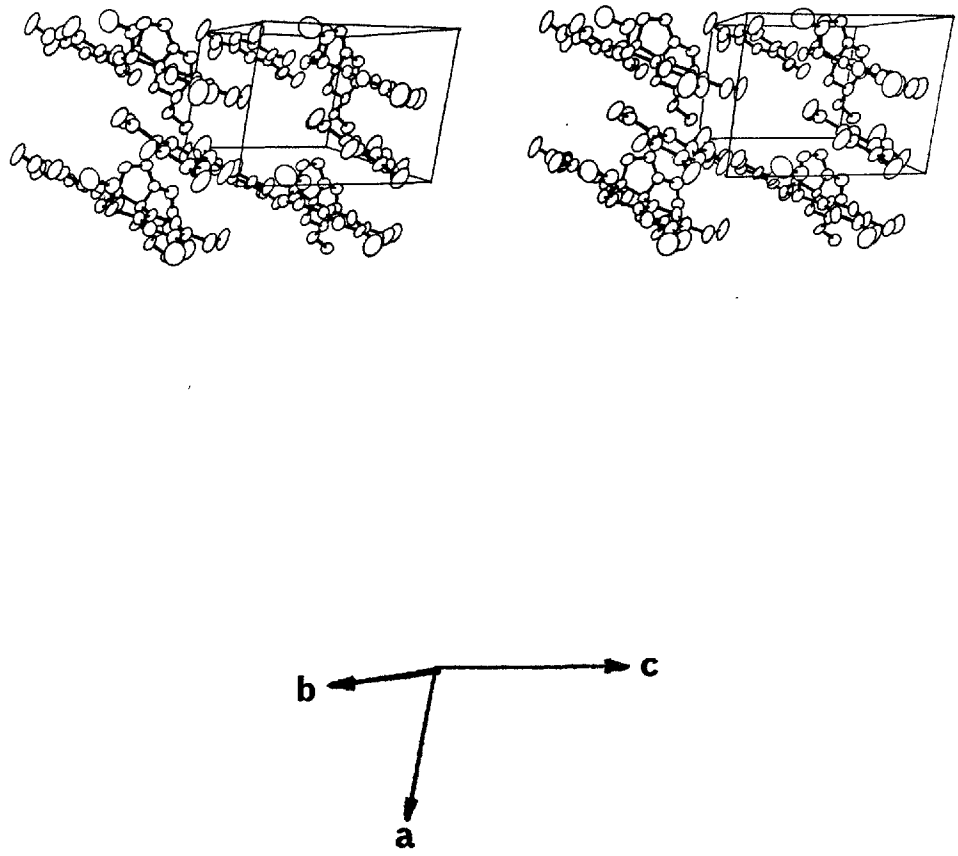


Figure 8 The crystal structure

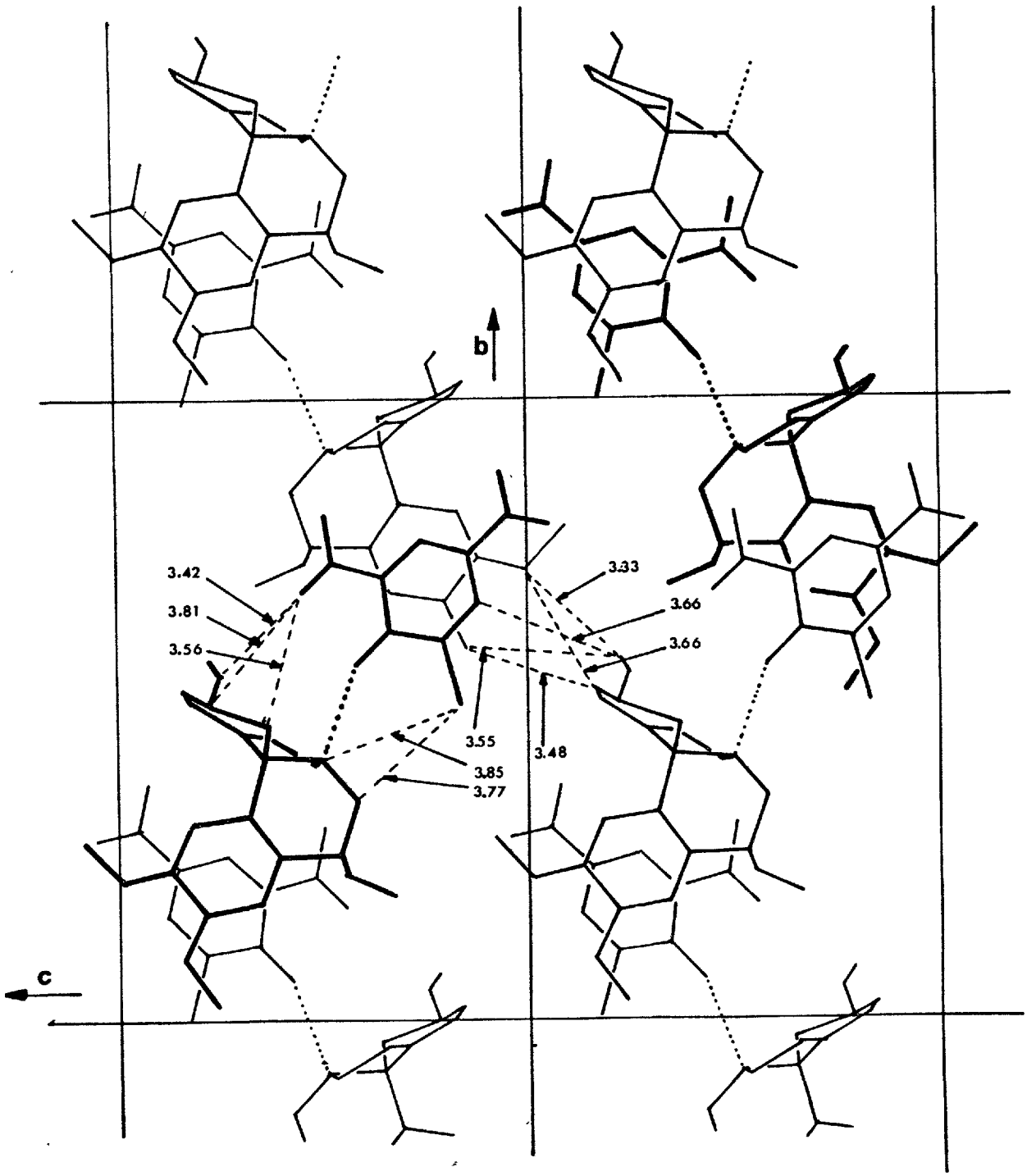


Figure 9

nitrogen atom (N^+ in vivo). They pointed out that the oxygen atom in the 3-methoxy group was also at this distance from the nitrogen atom. An examination of the nitrogen to oxygen distances in the three determined erythrina alkaloids, all of which are curare agents, seems to indicate that this is the more likely acceptor since it is the only oxygen atom common to all three structures which lies at about 4.7Å from the nitrogen atom:-

	Nature and Position of Oxygen Atom					
	3	16	16	15	15	11
Alkaloid	OMe	-O-	OMe	=O	OMe	OMe
HERTO	4.88	4.92		5.80		
ERTA	4.76		6.35		6.15	
ERTE	4.77		6.46		6.19	3.12

At first sight the presence of the 11-methoxy group in ERTE might suggest that the molecule should show muscarinic activity since it contains N - O separations of 4.7 and 3.1Å. However, models of ERTE and muscarine (fig. 1,d) show that this is unlikely since the relative positions of the nitrogen atom and the two suggested[74] H-bond acceptor groups in muscarine and its mimics ~~is~~ are considerably different from those in ERTE.

The Crystal Structure

Due to the shape of the erythristemine moiety, the packing of the molecules to form the crystal structure is difficult to illustrate without serious overlap. Figure 8 is stereoscopic diagram looking into the b axis from "below" the unit cell. It contains four asymmetric units related by unit-cell translations in the a and c directions and two 2BDNP moieties from molecules generated by the diad screw axes. The figure shows the most characteristic feature of the crystal packing, namely the alternating parallel arrangement of the ERTE moieties (ring D) and the 2BDNP moieties to form continuous stacks or columns of molecules. Figure 9 is a cross section through a number of these columns and shows the relationship between them. In this figure the components of two diad

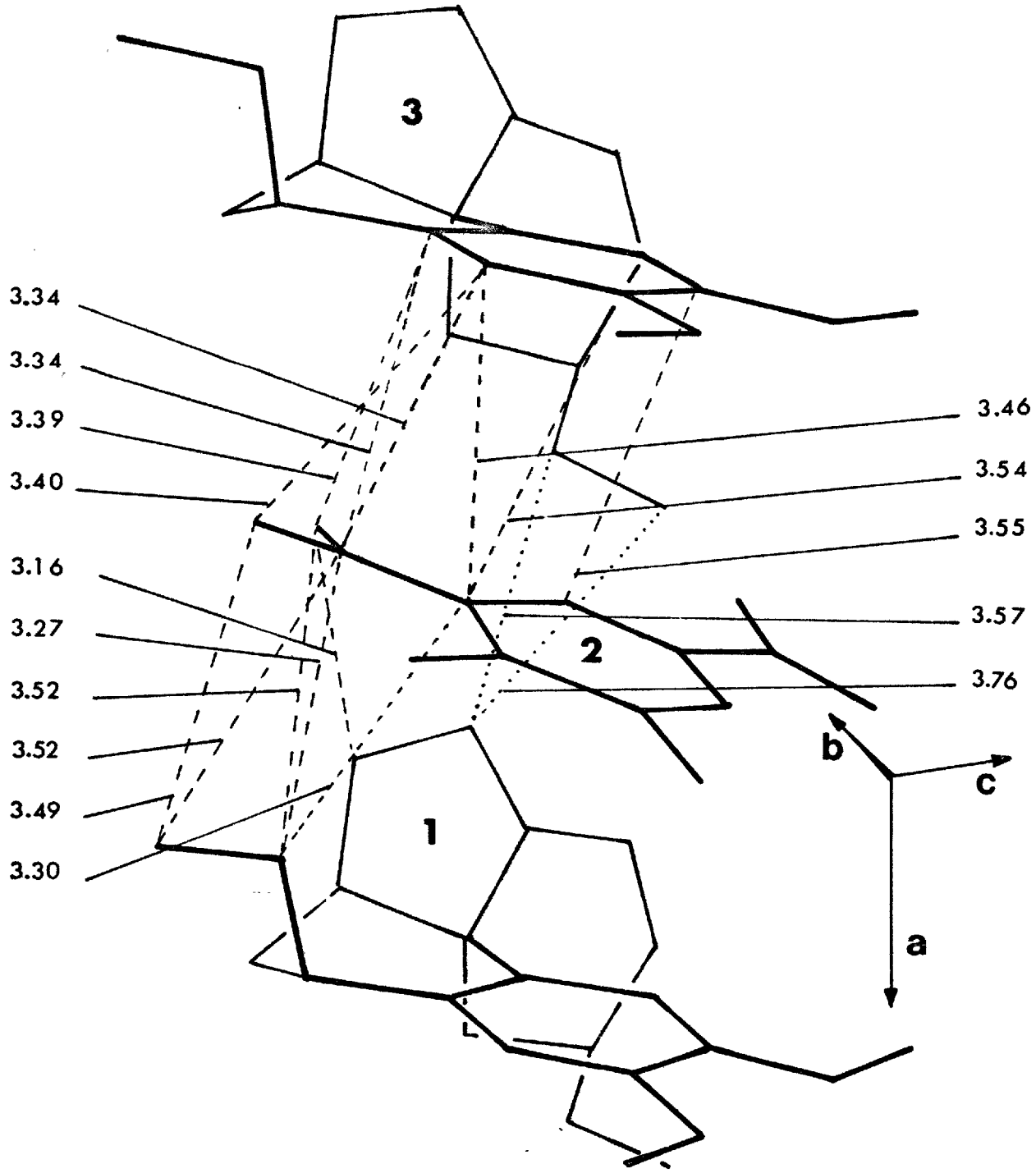

3
2
1
3.34
3.34
3.39
3.40
3.16
3.27
3.52
3.52
3.49
3.30
3.46
3.54
3.55
3.57
3.76
b
c
a
L

Figure 10

screw related molecules are drawn more heavily and show that the anion and cation (*i.e.* ERTE and 2BDNP) in a complete molecule are components of neighbouring columns. As a consequence of this, the columns are held together by the inter-molecular $NH^+ - O^-$ hydrogen bonds and inter-molecular contacts between Br33 and O29 in the 2BDNP anion and various atoms in the ERTE cation. These distances, together with other short contacts between the columns, are shown in figure 9.

Figure 10 shows the arrangement of molecules within the columns. The ERTE cations marked "1" and "3" in the figure are related by a unit-cell translation in the a direction and "2" is an intruding anion from a symmetry related molecule. The pattern shown in figure 10 is repeated continuously to form the columns. The figure also shows that the aromatic moieties are stacked obliquely to the column axis, a feature which is very common in crystals of aromatic compounds.

Figure 10 also contains the lengths of the shorter intramolecular contacts within the columns, and shows that these fall into two categories. The contacts in the first category are normal van der Waals contacts between sp^3 hybridised atoms in molecule 1 and sp^2 hybridised atoms in molecule 2. The contacts in the second category are between atoms in molecules 2 and 3 which are all sp^2 hybridised and may imply $\pi - \pi$ interaction.

Ring D, the aromatic ring in the ERTE moiety, carries two methoxy and two aliphatic substituent groups. Since methoxy groups are strongly and aliphatic groups weakly electron donating the aromatic ring should be comparatively rich in electrons and thus able to donate charge to the strongly electron accepting 2BDNP group to form a π-molecular complex. Compounds of this type have been reviewed by Prout and Wright[77] and very extensively by Herbstein[78]. In general the donor-acceptor process results in a coloured compound due to the formation of a "charge transfer" band in the visible region of the spectrum (*ca.* 460 nm. for nitro-aromatic

acceptors)[77] - crystals of the present compound were an intense yellow whereas erythristemine itself is white.

The contact distances between molecules 2 and 3 in figure 12 are typical C - N and C - O inter-molecular distances for π - π compounds with nitro-aromatic acceptors. The presence of the 11-methoxy group in the ERTE molecule and the C7 - O18 and C19 inter-molecular contacts prevent the close approach of ring D in molecule 1 to the 2BDNP molecule (2) so that the π - π interaction is not continuous through the columns of molecules.

The van der Waals volume of the complete ERTE - 2BDNP molecule, calculated using the increments tabulated by Bondi[79] is 465$\AA^3$ giving a packing coefficient of[80] 0.689. This is a typical value for simple aromatic compounds[80] and probably reflects the effect of the parallel packing of the aromatic moieties in the structure.

Table 1. Positions of the non-hydrogen atoms in fractional coordinates x 10^4 (Standard deviations in parentheses).

Atom	x	y	z
C1	2223(14)	5142(7)	8288(11)
C2	3842(15)	5280(6)	8520(11)
C3	5139(13)	5015(8)	7785(9)
C4	4407(13)	4709(6)	6405(9)
C5	2841(12)	4226(7)	6432(9)
C6	1590(14)	4663(6)	7133(10)
C7	4(14)	4613(8)	6522(11)
C8	3(15)	4130(7)	5308(11)
N9	1830(11)	4174(5)	5093(8)
C10	2427(15)	3525(8)	4211(11)
C11	2337(15)	2630(7)	4709(11)
C12	2977(13)	2609(7)	6141(10)
C13	3119(14)	3286(7)	6917(10)
C14	3673(16)	3163(8)	8254(10)
C15	4145(17)	2420(7)	8698(11)
C16	3994(17)	1736(7)	7886(12)
C17	3406(17)	1841(7)	6629(12)
O18	6400(10)	5563(5)	7578(7)
C19	7518(15)	5761(9)	8757(11)
O20	650(11)	2351(5)	4566(8)
C21	88(28)	2031(16)	3370(19)
O22	4762(15)	2239(5)	9950(9)
C23	4851(26)	2894(11)	10859(13)
O24	4502(14)	980(5)	8452(9)
C25	4327(26)	239(7)	7654(17)
C26	1855(15)	6251(7)	3392(11)
C27	972(16)	6097(7)	2133(11)

Table 1 (continued)

Atom	x	y	z
C28	628(16)	6693(7)	1201(12)
C29	1174(18)	7477(7)	1500(12)
C30	2044(16)	7679(7)	2659(11)
C31	2346(13)	7099(6)	3591(10)
O32	2114(13)	5671(5)	4209(8)
BR33	342(2)	5000(0)	1712(1)
N34	854(17)	8102(7)	488(10)
O35	-37(19)	7931(6)	-479(9)
O36	1524(18)	8788(6)	714(11)
N37	3246(12)	7343(5)	4801(9)
O38	3856(13)	6840(5)	5580(8)
O39	3386(11)	8093(5)	5019(8)

Table 2 Anisotropic Temperature Factors x 10^4 (Standard deviations in parentheses).

The anisotropic temperature factor is given by:-

$\exp-(\beta_{11}h^2 + \beta_{22}k^2 + \beta_{33}\ell^2 + \beta_{12}hk + \beta_{13}h\ell + \beta_{23}k\ell)$

ATOM	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}
C1	171(21)	24(5)	126(14)	- 4(9)	3(14)	2(7)
C2	177(22)	27(5)	97(13)	0(8)	10(14)	-10(6)
C3	188(20)	20(4)	77(10)	-11(10)	-13(11)	5(7)
C4	149(20)	26(5)	69(10)	- 2(8)	-11(11)	0(5)
C5	103(18)	40(5)	65(10)	0(8)	- 7(11)	10(6)
C6	15(20)	28(4)	89(12)	- 2 (8)	11(12)	1(6)
C7	143(20)	41(5)	101(13)	-11(9)	8(13)	- 6(7)
C8	140(20)	36(5)	102(14)	0(9)	-24(14)	- 5(7)
N9	125(16)	31(4)	70(9)	5(7)	-17(9)	- 1(5)
C10	185(23)	39(6)	79(11)	8(10)	6(13)	11(7)
C11	170(23)	34(5)	88(12)	- 5(9)	-26(13)	9(7)
C12	15(21)	30(5)	71(11)	- 5(8)	2(12)	7(6)
C13	157(21)	30(5)	78(11)	- 3(8)	-12(12)	11(6)
C14	216(26)	37(5)	75(12)	-12(10)	8(14)	7(7)
C15	238(27)	31(5)	75(12)	- 4(10)	-11(14)	20(7)
C16	234(27)	29(5)	10(13)	- 1(10)	-11(15)	24(7)
C17	229(27)	29(5)	99(13)	-14(10)	5(15)	3(7)
O18	17(16)	48(4)	89(9)	-29(7)	-44(9)	12(5)
C19	153(23)	65(8)	93(13)	-21(11)	-38(14)	-29(9)
O20)	221(19)	31(4)	124(11)	-16(7)	-17(11)	-19(5)
C21	397(52)	42(19)	186(26)	-92(26)	26(29)	-99(19)
O22	381(27)	41(5)	82(9)	6(10)	-43(12)	-17(6)
C23	485(52)	62(9)	81(15)	2(18)	-66(22)	- 5(9)
O24	382(27)	33(4)	123(11)	0(9)	-10(14)	27(6)
C25	497(54)	17(6)	200(24)	6(13)	8(29)	2(8)
C26	175(22)	30(5)	93(12)	- 2(9)	-29(13)	43(7)
C27	215(26)	24(5)	117(14)	-11(9)	-32(16)	-5(7)
C28	214(27)	34(6)	100(14)	0(10)	-27(15)	10(7)
C29	285(30)	22(5)	98(14)	- 3(10)	-25(17)	12(7)
C30	196(24)	31(5)	94(13)	6(9)	- 4(14)	-7(7)
C31	128(18)	26(4)	81(11)	- 5(8)	-13(11)	10(6)
O32	343(25)	36(4)	111(11)	-24(9)	-79(13)	17(6)

Table 2 (continued)

BR33	394(4)	30(1)	136(2)	-24(1)	-87(2)	1(1)
N34	381(34)	28(5)	118(14)	-12(11)	-31(17)	16(7)
O35	606(41)	47(5)	115(12)	-21(12)	-154(19)	20(7)
O36	530(39)	44(5)	151(14)	-18(12)	-86(19)	33(7)
N37	159(18)	26(4)	98(11)	- 1(7)	- 8(11)	0(6)
O38	320(24)	35(4)	112(11)	1(8)	-72(13)	6(6)
O39	215(18)	37(4)	118(10)	-13(7)	-32(11)	-14(5)

Table 3. Positions of the hydrogen atoms in fractional coordinates x 10^3. Each hydrogen atom is given the number of the carbon atom to which it is bonded. Atoms marked with an asterisk are shown with calculated coordinates (see text).

Atom	x	y	z
H1	132	528	890
H2	416	572	907
H3	552	452	803
H4A	528	450	593
H4B	440	504	576
H7	868	488	673
H8A	942	460	450
H8B	964	350	547
H9*	195	478	473
H10A	140	368	333
H10B	384	380	400
H11	348	234	420
H14	380	372	883
H17	328	148	623
H19A	768	504	873
H19B	672	598	960
H19C	912	552	860
H21A	868	180	367
H21B	133	220	263
H21C*	-22	177	251
H23A	558	280	173
H23B	360	340	80
H23C*	484	262	1171
H25A	480	984	817
H25B	340	8	683
H25C	520	16	680
H28*	0	656	34
H30	224	804	273

TABLE 4 Comparison of observed and calculated structure amplitudes for Erythristemine 2-Bromo-4,6-Dinitro-Phenolate

The data are listed in groups of constant h and k and consist of l, $10|F_o|$ and $10|F_c|$. Reflection marked "*" were classified as unobserved.

0,0,L
1 193 232
2 159 122
3 853 913
4 223 227
5 127 136
6 510 910
7 15* 31
8 89* 35
9 107 90
10 119 107
11 36 35

0,1,L
1 855 902
2 475 508
3 177 139
4 557 575
5 461 460
6 149 141
7 186 185
8 195 198
9 142 143
10 35 28
11 61 58

0,2,L
0 221 61
1 447 393
2 1223 1304
3 591 613
4 347 363
5 461 447
6 180 172
7 152 140
8 117 111
9 208 199
10 61 59
11 39 37

0,3,L
1 201 328
2 538 551
3 314 325
4 390 404
5 187 195
6 67 52
7 333 323
8 253 259
9 120 120
10 157 145
11 58 46

0,4,L
0 1635 1292
1 753 850
2 373 318
3 618 589
4 117 116
5 451 251
6 414 401
7 148 149
8 90 77
9 213 194
10 56 51
11 119 116

0,5,L
1 244 180
2 840 761
3 204 179
4 121 141
5 196 183
6 63 69
7 164 150
8 80 88
9 58 51
10 125 113
11 11* 5

0,6,L
0 426 295
1 508 431
2 630 600
3 212 243
4 167 170
5 288 297
6 231 227
7 122 124
8 267 258
9 132 128
10 12* 5
11 31 21

0,7,L
1 531 438
2 322 306
3 262 214
4 399 386
5 225 212
6 66 64
7 124 126
8 87 77
9 57 59
10 61 55

0,8,L
0 373 383
1 396 341
2 297 297
3 494 490
4 200 190
5 215 235
6 235 228
7 175 154
8 67 60
9 114 137
10 16 19

0,9,L
1 506 452
2 337 321
3 95 58
4 282 254
5 257 237
6 13 22
7'269 259
8 85 70
9 40 43
10 94 88

0,10,L
0 497 414
1 328 309
2 149 142
3 358 351
4 143 140
5 144 148
6 364 352
7 85 80
8 129 132
9 103 96

0,11,L
1 324 299
2 286 270
3 179 179
4 170 157
5 219 212
6 87 89
7 66 50
8 135 124
9 28 27

0,12,L
0 466 382
1 32 26
2 221 238
3 491 500
4 122 115
5 121 127
6 186 176
7 103 100
8 53 51

0,13,L
1 264 254
2 245 241
3 114 107
4 164 164
5 112 114
6 23 24
7 125 117
8 111 116

0,14,L
0 89 101
1 205 209
2 95 74
3 154 173
4 29 27
5 138 136
6 48 71
7 44 28

0,15,L
1 199 195
2 120 130
3 61 60
4 80 74
5 136 149
6 11 9

0,16,L
0 227 227
1 79 73
2 96 119
3 173 167
4 68 76
5 77 83

0,17,L
1 104 111
2 68 112
3 39 27
4 93 114

0,18,L
0 94 99
1 114 107

1,0,L
-11 11* 1
-10 42 30
-9 113 119
-8 225 242
-7 19 27
-6 150 169
-5 240 255
-4 247 255
-3 68 12
-2 179 125
-1 181 235
0 352 404
1 64 ' -3
2 54 39
3 518 498
4 93 91
5 292 283
6 266 273
7 59 69
8 227 231
9 224 223
10 93 96
11 80 79

1,1,L
-11 79 76
-10 43 44
-9 27 27
-8 157 173
-7 65 57
-6 54 58
-5 385 408
-4 293 289
-3 120 115
-2 391 461
1 103 223
2 550 541
3 195 193
4 776 742
5 152 157
6 315 317
7 204 275
8 43 55
9 184 108
10 74 89
11 33 31

1,2,L
-11 77 81
-10 151 158
-9 186 195
-8 61 63
-7 53 55
-6 422 447
-5 161 168
-4 275 254
-3 414 453
-2 334 388
-1 156 111
1 380 393
2 60 85
3 501 467
4 253 203
5 267 251
6 299 311
7 180 183
8 167 166
9 18 26
10 49 50
11 114 185

1,3,L
-11 68 81
-10 73 51
-9 93 93
-8 69 69
-7 172 185
-6 51 48
-5 192 168
-4 482 392
-3 184 154
-2 503 565
-1 299 336
0 268 270
1 884 918
2 516 505
3 239 242
4 394 397
5 452 447
6 292 293
7 214 210
8 216 224
9 149 134
10 122 119
11 67 59

1,4,L
-11 58 55
-10 125 127
-9 26 134
-8 221 232
-7 233 248
-6 173 191
-5 458 477
-4 347 346
-3 600 550
-2 318 341
-1 385 400
0 241 324
1 225 197
2 574 551
3 388 333
4 151 158
5 126 115
6 373 397
7 79 95
8 192 197
9 251 257
10 69 71
11 55 49

1,5,L
-11 71 68
-10 53 59
-9 53 45
-8 231 231
-7 129 138
-6 63 54
-5 358 345
-4 243 233
-3 348 357
-2 585 550
-1 123 146
0 192 193
1 724 722
2 455 416
3 325 285
4 792 741
5 213 223
6 43 38
7 349 350
8 134 133
9 72 56
10 131 128
11 30 40

1,6,L
-11 47 52
-10 31 25
-9 117 118
-8 13 21
-7 127 125
-6 329 353
-5 56 53
-4 295 275
-3 83 107
-2 172 173
-1 337 353
0 324 328
1 925 874
2 335 315
3 451 394
4 468 446
5 370 356
6 135 124
7 170 172
8 74 70
9 155 151
10 40 46

1,7,L
-10 53 55
-9 54 53
-8 63 65
-7 97 99
-6 164 187
-5 253 244
-4 321 296
-3 332 391
-2 644 558
-1 201 229
0 133 117
1 771 730
2 482 437
3 159 121
4 373 357
5 270 256
6 142 144
7 350 353
8 114 111
9 21 13
10 101 97

1,8,L
-10 55 58
-9 94 107
-8 201 224
-7 196 215
-6 134 143
-5 132 130
-4 223 199
-3 451 448
-2 403 408
-1 227 219
0 477 469
1 316 262
2 299 287
3 505 400
4 198 183
5 188 185
6 299 304
7 159 158
8 112 116
9 94 92
10 25 26

1,9,L
-10 32 28
-9 43 43
-8 54 58
-7 240 268
-6 106 117
-5 238 248
-4 246 234
-3 68 68
-2 329 363
-1 302 288
0 92 87
1 334 343
2 132 151
3 331 304
4 200 207
5 259 255
6 362 368
7 31 37
8 139 136
9 64 66

1,10,L
-9 92 86
-8 97 100
-7 76 77
-6 180 197
-5 45 37
-4 100 70
-3 231 218
-2 148 155
-1 332 330
0 273 273
1 226 227
2 268 258
3 177 150
4 130 137
5 232 237
6 117 115
7 101 95
8 83 72
9 87 94

1,11,L
-9 56 52
-8 84 92
-7 85 94
-6 38 51
-5 127 127
-4 262 249
-3 42 52
-2 258 274
-1 167 170
0 190 193
1 411 383
2 129 124
3 119 97
4 220 214
5 51 69
6 22 30
7 180 196
8 37 33
9 58 52

1,12,L
-8 64 71
-7 84 97
-6 53 46
-5 91 86
-4 225 201
-3 311 302
-2 86 70
-1 79 81
0 206 215
1 154 154
2 173 155
3 343 328
4 21 12
5 72 66
6 231 244
7 92 03
8 97 93

1,13,L
-8 55 47
-7 96 101
-6 22 26
-5 111 117
-4 66 64
-3 133 132
-2 55 73
-1 332 322
0 169 166
1 141 141
2 350 331
3 117 120
4 106 99
5 141 147
6 133 129
7 31 62

1,14,L
-7 40 40
-6 130 144
-5 38 33
-4 141 131
-3 243 230
-2 117 121
-1 180 181
0 231 232
1 130 120
2 71 83
3 254 239
4 91 109
5 111 123
6 173 168
7 35 32

1,15,L
-6 25 42
-5 114 118
-4 118 128
-3 57 50
-2 180 187
-1 196 208
0 74 48
1 246 241
2 137 121
3 29 76
4 229 237
5 24 16
6 49 38

1,16,L
-5 77 91
-4 86 86
-3 107 113
-2 49 46
-1 199 210
0 156 162
1 73 88
2 120 131
3 98 107
4 96 95
5 95 96

1,17,L
-4 109 110
-3 23 32
-2 94 96
-1 131 161
0 16 18
1 154 155
2 122 130
3 104 109

1,18,L
-1 102 117
0 102 113

2,0,L
-11 11* 13
-10 35 29
-9 31 49
-8 51 59
-7 203 200
-6 221 220
-5 27 17
-4 248 253
-3 100 66
-2 70 93
0 455 463
1 95 91
2 451 454
3 800 830
4 212 190
5 639 578
6 240 240
7 166 164
8 207 211
9 55 50
10 55 57
11 120 125

2,1,L
-11 28 30
-10 119 132
-9 12 6
-8 140 133
-7 146 138
-6 123 121
-5 355 350
-4 392 414
-3 286 306
0 694 753
1 608 630
2 503 502
3 511 520
4 576 562
5 208 212
6 150 138
7 194 201
8 169 200
9 65 66
10 103 107
11 33 43

2,2,L
-11 11* 5
-10 95 108
-9 118 115
-8 23 24
-7 85 86
-6 296 279
-5 66 67
-4 186 197
-3 328 375
-2 387 401
-1 526 593
0 423 497
1 215 188
2 624 646
3 422 446
4 338 287
5 435 408
6 243 239
7 121 118
8 95 110
9 107 113
10 26 17
11 89 94

2,3,L
-11 95 86
-10 62 64
-9 76 78
-8 200 203
-7 143 129
-6 109 91
-5 262 263
-3 155 168
-2 192 226
-4 457 474
-1 588 630
0 337 361
1 437 481
2 148 151
3 206 196
4 380 369
5 190 177
6 182 120
7 338 357
8 84 73
9 268 163
10 150 152

2,4,L
-11 29 34
-10 06 80
-9 118 128
-8 87 94
-7 67 60
-6 119 122
-5 52 39
-4 318 340
-3 80 65
-2 90 91
-1 412 441
0 625 578
1 348 359
2 322 300
3 216 211
4 89 102
5 425 393
6 265 270
7 196 204
8 284 294
9 89 96
10 24 17

2,5,L
-11 32 28
-10 75 88
-9 72 79
-8 193 194
-7 123 133
-6 143 129
-5 259 266
-4 197 210
-3 183 166
-1 75 58
0 97 80
1 445 478
2 389 300
3 122 118
4 328 304
5 54 47

2,5,L
6 149 136
7 185 202
8 212 213
9 160 163
10 73 70

2,6,L
-11 10* 10
-10 53 52
-9 79 80
-8 50 53
-7 78 70
-6 119 118
-5 125 118
-4 272 275
-3 310 331
-2 119 131
-1 283 293
0 495 537
1 163 164
2 274 291
3 329 315
4 182 184
5 205 208
6 205 215
7 197 202
8 26 23
9 102 109
10 20 28

2,7,L
-10 61 63
-9 17 17
-8 94 94
-7 191 186
-6 122 121
-5 216 209
-4 206 227
-3 100 96
-2 248 268
-1 230 211
0 141 142
1 290 295
2 147 115
3 176 180
4 459 450
5 121 116
6 95 90
7 267 290
8 56 49
9 75 74
10 121 123

2,8,L
-10 28 24
-9 31 36
-8 54 57
-7 33 25
-6 164 180
-5 51 52
-4 117 140
-3 324 348
-2 27 30
-1 257 257
0 312 346
1 118 102
2 455 444
3 301 271
4 73 66
5 269 260
6 218 225
7 128 112
8 112 110
9 80 78

2,9,L
-10 36 33
-9 13 14
-8 234 240
-7 31 43
-6 113 112
-5 240 219
-4 74 81
-3 157 159
-2 381 424
-1 217 226
0 57 64
1 394 398
2 201 192
3 46 58
4 345 331
5 123 111
6 268 273
7 166 177
8 101 122
9 90 92

2,10,L
-9 77 76
-8 74 77
-7 156 164
-6 124 127
-5 129 140
-4 147 160
-3 303 313
-2 106 95
-1 289 311
0 359 404
1 25 32
2 213 203
3 215 200
4 70 79
5 182 186
6 80 73
7 123 123
8 123 133
9 79 83

2,11,L
-9 46 31
-8 45 57
-7 199 211
-6 113 109
-5 214 209
-4 212 228
-3 254 261
-2 164 173
-1 252 251
0 180 173
1 240 242
2 145 145
3 160 155
4 258 245
5 226 227
6 97 108
7 146 145
8 61 55

2,12,L
-8 21 31
-7 64 51
-6 30 37
-5 28 30
-4 129 140
-3 267 287
-2 114 126
-1 265 287
0 149 154
1 84 37
2 254 253
3 237 235
4 59 51
5 200 198
6 138 158
7 31 29
8 123 127

2,13,L
-8 72 74
-7 16 22
-6 25 26
-5 299 282
-4 78 73
-3 91 107
-2 172 179
-1 57 54
0 85 90
1 382 373
2 145 137
3 124 121
4 158 152
5 166 157
6 93 106
7 112 122

2,14,L
-7 24 25
-6 68 58
-5 71 76
-4 58 56
-3 138 149
-2 117 127
-1 50 50
0 169 178
1 198 191
2 115 130
3 124 114
4 58 48
5 69 57
6 104 101

2,15,L
-6 49 49
-5 65 67
-4 91 132
-3 82 100
-2 122 117
-1 123 146
0 139 137
1 63 36
2 121 123
3 47 47
4 118 123
5 52 50

2,16,L
-5 39 45
-4 46 58
-3 62 71
-2 67 72
-1 138 153
0 113 112
1 12* 11
2 139 145
3 175 189
4 36 28

2,17,L
-3 32 59
-2 120 133
-1 47 38
0 38 39
1 158 162
2 39 35

3,0,L
-11 11* 13
-10 90 84
-9 79 53
-8 36 36
-7 48 33
-6 317 317
-5 10* 8
-4 88 93
-3 357 383
-2 228 195
-1 228 235
0 441 462
1 250 261
2 341 378
3 198 205
4 50 51
5 158 155
6 146 150
7 141 157
8 137 148
9 58 70
10 24 10

3,1,L
-11 72 59
-10 62 59
-9 14 25
-8 84 69
-7 22 13
-6 76 80
-5 76 58
-4 54 42
-3 97 119
-2 446 485
-1 111 100
0 9 20
1 382 403
2 316 317
3 292 309
4 338 341
5 215 194
6 251 242
7 218 225
8 110 122
9 131 133
10 117 114

3,2,L
-11 36 37
-10 90 93
-9 40 32
-8 23 24
-7 203 175
-6 172 155
-5 70 61
-4 286 285
-3 294 295
-2 96 93
-1 587 552
0 93 79
1 290 294
2 365 406
3 140 122
4 129 120
5 337 321
6 147 163
7 36 31
8 50 52
9 40 45
10 26 22

3,3,L
-11 33 22
-10 24 24
-9 34 34
-8 89 99
-7 150 140
-6 158 142
-5 204 197
-4 103 109
-3 54 48
-2 132 143
-1 358 382
0 326 335
1 406 454
2 154 144
3 137 113
4 436 439
5 166 179
6 135 195
7 185 168
8 51 76
9 177 187
10 115 134

3,4,L
-11 49 45
-10 64 58
-9 147 142
-8 45 38
-7 90 72
-6 223 216
-5 56 71
-4 465 430
-3 222 237
-2 210 225
-1 139 147
0 267 322
1 236 234
2 201 221
3 335 343
4 156 159
5 241 245
6 198 203
7 172 181
8 247 250
9 100 117
10 44 38

3,5,L
-11 52 43
-10 31 28
-9 65 90
-8 95 34
-7 79 92
-6 42 36
-5 118 128
-4 54 70
-3 184 179
-2 231 239
-1 356 393
0 238 261
1 399 416
2 129 136
3 323 337
4 414 424
5 128 128
6 241 233
7 258 278
8 36 46
9 59 53
10 87 91

3,5,L
-10 57 56
-9 58 59
-8 30 24
-7 155 136
-6 135 131
-5 125 132
-4 96 99
-3 225 229
-2 313 318
-1 293 327
0 375 404
1 41 38
2 453 472

3,7,L
-5 138 144
-4 153 155
-3 115 109

3,6,L
3 331 336
4 74 72
5 257 246
6 323 330
7 28 32
8 201 235
9 85 94

3,7,L
-10 35 37
-9 56 53
-8 13* 11
-7 87 83
-5 73 71
-2 148 152
-1 335 358
0 336 314
1 344 359
2 359 343
3 130 131
4 340 334
5 252 248
6 258 257
7 193 191
8 37 39
9 35 107

3,8,L
-10 17 22
-9 114 123
-8 46 45
-7 164 141
-6 61 62
-5 18 26
-4 314 318
-3 245 257
-2 135 145
-1 230 256
0 139 135
1 164 153
2 112 129
3 276 231
4 208 135
5 342 328
6 36 44
7 86 97
8 120 124
9 56 59

3,9,L
-9 28 22
-8 112 68
-7 79 75
-6 53 55
-5 100 65
-4 167 174
-3 92 69
-2 307 323
-1 175 189
0 91 114
1 324 326
2 26 26
3 167 147
4 209 205
5 143 135
6 54 72
7 108 121
8 13 17

3,10,L
-9 83 84
-8 84 64
-7 110 117
-6 147 123
-5 26 27
-4 97 110
-3 107 115
-2 120 126
-1 167 170
0 152 159
1 97 98
2 201 187
3 188 191
4 68 69
5 189 193
6 102 125
7 16 14
8 109 130

3,11,L
-9 37 39
-8 27 26
-7 35 44
-6 74 77
-5 152 151
-4 177 177
-3 189 100
-2 61 59
-1 178 208
0 179 193
1 162 159
2 169 174
3 213 216
4 166 171
5 72 81
6 164 173
7 129 136

3,12,L
-8 18 23
-7 55 55
-6 105 91
-5 97 100
-4 116 129
-3 84 98
-2 64 51
-1 175 182
0 205 199
1 120 124
2 116 114
3 146 141
4 148 143
5 209 228
6 86 91
7 70 86

3,13,L
-7 53 45
-6 47 49
-5 60 47
-4 84 91

Table 5. Bond lengths (Å) and angles (°) around C5 in the Erythrina Alkaloids

	Compound			
Angle	ERTE	ERTA	HERTO	COCC
C4-C5-C6	112.5°	113	111	
C4-C5-N9	110.9	109	110	
C4-C5-C13	114.9	115	113	
C6-C5-N9	100.1	100	98	
C6-C5-C13	111.2	110	113	
N9-C5-C13	105.9	108	110	
Bond	ERTE	ERTA	HERTO	COCC
C5-C4	1.48	1.56	1.59	1.61
C5-C6	1.49	1.49	1.53	1.46
C5-N9	1.53	1.52	1.48	1.55
C5-C13	1.61	1.57	1.50	1.47

Table 5a Carbon to nitrogen bond lengths (Å) in the Erythrina Alkaloids

	Compound			
Bond	ERTE	ERTA	HERTO	COCC
C5-N9	1.53	1.52	1.48	1.55
C8-N9	1.51	1.48	1.50	1.53
C10-N9	1.52	1.47	1.52	1.51

Table 6. Least-squares planes in the structure

Planes are expressed in the form $\underline{P}x + \underline{Q}y + \underline{R}z - \underline{S} = 0$; x, y, z being the fractional coordinates of Table 1. Deviations of the atoms from the planes are tabulated in $\text{Å} \times 10^3$. Atoms marked with an asterisk were not included in the calculation of the planes. The value shown with these atoms is their deviation from the planes through the other atoms.

1) Ring A and the first olefin system

$\underline{P}$ = -0.351	$\underline{Q}$ = 13.660	$\underline{R}$ = -5.495	$\underline{S}$ = 2.394
C1	-2	C5*	-255
C2	2	C7*	322
C3	-1	N9*	444
C6	1	C13*	-1816
C4*	364	O18*	816
C19*	399		

---

2) Ring B and the second olefin system

$\underline{P}$ = 1.686	$\underline{Q}$ = 13.627	$\underline{R}$ = -5.446	$\underline{S}$ = 2.737
C5	-1	C4*	935
C6	2	N9*	485
C7	-2	C10*	183
C8	1	C13*	-1450
C1*	132		

---

3) Ring C

$\underline{P}$ = 7.846	$\underline{Q}$ = 1.126	$\underline{R}$ = -3.187	$\underline{S}$ = 0.642
C5	13	N9*	-360
C11	-13	C10*	317
C12	30	C14*	-35
C13	-30	C17*	125
C4*	1305	O20*	-1322
C6*	-1427		

Table 6 (contd.)

4) Ring D

P = 7.756	Q = 2.398	R = -3.307	S = 0.903
C12	1	C5*	187
C13	16	C11	-17
C14	-25	O24*	29
C15	16	C25*	-21
C16	3	O22*	37
C17	-11	C23*	-38

5) The Complete 2BDNP Moiety

P = 7.376	Q = -3.147	R = -4.729	S = -2.180
C26	-22	Br33	49
C27	-31	N34	29
C28	-31	O35	-117
C29	-17	O36	201
C30	14	N37	-7
C31	-22	O38	232
O32	-35	O39	-243

6) The Benzene Ring in the 2BDNP Moiety

P = 7.348	Q = -3.334	R = -4.754	S = -2.337
C26	3	Br33*	108
C27	5	N34*	32
C28	-3	O35*	-107
C29	-6	O36*	187
C30	15	N37*	-8
C31	-14	O38*	237
O32*	1	O39*	-259

Table 7. The torsion angles (degrees)

1) Ring A

C6 - C1 - C2 - C3	-0.5
C1 - C2 - C3 - C4	15.0
C2 - C3 - C4 - C5	-39.6
C3 - C4 - C5 - C6	50.2
C4 - C5 - C6 - C1	-37.7
C5 - C6 - C1 - C2	11.5

2) Ring B

C5 - C6 - C7 - C8	-0.4
C6 - C7 - C8 - N9	-18.9
C7 - C8 - N9 - C5	30.4
C8 - N9 - C5 - C6	-30.0
N9 - C5 - C6 - C7	19.1

3) Ring C

C5 - N9 - C10 - C11	61.4
N9 - C10 - C11 - C12	-42.8
C10 - C11 - C12 - C13	17.9
C11 - C12 - C13 - C5	-7.3
C12 - C13 - C5 - N9	19.2
C13 - C5 - N9 - C10	-44.7

4) Ring D

C12 - C13 - C14 - C15	4.7
C13 - C14 - C15 - C16	-4.6
C14 - C15 - C16 - C17	1.9
C15 - C16 - C17 - C12	0.8
C16 - C17 - C12 - C13	-0.6
C17 - C12 - C13 - C14	-2.0

Table 7 (contd)

5) Torsion angles about the C-N bonds in the 2BDNP moiety

C28 - C29 - N34 - 035	9.3
C30 - C29 - N34 - 036	5.5
C30 - C31 - N37 - 038	14.7
C26 - C31 - N37 - 039	12.8

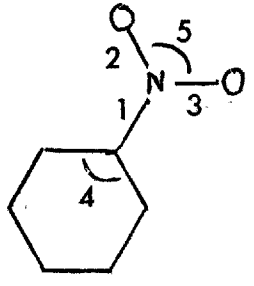
TABLE 8 The Geometry of Nitro Groups in Nitro-phenols and Nitro-phenolates

Compound	1	2	3	σ_d	4	5	σ_a	Ref.
α -p-nitrophenol	1.442	1.232	1.236	.006	122.0	122.300	.4	44
β-p-nitrophenol	1.450	1.243	1.241	.003	122.8	122.000	.3	45
p-nitrophenol adduct	1.470	1.220	1.230	.020	123.9	123.200	1.5	46
Picric acid - complex	1.460	1.150	1.160	.020	124.4	123.500	1.0	47
	1.440	1.200	1.220	.020	124.4	123.100	1.0	
	1.510	1.170	1.190	.020	125.3	122.300	1.0	
K-o-nitrophenolate complex	1.433	1.230	1.224	.004	119.8	121.700	.4	48
Potassium picrate	1.472	1.251	1.194	.010				49
	1.423	1.228	1.228	.011				
Potassium picrate	1.457	1.232	1.229	.005	122.6	124.900	.4	50
	1.436	1.228	1.228	.006	123.3	122.000	.4	
Potassium picrate	1.459	1.222	1.230	.003				51
	1.440	1.224	1.224	.003				
Ammonium picrate	1.461	1.237	1.206	.005	123.3	124.200	.3	50
	1.457	1.212	1.212	.007	124.0	121.900	.2	
Serotonin picrate	1.449	1.229	1.225	.004	123.3	120.800	.3	52
	1.438	1.226	1.239	.004	120.3	123.900	.3	
	1.459	1.231	1.224	.004	122.4	125.300	.3	
K-o-nitrophenolate.$\frac{1}{2}H_2O$	1.390	1.260	1.170	.010	122.0	122.000	2.0	53
K-o-nitrophenolate.$\frac{1}{2}H_2O$	1.424	1.251	1.239	.003	118.8	122.400	.3	54
Diazepin picrate	1.458	1.203	1.216	.008	123.1	122.400	.9	55
	1.442	1.236	1.222	.008	120.7	124.400	.9	
	1.458	1.234	1.233	.008	123.4	124.800	.9	

TABLE 8 (contd.)

Compound	1	2	3	σ_d	4	5	σ_a
2BDNP	1.430	1.240	1.210	.010	120.8	122.300	.9
	1.470	1.240	1.200	.020	124.3	122.700	1.2
Mean Values (Weighted)	1.445	1.232	1.228	.001	122.4	122.795	0.1

The columns refer to



Nomenclature for

Tables 9a and b

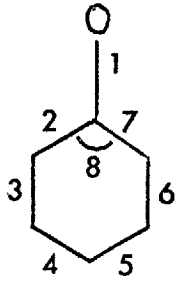


Table 9a The Geometry of Phenol Rings in Nitrophenols

Bonds and Angles										
1	2	3	4	5	6	7	σ_d	8	σ_a	Ref.
1.351	1.393	1.380	1.383	1.388	1.377	1.387	.006	120.9	.4	44
1.361	1.396	1.382	1.393	1.394	1.378	1.399	.003	120.5	.3	45
1.340	1.410	1.390	1.370	1.370	1.390	1.370	.020	120.2	1.5	46
1.330	1.410	1.390	1.370	1.410	1.350	1.350	.020	119.0	1.0	47
Weighted Mean values										
1	2	3	4	5	6	7				
1.358	1.396	1.382	1.390	1.393	1.378	1.395	.003	120.5	.2	

Table 9b The Geometry of Phenolate Rings in Nitrophenolates

Bonds and Angles										
1	2	3	4	5	6	7	σ_d	8	σ_a	Ref.
1.298	1.421	1.393	1.347	1.383	1.372	1.414	.006	114.7	.4	48
1.282	1.467	1.373	1.407	1.407	1.373	1.467	.010			49
1.243	1.452	1.372	1.382	1.382	1.372	1.452	.006	111.1	.4	50
1.245	1.455	1.370	1.389	1.389	1.370	1.455	.003			51
1.239	1.450	1.372	1.368	1.368	1.372	1.450	.005	111.5	.2	50
1.240	1.457	1.375	1.378	1.389	1.366	1.451	.004	111.1	.3	52
1.260	1.430	1.380	1.290	1.400	1.420	1.460	.010	113.0	2.0	53
1.281	1.424	1.409	1.375	1.393	1.360	1.438	.003	115.2	.3	54
1.245	1.436	1.402	1.379	1.384	1.356	1.485	.008	111.8	.9	55
Weighted Mean Values										
1	2	3	4	5	6	7		8		
1.258	1.443	1.385	1.376	1.388	1.368	1.448	.002	112.4	.1	

(The nomenclature is shown on page 179)

References

1. R. Fried, J. Chem. Ed., 1968, 45, 322.
2. H.H. Dale, J. Pharmacol., 1914, 6, 147.
3. W.H. Beers and E. Reich, Nature; 1970, 228, 917.
4. L.E. Craig, Alkaloids, 1955, 5, 265.
5. B.C. Brodie, Phil. Trans. Roy. Soc. (London), 1811, 194; 1812, 212.
6. C. Bernard, Compt. Rend., 1850, 31, 533; 1856, 43, 824.
7. A. Crum Brown and T. Fraser, Trans. Roy. Soc. Edin., 1868, 25, 151 and 693.
8. R. Boehm, Hefters Handbk. Exptl. Pharmakol., 1920, 2 part 1, 179.
9. L.E. Craig, Chem. Revs., 1948, 42, 285.
10. H. Hoffmann, Berl. Klin. Wschr., 1879, 16, 637.
11. H. King, Nature, 1946, 158, 1515.
12. M. Greshoff, Ber., 1890, 23, 3540.
13. K. Folkers and K. Unna, J. Amer. Pharm. Assoc., 1938, 27, 693.
14. V.H. Cicardo and E. Hug, Compt. Rend. Soc. Biol., 1937, 126, 154.
15. K. Folkers and R.T. Major, J. Amer. Chem. Soc., 1937, 59, 1580.
16. K. Folkers and K. Unna, J. Amer. Pharm. Assoc., 1939, 28, 1019.
17. R. Pichard and J.V. Luco, J. Pharmacol., 1944, 80, 62.
18. L. Marion, Alkaloids, 1952, 2, 499.
19. A. Mondon, Chemistry of the Alkaloids, ed. S.W. Pelletier, Van Nostrand Reinhold Co., 1970.
20. V. Boekelheide, Alkaloids, 1960, 7, 201.
21. R.K. Hill, Alkaloids, 1967, 9, 483.
22. K. Folkers and F. Koniuszy, J. Amer. Chem. Soc., 1940, 62, 436.
23. K. Folkers, F. Koniuszy and J. Shavel, J. Amer. Chem.Soc.,1942,64, 2146.
24. M. Carmick, B.C. McKusick and V. Prelog, Helv. Chim. Acta, 1951, 34, 1601.
25. W. Nowacki and G.F. Bonsma, Z. Kryst., 1958, 110, 89.
26. F. Koniuszy and K. Folkers, J. Amer. Chem. Soc., 1950, 70, 5579.

27. V. Boekelheide, J. Weinstock, M.F. Grundon, G.L. Suavage and E.J. Agnello, J. Amer. Chem. Soc., 1953, 75, 2550.

28. A.W. Hanson, Acta Cryst., 1963, 16, 399.

29. F. Koniuszy and K. Folkers, J. Amer.Chem. Soc., 1951, 73, 333.

30. K. Wada, S. Marumu and K. Munakata, Tet. Lett., 1966, 42, 5179.

31. V. Boekelheide and G.R. Wenzinger, J. Org. Chem., 1964, 29, 1307.

32. W. Weiss and H. Ziffer, Experientia, 1963, 19, 108.

33. D.H.R. Barton, R.B. Boar, D.A. Widdowson, J. Chem. Soc. (C); 1970, 1213.

34. R. Razakov and S.U. Yunusov, J.C.S. Chem. Commun., 1970, 150.

35. P. Coppens, L. Leiserowitch and D. Rabinowitch, Acta Cryst., 1965, 18, 1035.

36. NUCLS, full-matrix least-squares program, a revised and rewritten version of the Doedens-Ibers NUGLS, and basically a highly modified version of the Martin-Busing-Levy ORFLS; adapted for use on the University of London CDC 6600 computer by D. Bright and F.A. Stephens.

37. L.E. Sutton, Interatomic Distances, (suppl.), 1964, The Chemical Society, London.

38. A. Bondi, J. Phys. Chem., 1964, 68, 441.

39. W. Haugen and M. Trættenberg, Acta Chem. Scand., 1966, 20, 1726.

40. T. Hahn, Z. Kryst., 1957, 109, 438.

41. J.A. Hamilton, T.A. Hamor, J.M. Robertson and G.A. Sim, J. Chem. Soc., 1962, 5061.

42. G.I. Birnbaum, Acta Cryst., 1967, 23, 526.

43. A.T. Christensen and K.O. Strømme, Acta Cryst., 1969, B25, 657.

44. W.R. Krigbaum and G.C. Wildman, Acta Cryst., 1971, B27, 2353.

45. C.S. Choi and J.E. Abel, Acta Cryst., 1972, B28, 193.

46. P. Coppens and G.M.J. Schmidt, Acta Cryst., 1965, 18, 62.

47. P. Coppens and G.M.J. Schmidt, Acta Cryst., 1965, 18, 654.

48. E.N. Baker, D. Hall and T.N. Waters, J. Chem. Soc. (A), 1970, 400.

49. E. Carstensen-Oeser, S. Gottlicher and G. Habermehl, Chem. Ber., 1968, 101, 1648.

50. M.A. Bush and M.R. Truter, J. Chem. Soc. (A), 1971, 745.

51. M. Bettman and E.W. Hughes, Structure Reports for 1954, 1961, 18, 715.

52. K. Maarrman-Moe, Acta Cryst., 1969, B25, 1452.

53. G.J. Palenik, Acta Cryst., 1972, B28, 1633.

54. U. Thewalt and C.E. Bugg, Acta Cryst., 1972, B28, 82.

55. J.P.G. Richards, Z. Kryst., 1961, 116, 468.

56. E. Krogh Andersen, Molekylstruktur og Syrestyrke, Odense Univ. Forlag, 1971.

57. R. Gerdil, Helv. Chim. Acta, 1972, 55, 2159.

58. J.V. Barve and L.M. Pant, Acta Cryst., 1971, B27, 1158.

59. J. Trotter, Canad. J. Chem., 1961, 39, 1638.

60. P. Coppens, J. Chem. Phys., 1962, 36, 2523.

61. A.L. Mcdonald and J. Trotter, J.C.S. Perkin II, 1973, 476.

62. J. Trotter, Acta Cryst., 1960, 13, 86.

63. Handbk. of Chem. and Physics, 51st edition, Chemical Rubber Co., 1971.

64. G.C. Pimentel and A.L. McClellan, The Hydrogen Bond, Freeman and Co., 1960.

65. W. Fuller, J. Phys. Chem., 1959, 63, 1705.

66. S.C. Wallwork, Acta Cryst., 1962, 15, 758.

67. W.C. Hamilton and J.A. Ibers, Hydrogen Bonding in Solids, Benjamin, New York, 1968.

68. P-G. Jonsson, Hydrogen Bond Studies, Acta Universitatis Upsaliensis, Uppsala, 1973.

69. P-G. Jonsson and W.C. Hamilton, Acta Cryst., 1970, B26, 536.

70. Å. Kvick, P-G. Jonsson and R. Liminga, Acta Chem. Scand., 1972, 26, 1087.

71. J.M. Bijvoet, A.F. Peerdeman and A.J. van Bommel, Nature, 1951, 271.

72. P.G. Waser, Ciba Fndn. Symp. Enz. and Drug Action, p.206, Churchill (London), 1962.

73. P.G. Waser, Ciba Fndn. Symp. Mol. Properties of Drug Receptors, p. 59. Churchill (London), 1970.

74. Kier, Molecular Orbital Theory in Drug Res., p.162, Acad. Press, 1971.

75. C. Chothia, Nature, 1970, 225, 36.

76. E. Shefter and D.J. Triggle, Nature, 1970, 227, 1354.

77. C.K. Prout and J.D. Wright, Angew. Chem., 1968, 80, 688.

78. F.H. Herbstein, Perspectives in Struct. Chem., ed. J.D. Dunitz and J.A. Ibers, 1971, 4, 166.

79. A. Bondi, Phys. Props. of Mol. Crystals and Glasses, Wiley and Sons, New York.

80. A.J. Kitaigorodski, Organic Chemical Crystallography, Consultants Bureau, 1957.

CHAPTER 4

The crystal structure of Ferri-mycobactin P.

INTRODUCTION

1. The Mycobacteria[1]

The bacterial order Actinomycetales, which contains at present 10 genera including the Actinomycetaceae, the Mycobacteriaceae, the Nocardiaceae and the Streptomycetaceae, is one of the most highly developed orders. It includes a number of species which are of great importance, both as pathogens and as sources of powerful antibiotics. (A compilation of the microbial sources of antibiotics discovered in the U.S.A. and Japan between 1953 and 1970 revealed that 85% were produced by actinomycetes, 11% by fungi and 4.5% by other bacteria[2].) The order has been a source of taxonomic controversy until recently, bacteriologists claiming them as bacteria and mycologists as fungi, but the problem is now held to be resolved.

The genera in the actinomycetes all produce hyphae, although these may be difficult to observe, and generally reproduce asexually, yielding free spores. With the exception of the actinomycetaceae, the actinomycetes are aerobic and are found in a wide variety of habitats ranging from the soil to sea water and the higher mammals.

The family, mycobacteriaceae, contains two genera, the mycobacteria and the mycoccocci, although the latter are still surrounded by confusion and are held by some workers to be morphological variants of the former. The mycobacteria are acid-fast, Gram positive, non-motile bacteria which grow in slender filaments or irregular branched rods. They have a waxy exterior, agglomerate in aqueous media and have an unusually high lipid content in the cell walls.

The genus mycobacterium falls roughly into two parts based on the rate of growth of the species, the slowly growing organisms taking some 2 to 6 weeks and requiring complex culture media. The fast growing species produce large cultures within one week at 28^{o}C and on simple media. The division of the genus is shown in Table 1 which also contains some details of the habitat, pathogenicity and mycobactin production of some of the

TABLE 1

The Genus Mycobacterium

Species	Habitat	Pathogenicity	Mycobactin Production	Growth Rate
M.tuberculosis	Man and higher mammals	TB. in man and the higher mammals	1.9	Slow
M.avium	Fowl	TB. in birds		"
M. bovis	Cattle	Bovine tuberculosis		"
M.paratuberculosis	Intestinal mucosa	Chronic enteritis in cattle		"
M. leprae	Man (skin)	Leprosy		"
M. ulcerans		Chronic skin lesions		"
M. lepraemurium	Skin and lymph nodes (rat)	Rat leprosy		"
M. marinum	Sea fish,	Marine fish TB., cutaneous lesions in mammals	3.4	"
M. terrae	Soil, sputum, radishes		2.3	"
M. kansasii	Fresh water	Skin and pulmonary lesions	0.5	"
M. piscium	Freshwater fish	Freshwater fish TB.	7.3	Rapid
M. phlei	Soils, grass and hay		10.1	"
M.smegmatis			17.3	"
M.thermoresistibile	Soils		11.8	"
M.fortuitum	Soils	Abscesses, cervical and pulmonary diseases	19.5	"
M.aurum	Soils		0.7	"
M.chelonii	Turtles	Reptilian TB.		"
M.thamnopheos[1]	Snakes	Reptilian TB.	0.9	"
M.balnei[2]			5.4	Slow

footnotes 1 This organism has recently been transferred to the genus Nocardia.

2 Generally considered to be the same organism as M.marinum.

species. The genus contains several species which are highly pathogenic to mammals, notably the causative organisms of tuberculosis, the serious cattle enteritis known as Jöhnes disease, and leprosy. In general, mycobacterial infections produce granulomatous lesions which contain vast numbers of the bacteria. The swellings which are characteristics of leprosy are typical, containing cells which are infected to bursting point with M.leprae.

2. The Discovery and Isolation of the Mycobactins

At the beginning of the twentieth century all but two of the then known mycobacteria had been cultivated in the laboratory on normal media. The two exceptions were M. paratuberculosis and M. leprae and it was in a search for suitable media for these organisms that Twort and Ingram[3] made the important suggestion that their failure to grow on normal media "must be due to the absence of some necessary foodstuff" and that this substance was most likely to be produced by the other members of the genus. They were able to confirm the suggestion, by demonstrating the growth of M. paratuberculosis on media containing dead tuberculosis bacilli or glycerol and ethanol extracts from living cultures of this organism. They concluded[4] that the mycobacteria needed an essential substance which most of the species could produce for themselves and that this substance could activate the growth of M. paratuberculosis. M. leprae has so far resisted attempts to cultivate it in vitro, a situation which has seriously hindered the progress of investigations into an effective treatment for leprosy.

No particular significance was attached to this work until after the rise of the science of antibiotics under the impetus of the second world war. In 1945, J. Francis, working in the pharmaceutical division of Imperial Chemical Industries, pointed out that a modified form of Twort and Ingram's "essential substance" might serve as a highly specific antibiotic for the treatment of tuberculosis. Since mammals neither make nor are

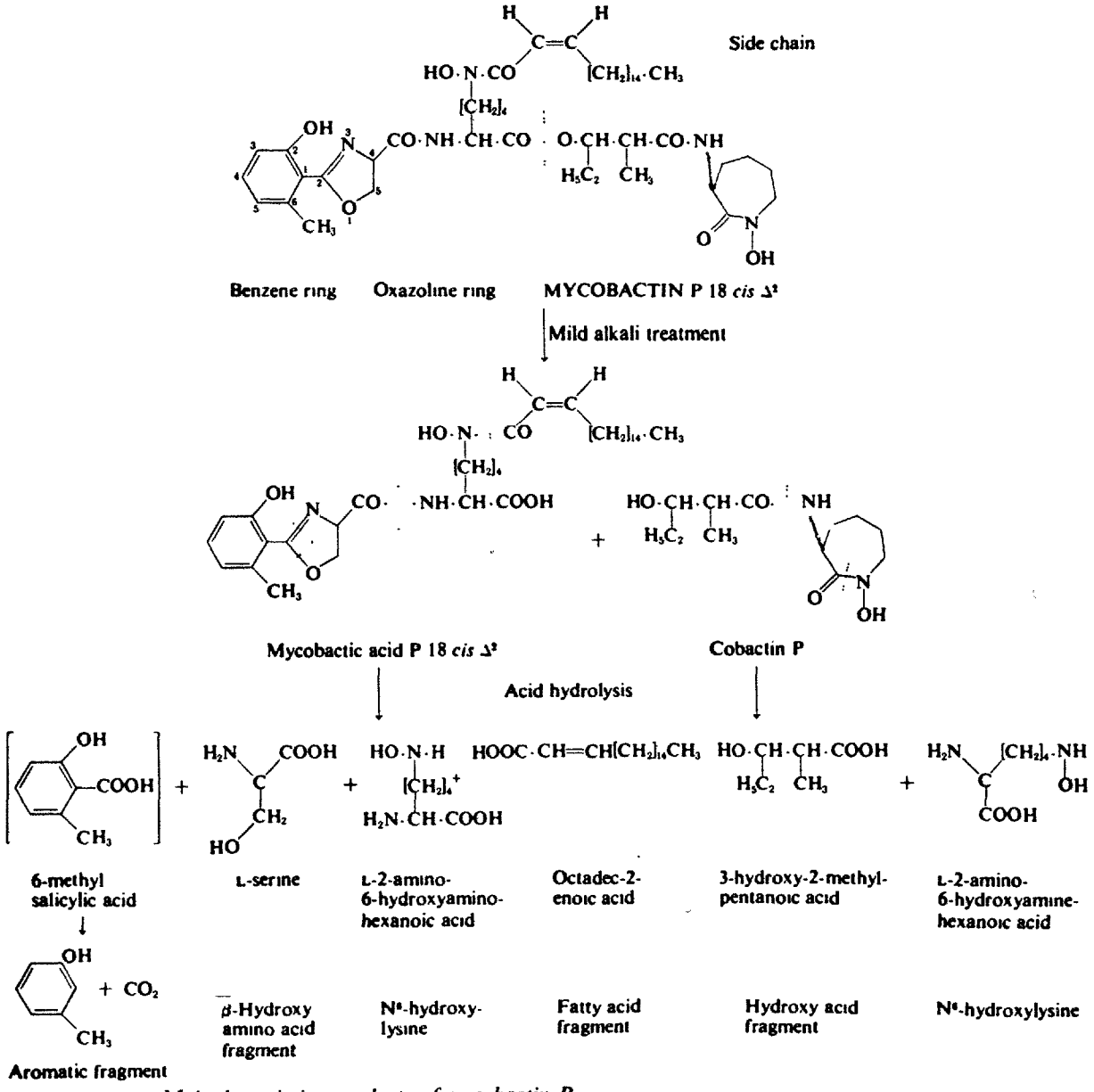
dependent on this substance such an antibiotic could have no foreseeable toxicity. As a result of this proposal, a course of research which has been excellently summarised by Dr. G.A. Snow[5] of I.C.I. was embarked upon. It was hoped to isolate and characterise the substance and to synthesise an antibiotic from it.

Twort and Ingram had reported that the growth promoting substance could be extracted from cultures of M. tuberculosis by organic solvents and work was therefore concentrated on isolation of the active component from solvent extracts using the non-pathogenic and easily cultured soil bacterium M. phlei. The work was greatly facilitated by the accidental formation of a crystalline aluminium complex of the growth factor in an attempt to isolate it by chromatography in an alumina column.[6] An X-ray examination of the crystalline powder gave the following crystal parameters:

Monoclinic, $a = 15.81$, $b = 12.31$, $c = 13.54$Å, $\beta = 101.5^{\circ}$, $D_m = 1.175$ $g.cm^{-3}$.

This gives a calculated molecular weight of 1828 ± 73 per cell. However, titration with perchloric acid (see below) indicated a molecular weight of half the crystallographic value, implying 2 molecules in the unit cell and a molecular weight of 914 ± 36.

In the early stages of the work[7], acetone extracts from moist M. phlei cells were found to promote the growth of M. paratuberculosis. Addition of light petroleum ether to this extract gave a precipitate with greatly enhanced biological activity and the crystalline aluminium complex was obtained from this fraction. The product was always reddish-brown in colour due to contamination with ferric iron (from the alumina) so a more complex isolation procedure was devised which made use of the insolubility of the cupric complex in ethanol and its solubility in chloroform. The metal was finally removed by treatment with hydrogen sulphide to yield the purified growth factor. (This method has been superceded by a procedure using the ferric complex.)



Main degradation products of mycobactin P.

Figure 1

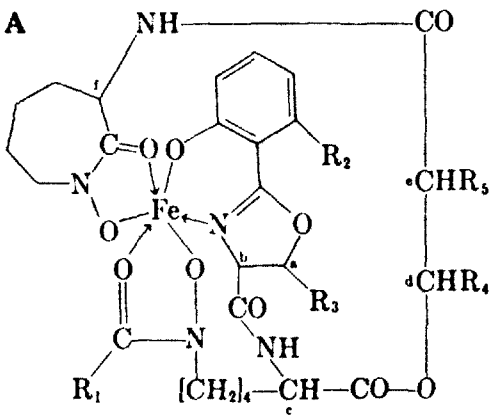
The product, a white, micro-crystalline, optically active powder, M.Pt. 165-166°C, α_D^{25} (c = 4.9) in $CHC\ell_3$ = -19°, was named "mycobactin". It was shown by electromeric titration to have one weakly basic and two weakly acid functions. The former permitted the determination of the equivalent weight by titration with perchloric acid as 870±4. A chloride and picrate were also reported and elemental analysis gave the formula $C_{47}H_{75}O_{10}N_5$. The solid exhibits a pale green fluorescence and, in methanol solutions, a blue fluorescence. It is stable to heat up to 150°C and forms complexes with a number of metals (see below).

The structure of the compound was established by a series of degradation studies which are shown in Figure 1, combined with spectroscopic and chromatographic investigations[8]. It is of interest that neither of the two major fragments obtained by mild alkaline hydrolysis show any biological activity. The molecule is derived from six basic units which arise from several different biosynthetic pathways and includes three amino-acid-derived moieties.

The successful isolation of a mycobactin from M.phlei led to an attempt to isolate the same compound or its analogue from M. tuberculosis but this work was hampered by the different solubility properties of the new compound. The mycobactin from M. phlei had been found to occur in the metal-free form but in M. tuberculosis it is present as the ferric complex and a new extraction procedure using an initial ethanol extraction was developed. When isolated[9], the new compound was found to differ in a number of ways from the M. phlei form and an extended nomenclature was introduced containing a postcript which denotes the bacterial origin of the compound, the mycobactin from M. phlei being termed mycobactin P and that from M. tuberculosis, mycobactin T. Mycobactins have now been isolated from some 9 different species and Snow[5] has suggested that they could serve as factors in the classification of the mycobacteria.

Figure 2 shows the structures of the mycobactin compounds which have

A

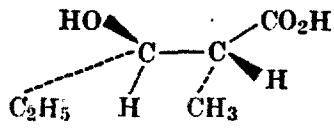


B

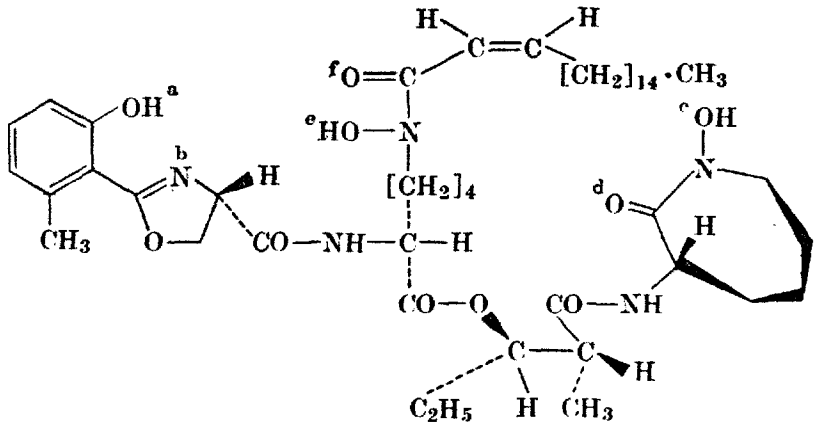
Mycobactin	R_1	R_2	R_4	R_5	R_6	a	b	c	d	e	f
A^+	**13**Δ	CH_3	H	CH_3	H						
F*	**17**, 15, 13, **11**, 9Δ	H	CH_3	CH_3	H	Threo			*S*	( )	L
H	**19**, **17**Δ	CH_3	CH_3	CH_3	H	*R*	L	L	*S*	( )	L
M	1	H	CH_3	18, **17**, 16, 15‡	CH_3				*R*§	*S*§	
N^+	2	H	CH_3	18, **17**, 16, 15‡	CH_3						
P	19, **17**, 15 *cis*Δ[1]*n*	CH_3	H	C_2H_5	CH_3	( )	L	L	*S*	*R*	L
R	**19**Δ	H	H	C_2H_5	CH_3	( )	L	L	*R*	*S*	L
S	19, **17**, **15**, 13*cis*Δ	H	H	CH_3	H	( )	L	L	*S*	( )	L
T	20, **19**, 18, 17 }	H	H	CH_3	H	( )			*R*	( )	L
	20, **19**, 18, 17Δ }										

*(A) General structure of the ferric mycobactins. Side chains R_1 are alkyl groups having the number of C atoms shown; double bonds are indicated where they are known. (B) Figures show the main types of side chain; those of greatest abundance are shown in bold type. Asymmetric centers are labeled a–f. Blanks indicate that the configuration has not been determined. Key: ( ), lack of asymmetric center; *, structure inferred from mixture with mycobactin H; †, tentative structures; ‡, saturated alkyl groups having the number of C atoms shown; §, relative configuration at d and e is erythro—absolute configuration is uncertain.*

Figure 2



(I)



(II)

Figure 3

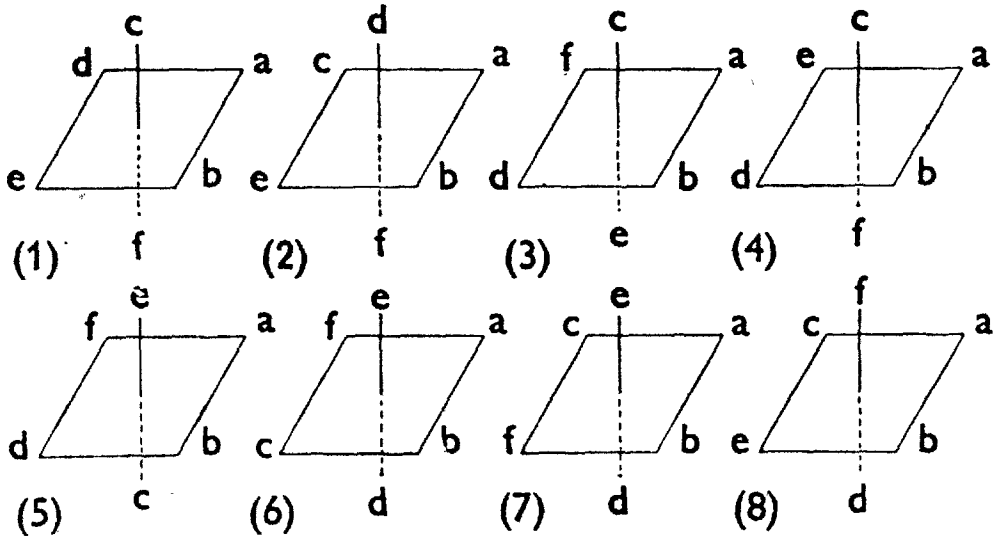


Figure 4

so far been isolated.[5] It is apparent that they have a common nucleus but differ in a number of ways. They fall into two classes differentiated by the position of the long chain fatty acid residue. Most mycobactins carry this at R_1 but in mycobactins M and N the fatty acid side chain is at R_4. These two molecules are also unusual in that they have growth inhibiting properties which are competitive with the other mycobactins. The other differences are smaller, *viz.* the presence or absence of a methyl group on (i) the salicylate derived ring, (ii) the oxazoline ring and (iii) the β-hydroxy acid residue, and differencesin the chiralities at the asymmetric centres in this moiety.

An important characteristic of the mycobactins is their ability to form complexes with metals.[10] So far complexes with Al^{3+}, Ga^{3+}, Cr^{3+}, Cu^{2+}, V^{3+} and Fe^{3+} have been reported but the last of these is by far the most stable and treatment of all but the chromium complex with Fe^{3+} ions leads to an immediate exchange of the metal to give the ferric complex. The chromic complex has been shown to act as a competitive inhibitor for the growth of M. tuberculosis.

The number of different coordination isomers which can be obtained from 6 different ligands octahedrally disposed about a central atom is 90 but for the mycobactins the ligands are not independent and consist of the three pairs shown coordinated to the iron atom in Figure 2. This restriction allows 16 possible isomers but the number is reduced further by the non-independence of the pairs themselves. Snow[11] was able to demonstrate from model building experiments that only 8 of the possibilities are stereochemically feasible. The ligand atoms themselves are marked a to f in Figure 3 and the coordination pattern in the 8 structures for which models could be built are shown in Figure 4. Snow preferred numbers 5, 6 and 7 in this figure as being most compact and least subject to strain.

Ligands c, d, e and f are components of the two hydroxamate groups present in the molecule and their involvement in the coordination of the

metal ion suggesta a connection with the large group of metabolites known as the siderochromes.[12] These are water soluble, iron complexing hydroxamate compounds which are produced by almost all micro-organisms. In most they act as growth factors but a number of iron-containing metabolites, notably albomycin, ferromycin and sideromycin, are known.

3. Iron in Biological Systems

In a recent review article on the evolution of biological iron binding centres J.B. Nielands[13] stated that "life, in any form, without iron is in all likelihood impossible." (The element is involved in the transport and storage of elemental oxygen, in almost all biological redox processes and in electron transport in the cytochrome sequences.) In the animal kingdom, the first of these roles is played by the haemoproteins myoglobin and haemoglobin and a similar task in root nodules of leguminous plants is carried out by closely related compounds. Electron transport haemoproteins (cytochromes) are found in plants, animals and micro-organisms and non-haem "iron-sulphur" proteins such as the ferredoxins and rubredoxins are similarly distributed and are also involved in electron transport. In addition to the large number of haem containing enzymes which are concerned with redox reactions, iron is found in a wide range of flavoprotein oxidases and dehydrogenases. The non-proteinaceous substances ferroverdin and pyrimine are unique in that they preferentially form ferrous complexes, in all other compounds the iron is found in the higher (ferric) oxidation state.

In order to satisfy the demand for iron a wide range of compounds has been evolved which are concerned with the acquiring, transport and storage of the metal. At physiological and environmental pH's iron normally occurs as the insoluble ferric hydroxide and the first stage in the utilisation of the element must be to render it soluble. This is achieved by chelation of the metal to form a water soluble complex which can be utilised in the cell. In mammals, iron transport is carried out by proteins such as transferrin and conalbumin but in micro-organisms the task is carried out

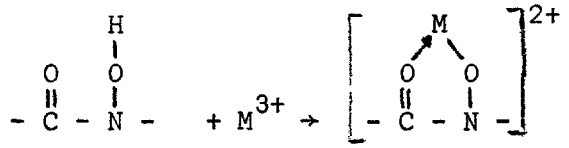
Table 2 Stability constants of cation complexes of acethydroxamic acid at 20°C and $\mu = 0.1$. (From reference 14)

Cation	$\log K_1$	$\log \beta_2$	$\log \beta_3$
H^+	9.35		
Mn^{2+}		6.9	
Fe^{2+}		8.5	
Co^{2+}		8.9	
Zn^{2+}		9.6	
La^{3+}			11.9
Al^{3+}			21.5
Fe^{3+}			28.3

by a range of low molecular weight (<1000) compounds known as the siderochromes. Iron storage is carried out in mammals by the protein ferritin and also by a granular ferric hydroxide - polysacharide-protein (hemosiderin) complex. Ferritin-like compounds have been reported in moulds.

The ability of the biological iron binding centres mentioned above to chelate metals is, with some notable exceptions such as vitamin B_{12}, highly specific for ferric iron. This is particularly the case for the siderochromes which are among the strongest known chelating agents for Fe^{3+}. The term originally covered a range of water soluble hydroxamate compounds of microbial origin[13] but has now been extended[12] to cover compounds in which the chelation makes use of other ligands.

This specificity is characteristic of hydroxamic acids, complex formation occurring by stepwise association of the hydroxamate anion around the metal ion:-

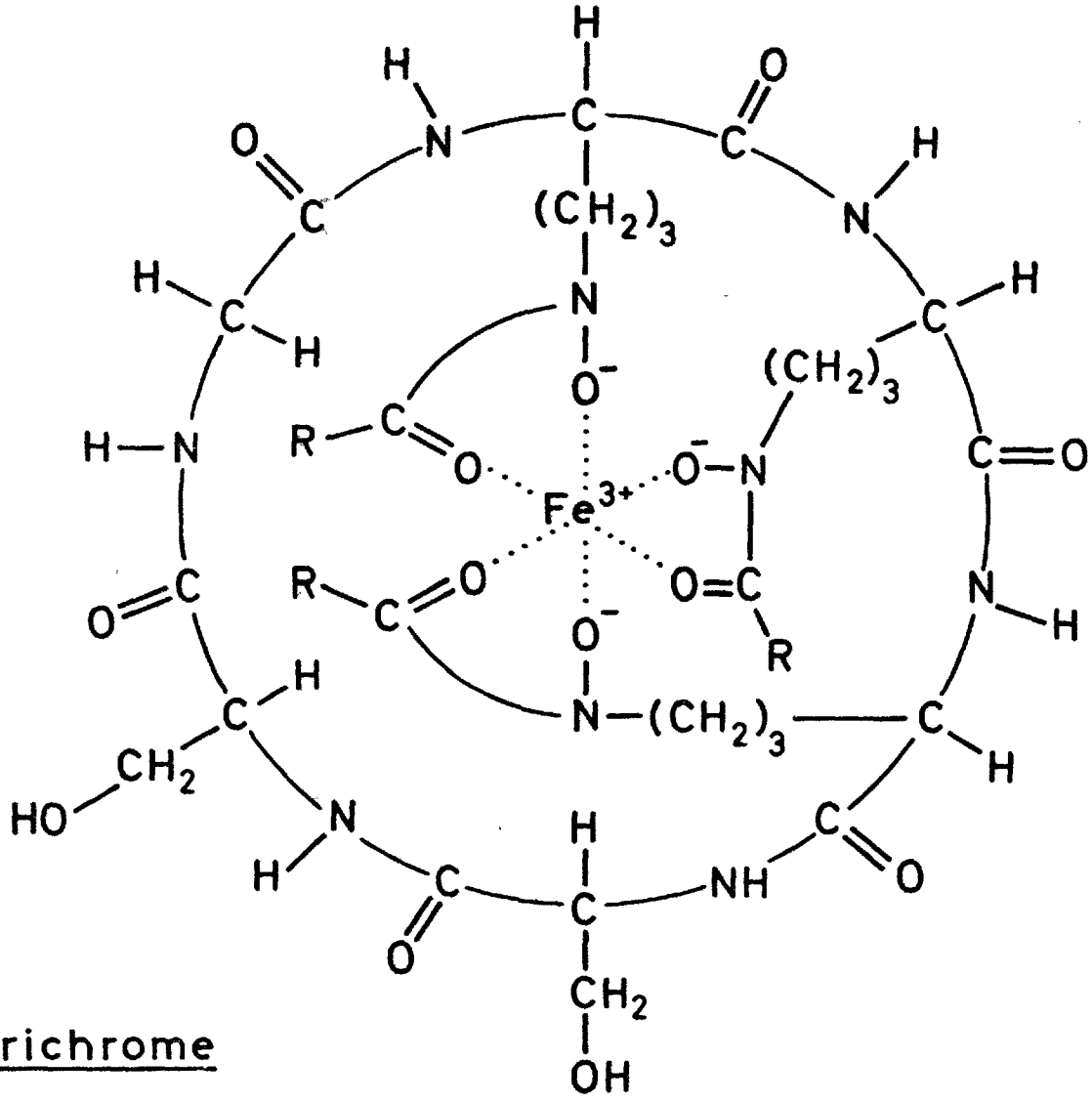


etc.

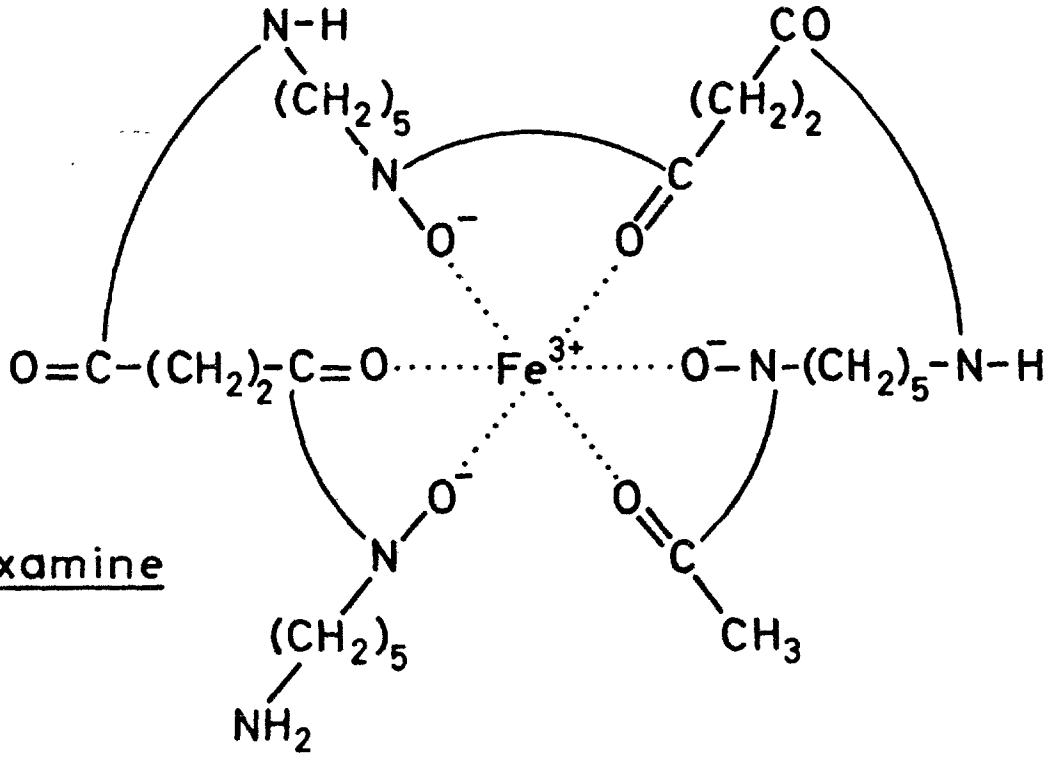
The stability constants[14] of a number of cation complexes with acethydroxamic acid ($CH_3CON(OH)H$) are given in Table 2 and illustrate the preference for the Fe^{3+} ion. It is also notable that Fe^{2+} ions form relatively weak complexes.

4. The Siderochromes and their Role in Nature

In 1960, Prelog *et al.*[15] suggested that the term "siderochrome" be applied to all the red-brown iron-containing metabolites with characteristic absorption maxima in the range 420 - 440mμ at pH 7. (λ_{max} for tris-acethydroxamato-iron(III) at this pH is in the range 425 - 440 mμ depending on concentration.) They subdivided the siderochromes into two groups based on their biological activity. The metabolites with growth promoting activity were termed sideramines and those with antibiotic activity, the



Ferrichrome



Ferrioxamine

Figure 5

sideromycins. The name "siderochrome" was thus originally reserved for hydroxamate compounds but has now been applied to all naturally occurring microbial iron-transport compounds.[12] However, since the development of understanding of the function of the mycobactins is better seen against the older definition of the siderochromes this term will be restricted to naturally occurring, iron chelating, water soluble trihydroxamate compounds.

These compounds have been isolated from a wide range of micro-organisms and the chemical structures of most of the compounds known today have been determined. The sideramines fall into two classes, the ferrichromes and the ferrioxamines and the chemical nuclei of these two series are shown in Figure 5. The sideromycin group contains members of both classes but they are chemically differentiated from the sideramines by the presence of an extremely bulky pendant group which probably prevents the entry of the molecule into the living cell to exchange its iron atom with receptor molecules. The production of siderochromes is invariably found to be stimulated by iron deficiency in the culture medium and the action of the antibiotic sideromycins is competitively reversed by the sideramines.

The biological role of the sideramines is summarised in figure 6 and may be broken down into the following stages:-

1. The synthesis of the trihydroxamate compound.
2. Diffusion of this compound out of the cell to complex with ferric ions in the culture medium.
3. Transfer of the ferri-trihydroxamate complex into the bacterial cell.
4. Presentation of the Fe^{3+} ion to a reducing enzyme.
5. Reduction of the Fe^{3+} ion to Fe^{2+} and release of this ion to enzyme systems responsible for iron metabolism.
6. Migration of the desferri-trihydroxamate out of the cell to begin the cycle again.

Since the iron complexes of the mycobactins show absorption maxima which are characteristic of hydroxamate complexes and their production is

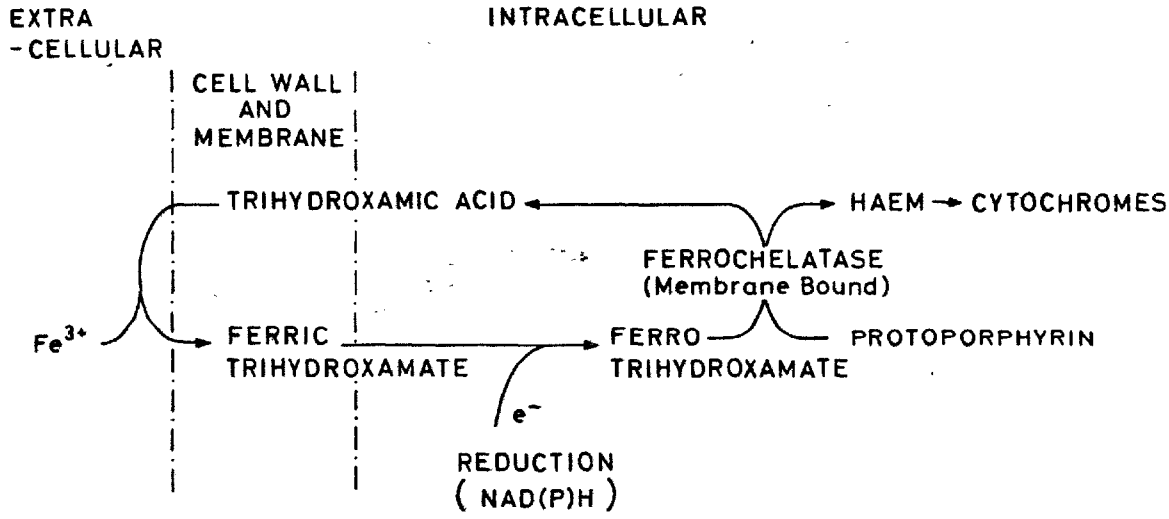


Figure 6 Iron transport in Ustilago Sphaerogena

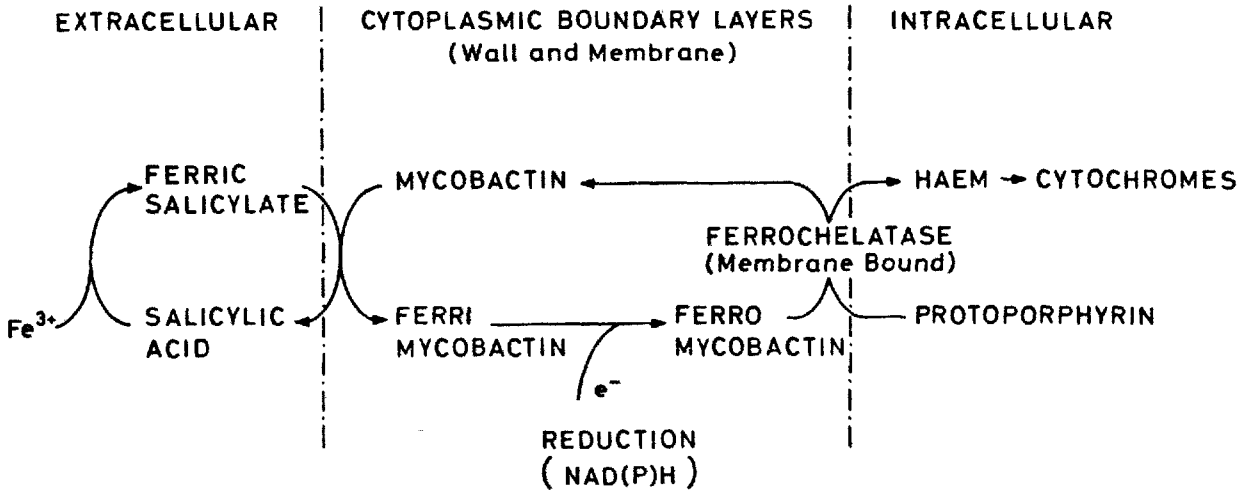


Figure 7 Iron transport in the Mycobacteria

stimulated by iron deficiency, it was natural to assume that their biological role was similar to that of the sideramines. The similarities between the two species of compounds are:-

1. Both are natural products with highly specific iron chelating properties.
2. The production of both is stimulated by iron deficiency in the culture medium.
3. Both will competitively reverse the action of iron containing antibiotics.
4. Both are believed to be involved in iron transport.

There are, however, a number of differences:-

1. In the sideramines the iron chelation is by three hydroxamate groups whereas in the mycobactins the chelation involves two hydroxamate and one phenyl-oxazoline group.
2. The mycobactins can replace the sideramines as growth promoting substances but the sideramines do not promote the growth of the mycobacteria.
3. The sideramines are water soluble whereas the mycobactins are highly insoluble in water (5 - 15 μg/ml at 20°C), both as the iron complex and in the desferri form. The mycobactins are, however, soluble in lipids.

The reason for the assumption of a common role for the mycobactins and the sideramines have been noted above, but Snow pointed out that the former compounds have a very low solubility in water and although they chelate iron readily this would impose a severe constraint on iron transport.

In a preliminary publication in 1970 Ratledge[16] proposed a new mechanism for the transport of iron in the mycobacteria based on the assumption that mycobactins are located in the lipid areas* of the bacterial cell wall, and in 1972, Ratledge and Marshall published[17] the results of an elegant series of experiments based on the new theory. They were able to establish that more than 95% of the total mycobactin in a growing culture of M. smegmatis

---

Footnote:

The genus Mycobacterium is unusual among the bacteria in the exceptionally high lipid content of the cell walls. The cell walls of M. tuberculosis contain up to 65% of their dry weight as lipids[18] and lipids make up some 34% of the total dry weight of the cells of M. smegmatis.[19]

was located in the bacterial cell walls, that virtually all of the remainder was present in the cytoplasm and that the concentration in the culture medium was less than 2×10^{-13} g/ml - lower than the solubility of ferri or desferri mycobactin.

Using $^{55}Fe^{3+}$ compounds they demonstrated that the uptake of iron into normal rapidly growing cells was very fast and that little of the iron was present as ferrimycobactin. In starved cells, however, where metabolic activity was slower and the concentration of receptors for iron should be low, the uptake was still very rapid but the iron remained as ferrimycobactin until the metabolic processes were underway.

The uptake of iron was found to be non-temperature dependent and therefore, not enzyme mediated. In addition to this, treatment of the cells with trifluoro-acetic acid - a powerful enzyme deactivating agent - was found to have no effect on iron uptake. In the light of this evidence Ratledge and Marshall concluded that iron uptake must be a straightforward chelation of Fe^{3+} ions by desferrimycobactin.

The relatively low stability of hydroxamate complexes with the Fe^{2+} ion suggested that release of the metal atom from ferrimycobactin would be most likely to occur by an enzymic reduction process to ferrimycobactin. Using homogenised cell extracts it proved possible to demonstrate that such a reduction does occur and that under anaerobic conditions the process requires addition of the reduced cofactor NADH.

Finally, Ratledge and Marshall showed by measurement of the interfacial tensions of ferri- and desferrimycobactin at an ethyl-n-decanoate/water boundary that they adsorbed equally strongly at the solvent/water surface.

Some recent work on deficient strains of M. smegmatis[15] has shown that salicylate is also required for the transport of iron in the mycobacteria. Since this is soluble in water and is found in the culture medium it seems reasonable to suggest that this molecule serves as a scavenger for Fe^{3+} ions from the environment.

The mechanism of iron transport in the mycobacteria is shown in figure 7 and may be broken down into the following steps:-

1. Secretion of salicylic acid into the growth medium where it chelates Fe^{3+} ions.

2. Diffusion of the resulting salicylate complex to the outer surface of the bacterial cell wall.

3. Exchange of the Fe^{3+} ions with des ferri-mycobactin molecules concentrated on the outer surface of the cell wall.

4. Diffusion of the ferri-mycobactin complex through the lipid areas in the cell wall to the inner surface.

5. NAD(P)H dependent, enzyme mediated reduction of ferri-mycobactin to ferro-mycobactin.

6. Loss of the ferrous ion from the relatively unstable ferro-mycobactin to ferro-chelatase, a ferrous ion chelating enzyme, which catalyses its insertion into protoporphyrins to form haem compounds.[21]

7. Diffusion of desferri-mycobactin to the outer surface of the cell wall to begin the cycle again.

This mechanism, termed a "facilitated diffusion" process by Ratledge and Marshall[17], is dependent on the removal of Fe^{2+} ions into the cytoplasm by an enzymic reduction process and this enables the cell to avoid an iron overload. It also permits the storage of Fe^{3+} ions as ferri-mycobactin when iron is not required by the cellular metabolism.

In summary, research on the mycobactins under the guidance of Dr. G.A. Snow of I.C.I. and, subsequently Dr. C. Ratledge in the University of Hull has resulted in the development of an elegant and firmly based understanding of iron transport in the mycobacteria.

Subsequent to the work outlined above Snow, and Ratledge have demonstrated the production of a lipid soluble iron transporting compound which is very similar to mycobactins M and N (see figure 2) by species in the genus Nocardia[22]. This genus is also characterised by a high

lipid content. The compound, named Nocobactin NA (from nocardia asteroides) contains a phenyl-oxazole system in place of the phenyl-oxazoline system in the mycobactins.

The Crystallography of Mycobactin P

The mycobactins are all white powders which show a microcrystalline structure under X-ray examination but it has so far proved impossible to obtain useable single crystals of the metal-free compounds. Mycobactin P forms crystalline complexes with Al^{3+}, Ga^{3+}, Cr^{3+}, and Fe^{3+} [7,10]. The gallium and chromium complexes have not yet been subjected to an X-ray study but the determination of the unit-cell dimensions of the aluminium complex by X-ray powder method was reported in 1953[7]. Ferric and aluminium mycobactin P readily give thin hexagonal plates which show extinction of polarised light parallel to the major direction of development. In a preliminary study carried out at Imperial College in 1966, F.H. Allen[23] was able to measure the unit-cell dimensions of the ferric complex by single-crystal methods. He demonstrated that it was isomorphous with the aluminium complex and crystallised in the space group $P2_1$. Unfortunately the size and quality of the crystals precluded a structure determination. Some two years later Dr. Snow succeeded in growing a batch of far larger crystals from a methanolic solution at $3^{\circ}C$. These proved to be far superior to any crystals obtained previously and were suitable for further investigation.

Ferric mycobactin P (referred to hereafter as "ferri-mycobactin") crystallises as irregular hexagonal plates with prominent (0 1 0) faces and shows extinction of polarised light parallel to the major axis of the hexagon. A photographic examination confirmed Dr. Allen's results and showed that the crystals were suitable for the collection of intensity data. Most of the crystals were, in fact, too large for this purpose and smaller specimens had to be prepared by "micro-surgery".

The unit-cell parameters were determined by least-squares refinement from the 2θ values of 46 axial, zonal and general reflections measured on a Siemens A.E.D. diffractometer with the crystal mounted about the b axis. These values, together with the other main crystallographic parameters are summarised below:-

$C_{47}H_{72}N_5O_{10}Fe$, M = 922.9, M.Pt. = 199-201°C; monoclinic, a = 16.061(2), b = 12.193(1), c = 13.282(3)Å, β = 101.39(2)°; U = 2550.6Å^3; D_m = 1.201(1), D_c = 1.201 g.cm^{-3} for Z = 1; Space group $P2_1$; F(000) = 984; μ(Cu-K_α) = 27.89 cm^{-1}.

The absent spectra (0 k 0, k = 2n) are consistent with space groups $P2_1$ and $P2_1/m$ but the optical activity of the compound eliminates the second of these possibilities. The density of the compound was originally determined by Dr. Allen to be 1.16 g.cm^{-3}, a value which gives a calculated molecular weight of 891, significantly lower than that quoted above. A later, and very careful, determination gave a value which is the same as the theoretical value for the generally accepted formula for ferrimycobactin P.

Intensity data out to θ = 50° were collected on the automatic diffractometer using a crystal mounted to rotate about the b axis on a quartz fibre. The data collection was carried out in the normal way using the 0 4 0 reflection as an intensity reference and the 5 0 0 and 0 0 8 reflections as checks against crystal movement. The data were processed in the normal way giving a total of 2786 reflections of which 2528 were classified as significantly above the background radiation level. The intensity of the reference reflection remained constant throughout the experiment.

In view of the presence of an iron atom in the structure it was hoped to solve the phase problem by the heavy-atom method and a sharpened origin-removed Patterson map was therefore computed. This gave the encouraging result that the Harker section at 0 ½ 0 was found to contain a single large

peak at u = 0.12 and w = 0.35. This peak was surrounded by approximately symmetrically distributed smaller peaks. The large peak was assumed to be the Fe - Fe Harker vector and the four smaller peaks to be Fe - ligand vectors. The positions of the five atoms thus obtained from the Patterson map were:-

Atom	x	y	z
Fe	0.06	0.00	0.175
L1	0.14	0.00	0.30
L2	-0.04	0.00	0.27
L3	-0.03	0.00	0.05
L4	0.16	0.00	0.125,

and two further Fe - ligand vector peaks were found in the Patterson map in mirrored positions vertically above and below the Fe - Fe peak.

In the space group $P2_1$, a Fourier synthesis using phases calculated from the position of one atom only or from a group of atoms all having the same y coordinate contains a mirror plane perpendicular to the y axis and at the same y coordinate as the single atom or group of coplanar atoms. It was therefore, unfortunate that four of the atoms comprising the coordination shell around the iron atom lay at the same height in the y direction as the iron atom and that the remaining two ligands were mirrored through that plane, since none of them can contribute to the breaking of the mirror symmetry.

A difference Fourier synthesis was computed based on the iron atom (R = 55%) and from this map the positions of the ligand atoms tentatively identified in the Patterson map were confirmed. Five of these atoms were then included in the calculation of a new difference Fourier synthesis. The sixth ligand, that lying vertically beneath the iron atom in the y direction, was omitted in the hope that the pseudo-symmetry would be broken sufficiently to enable the selection of the correct y coordinates for any new atoms. The inclusion of the five ligands (as oxygen atoms) resulted

in a small improvement in the R factor to 54% but unfortunately no chemically sensible part of the rest of the structure could be identified in the map. The omission of the one ligand atom did not result in any significant breaking of the pseudo-symmetry, all of the new peaks having approximately equal intensities at the + and - y positions. A three-dimensional model was built of the major peaks but again nothing was recognisable and its complexity precluded the use of a "distance" scan to detect bonded atoms.

There was, however, one optimistic sign in the difference synthesis, namely the general shape of the electron-density "cloud" around the iron atom. In his first discussion of the shape of the mycobactin molecule, Snow pointed out that in all the possible isomers the iron atom was displaced well to one side of the molecule and that space filling models of the most favourable isomers suggested a flattish surface on the side of the structure occupied by the iron atom. The electron-density contours in the difference map were copied on to perspex sheets to reveal a general form which was close to that found in the models. This was felt to imply that the heavy atom phases did contain enough information to enable the structure determination if only the pseudo-symmetry could be broken.

In view of the dependence of the heavy-atom method on the correct location of the heavy atom, it was decided to carry out a direct-methods phase determination as a cross check on the interpretation of the Patterson map. In parallel with this work, F_o, $(F_o - F_c)$ and weighted Fourier syntheses were calculated, and the results of these were compared with an E-map obtained by direct methods. The application of the latter technique was somewhat hindered by the distribution of the strongest $|E|$ values. It is outlined below:-

In the space group $P2_1$, the origin is located by the assignment of arbitrary phases to two linearly independent zonal reflections of the type h 0 ℓ and to one h 1 ℓ reflection. In the present case the selection

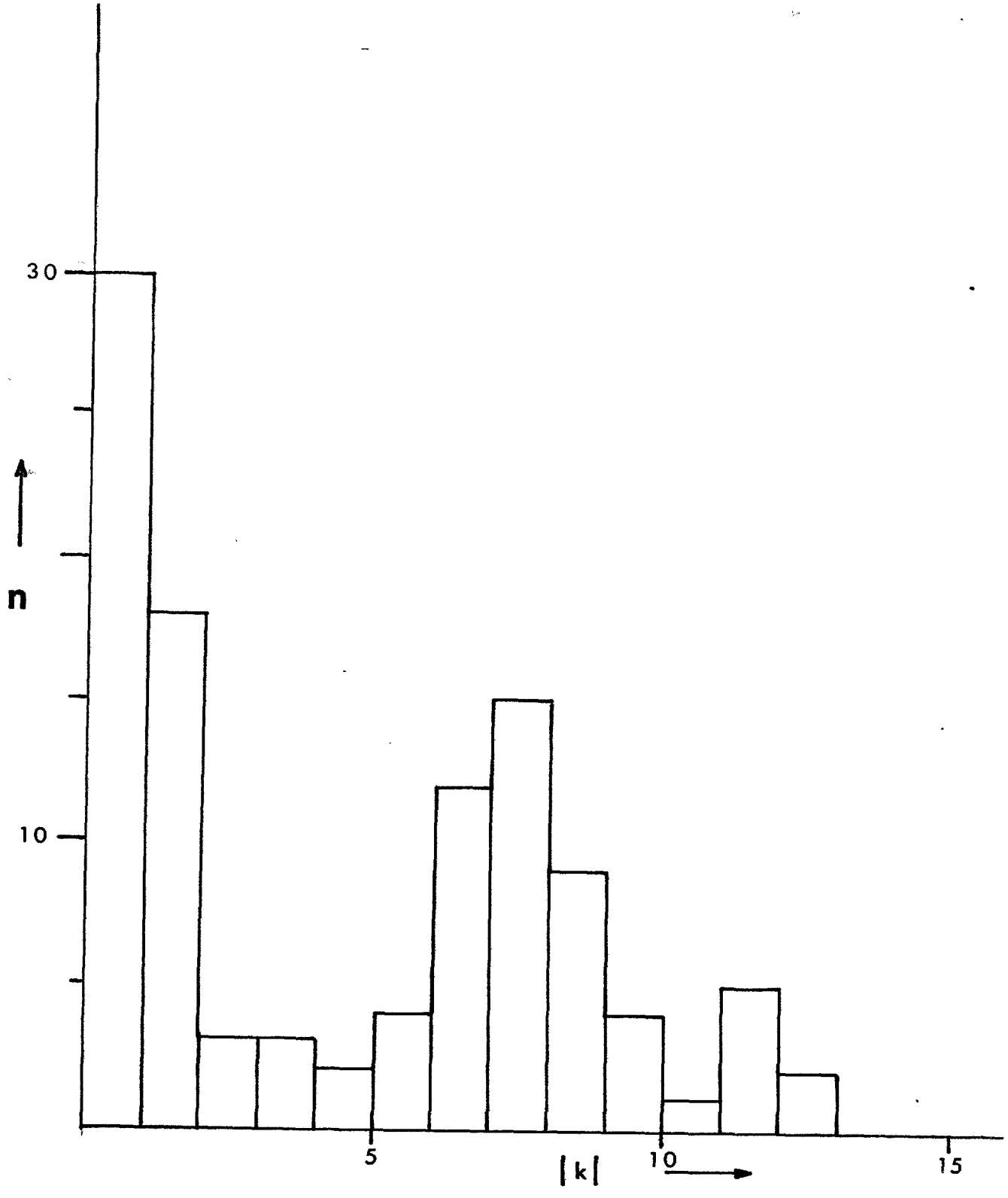


Figure 8 Distribution of the 120 highest |E|'s with respect to k

of these origin defining reflections was restricted by the distribution of the |E| values to a small number of possibilities. Figure 8, a histogram of the distribution of the 120 highest |E|'s with respect to the value of their $\underline{k}$ index, suggests that there should be no difficulty, but only 6 of the 18 |E|'s in the $\underline{h}$ 1 $\underline{\ell}$ group have |E| > 2.25; the rest are well down the list. In the first runs the reflections used to specify the origin, and the number of Σ_2 relationships that they are involved in, are:-

$\underline{h}$	$\underline{k}$	$\underline{\ell}$	|E|	$n\Sigma_2$	Phase
3	0	5	3.05	22	0
5	0	4	2.78	24	0
$\bar{7}$	1	4	2.45	24	0

In addition, 8 reflections of the type $2\underline{h}$ 0 $2\underline{\ell}$ were found whose phases could be determined by the Σ_1 formula:-

$$E_{2\underline{h}\ 0\ 2\underline{\ell}} = \frac{\sigma_2^{3/2}}{\sigma_3} < (-1)^{\underline{k}}(|E_{\underline{h}\ \underline{k}\ \underline{\ell}}|^2 - 1) >_{\underline{k}}$$

where the symbols have their usual meaning (see chapter $\frac{2}{3}$ of this thesis). The probabilities of two of these reflections, calculated using an appropriately re-arranged form of equation 21 in Chapter $\frac{2}{3}$, were high enough for their inclusion in the starting set, and this was further extended by the inclusion of three reflections which were assigned the symbolic phases $\underline{x}$, $\underline{y}$ and $\underline{z}$. The starting set was therefore:-

$\underline{h}$	$\underline{k}$	$\underline{\ell}$	E	$n\Sigma_2$	Phase	Probability
3	0	5	3.05	22	0	
5	0	4	2.78	24	0	
$\bar{7}$	1	4	2.45	24	0	
4	0	4	2.75	25	0	$P_+ = 0.989$
2	0	2	2.22	22	180	$P_+ = 0.067$
$\bar{6}$	7	3	3.50	30	$\underline{x}$	
$\bar{7}$	4	5	2.29	15	$\underline{y}$	
$\bar{4}$	11	3	2.20	12	$\underline{z}$	

The application of the Σ_2 relationship to the 120 strongest |E| values using this starting set gave phase indications for a further 14 reflections but their variations (see Chapter 2, page 82) were all considerably greater than the 0.5 recommended by Karle and Karle[25] as the maximum value that should be accepted. All these indications were obtained from single Σ_2 triples and, in the hope that a more extensive propagation might occur, the four best indications were accepted as correct. Although this resulted in a small improvement, the propagation of phases was still very inadequate and the symbolic phases $\underline{x}$ and $\underline{y}$ were reassigned several times. Unfortunately no progress was achieved and a completely new starting set was selected; again without progress. In view of this lack of success, a tangent refinement was attempted using 6 reflections from the original starting set and 8 reflections which had been determined from Σ_2 relationships. Since the $\underline{x}$ and $\underline{y}$ reflections were not involved in any useful Σ_2 relationships they, together with the $\bar{5}$ 7 6 reflection (this had the highest variance among the new phase indications included in the starting set) were omitted. This left only one reflection with a symbolic phase and since this must be used to define the enantiomorph, the number of possible tangent runs is reduced to two. Tangent refinements were therefore carried out with the phase $\underline{x}$ equal to 45° and 135°, allowing the phase development to extend over the 481 reflections having |E| greater than 1.25. In the first case 303 phases were obtained and in the latter 308. E-maps were calculated for both sets and were found to be essentially similar. Both contained a large peak at $\underline{x}$ = 0.95 and $\underline{z}$ = 0.83 but the y coordinate differed. In the first map (phase x = 45°) the peak was at y = 0.19 and in the second at 0.22. (These peaks are related to the iron position found from the Patterson Harker section by a twofold operation and different origin shifts in the y direction. They therefore confirm the selection of the largest peak in the Harker section as the iron-iron vector peak.)

In both E-maps the iron atom was surrounded by four peaks at the same y coordinate as the iron atom in a pattern similar to that found around the iron - iron vector peak in the Patterson map and around the iron peak in the Fourier syntheses. In addition, the remaining ligand atoms were found lying above and below the iron atom (in the y direction) but in both E-maps that lying above the iron atom was disproportionately electron dense (approximately 70% of the iron peak) and was surrounded by four satellite peaks resembling those around the iron atom itself. There was thus a strange repetition of the peak pattern around the iron atom vertically displaced some 2Å from that atom.

Both E-maps contained traces of pseudo-symmetry but the heights of the mirror-image peaks differed considerably. It was unfortunately not possible to identify any further chemically sensible features in either map but, since they both resembled the various heavy-atom Fourier syntheses, the departure from the perfect or near-perfect pseudo-symmetry found in the latter maps, could, it was hoped, be used to overcome this problem.

A difference Fourier synthesis based on the iron atom alone ($\underline{R}$ = 55%) and a weighted Fourier synthesis based on the iron atom plus five "ligand" atoms (included as oxygen atoms - $\underline{R}$ = 54.3%) were calculated and found to contain some 17 new peaks in common. The six strongest peaks were also found to occur in the E-maps and were therefore used as carbon atoms. (In each case the mirror image favoured in the E-maps was selected.) A new weighted Fourier synthesis gave an $\underline{R}$ factor of 51.7% and from this map seven new peaks were chosen (they were also found to be present, though weak, in the E-maps) for inclusion in a further Fourier synthesis ($\underline{R}$ = 49.6%) and were also subjected to three cycles of full-matrix least-squares refinement which reduced $\underline{R}$ to the lowest value obtained thus far, 38.8%.

Although the $\underline{R}$ factor from this refinement was fairly low, the general appearance of the parameter changes was not very encouraging and there

was no indication of a significant convergence. The temperature factors of several of the atoms were high (15 - 27$Å^2$) and neither of the difference Fourier maps gave usable new atoms. The structure determination was thus, again, at an impasse with no recognisable part of the molecule identified.

In view of the considerable importance of the structure of ferrimycobactin to an understanding of the mechanism of iron transport in the mycobacteria it was decided not to abandon the work but to attempt to solve the problem by one of the remaining possible methods. These were:-

a) Patterson Superposition.

b) Tangent refinement of phases calculated from a small non-centrosymmetric part structure.

c) Isomorphous replacement.

d) Methods based on the effect of anomalous dispersion.

e) A combination of methods c and d.

A trial hand application of the first of these methods by the "minimum function" method was unsuccessful and, in view of the failure of the direct-methods approach outlined above, it seemed unlikely that method b, tangent refinement of a set of phases calculated from a small group of atoms, would be successful. The method has been applied in a number of instances[26] to solve unknown structures and has been discussed in connection with the refinement of protein structures. Successful application depends partly on the use of an initial set of atoms which is not pseudo-symmetric and it was felt that the dominance of the iron atom in ferrimycobactin molecule would make this difficult to achieve.

The existence of an aluminium chelate which is isomorphous with the ferric complex has already been mentioned and this should enable the application of method c above. This method is used primarily for the solution of the phase problem in protein structures and ideally requires more than two isomorphs to avoid an ambiguous solution, although some workers

have proposed that this uncertainty may be resolved using special Fourier syntheses. The application of the method in the present case would require the preparation of good single crystals of the aluminium complex and possibly the chromium complex. Since this had not yet been achieved, the use of this method would inevitably have necessitated a considerable delay.

Method d, the use of anomalous dispersion effects seemed the most likely to yield a rapid structure solution. A number of structures[27,28] comparable in size to ferrimycobactin have been solved in this way and in each case the method gave a set of phases which contained enough information to enable the location of almost all of the atoms in the structures from a single Fourier map.

The method may be applied in two ways, both of which require intensity data from a single derivative. In both cases the position of the fluorescent atom must be known. The first method requires two sets of intensity data, a set of "Bijvoet" pairs measured with an X-radiation which excites fluorescence and a second set in which the asymmetric part of the same volume in reciprocal space is measured with a radiation which does not excite fluorescence. From these data, combined with a knowledge of the position of the anomalous scatterer (*e.g.* from a Patterson synthesis) a unique set of phases may be readily calculated.

The second method requires only a set of Bijvoet pairs but gives two possible values for each phase angle. In spite of this ambiguity, however, this method is the most widely used since the assumption that the correct phase is that which lies closest to the phase of the heavy atom contribution has been found to be sufficiently valid to enable the structure to be solved.

This latter method requires the measurement of a complete set of anomalous intensities from a single crystal and is therefore less demanding on time than the first. In the present case the position of the iron atom is known with certainty from Patterson and direct methods and the imaginary component of the fluorescence is fairly large ($\Delta f''$ is 3.3 [29]).

An examination of the literature showed that of the small number of structures which have been solved by this method, the majority are in the monoclinic space group $P2_1$ and that several of the problems are of the same order of magnitude as ferri-mycobactin. The first-row transition metals cobalt and chromium have been used as the fluorescent atom and therefore iron should be eminently suitable. The most noticeable feature of these applications of the method is that in each case the great majority of the structure was obtained from the Fourier synthesis using the phases obtained from fluorescence effects.

The theory of phase calculation from fluorescence effects, together with the "extended Fortran" computer program "BIJVOET" which was written for this purpose is summarised in the Appendix.

In order to carry out the phase determination a hemisphere of data of the type $\pm h \pm k \ell$ was measured using a crystal mounted about the c^* axis, and out to $\theta = 50^\circ$. This comprised 5079 reflections of which 278 were classified as unobserved. A new Patterson synthesis was computed using the $\pm h + k \ell$ quarter sphere to check that the iron position was unchanged. This map was found to be essentially identical to the previous map.

Since the phase determination involves a calculation containing both observed and calculated structure factors, the intensity must be placed on a correct absolute scale and a Wilson plot calculation was therefore carried out. This gave a scale factor of 0.55 and an overall temperature factor of 6.5.

The rejection ratio in the program was set at 0.01 to ensure that only reflections having a heavy-atom contribution less than 0.01 times the mean observed intensity for a given intensity pair would be rejected from the phase calculation. Pairs where one member was classified as unobserved were included in the calculation. The h 0 ℓ zonal reflections were also included, being assigned the phases of their heavy-atom contribution.

(It must be emphasised that this was intended to be a trial application of the computer program and the structure solution which resulted was somewhat of a surprise.) The effect and validity of these selections is discussed in the Appendix.

Using these criteria and the complete set of 5079 intensities as input data a phase calculation was carried out. 2179 Bijvoet pairs were accepted by the sorting procedure and were used for the phase calculation. Together with the ±h 0 ℓ zonal reflections this resulted in a total of 2477 phases which were input into the X RAY-67 Fourier program and Fourier syntheses was carried out.

A rapid examination of this map showed that the iron atom was present in its correct position and that there was no pseudo-symmetry. The map contained a large number of other peaks, ranging in height from +85 to -140, but the majority were positive with an average height of *ca.* 35. The map was quickly found to contain the linked five-and six-membered rings of the 6-methyl salicyclate and oxazoline systems and the majority of the structure was located. In all 41 atoms in addition to the iron atom were found in chemically sensible positions and it was clear that the formula shown in figure 2 was correct.

38 of these atoms were subjected to 3 cycles of full-matrix least-squares refinement giving an R factor of 31% and from a difference Fourier synthesis 11 further atoms were identified. Since the structure found was in complete agreement with that proposed by Snow the atoms were given their correct chemical identity and the whole structure refined for three more cycles to an R factor or 22%. 8 more atoms were located from a new difference Fourier map taking the total up to 57 and leaving 6 atoms unlocated. The iron atom and its six ligands were then assigned anisotropic temperature factors and the structure refined further. This reduced R to 13.5% and a further 0.5% improvement was obtained by including a correction for anomalous dispersion.

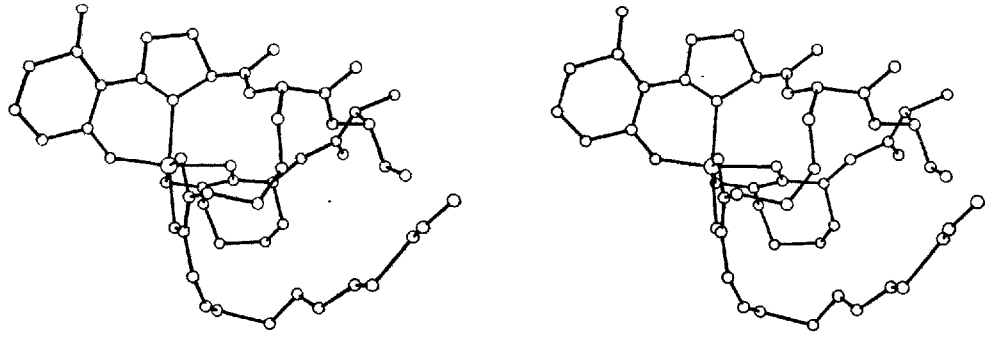


Figure 9 Stereoscopic perspective view of the molecule

The 6 unlocated atoms were the terminal members of the C_{15} aliphatic side chain (see Fig. 3). The temperature factors for the 9 identified atoms in this system showed a progressive increase along the chain from 14Å^2 for the first methylene group (C49) to 29Å^2 for C57 and the standard deviations in the coordinates of the latter atom were some 20 times greater than those for the iron atom. It thus appears that the terminal atoms in the side chain are subject to severe thermal movement and/or disorder.

In order to check the location of the aliphatic side chain, a difference Fourier synthesis was calculated omitting this part of the structure. The map showed that, apart from some small adjustments, the nine atoms were correct. Two additional atoms were located, but the electron density decreased rapidly along the chain so that they were very diffuse. The inclusion of these atoms brought the total to 59 and resulted in an improvement of the R factor after refinement to 12.1%. A further difference map was then calculated but there was no sign of the remaining 4 atoms.

It is hoped that further refinement will reduce the R factor but it seems unlikely that the last atoms will be located. The positional and thermal parameters for the 59 atoms are given in Tables 3 and 4. Table 5 contains a list of the observed and calculated structure factors at R = 12.1%, and figure 9 shows the molecular structure.

Results and Discussion

1) General Aspects

Comparison of figure 3, the structure proposed by G.A. Snow in 1965[11], with figure 9, the results of the present crystal structure determination, shows that the former is correct in all the chemical details. It is also clear that Snow chose the correct ligand atoms (a - f in fig. 3) and that the coordination pattern is that numbered 5 in figure 4, one of the three he selected as the most compact and least strained of the 8 isomers where models could be constructed.

The stereochemistry of the molecule differs from that shown in figure 3 at two sites but one of these is due to a drawing error in the original paper and the second is of minor importance. The first of these instances concerns C18, the bridgehead atom between the three principal chains in the structure and the α-carbon atom in one of the two N^6-hydroxy-lysine residues in the molecule. In figure 3 this amino acid residue is drawn in the D configuration although Snow reported the isolation of two moles of L-N^6-hydroxy-lysine from the degradation products of mycobactin P (see figure 1). The second difference is in the conformation of the seven-membered ring which Snow assumed to be in a "staggered conformation with the amide group (N29) equatorial". In the figure he drew the ring in an approximately "twist-chair"[30] conformation, whereas the crystal-structure determination shows it to be in the "chair"[30] form. Essentially, these two conformers are related by the "flipping" of one atom, in this case C33. The conformation of the seven-membered ring will be considered in greater detail below.

The crystal-structure determination shows that for all the five dissymmetric carbon atoms in the molecule the conformation is that proposed by Snow from chemical studies. In Cahn-Ingold-Prelog notation the structure is 12S, 18S, 22S, 25R, 30S. The three carbon atoms which originated as α-carbon atoms in amino acids (C12, C18, and C30) are thus in the L (≡ S) conformation. This is in agreement with the results of the degradation studies shown in figure 1 and is normal for naturally occurring amino acids and their derivatives. The octadec-2-enoic moiety (see fig. 2) contains a cis double bond, again in agreement with the results of the chemical studies.[11]

The ferrimycobactin molecule is, apart from the phenyl-oxazoline system, roughly spherical in shape with an average diameter of about 10Å. The phenyl-oxazoline moiety lies almost perpendicular to the rest of the molecule but the fatty acid side chain is folded so that it lies

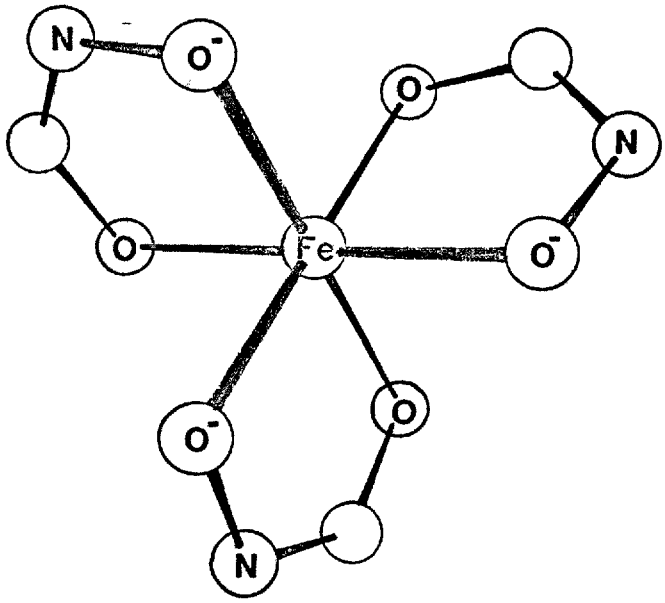


Figure 10a

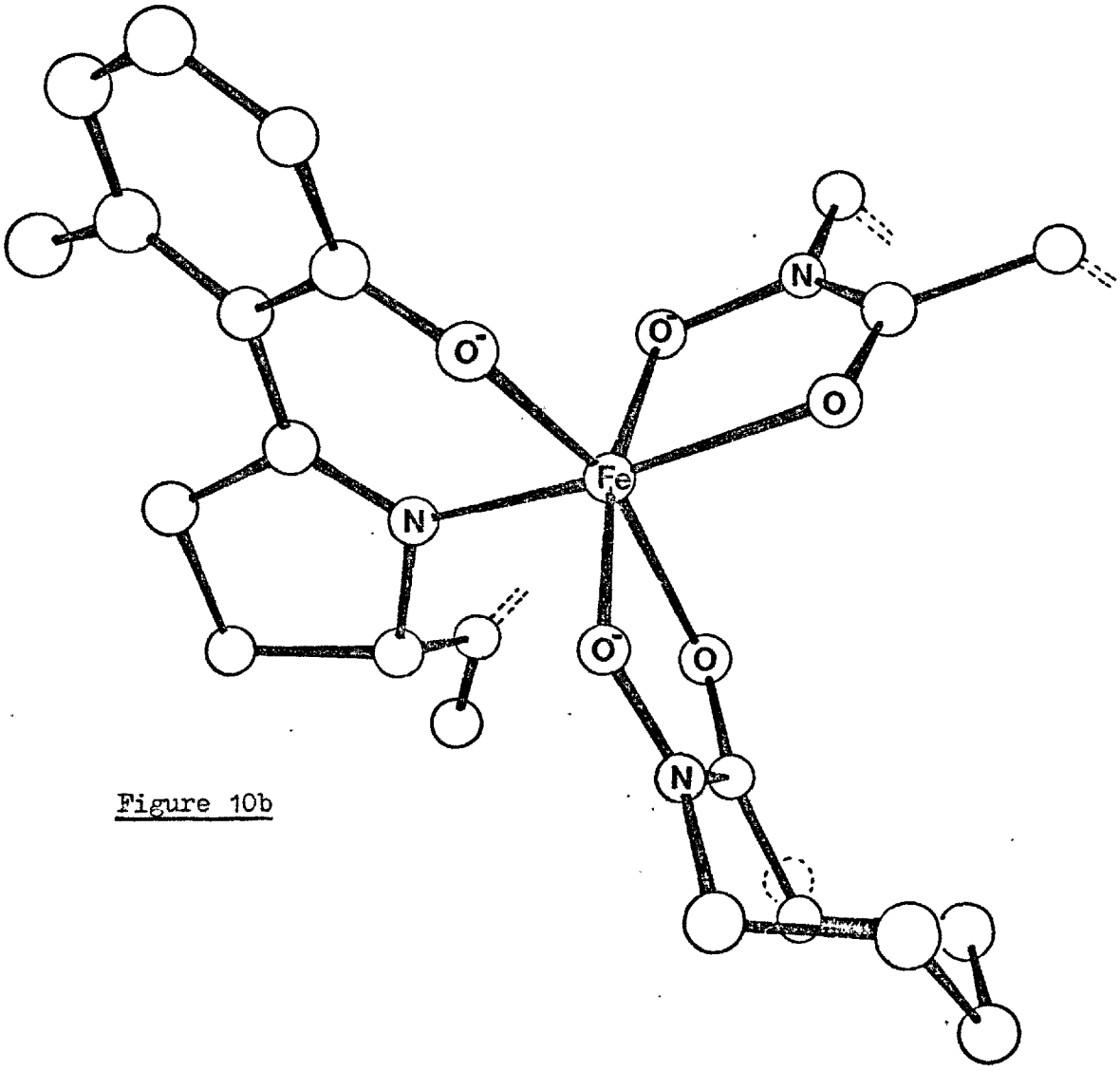


Figure 10b

between the C39 to C42 methylene chain and the chain of atoms from C24 *via* C21 to C33. The majority of the external surface of the molecule consists of the aromatic, salicylate derived, ring, the fatty acid side chain, methyl groups on C7 and C25, the ethyl group on C22 and methylene groups in the C39 to C42 system and in the seven membered ring. All these groups are hydrophobic and, in spite of the presence of three carbonyl groups (O16, O20 and O28), the result must be a molecule with an almost (see below) completely hydrophobic exterior and a consequent low solubility in aqueous media.

The coordination of the Fe^{3+} ion is effected by the six atoms which Snow[11] selected as ligands (a-f in Fig. 3) forming one six-membered and two five-membered chelate rings. The chirality of the chelation is shown in figure 10, together with that in the naturally occurring ferric tri-hydroxamate Ferrichrome A[31] (*cf.* Fig. 5). Both these structures are in what Zalkin *et al.*[31] termed a "left-handed propellor" configuration. A similar coordination was also encountered in the synthetic ferric hydroxamate complex tris-benzhydroxamato-iron(III)-trihydrate.[32] In ferrichrome A and the benzhydroxamate complex the ligands are bonded to the Fe^{3+} ion such that the three negatively charged oxygen atoms form one face of the coordination octahedron (*ie.* the hydroxamate groups are "cis") but this is not the case in Ferrimycobactin where they lie equatorially.

The atomic chains which link the chelate rings in ferrimycobactin all lie on one side of the molecule, meeting at the bridgehead atom C18. As a consequence of this the iron atom lies near the surface of the molecule surrounded by four of the oxygen ligands (a,c,e and f in fig. 3) and in the centre of a comparatively hydrophilic region. The linking of the chelate rings also results in their being pulled back from the metal ion toward C18 so that its exposed position is accentuated.

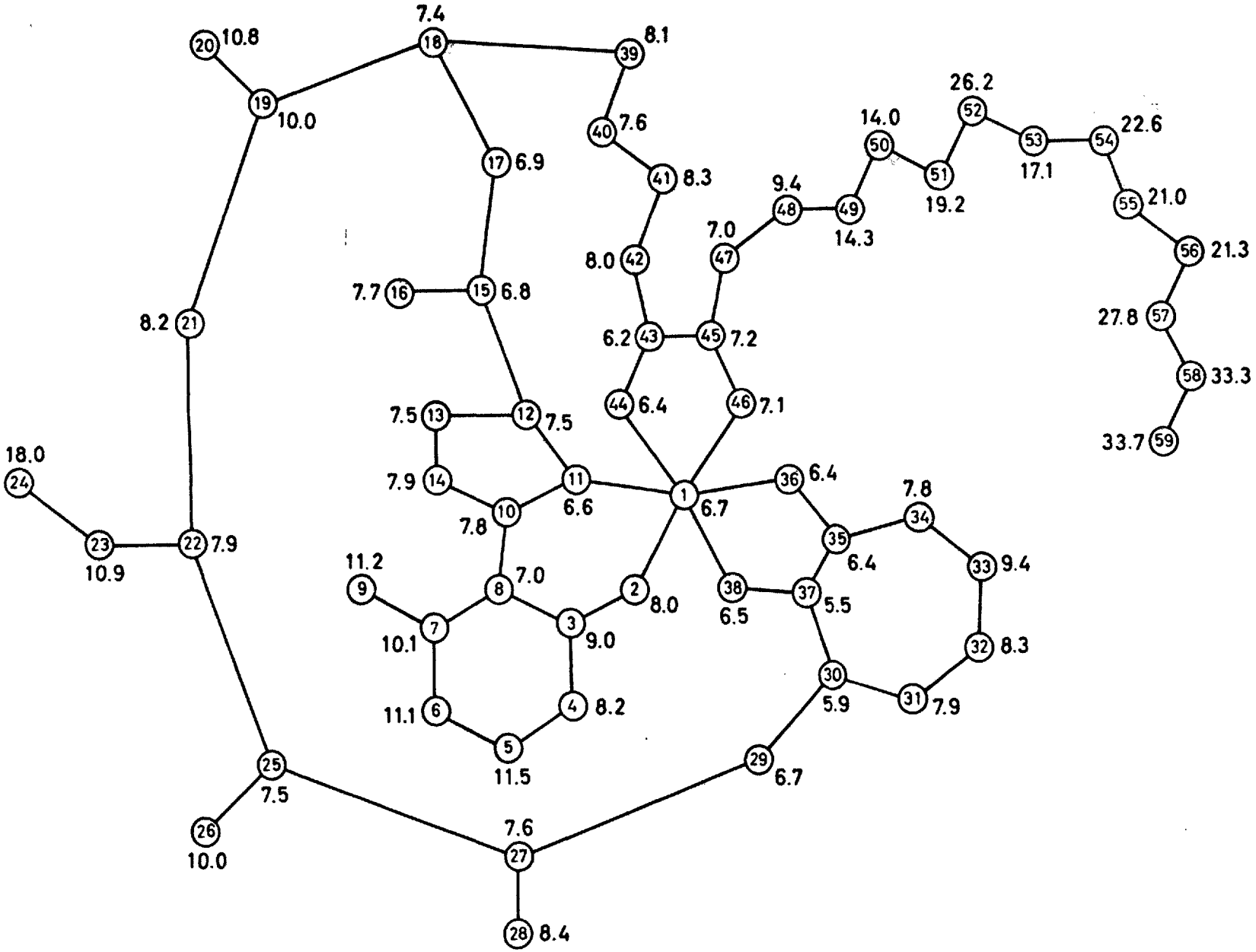


Figure 11 The atom numbering scheme (numbers inside the circles) and isotropic temperature factors (A^2)

In spite of the somewhat "spidery" appearance of the mycobactin molecule the chelation of the ferric ion results in a fairly rigid structure which is stiffened by intra-molecular hydrogen bonds between N17 and O44 (3.00(3)Å) and O21 and N29 (2.84(3)Å). The positions of the first three atoms of the fatty acid side chain (C47 - C49) are restricted by double bonds but the remainder of the chain is only restrained by weak conformational forces and crystal lattice effects. From figure 11, which shows the isotropic temperature factors for the atoms in the structure, it is clear both that the general level of thermal movement is rather high (average B value is 10.4Å^2) and that it increases rapidly as one moves away from the ferric ion. The latter effect is particularly marked along the fatty acid side chain and is responsible for the large standard deviations in the positional parameters of the last few atoms and the absence of the 4 terminal atoms in difference Fourier syntheses. In the crystal structure (see fig. 18 below) ferrimycobactin molecules are packed with the fatty acid side chains lying in a zone which bisects the unit cell in the a direction and which is occupied exclusively by hydrocarbon residues. The packing in this region is very loose and results in prominent {1 0 0} crystal faces and easy cleavage in these planes.

2) The Structure and its Biological Activity

It was demonstrated above that, in the crystalline state, the ferrimycobactin molecule has an external surface which, with the exception of a small area surrounding the iron atom, is almost entirely hydrophobic in nature. On this basis the compound's very low solubility in aqueous media, high solubility in lipids and lipid-like solvents and its specific action as an iron transporting agent in the mycobacteria are not surprising. The biochemical studies of Ratledge and his colleagues which were outlined in the introduction have shown that mycobactins are located only in the mycobacterial cell walls and that they accumulate at lipid/aqueous system

interfaces. The observed structure of ferrimycobactin P fits this evidence remarkably well if one assumes "soap like" behaviour with the hydrophilic region of the molecule in direct contact with the culture medium and the hydrophobic remainder embedded in the lipid regions of the cell wall which in the mycobacteria are unusually thick (270Å against the normal 60 - 70Å). In the present case, the "soap" action functions in reverse since it results in the solution of the normally hydrophilic ferric ion in a lipid medium.

The comparatively rigid and compact structure of the central part of the ferrimycobactin molecules results from intra-molecular hydrogen bonds (see above) and from the chelation of the ferric ion. Removal of the latter would therefore, at first sight, be expected to result in the loss of this rigidity to give a loose open-structured molecule with considerably different physical properties. Snow[5] attributed the differences in the optical rotation of desferri-mycobactin P ($[\alpha]_D^{15} = -19^\circ$ in $CHCl_3$) and its aluminium chelate ($[\alpha]_D^{15} = +376^\circ$ in $CHCl_3$) to this effect. (In ferrichrome the difference between the compact iron chelate and the open desferri form is believed to be recognised by the membrane transport system in Ustilago sphaerogena to permit free egress of the ligand and retention of the chelate.[33]) However Ratledge, commenting on the remarkably similar interfacial tensions of desferri- and ferrimycobactin S at a water/ethyl n-decanoate interface (20.6 ± 0.3 and 20.9 ± 0.9 dynes/cm. respectively) stated that this "must be due to intramolecular hydrogen bonds in desferri mycobactin making it approximate to ferri-mycobactin".

The similarity in the solubility properties of ferri- and desferri-mycobactin suggests that they have similar external surfaces and that the hydrophilic hydroxamic acid groups which would be exposed in an open structured desferri-mycobactin molecule are probably still "hidden" within the structure. The simplest way in which this can be achieved

is by the formation of intra-molecular hydrogen bonds between these groups so that the compact conformation of the molecule is largely unaltered by the removal of the metal ion.

No detailed spectroscopic study of this possibility has been carried out, but Greatbanks and Bedford in their paper on the identification of mycobactins from NMR spectra[34] stated that "the NH and OH protons... appeared between $\tau = 4$ and $\tau = -2$" which suggests the possibility of strong hydrogen bonding. (Hydrogen bonded protons often give NMR signals with negative τ values.)

A detailed prediction of hydrogen bonding in desferri-mycobactin from the intermolecular distances in ferri-mycobactin is of questionable value since it is possible to make quite large movements of the potential donor and acceptor atoms without seriously altering the overall conformation of the molecule. However, assuming a neutral desferri-mycobactin molecule, the following atoms can act as hydrogen bonding donors: O2, N17, N29, O36 and O44 and the following as acceptors: N11, O14, O16, O20, O21, O28, O38 and O46. All the donor atoms, with the possible exception of O6, and four of the potential acceptors (N11, O21, O38 and O46) "look" toward the centre of the molecule, whereas the remaining acceptors are on the external surface (see above). The formation of a hydrogen bonded "core" which serves the same conformational purpose as the ferric ion thus seems possible.

The distances (in Å) between the potential donors and acceptors in a postulated desferri-mycobactin which retains the exact structure of the metal chelate are:-

	Acceptor			
Donor	N11	O21	O38	O46
O2	2.78*	>5.0	3.98	3.10
N17	2.82	2.87	3.08	4.90
N29	4.88	2.84	2.57	>5.0
O36	3.26	*5.0	2.59*	2.94
O44	2.78	5.02	3.02	2.54*

and it is plain from these data that several hydrogen bonding systems can be achieved without loss of the overall shape of the ferri-mycobactin molecule. The three interatomic distances which are marked with an asterisk in the table do not contribute to the maintenance of conformation since they merely retain the conformation of the individual chelate rings and do not bridge between them.

In a detailed study of the infra-red spectra of a series of hydroxamic acids Hadži and Prevoršek[35] have shown that they form both intra- and inter-molecular hydrogen bonds and, from the effect of temperature on the OH absorption bands, concluded that the latter are the stronger. Only one crystal structure of a hydroxamic acid appears to have been carried out so far[36] and in this case the molecules were found to be linked by inter-molecular hydrogen bonds.

In a paper on the structure of the mycobactin analogue Nocobactin NA (see page 203) Ratledge and Snow have attributed fine structure in the ultra-violet spectrum of the compound to strong hydrogen bonds between the phenolic hydroxyl group (corresponding to O2 in ferri-mycobactin P) and the oxazole nitrogen atom (corresponding to the oxazoline nitrogen atom, N11, in ferri-mycobactin P). In nocobactin the latter atom is less basic than its equivalent in mycobactins and one would therefore expect an even stronger O2 ... N11 hydrogen bond in desferri-mycobactin P. It therefore seems likely that the desferri-mycobactin P molecule is held in a compact conformation which is close to that of the metal chelate by a system of hydrogen bonds between some of the atoms listed above, but that O2 and N11 may be excluded from this network by the formation of a strong "intra-phenyl-oxazoline" hydrogen bond.

The possible exclusion of O2 and N11 from a hydrogen bond network within the core of the desferri-mycobactin molecule may be the key to an understanding of the mechanism of the insertion of the ferric ion into the ligand molecule and its subsequent reductive removal.

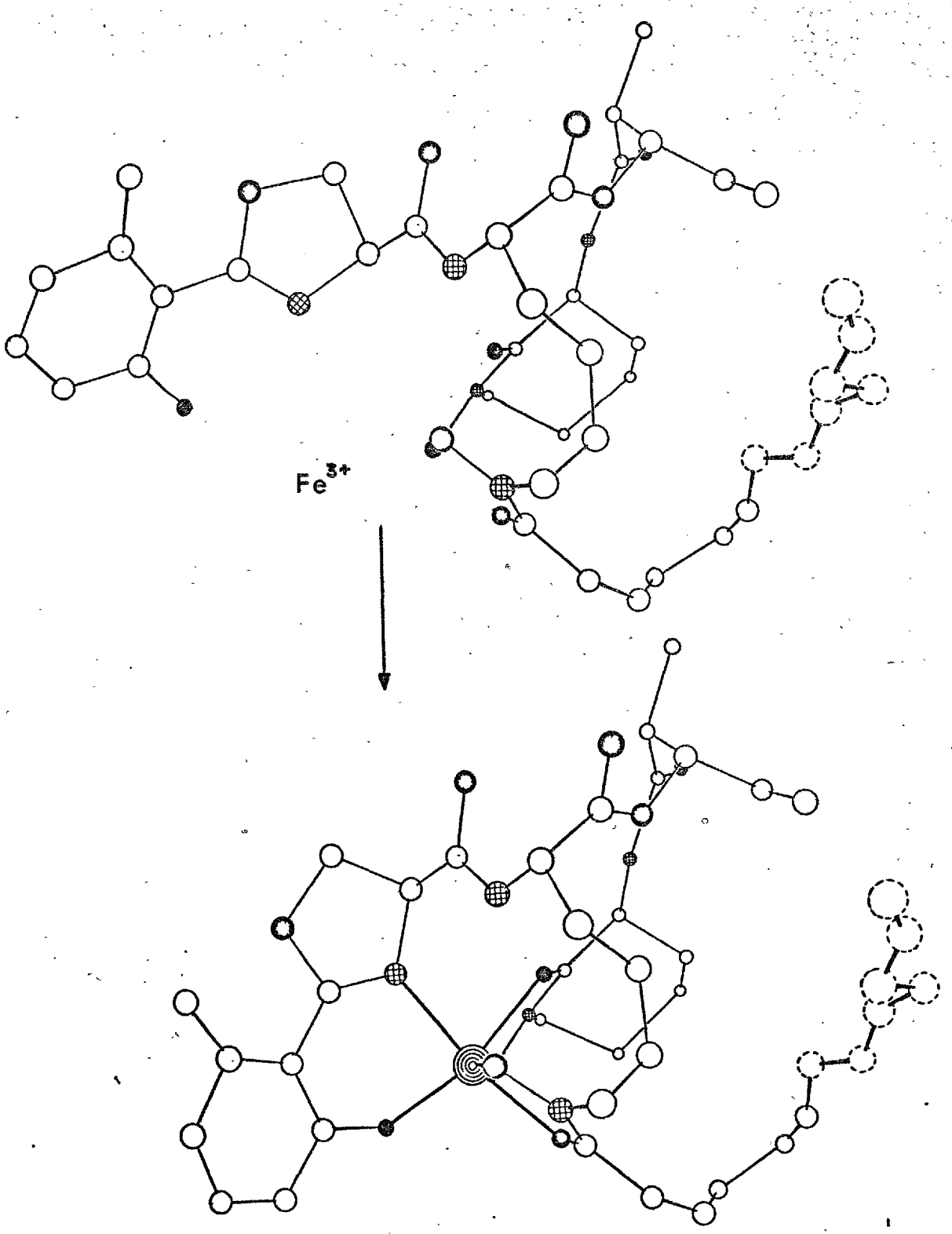


Figure 12

If one assumes that the "peptide" nitrogen atom N17 is hydrogen bonded to an acceptor within the core of the postulated desferri-mycobactin molecule (*eg.* O38) and that the peptide link between C15 and N17 remains in the "planar trans" conformation, the formation of an O2 ... N11 hydrogen bond frees the whole of the phenyl-oxazoline moiety to rotate about the C12 - C15 bond. The effect of this rotation is shown in figure 12 which is an [0 1 0] projection of the ferri-mycobactin P molecule with the ferric ion omitted. From this figure it is clear that the displacement of the phenyl-oxazoline system results in a) an increase in the area of the hydrophilic part of the molecular surface, b) the appearance of an additional hydrophilic atom at this surface (N11), c) the exposure of most of the atoms which serve as ligands in the chelation of the ferric ion, and d) retention of the predominantly hydrophobic nature of the molecule.

In view of the overall size of the mycobactin molecule it seems unlikely that points a and b would seriously affect the solubility properties of the molecule although one might expect an increase in the inter-facial tension of the desferri form from point b. Since it has been argued that the desferri form probably retains the overall conformation of the metal chelate, the most important consequence of a rotation of the phenyl-oxazoline moiety around the C12 - C15 bond is point c above, the exposure of most of the ligand atoms on the surface of the molecule, thus making them accessible to an approaching metal ion.

From the biochemical considerations detailed in the introduction, it seems likely that Fe^{3+} ions in the culture medium are retained in solution as a salicylate complex and an exchange process must, therefore, occur at the cell wall lipid/culture medium interface. It is possible that this occurs *via* an initial capture of the metal ion by the extremely strongly complexing hydroxamate groups followed by the return of the more weakly complexing phenyl-oxazoline system to complete the coordination sphere around the iron atom. A detailed discussion of the mechanism of

exchange of the ligand atoms in the salicylato iron (III) complex to give ferri-mycobactin is beyond the scope of this thesis but the sequence proposed above is in accordance with the complexing ability of the ligands involved in the process.

Spectroscopic and potentiometric studies by Agren[37], Park[38], and Ogawa and Tobe[39] have shown that iron (III) is complexed by salicylic acid in three stages depending on pH. Below pH 2.5 the mono-salicylato complex is formed exclusively but as the pH is increased the bis- and tris- complexes appear. At pH 6, the lower limit in normal physiological and environmental situations, all three species are present but the mono-salicylato complex predominates even in a large excess of salicylic acid. The stability constant for this complex is approximately 10^{16}, a value which is drastically lower than that of ferri-mycobactin ($> 10^{31}$) so that an exchange equilibrium of the type:-

$$(\text{Sal Fe})^{2+} + \text{MycH}_3 \rightleftharpoons \text{MycFe} + \text{Sal}^- + 3\text{H}^+$$

will strongly favour the formation of the mycobactin complex.

Biochemical studies described in the introduction have shown that iron is removed from ferri-mycobactin *via* an enzyme-mediated reduction to iron (II). This may occur in two ways, the first requiring a re-arrangement or replacement of part of the coordination sphere around the ferric ion so that it is brought into close contact with the reductase active site before reduction, and the second requiring reduction of the metal ion in an undisturbed ferri-mycobactin by an electron-transfer process which utilises one (or more) of the ligand atoms. Since the overall result is the loss of the metal atom, both of these possibilities require the breakdown of the ferri-mycobactin chelation pattern. In the first mechanism this would partially occur before reduction and in the second after reduction.

In the first mechanism comparison of the stability constants of ferric hydroxamates ($K_1 \approx 10^{11}$, $K_2 \approx 10^9$)[14] with those of ferric phenolates

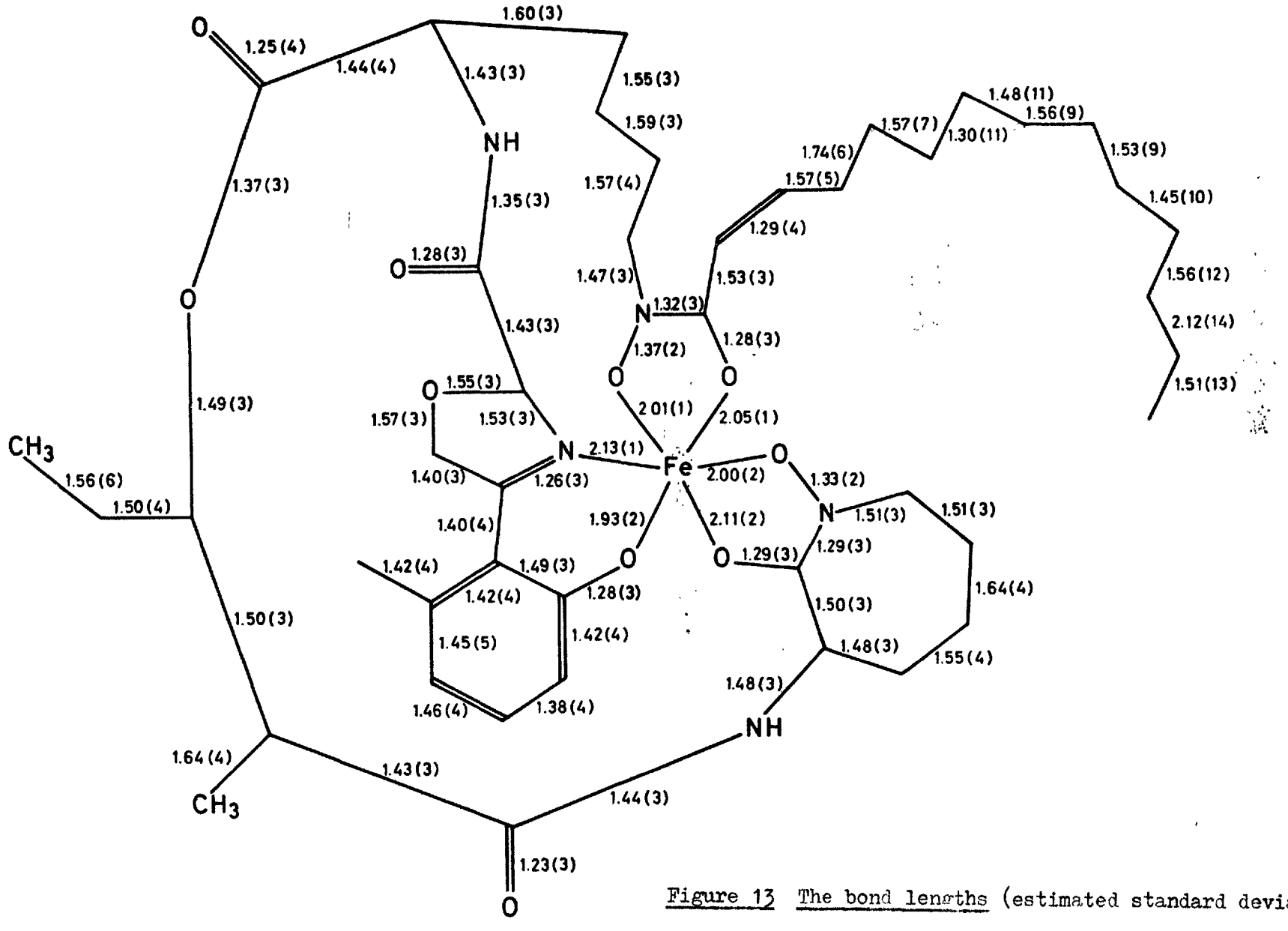


Figure 13 The bond lengths (estimated standard deviations in parentheses)

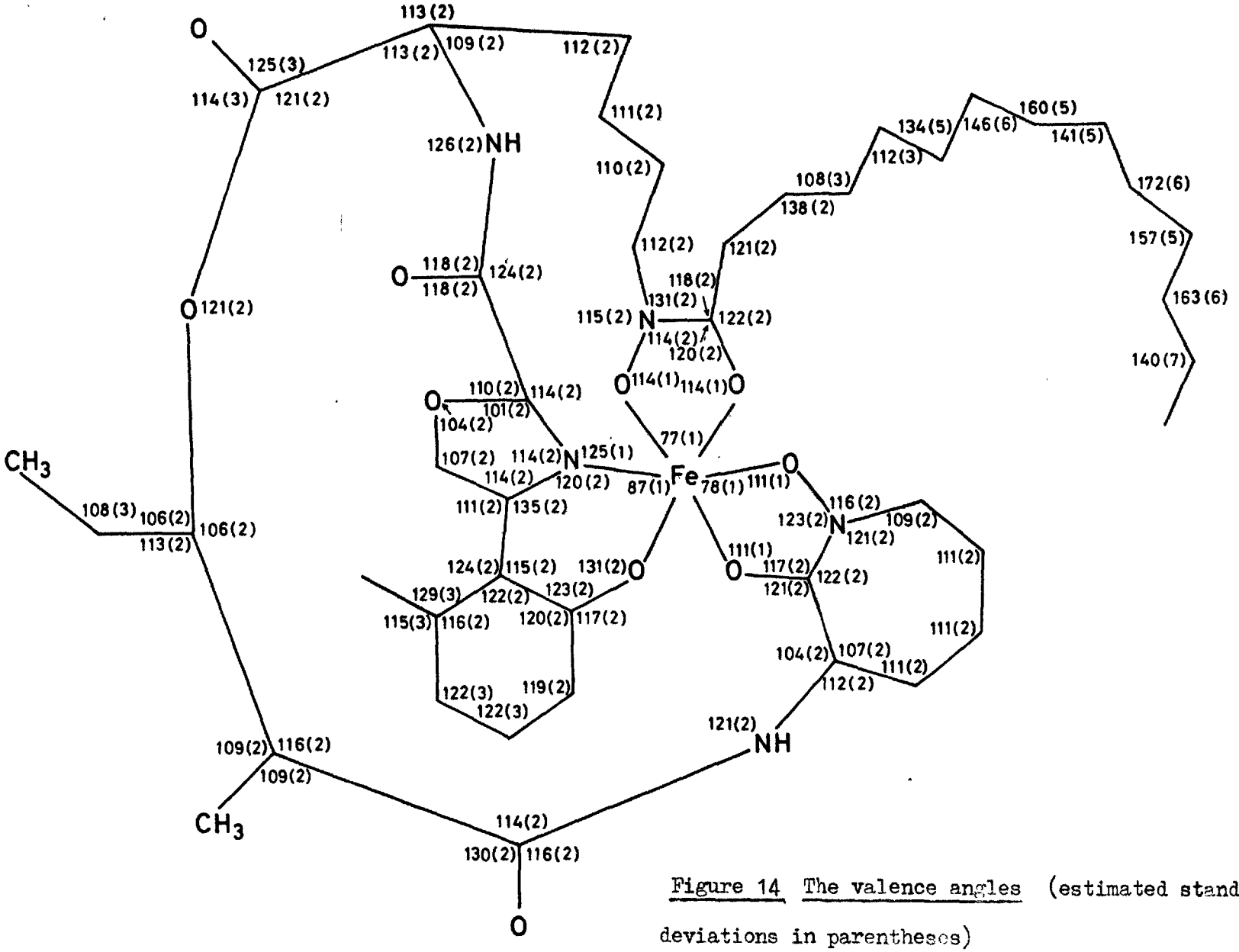


Figure 14 The valence angles (estimated standard deviations in parentheses)

($K_1 \approx 10^8$)[40], combined with the fact that nitrogen is generally a weaker donor to iron(III) than oxygen suggests that the necessary opening of the iron coordination sphere is likely to occur by loss of the phenyl-oxazoline system. The situation is less clear for the second possible mechanism since stability constant data are not available for iron (II)/phenol complexes. However, in view of the possible conformation of desferri-mycobactin it seems likely that the loss of the metal ion after reduction would again occur by initial departure of the phenyl-oxazoline moiety.

In summary, it seems from the geometry of the ferri-mycobactin P molecule combined with spectroscopic, solubility, interfacial tension and stability constant data that the desferri-mycobactin molecule may retain the same overall conformation as the metal chelate and that the phenyl-oxazoline moiety may serve as a "door" which seals a newly captured Fe^{3+} ion in place in the metal complex and which opens either to allow its reduction and release or its release after reduction.

3) Details of the Ferri-mycobactin Structure

The atomic numbering scheme, bond length and valence angles in ferrimycobactin P are shown in figures 11, 13 and 14 respectively. (The complete list of angles at the iron atom are given in figure 17). With one exception (C57 - C58, 2.12(14)Å) these parameters are within 3 standard deviations of the normal values although the effect of thermal movement is very marked in the fatty acid side chain. The size of the standard deviations of the majority of the geometrical parameters precludes any detailed consideration of the bonding within the molecule but this will be discussed in those places where it is important to the biological function of the compound or in connection with conformational arguments raised in the previous section. The major organic parts of the molecule will be considered first, followed by the iron coordination. Least squares planes and torsion angles in the

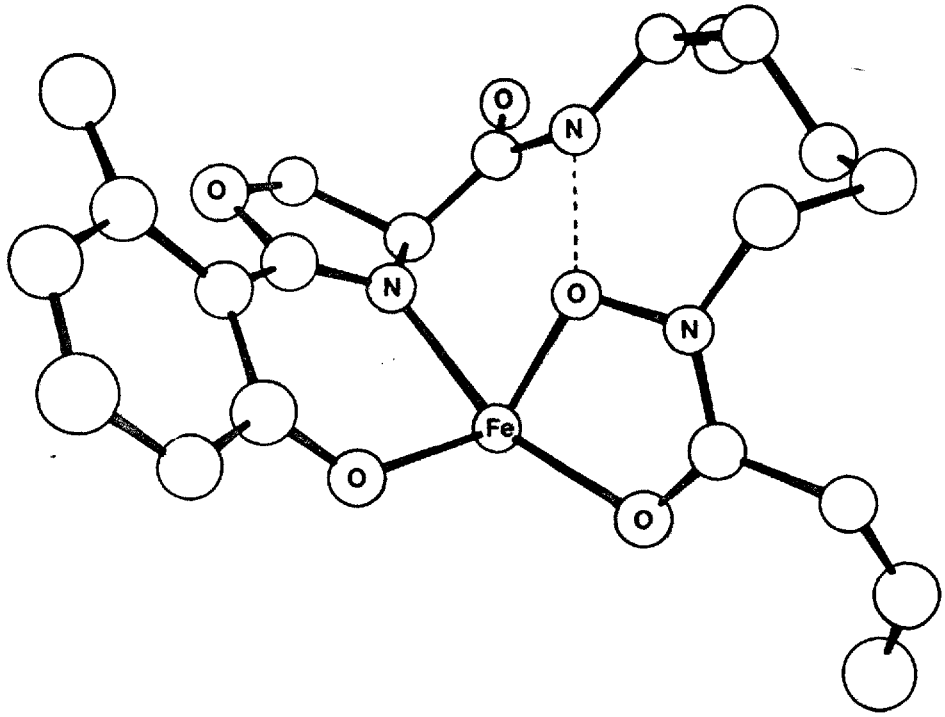


Figure 15

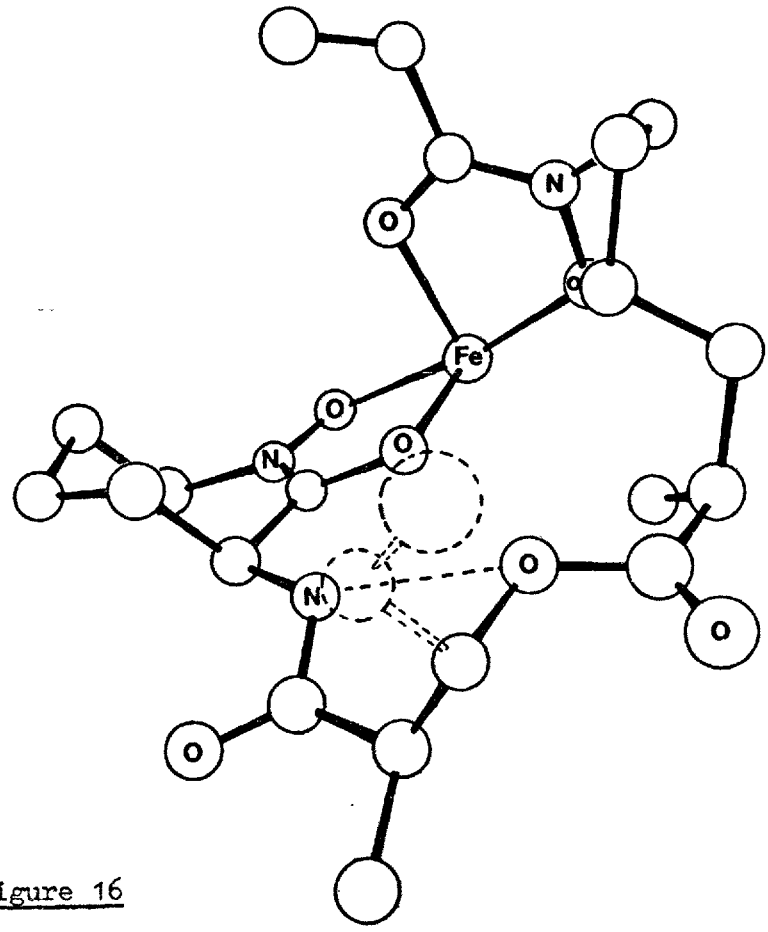


Figure 16

molecule are given in tables 6 and 7 respectively.

Figure 15 shows the conformation of the phenyl-oxazoline system and the chain of atoms through C18 and C42 which links the O2 to N11 and the O44 to O46 chelate ring systems. It is clear from this figure that the phenyl-oxazoline system consists of two approximately planar rings which are linked by the C8 - C10 bond, and that the oxazoline ring is rotated some 30° out of the plane of the benzene ring. The deviations of the atoms in the latter ring from their least-squares plane (σ = 0.03Å) are of the same order as the standard deviations in their positional coordinates, and χ^2 [41] for the system is 5.56. The probability[42] that a "planar" set of six atoms will have this χ^2 value is *ca.* 36%, a value which is too high to justify the statement that the benzene ring is nonplanar.

The departure of the atoms which comprise the oxazoline ring from their best plane are more widely spread (see table) but the standard deviations of the atoms from the plane is again 0.03Å. In this ring, however, the χ^2 value (8.4) gives a far lower probability for planarity (*ca.* 8%) and it is, therefore, reasonable to describe the ring as nonplanar.

The best plane in the phenyloxazoline system is that centred around the C10 - N11 double bond. This comprises C8, C10, N11, C12 and O14. The standard deviations of these atoms from their least-squares plane is 0.004Å for which the χ^2 value (0.14) gives a probability for planarity which is greater than 99%. The remaining atom in the oxazoline ring (C13) lies 0.10Å out of this plane so that the ring has a shallow "envelope" conformation.

Only one oxazoline compound, perkinamine, appears to have been described in full in the crystallographic literature[41]. The oxazoline ring system in this compound carried a methyl group in the 2 position (*ie.* replacing the phenyl group in mycobactin) and the complete methyl-oxazoline system is approximately planar.

A bonding system which is similar to that involving C3, C8, C10, N11, C12 and O14 also is found in two phenyl oxazepines [44,45] (a seven-membered ring system containing nitrogen and oxygen) and the system around the C to N double bond is planar in both cases.

The non-coplanarity of the two ring systems is attributable to repulsion effects between the C9 methyl group and O14 and the tendency of a 6-coordinate iron atom to attain a symmetry which is as close as possible to octahedral.

A graphical construction of the phenyl-oxazoline moiety, drawn with the observed bond lengths, with the "idealised" bond angles:

C7-C8-C3	120°	
C8-C3-O2	120°	
O14-C10-N11	114°	(reference 43)
C8-C10-N11	123°	
C8-C10-O14	123°	

and assuming the two rings to be coplanar gives C9-O14 and O2-N11 distances of 2.5 and 2.55Å respectively, both of which are considerably shorter than the sum of the relevant van der Waals radii (3.22 and 3.07Å respectively)[46] and imply severe repulsion forces.

From a series of phenyl-isoxazolines it is clear that the torsion angle about the bridge between the two rings is, as expected, related to the nature of the atoms alpha to the bridge. In 3-hydroxy-5-phenyl isoxazole, where the alpha positions are occupied by hydrogen atoms, the two rings are almost coplanar (torsion angles of 2.4° in the α form[47] and 1.6° in the β form[48]). In N-methyl-4-phenyl-isoxazolin-5-one[49], where one alpha position is occupied by a carbonyl oxygen atom, the torsion angle is 14°. (This is probably considerably reduced from the value which might be expected from interatomic repulsion forces alone by conjugation through the majority of the molecule). In N-methyl-3-phenyl-4-bromo-oxazoline-3-one where one position is occupied by a methyl group, the

torsion angle between the rings is 54°.[50] One might, therefore, expect that in the present structure, which contains both methyl (C9) and hydroxyl (O2) substituents α to the ring bridge, the two rings would be virtually perpendicular to each other. That this is not the case reflects the effect of the second factor which controls the overall conformation, namely the complexing of the ferric ion to complete a six-membered chelate ring.

From the graphical construction mentioned above, a planar phenyl-oxazoline gives an O2 to N11 distance of 2.55Å, and, assuming N to Fe and O to Fe distances of 2.12 and 1.93Å respectively, (see table 6 below) this gives an N-Fe-O angle of 78°.

Rotation of the oxazoline ring about the C8-C10 bond increases the O2-N11 distance and consequently the chelate angle at the iron atom. From simple trigonometric considerations, an O2-Fe-N11 angle of 90° is achieved when the O2-N11 separation is 2.86Å and this arises when the torsion angle about the C8-N11 bond is *ca.* 35°, *ie.* is only slightly higher than the observed value. A C9-O14 separation which is greater than 3.22Å (*ie.* the van der Waals separation) is not reached until the C8-C10 torsion angle is *ca.* 70°. This corresponds to an O2-Fe-N11 angle of 104°, so that it is clear that the formation of an unstrained chelate ring takes precedence over the relief of van der Waals forces between C9 and O14.

Five of the bond lengths in the benzene ring are within three standard deviations of the normal value (1.39Å) but the C3-C8 bond is significantly long (1.49(3)Å). The ring as a whole is larger than normal (mean C-C distance 1.44Å) but this is not uncommon in complexes which contain an atomic arrangement similar to that in the present case. (In μ-oxo-bis (bis-N-n-propyl-salicylidineimato)iron(III), for example, the mean C_{arom}-C_{arom} bond length is 1.42Å).[51]

The bond lengths in the series of compounds (mostly complexes) which contain a bond system similar to that involving O2 to N11 are given in Table 8. From these data it appears that the bond lengths in the O2, C3, C8, C10, N11 system in ferri-mycobactin P are respectively shorter, longer, shorter and normal than those in the table, but that the differences are not significant.

The bond lengths and angles in the oxazoline ring are compared below with those in perkinamine[43], the only other fully described structure which contains an oxazoline ring, and with the relevant parts of the two oxazepine compounds[44,45] (the mycobactin atomic numbering scheme is used for all four compounds).

Length or angle	Mycobactin	Perkinamine[43]	Oxazepine[44]	Oxazepine[45]
C8-C10	1.40(4)	1.47(2)	1.45(3)	1.46(1)
C10-N11	1.26(3)	1.29(1)	1.25(3)	1.27(1)
N11-C12	1.53(3)	1.53(1)		
C12-C13	1.55(3)	1.59(2)		
C13-O14	1.57(3)	1.52(1)		
O14-C10	1.40(3)	1.33(1)	1.40(3)	1.43(1)
O14-C10-N11	114(2)	115(2)	127*(2)	120(1)
C10-N11-C12	114(2)	112(2)	127*(2)	124(1)
N11-C12-C13	101(2)	100(1)		
C12-C13-O14	104(2)	103(1)		
C13-O14-C10	107(2)	109(2)	118*(2)	107(1)

Both the oxazepine compounds contain a seven-membered heterocycle which makes the bond angles marked with an asterisk incompatible with those in mycobactin. With this exception, however, it is clear that the oxazoline ring in mycobactin is very similar to that in perkinamine, and that the C8-C10, C10-N11 and C10-O14 bonds in all four compounds listed above are virtually identical.

A second major feature of the ferri-mycobactin P molecule is the N^6-hydroxy-lysine derived ring (C30 to C37). This ring, together with the chain of atoms from N29 to the C18 bridgehead and further to the O44, O46-containing chelate ring is shown in figure 16. The deviations from the least-squares plane through C30, C31, C33 and C34 and the torsion angles in the ring are given in tables 6 and 7. From these data and figure 16 it is clear that the ring is in a conformation which is close to a perfect chair. The standard deviation of C30, C31, C33 and C34 from their least-squares plane is less than 0.01Å, and, apart from a 10° torsion about the N35-C37 bond, and a compensating increase in the torsion angle about the C37-C30 bond, the whole system shows a close to perfect mirror symmetry about the line from C32 to the mid point of the N35-C37 bond and indeed through the adjacent five-membered ring to the iron atom. It can be seen from figures 13 and 14 that this symmetry also applies to the bond lengths and angles in both rings.

Theoretical studies on the conformations of seven-membered rings were summarised on pages 32-34 of this thesis in connection with the seven-membered ring in the fusicoccin aglycone structure. It was concluded that, although the "twist chair" form is energetically preferable, most observed structures are in conformations which are close to the "chair" form. The torsion angles in the present seven-membered ring are compared with the theoretical values calculated by Hendrickson[60] and Bixon and Lifson[61] for the "chair" form:-

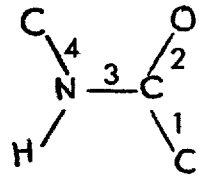
Bond	Ferri-myco P	Hendrickson	Bixon and Lifson
C32 - C33	+67°	+64°	+67°
C33 - C34	-84	-83	-90
C34 - N35	+65	+66	+55
N35 - C37	+10	0	+20
C27 - C30	-79	-66	-56
C30 - C31	+87	+83	+81
C31 - C32	-66	-64	-82

It is clear from these data that the seven-membered ring in ferri-mycobactin P is closer to Hendrickson's strictly symmetric chain than to Bixon and Lifson's slightly twisted chair.

The ferri-mycobactin P molecule contains two amide linkages, between C15 and N17 and between C27 and N29. The first is more correctly termed a peptide bond since it connects an L-serine residue to an N^6-hydroxylysine residue. The second connects the 3-hydroxy-2-methyl pentanoic acid residue (O21-O28) to the second of the N^6-hydroxylysine residues (N29-O38). The conformations of these links are shown in figures 15 and 16, where it is plain that both are in the trans conformation and are approximately planar. The torsion angles about the C-N bonds are:-

C12-C15-N17-C18	-174°
O16-C15-N17-C18	7°
C25-C27-N29-C30	163°
O28-C27-N29-C30	-8°

The bond lengths, using the nomenclature

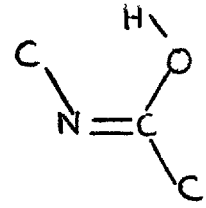


are compared with those in four typical cyclic peptides and with those in the iron transporting hydroxamate complex, ferrichrome A, (see figure 5) below:-

	Bond					
Compound	1	2	3	4	N	Reference
Cyclo sarc ala_4	1.53	1.22	1.35	1.46	5	62
Cyclo $sarc_2$	1.51	1.23	1.35	1.46	2	63
Cyclo $sarc_4$	1.53	1.23	1.36	1.46	4	64
Cyclo $sarc_8$	1.53	1.23	1.34	1.46	8	65
Ferrichrome A	1.52	1.24	1.36	1.46	7	31
Fe-Mycobactin P	1.43(3)	1.28(3)	1.53(3)	1.43(3)		
	1.43(3)	1.23(3)	1.44(3)	1.48(3)		

where N is the number of peptide links in each compound.

The lengths of bond 3 in the two amide groups in ferri-mycobactin P are significantly different, one (C15-N17) being the normal partial double bond (order *ca.* 1.6) and the second (C27-N29) virtually a single bond. This effect is also apparent in the lengths of the carbonyl bonds (Number 2 above) where the C27-028 distance implies more double-bond character than that in the C15-016 bond. These effects suggest a smaller contribution from the "imine" form:-



The effect of this is apparent in the deviation from planarity in the two systems and in the torsion angles (see above). In the first case, where a contribution from the "imine" form is greater, the standard deviation from the least-squares plane is 0.03Å, but in the second system this value is 0.09Å.

The C12-C15 and C25-C27 bonds in ferri-mycobactin P are also significantly shorter than the equivalent bonds in the peptides shown above but it is difficult to offer an explanation for this since there are no possible resonance structures which can shorten these bonds.

The ferri-mycobactin P molecule contains an ester linkage between the hydroxyl group in the 3-hydroxy-2-methyl pentanoic acid residue and an N^6-hydroxy lysine residue. The relevant parameters which are contained in figures 13 and 14 and tables 6 ~~and 7~~, show that the system from C18 to C22 is approximately planar (σ = 0.06Å) although there is a torsion angle of 14° about the C19-021 bond in the 020-C19-021-C22 system. The carbonyl bond is normal[66] and the C19-021 bond (1.37(3)Å) is similar to those in the dimethyl ester of meso-tartaric acid (1.34(1)Å)[67] and in the indole compound described in chapter two of this thesis (1.33(1)Å).

The least precisely determined fragment of the ferri-mycobactin P molecule consists of the long polymethylene side chain derived from cis-octadec-2-enoic acid (see figure 1). Due to severe thermal movement the last four atoms in this chain are still unlocated, and the general geometry of the system is poor although the bond lengths are within three standard deviations of the normal value (see below).

The side chain appears in difference Fourier synthesis as an elongated electron density cloud with clear maxima for the majority of the atoms but becoming increasingly diffuse as one approaches the end of the chain, and merging into the general background after C59. C59 was located as a shallow peak with electron density $0.4e\text{Å}^{-3}$ against a background level of $0 \pm 0.2e\text{Å}^{3}$. The isotropic temperature factor (B) for this atom is 33.7Å^{2} corresponding to a root-mean-square amplitude ($\bar{U}$) of 0.41Å.

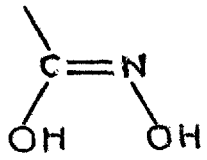
The weighted mean bond length and angle for the C_{sp^3}-C_{sp^3} bonds in the chain (*ie.* C49-C59) are 1.59(3)Å and 133(2)°, both of which are larger than is normal[68] in n-alkanes. The C57-C58 bond length is very long (2.12(14Å), but there is no evidence in difference Fourier maps to suggest that these atoms are incorrectly positioned. The length of the fatty acid side chain from C48 to C59 is 12.1Å, some 2Å shorter than that for a linear trans 12 carbon chain (assuming a C-C repeat distance of 2.54Å [68]). However, inspection of figure 9 and the torsion angles in the chain shows that it is folded back at the C49-C50 bond to lie in the cleft between the seven-membered ring and the C39-C42 methylene chain. This would result in an approximately 2.5Å decrease in the expected overall chain length of a linear C_{12} system to give a length close to that observed in the present case. It is thus highly unlikely that any of the unlocated atoms could be contained within the observed electron-density cloud.

The crystal structures of several saturated long chain hydrocarbons have been reported in the literature, *eg.* [69,70] and in all these instances the chain is in a linear "trans" conformation. This results in torsion angles about the C-C bonds of 180°, the C-C-C repeat distance mentioned above (*ie.* 2.54Å [67]) and maximisation of the hydrogen to hydrogen contact distances.

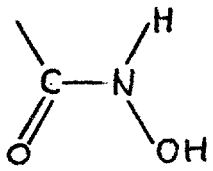
With two exceptions (C50-C51 and C57-C58), the torsion angles in the present system are close to those for staggered C-C conformation (±60° and 180°). The two exceptions (C50-C51 and C57-C58 with torsion angles of 125° and 108° respectively) are, however, close to those for the eclipsed situation (0° and±120°) and must imply some hydrogen - hydrogen repulsion. The fatty acid chain in ferri-mycobactin P is thus unusual in being non-linear and not all trans, but there are no obvious reasons for this since the side chain is packed into a loosely occupied zone in the crystal structure and does not contain any polar groups which could serve to "tie it down". The early difficulties encountered in crystallising this compound doubtless have a lot to do with finding the right conditions to "discipline" this aliphatic chain.

The molecule contains a second methylenic system comprising C39-C42. Since this group links one of the chelate rings to the C18 bridgehead atom it is conformationally restricted. In spite of this, all of the C-C bonds have torsion angles close to 60° and are thus staggered. The mean C-C bond length in this system, 1.58(3)Å, is again slightly longer than normal, but the C-C-C angles are close to the normal value.

Possibly the most important parts of the ferri-mycobactin P molecule are the hydroxamate groups comprising N35 to O38 and N43 to O46. These groups are responsible for the strong iron-chelating ability of the compound and thus its biological role. The structures of hydroxamic acids have been a subject of debate for a long time, Exner claiming that the "hydroximic" form:-



predominates in solution[71] whereas spectroscopic[35,72] and crystallographic evidence[31,32,36] provides strong support for the hydroxamic form:-



However, the problem does not arise in the mycobactins (or in the ferrichromes) since they are N-substituted and the hydroximic form is impossible. The bond lengths in the hydroxamate groups in ferric-mycobactin P are compared with those in the three other determined hydroxamates below ($Febz_3$ refers to tris-benzhydroxamato-iron(III)-trihydrate,[32] acet to acethydroxamic acid hemihydrate[36], ferri to ferrichrome A[31], and FemycP to the present compound):-

			Bond		
Compound	C=O	C-N	N-O	O_c-Fe	O_n-Fe
Acet	1.24(1)	1.51(1)	1.40(1)	-	-
$Febz_3$	1.28(1)	1.32(1)	1.37(1)	2.06(1)	1.98(1)
Ferri	1.28(2)	1.32(2)	1.38(1)	2.04(1)	1.98(1)
FemycP	1.29(3)	1.29(3)	1.33(2)	2.11(2)	2.00(2)
	1.28(3)	1.32(3)	1.37(2)	2.05(1)	2.01(2)

Comparison of the ferri-mycobactin values with those for the other iron complexes in the above table shows that they are identical to within two standard deviations. The C-N bond lengths in the iron compounds are, however, significantly different from that in acethydroxamic acid and the overall effect of the iron chelation is a more even distribution of the bond lengths with the C-N bond acquiring considerable double-bond character (order *ca.* 1.8), the N-O bond some double-bond character (order *ca.* 1.2) and the carbonyl bond less double-bond character (order *ca.* 1.7).

4) The Intra-molecular Hydrogen Bonds

In the general description of the ferri-mycobactin P structure it was stated that the molecule is "stiffened by intra-molecular hydrogen bonds which involve N17 and N29". These two atoms are in fact the only hydrogen bonding donor atoms in the molecule and although the structure contains 13 oxygen or nitrogen atoms which could act as acceptors only four of them are within potential hydrogen bonding "range". The relevant atoms and distances are:-

	Acceptor			
Donor	N11	021	038	044
N17	2.82*	2.87*	3.08*	3.00
N29	>4.5	2.84	2.57	>4.5

Examination of a model of the structure shows that the contacts marked with an asterisk are geometrically unfavourable for hydrogen bonding. The most likely hydrogen bonds are therefore between amide nitrogen N17 and hydroxylic oxygen atom 044, and between amide nitrogen N29 and ester oxygen 021.

5) The Iron Coordination

Figure 10b illustrates the coordination of the ferric ion by the phenyl-oxazoline system and the two hydroxamate groups in the molecule to form one six-membered and two five-membered chelate rings. The best planes and the torsion angles in these rings, given in tables 6 and 7, show that both of the five-membered rings are in a shallow "envelope" confirmation with the iron atom 0.21Å out of the plane through the 036 - 038 hydroxamate system and 0.08Å out of the plane through the 044 - 046 hydroxamate system. The geometry of the six-membered ring is best defined by a plane about the C10-N11 double bond and including the iron atom. Due to the non-coplanarity of the phenyl and oxazoline rings (see above) C3 and 02 lie 0.33 and 0.60Å out of this plane so that the ring as a whole is non-planar.

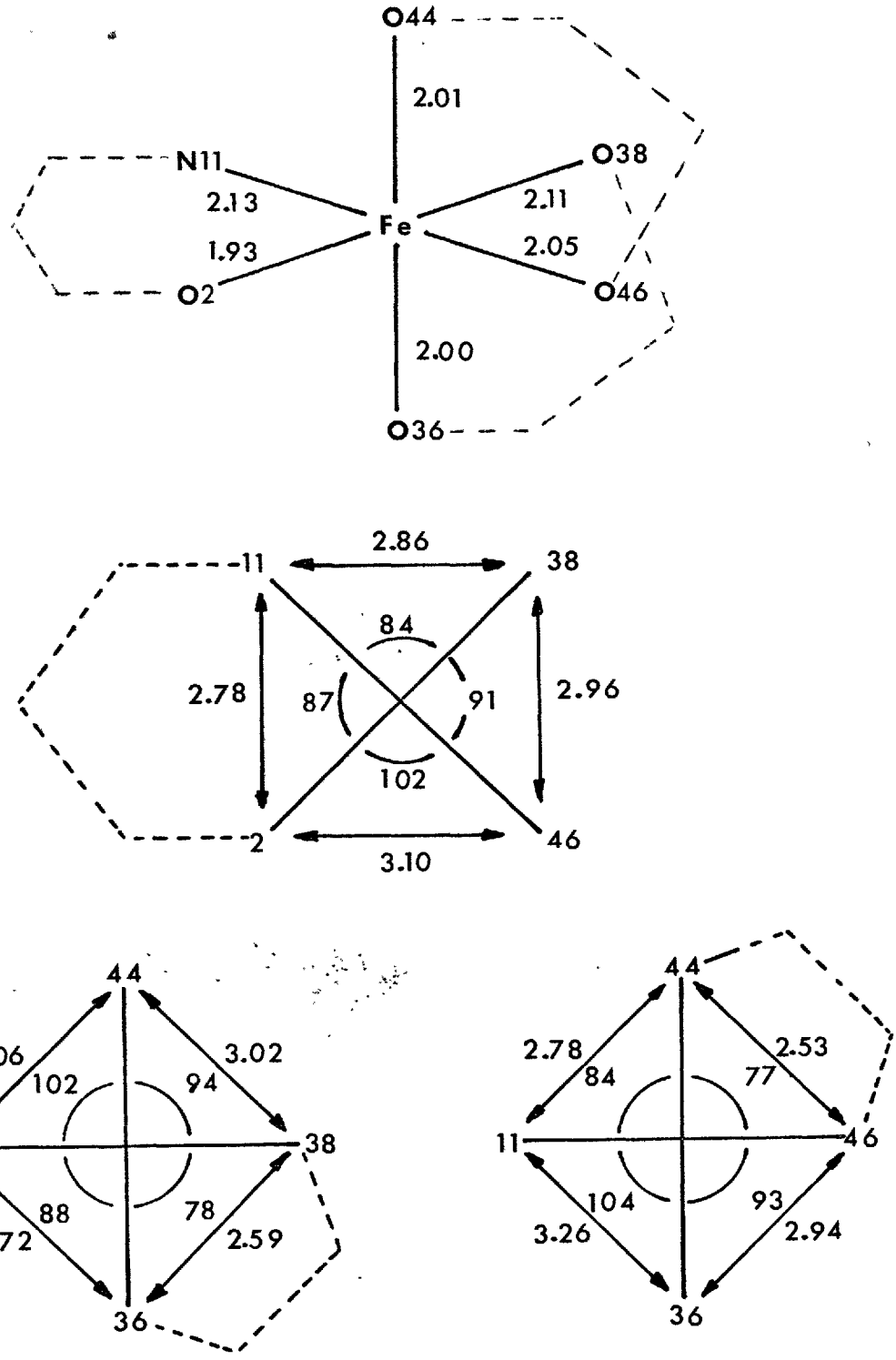


Figure 17 Geometry of the iron coordination sphere

The bond lengths, ligand - ligand distances and the bond angles around the metal atom are shown in figure 17. The most important aspects of the coordination geometry are (I), that all three trans ligand-Fe-ligand angles are considerably larger than 180° on the same side of the coordination octahedron and (II), that two of the largest ligand-ligand separations (O2-O44 and O2-O46) and the associated angles (*ie.* O2-Fe-O44 and O2-Fe-O46) lie adjacent to one another. As a consequence of these effects the iron atom is thrust away from the "centre of gravity" of the ligand atoms to lie exposed in a cleft in the coordination sphere flanked by O2, O44 and O46. Since O44 and O46 are members of the same hydroxamate group, the effect may also be described as a "bending back" of this chelate ring. The ring is connected to the remainder of the molecule by the C39-C42 methylene chain and by the probable N17 - O44 hydrogen bond and both of these pull it away from O2.

The O-Fe-O angles in the hydroxamate rings (77 and 78°) are the smallest in the system but are typical for five-membered chelate rings. (see Table 9). In both cases the angle which is trans to the chelate ring is large (102 and 104° - see figure 17) so that the sum of the angles is *ca.* 180°. This effect is, in part, responsible for the large O2-O44 distance mentioned above. The angle at the iron atom in the six-membered ring is again typical (see table 9), but the smallest angle in the coordination plane containing this ring is N11-Fe-O38 (84°). This is probably attributable to the N29-O21 hydrogen bond.

Table 9 contains the Fe-O and Fe-N distances and the chelate angle in a series of iron(III) complexes. (The compounds marked with an asterisk contain chelate rings which are similar to that involving the phenyl-oxazoline system in ferri-mycobactin P and they are therefore also shown in table 8.) It is clear from these data that the coordination distances and chelate angles in ferrimycobactin P are entirely typical.

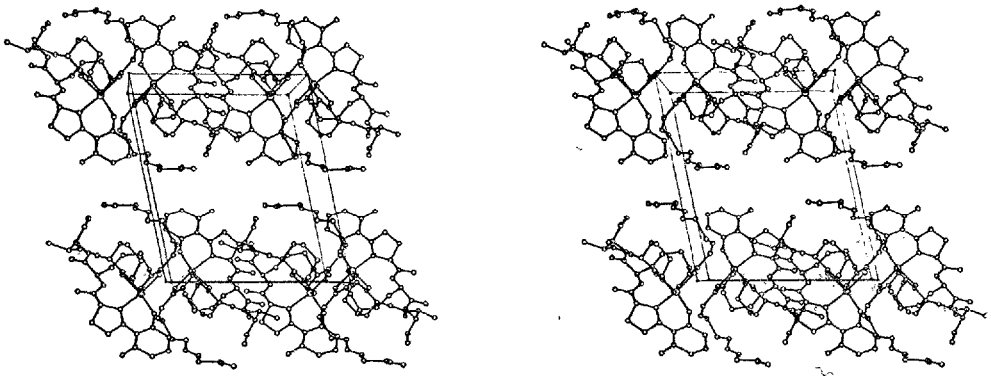


Figure 18 The crystal structure

The sums of the trans bond lengths are:-

O2.....O38	4.04Å
N11....O46	4.18Å
O36....O44	4.01Å

so that the two oxygen-oxygen trans distances are similar in spite of the short O2-Fe bond. The longer N11-O46 sum, however, merely reflects the longer nitrogen to iron coordination distance.

6) The Crystal Structure

Figure 18 shows a stereoscopic drawing of the crystal structure and figure 19 the arrangement of the molecules along the screw axes. The structure consists of layers of molecules lying parallel to the 1 0 0 planes which are separated by a very loosely occupied zone which bisects the unit-cell in the a direction. This region is occupied almost exclusively by hydrocarbon residues with high temperature factors. The intermolecular contact distances across the zone are all greater than 4.0Å so that the intermolecular forces which hold the layers together are very weak. This explains the ready cleavage of the crystals in the {1 0 0} planes.

The molecules are held together in the screw-axis direction (*ie.* b) by a number of van der Waals contacts but all of the distances are greater than the sums of the appropriate van der Waals radii and must again be weak. The only intermolecular distance which is significantly shorter than a normal van der Waals contact (3.22Å [46]) is 3.06(3)Å between C12 and O20 in molecules which are related by a screw operation combined with a one unit-cell translation in the c direction. The same operation also produces short contacts between C13 and O28 (3.24(2)Å) and between O14 and O28 (3.30(2)Å).

From the observations above it is not surprising that it was difficult to obtain useable single crystals of ferri-mycobactin P. The melting point of the compound, 199-201°C, is unexpectedly high but this is probably due to the large overall size and mass of the molecule.

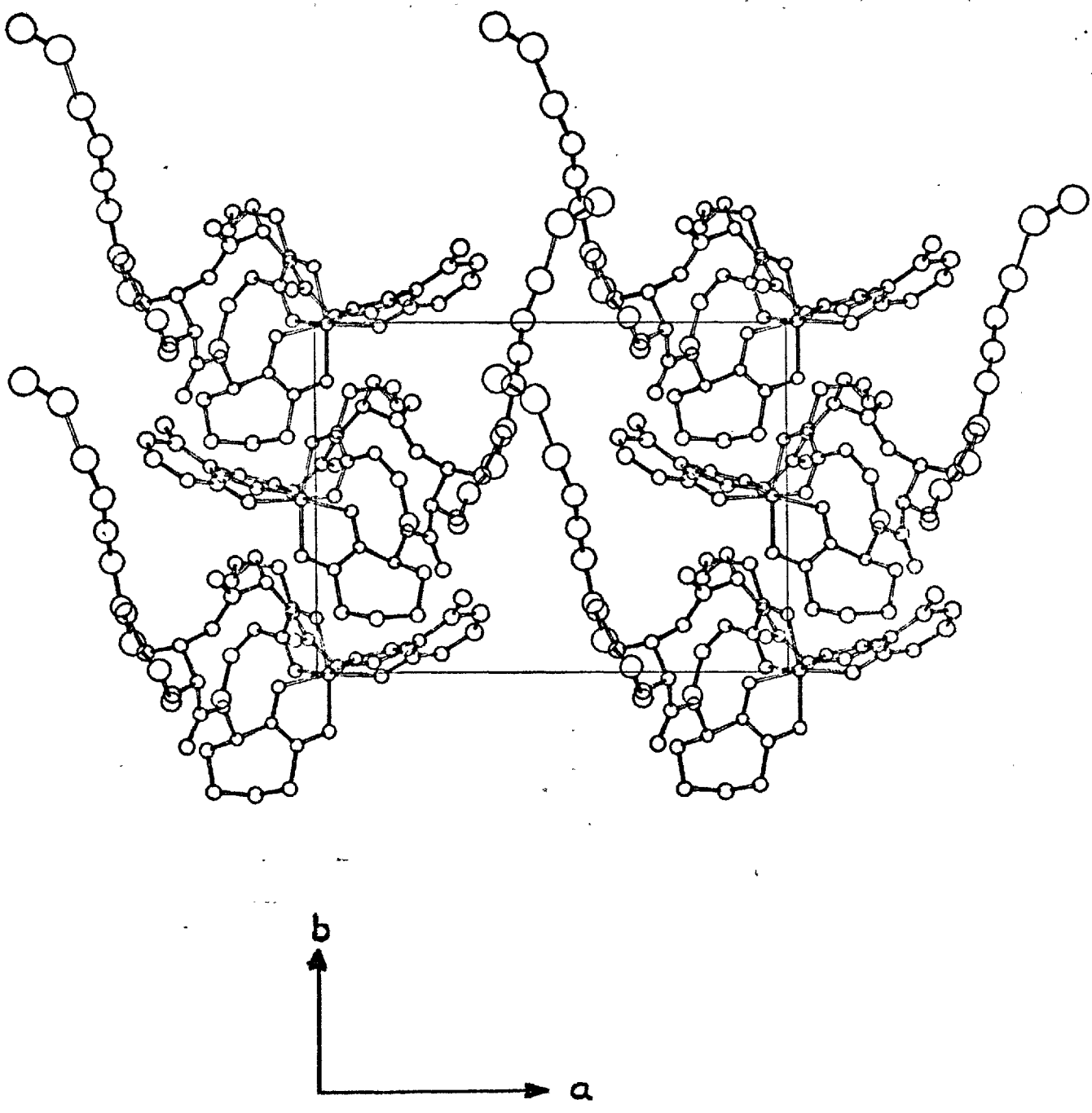


Figure 19

Table 3 Atomic positions (fractional coordinates x 10^4)

Standard deviations in parentheses.

Atom	x	y	z
Fe1	567(2)	0	1755(2)
O2	1591(8)	-67(14)	1206(10)
C3	2285(17)	473(22)	1434(19)
C4	2782(15)	550(20)	666(17)
C5	3514(19)	1172(27)	857(22)
C6	3810(18)	1705(27)	1851(23)
C7	3341(18)	1617(26)	2678(22)
C8	2609(13)	943(20)	2473(15)
C9	3728(19)	2128(26)	3613(23)
C10	2125(16)	678(21)	3211(17)
N11	1362(11)	427(13)	3180(11)
C12	1146(13)	211(22)	4235(15)
C13	2032(13)	309(19)	4947(16)
O14	2622(9)	690(13)	4206(11)
C15	579(13)	999(19)	4521(16)
O16	448(9)	966(13)	5438(11)
N17	158(10)	1765(15)	3878(12)
C18	-355(14)	2623(20)	4158(16)
C19	-989(18)	2239(25)	4703(22)
O20	-1199(12)	2723(17)	5440(15)
O21	-1468(9)	1328(13)	4374(22)
C22	-2032(15)	832(21)	5013(17)
C23	-2872(18)	630(27)	4306(21)
C24	-3325(27)	1757(42)	4045(33)
C25	-1607(13)	-205(19)	5446(16)
C26	-1971(16)	-559(24)	6471(20)
C27	-1675(14)	-1117(21)	4764(17)
O28	-1937(9)	-2049(15)	4862(11)
N29	-1339(10)	-920(15)	3855(12)
C30	-1161(12)	-1843(17)	3205(14)
C31	-1921(15)	-2144(21)	2427(17)
C32	-1839(15)	-3328(21)	2030(17)

Table 3 (continued)

Atom	x	y	z
C33	-1071(16)	-3412(24)	1389(19)
C34	-222(14)	-3234(20)	2092(17)
N35	-51(10)	-2019(14)	2202(12)
O36	563(8)	-1638(13)	1754(9)
C37	-514(12)	-1387(18)	2639(15)
O38	-349(8)	-353(10)	2648(10)
C39	-740(15)	3330(20)	3160(18)
C40	-1382(14)	2663(20)	2373(17)
C41	-1473(15)	3175(22)	1254(18)
C42	-616(14)	3050(21)	865(17)
N43	-349(10)	1893(14)	866(13)
O44	299(9)	1613(11)	1660(10)
C45	-651(14)	1065(22)	261(17)
O46	-357(8)	96(15)	455(9)
C47	-1392(13)	1295(19)	-632(16)
C48	-1946(17)	541(24)	-958(20)
C49	-2113(23)	-684(36)	-693(27)
C50	-3102(22)	-757(37)	-355(27)
C51	-3206(32)	140(54)	454(37)
C52	-3758(49)	915(80)	492(56)
C53	-3928(29)	1975(47)	953(36)
C54	-4095(37)	3226(57)	1046(44)
C55	-4089(32)	4117(57)	1859(41)
C56	-4048(35)	5059(68)	2529(44)
C57	-4353(42)	6165(73)	2954(51)
C58	-4787(55)	7793(85)	3081(64)
C59	-5575(51)	8398(70)	3214(58)

Table 4a Anisotropic temperature factors x 10^4.

Standard deviations in parentheses.

The temperature factor has the form:

$$\exp[-2\pi^2(u_{11}a^{*2}\underline{h}^2+u_{22}b^{*2}\underline{k}^2+u_{33}c^{*2}\underline{\ell}^2+u_{12}a^*b^*\underline{hk}+u_{13}a^*b^*\underline{h\ell}+u_{23}b^*c^*\underline{k\ell})]$$

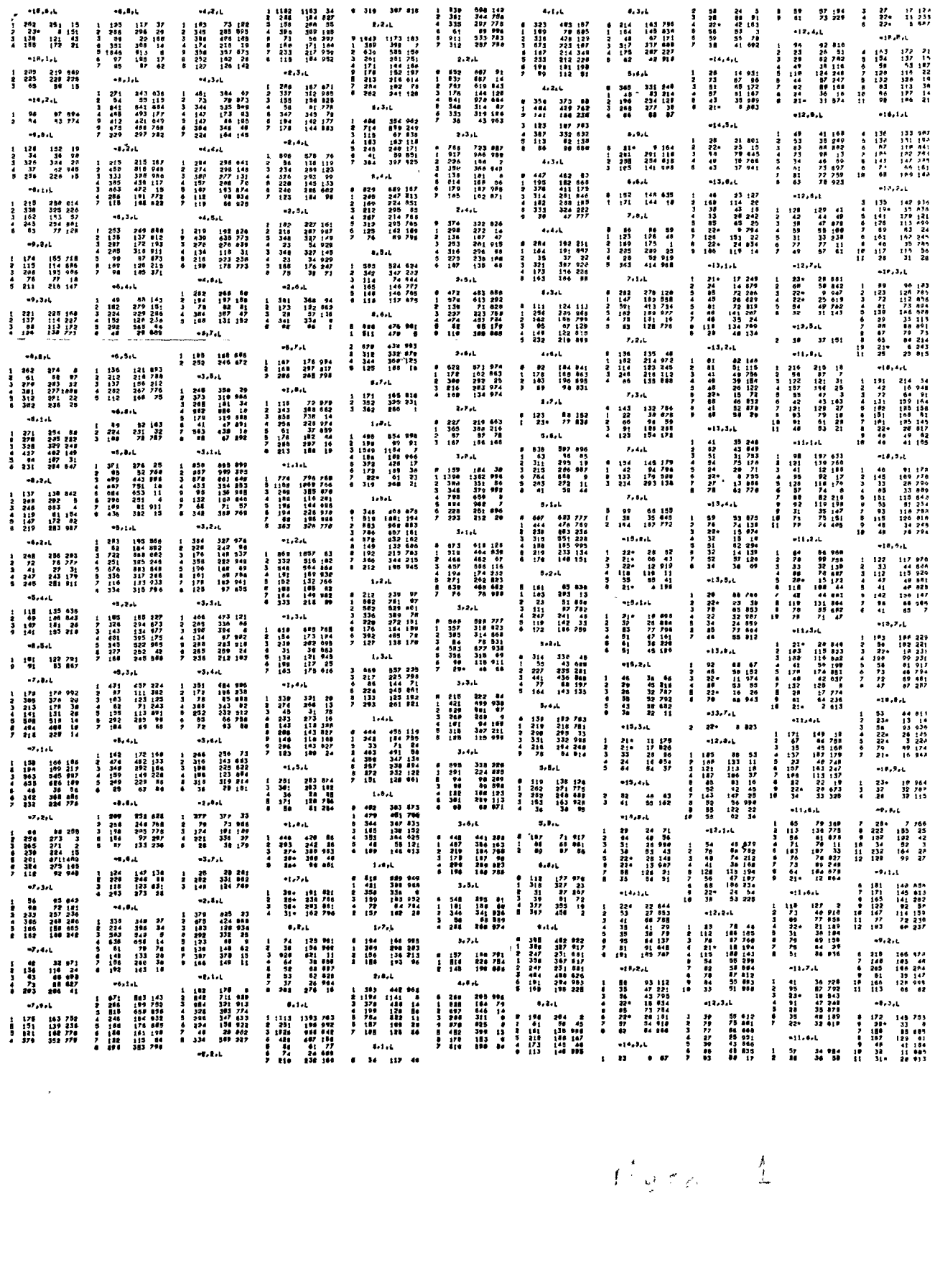
Atom	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
FE	106(0)	70(0)	81(0)	-6(0)	39(0)	-7(0)
02	88(0)	97(0)	120(0)	-17(0)	44(0)	1(0)
N11	111(0)	74(0)	64(0)	-7(0)	18(0)	-3(0)
036	84(0)	108(0)	64(0)	-8(0)	38(0)	-6(0)
038	104(0)	60(0)	83(0)	-5(0)	59(0)	0(0)
044	95(0)	64(0)	80(0)	-13(0)	39(0)	-10(0)
046	101(0)	80(0)	87(0)	-8(0)	22(0)	-22(0)

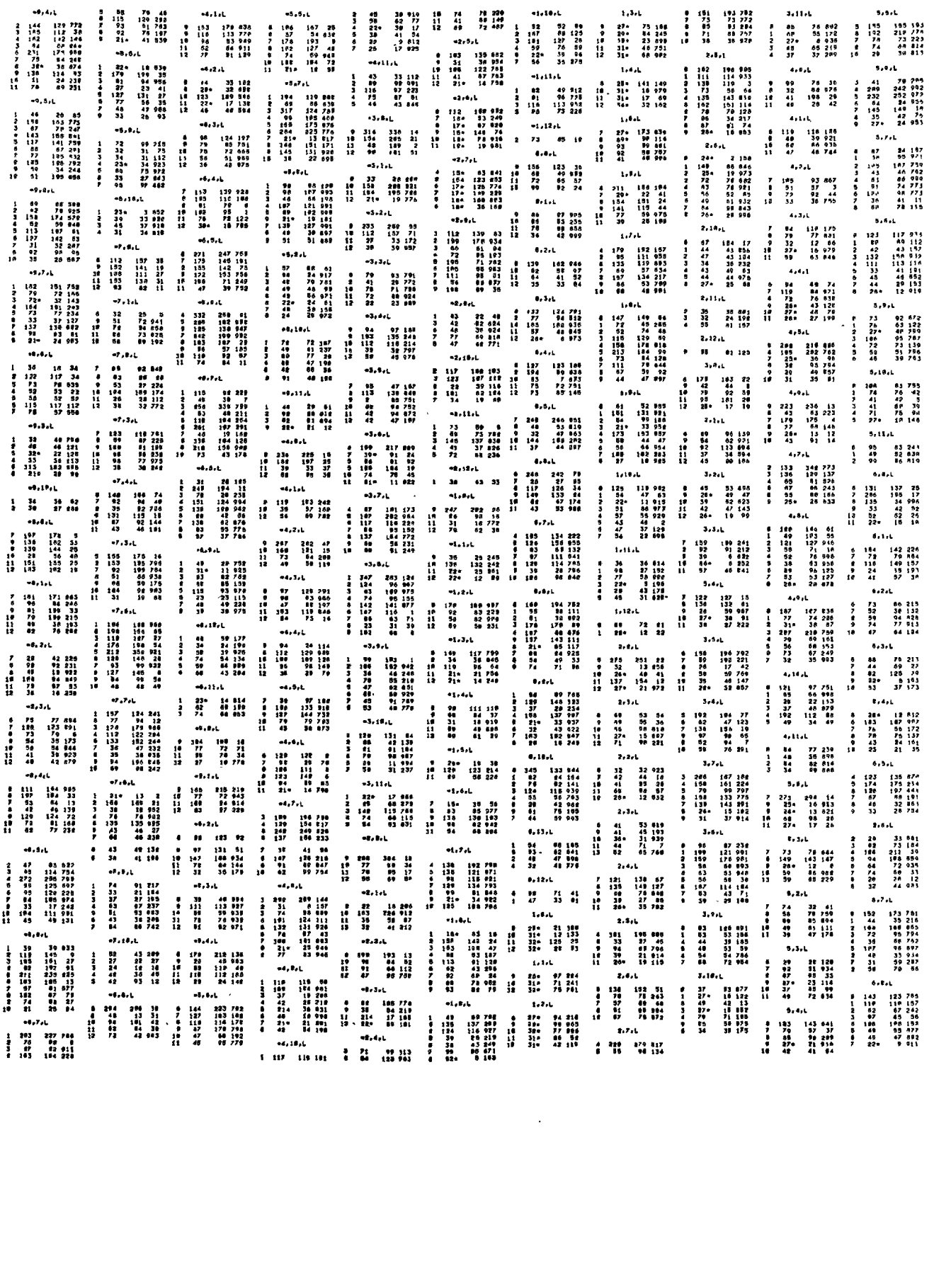
Table 4b Isotropic temperature factors.

Atom	B(Å^2)	Atom	B(Å^2)
C3	8.99	C30	5.95
C4	8.22	C31	7.92
C5	11.51	C32	8.27
C6	11.07	C33	9.38
C7	10.07	C34	7.76
C8	6.96	N35	6.41
C9	11.21	C37	5.46
C10	7.83	C39	8.07
C12	7.50	C40	7.58
C13	7.54	C41	8.32
O14	7.88	C42	7.96
C15	6.84	N43	6.16
O16	7.71	C45	7.23
N17	6.94	C47	6.99
C18	7.38	C48	9.37
C19	9.99	C49	14.28
O20	10.77	C50	14.05
O21	8.19	C51	19.24
C22	7.91	C52	26.18
C23	10.87	C53	17.11
C24	18.00	C54	22.63
C25	7.57	C55	20.96
C26	10.01	C56	21.35
C27	7.55	C57	27.84
O28	8.43	C58	33.26
N29	6.70	C59	33.69

TABLE 5 Comparison of observed and calculated structure amplitudes for Ferri-Mycobactin P

The data are listed in groups of constant h and k and consist of ℓ, $10|F_o|$, $10|F_c|$ and the phase in millicycles. Reflections marked "*" were classified as unobserved.





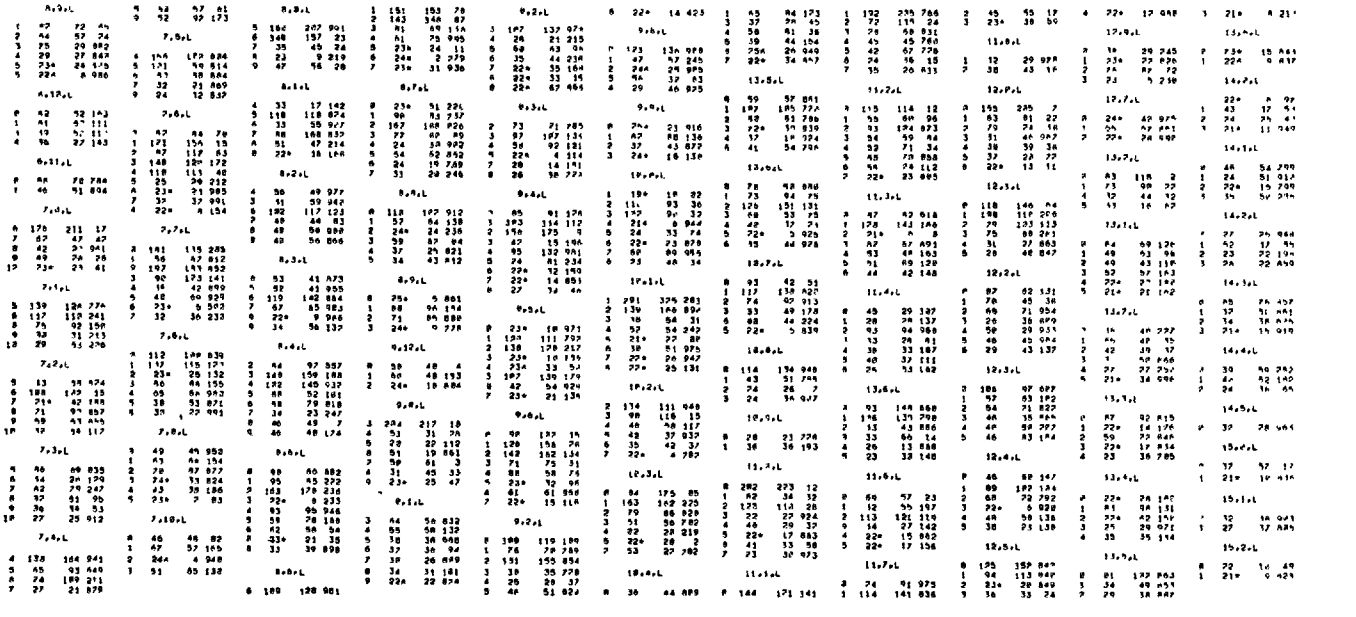


Table 6. Least-squares planes in the structure

Planes are expressed in the form $Px + Qy + Rz - S = 0$; x, y, z being the fractional coordinates of Table 3. Deviations of the atoms from the planes are tabulated in $\text{Å} \times 10^3$. Atoms marked with an asterisk were not included in the calculation of the planes. The value shown with these atoms is their deviation from the plane through the other atoms.

1) The phenyl ring

P = -7.512	Q = 10.010	R = -3.028	S = -1.721
C3	44	C6	5
C4	-20	C7	20
C5	-5	C8	-43

2) The oxazoline ring

P = -3.802	Q = 11.759	R = 2.127	S = .680
C10	-7	C13	-37
N11	-19	O14	29
C12	33	C8*	-36

3) The C10 - N11 double-bond system

P = -3.839	Q = 11.792	R = 1.723	S = .535
C8	2	C12	4
C10	2	O14	-3
N11	-6	C13*	-98

4) In the seven-membered ring

P = -4.531	Q = -9.517	R = 7.999	S = 4.848
C30	-4	C33	-4
C31	4	C34	4

5) The ester system

P = 9.740	Q = -6.237	R = 6.333	S = .607
C18	44	021	-96
C19	11	C22	69
020	-29		

---

6) The chelate rings

a)

P = 8.543	Q = -.580	R = 9.612	S = 2.214
Fe1	-42	038	54
N35	-24	C30*	-18
036	48	C34*	-205
C37	-36		

---

b)

P = 13.077	Q = 1.889	R = -9.418	S = .893
Fe1	-18	046	16
N43	-21	C42*	-151
044	26	C47*	-87
C45	-3		

---

c)

P = -5.056	Q = 11.527	R = -.243	S = -.316
Fe1	-13	02*	-594
C8	24	C3*	-329
C10	-54	C12*	-127
N11	43	014*	-316

Table 7 The torsion angles (degrees)

1) The phenyl ring

C3 - C4 - C5 - C6	3.2
C4 - C5 - C6 - C7	-.9
C5 - C6 - C7 - C8	2.9
C6 - C7 - C8 - C3	-7.4
C7 - C8 - C3 - C4	10.0
C8 - C3 - C4 - C5	-7.5

2) The oxazoline ring

N11 - C12 - C13 - O14	5.7
C12 - C13 - O14 - C10	-5.6
C13 - O14 - C10 - N11	3.1
O14 - C10 - N11 - C12	1.0
C10 - N11 - C12 - C13	-4.5

3) The polymethylene side chain

C47 - C48 - C49 - C50	120.3
C48 - C49 - C50 - C51	-48.9
C49 - C50 - C51 - C52	124.8
C50 - C51 - C52 - C53	-157.8
C51 - C52 - C53 - C54	108.6
C52 - C53 - C54 - C55	-163.5
C53 - C54 - C55 - C56	147.5
C54 - C55 - C56 - C57	78.2
C55 - C56 - C57 - C58	-53.9
C56 - C57 - C58 - C59	107.7

4) The C39 - C42 system

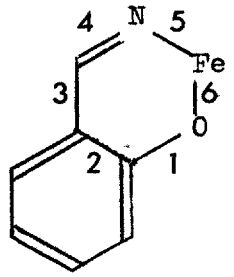
C18 - C39 - C40 - C41	-157.6
C39 - C40 - C41 - C42	65.9
C40 - C41 - C42 - N43	57.1

5) The chelate rings

a)	Fe1 - O2 - C3 - C8	28.0
	O2 - C3 - C8 - C10	4.5
	C3 - C8 - C10 - N11	-26.2
	C8 - C10 - N11 - Fe1	13.3
	C10 - N11 - Fe1 - O2	11.7
	N11 - Fe1 - O2 - C3	-31.9
b)		
	Fe1 - O38 - C37 - N35	7.0
	O38 - C37 - N35 - O36	-1.4
	C37 - N35 - O36 - Fe1	-5.3
	N35 - O36 - Fe1 - O38	6.4
	O36 - Fe1 - O38 - C37	-7.4
c)		
	Fe1 - O44 - N43 - C45	-4.0
	O44 - N43 - C45 - O46	2.0
	N43 - C45 - O46 - Fe1	1.0
	C45 - O46 - Fe1 - O44	-2.4
	O46 - Fe1 - O44 - N43	3.4

Table 8

Bond nomenclature :-



Compound	Bond 1	2	3	4	5	6		Ref.
A	1.33	1.41	1.44	1.29	2.10	1.93	0+03	51
B	1.28	1.42	1.48	1.31	2.14	1.92	0.03	52
C	1.31	1.41	1.51	1.33	2.09	1.94	0.02	53
D	1.33	1.44	1.44	1.26			0.02	54
E	1.36	1.40	1.45	1.25			0.02	55
F	1.33	1.43	1.44	1.28			0.01	56
G	1.33	1.42	1.42	1.34			0.01	57
H	1.33	1.42	1.43	1.29			0.02	58
I			1.47	1.29			0.01	43
J		1.42	1.45	1.26			0.03	44
K		1.40	1.46	1.27			0.01	45
Fe-myc P	1.28	1.49	1.40	1.26	2.13	1.92	0.03	

Compound identities :-

A μ-oxo-bis(NN'-ethylenebis(salicylideneimato)-iron(III)) . CH_2Cl_2.

B μ-oxo-bis(bis-N-n-propylsalicylideneimato)-iron(III) .

C Chloro-bis-(N-n-propyl-salicylaldiminato)-iron(III) .

D bis-(5-chloro-salicylaldoximato)-copper(II) .

E bis-(salicylaldoximato)-copper(II) .

F N-(carbamoyl-methyl)-salicylideneiminato-aquo-copper(II)-sulphate.H_2O.

G bis-salicylaldoximato-nickel(II) .

H bis-salicylaldoximato-palladium(II) .

I Perkinamine

J 2-phenyl-7-bromo-benz[d]-[1,3]-oxazepine .

K 2,4,5,7-tetra-phenyl-6-(4-bromo-phenyl)-1,3-oxazepine.

Table 9 Fe - O and Fe - N coordinate bond lengths (Å) and chelate angles (degrees) in some iron(III) complexes.

Compound	O - Fe	N - Fe	Fe	Ring size	Ref.
A	2.10(1)	1.92(1)	87	6	51
B	1.93(1)	2.14(2)		6	52
C	1.89(2)	2.09(2)	88(1)	6	53
D	1.95(1)			6	54
E	1.86(2)	2.09(2)	87.5(7)	6	73
F	1.99		87	6	74
G	2.01		77.8	5	75
H	1.97(1)	2.21(1)		6	76
I	2.00		75.6	5	77
J	1.99		87.1	6	78
K	2.02		78.7	5	32
L	2.01		78.0	5	31
Fe-myc P	1.92(2)	2.13(2)	78/87	5/6	
Mean	1.97	2.10	78/87	5/6	

Compound identities :-

A - D as in table 8

E Chloro-(NN -bis-salicylidene-ethylenediamine)-iron(III) .

F Tris-acetyl-acetonato-iron(III)-$AgC\ell O_4$ adduct .

G Tris-(tropolonato)-iron(III) .

H $(enH_2)[(HEDTA)FeOFe(HEDTA)].6H_2O$.

I Iron cupferron .

J Tris-acetyl-acetonato-iron(III) .

K Tris-benzhydroxamato-iron(III)-trihydrate .

L Chlorobispentane-2,4-dianato-iron(III) .

References

1. G. Sykes and F.A. Skinner, Actinomycetales, Soc. for Appl. Bacterial, Symp. Ser. 2, 1973, Acad. Press.
2. D. Gottlieb, in reference 1.
3. F.W. Twort and G.L.Y. Ingram, Proc. Roy. Soc. Ser. B, Biol. Sci., 1912, 84, 517.
4. F.W. Twort and G.L.Y. Ingram, Zentralblatt Bakteriol. Parasit. Infektionskr. Abt. Orig., 1914, 73, 277.
5. G.A. Snow, Bact. Rev., (1970), 99.
6. J. Francis, H.M. Macturk, J. Madinaveita, and G.A. Snow, Nature, 1949, 463, 365.
7. J. Francis, H.M. Macturk, Biochem. J., 1953, 55, 596.
8. G.A. Snow, J. Chem. Soc. (1954), 4080.
9. G.A. Snow, Biochem. J., (1965), 97, 166.
10. G.A. Snow, Biochem. J., (1969), 115, 199.
11. G.A. Snow, Biochem. J., (1965), 94, 160.
12. J.B. Nielands, Structure and Bonding, 1966, 1, 59.
13. J.B. Nielands, Structure and Bonding, 1972, 11, 145.
14. G. Anderegg, F. L'Eplattenier and G. Schwarzenbach, Helv. Chim. Acta, 963, 46, 1400.
15. H. Bickel, E. Gaumann, W. Keller-Schierlein, V. Prelog, E. Vischer, A, Wettstein and H. Zahner, Experientia, 1960, 16, 129.
16. C. Ratledge, Biochem. Biophys. Res. Commun., 1971, 45, 856.
17. C. Ratledge and B.J. Marshall, Biochim. Biophys. Acta, (1972), 279, 58.
18. S. Kotani, T. Kitaura, T. Hirano, and A. Tanaka, Biken's J. 1959, 2, 129.
19. M. Barbier, These Universite de Paris - see ref. 16.
20. C. Ratledge and M.J. Hall, J. Gen. Microbiol., 1972, 72, 143.
21. R.J. Porra and O.T.G. Jones, Biochem. J., 1963, 87, 186.
22. C. Ratledge and G.A. Snow, Biochem. J., 1974, 139, 407.

23. F.H. Allen and D. Rogers, personal commun., 1969.

24. F.H. Allen, D. Rogers and P.G.H. Troughton, Acta Cryst., 1971, B27, 1325.

25. J. Karle and I.L. Karle, Acta Cryst., 1966, 21, 849.

26. J. Karle in Crystallographic Computing, ed. F. Ahmed, Munksgaard, Copenhagen, 1969.

27. D. Dale, D.C. Hodgkin and K. Venkatesan, Crystallography and Crystal Perfection, ed. G.N. Ramachandvan, (1963), Acad. Press.

28. R. Bonnett, J.M. Godfrey, V.B. Math, E. Edmond, H. Evans, and O.J.R. Hodder, Nature, (1971) 229, 473.

29. International Tables -or X-ray Crystallography, Vol. III, The Kynoch Press, Birmingham, 1962.

30. J.B. Hendrickson, J. Amer. Chem. Soc., 1962, 84, 4537.

31. A. Zalkin, J.D. Forrester and D.H. Templeton, J. Amer. Chem. Soc., 1966, 88, 1810.

32. H.J. Lindner and S. Gottlicher, Acta Cryst., 1969, B25, 832.

33. C.E. Lankford, CRC Crit. Rev. Microbiol., 1973, 2, 273.

34. D. Greatbanks and G.R. Bedford, Biochem. J., 1969, 115, 1047.

35. D. Hadzi and D. Prevorsek, Spectrochim. Acta, 1957, 10, 38.

36. B.H. Bracher and R.W.H. Small, Acta Cryst., 1970, B26, 1705.

37. A. Agren, Srensk kem Tidskr., 1956, 68, 189.

38. M.V. Park, J. Chem. Soc. (A), 1966, 816.

39. K. Ogawa and N. Tobe, Bull. Chem. Soc. Jap., 1966, 39, 227.

40. W.A.E. McBryde, Canad. J. Chem., 1968, 46, 2385.

41. International Tables for X-ray Crystallography, Vol. II, The Kynoch Press, Birmingham, 1962.

42. In reference 41.

43. B. Rerat and C. Rerat, Acta Cryst., 1964, 17, 1119.

44. B. Jensen, Acta Cryst., 1972, B28, 771.

45. B. Jensen, Acta Cryst., 1972, B28, 774.

46. A. Bondi, J. Phys. Chem., 1964, 68, 441.

47. S. Biagini, M. Cannas and Marongiu, J. Heterocycl. Chem., 1969, 6, 901.

48. S. Biagini, M. Cannas and Marongiu, Acta Cryst., 1969, B25, 2108.

49. C. Sabelli and P.F. Zanazzi, Acta Cryst., 1969, B25, 182.

50. C. Sabelli and P.F. Zanazzi, Acta Cryst., 1969, B25, 193.

51. P. Coggon, A.T. McPhail, P.M. Gross, F.E. Mabbs, and V.N. McLachlan, J. Chem. Soc. (A), 1971, 1014.

52. J.E. Davies and B.M. Gatehouse, Cryst., Struct. Commun., 1972, 1, 115.

53. J.E. Davies and B.M. Gatehouse, Acta Cryst., 1972, B28, 3541.

54. P.L. Orioli, E.C. Lingafelter and B.W. Brown, Acta Cryst., 1964, 17, 113.

55. M.A. Jarski and E.C. Lingafelter, Acta Cryst., 1964, 17, 1109.

56. H. Tamura and K. Ogawa, Cryst., Struct. Commun., 1973, 2, 103.

57. R.C. Srivastava, E.C. Lingafelter and P.C. Jain, Acta Cryst., 1967, 22, 922.

58. C.E. Pfluger, R.L. Harlow and S.H. Simonsen, Acta Cryst., 1970, B26, 1631.

59. F. Bachechi and Zambonelli, Acta Cryst., 1972, B28, 2489.

60. J.B. Hendrickson, J. Amer. Chem. Soc., 1964, 86, 4854.

61. M. Bixon and S. Lifson, Tetrahedron, 1967, 23, 769.

62. P. Groth, Acta Chem. Scand., 1974, A28, 449.

63. P. Groth, Acta Chem. Scand., 1969, 23, 3155.

64. P. Groth, Acta Chem. Scand., 1970, 24, 780.

65. P. Groth, Acta Chem. Scand., 1973, 27, 3217.

66. L.E. Sutton, Interatomic Distances, (Suppl.), 1964, The Chemical Society, London.

67. J. Kroon and J.A. Kanters, Acta Cryst., 1973, B29, 1278.

68. H. Mathisen, N. Norman and B.F. Petersen, Acta Chem. Scand., 1967, 21, 135.

69. S.C. Nyburg and J.A. Potworowski, Acta Cryst., 1973, B29, 347.

70. P.W. Teare Acta Cryst., (1959), 12, 294.

71. O. Exner and W. Simon, Collection Czech. Chem. Commun., (1965), 30, 4078.

72. P.A.S. Smith, Open-Chain Nitrogen Compounds, (1966), Benjamin, New York.

73. M. Gerloch and F.E. Mabbs, J. Chem. Soc. (A), 1967, 1900.

74. L.R. Nassimbeni and M.M. Thackeray, Inorg. Nucl. Chem. Lett., 1973, 9, 539.

75. T.A. Hamor and J.D. Watkin, Chem. Commun., 1969, 440.

76. S.J. Lippard, H. Schugar and C. Walling, Inorg. Chem., 1967, 6, 1825.

77. D. van der Helm, L.L. Merritt, R. Degeilh and C. McGillavry, Acta Cryst., 1965, 18, 355.

78. J. Iball and C.H. Morgan, Acta Cryst., 1967, 23, 239.

APPENDIX

Phase determination using anomalous dispersion effects.

The absorption of X-rays by a specimen increases rapidly as the X-ray wavelength approaches the absorption edge of the element in the specimen. As a consequence of this the specimen emits large amounts of incoherently scattered radiation at its characteristic wavelength. This phenomenon is known as X-ray fluorescence or anomalous dispersion and is generally a nuisance in single crystal diffraction work since it results in a high background intensity.

In the presence of anomalous dispersion the normal atomic scattering factor f^0 requires the two wavelength dependent correction factors $\Delta f'$ and $\Delta f''$ such that :-

$$f = f^0 + \Delta f' + i\Delta f''$$

When anomalous scattering exists with a finite $\Delta f''$ component the contribution arising from this component is always $\pi/2$ ahead of the real components and is independent of the choice of axes or orientation of the scattering atoms. As a consequence, the intensities of reflections $\underline{hk\ell}$ and the inverse $\underline{\bar{h}\bar{k}\bar{\ell}}$ become unequal for a non-centrosymmetric crystal and Friedel's law is violated. This is shown on an Argand diagram below :-

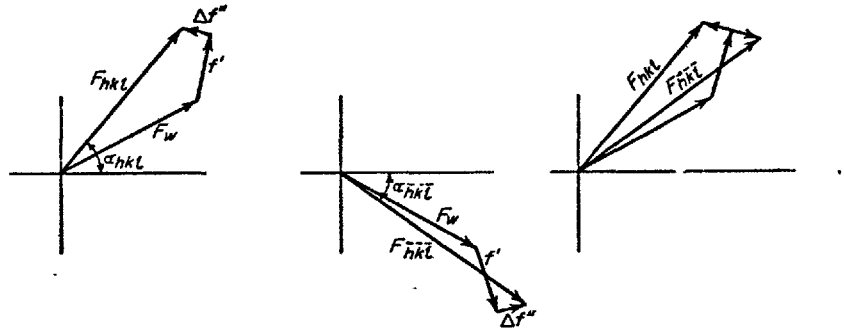


Figure 1 Vector representation showing anomalous dispersion effects for a Friedel pair in a non-centrosymmetric structure ; $|F_{\underline{hk\ell}}| \neq |F_{\underline{\bar{h}\bar{k}\bar{\ell}}}|$.

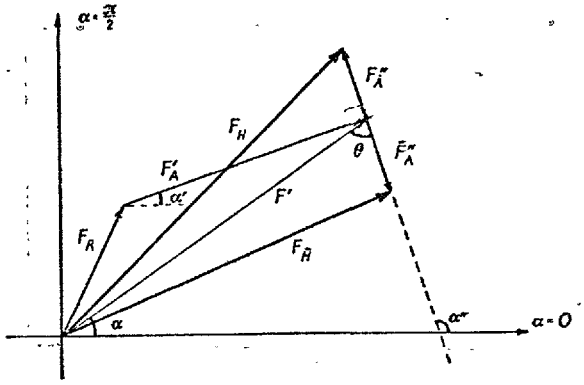
This effect was first observed by Coster, Knol and Prins[1] using a zinc sulphide crystal.

In 1949 J. M. Bijvoet[2] demonstrated that the effect can be used to

determine the absolute configuration of a molecule. This was first applied in 1951 when the absolute structure of sodium rubidium tartrate was determined[3]. Some 400 absolute structure determinations have now been carried out using the Bijvoet method.

A second and less common use of the intensity differences which arise from anomalous dispersion is in the solution of the phase problem for an unknown structure. This can be achieved by two methods, the first being the Patterson function method developed by Pepinsky and his co-workers[4] and the second, the Fourier approach of Ramachandran and Raman[5]. Since the structure of Ferri-mycobactin P was solved by the latter technique, this will be briefly described below.

The "Bijvoet - Ramachandran - Raman" method has so far been used to solve the structures of some eight compounds although several of these were small structures which were used primarily as test problems. In all the cases the heavy (i.e. fluorescent) atoms are in a centro-symmetric array but the light atoms are not. This is also the situation for ferri-mycobactin P which crystallises in the space group $\underline{P}2_1$ with one molecule (i.e. one iron atom) in the asymmetric unit. In the general case the structure factor phase diagram is :-



where F_H and $F_{\bar{H}}$ are the observed structure factors for reflections which comprise a "Bijvoet pair", F' is the mean intensity for the pair with the anomalous dispersion components removed and is given by :-

$$F' = \sqrt{\left(1/2\left(F_H^2 + F_{\bar{H}}^2\right) - F''^2_A\right)} \quad \text{- - - - - - - - - - - -} \quad 1,$$

F'_A is the real component (i.e. $f^0 + \Delta f'$) and F''_A the imaginary component of the scattering by the fluorescent atom and F_R is the total scattering by the light atoms in the structure.

The required phase angle α is given by the expression :-

$$\alpha = \alpha'' - \theta \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 2$$

where α'' is the phase of F'' and θ is the difference in phase between F''_A and F'. When only one anomalous scatterer is present :-

$$\alpha'' = \alpha' + \pi/2 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 3$$

where α' is the phase of F'_A.

The phase difference θ is given by :-

$$\cos \theta = \Delta F_o^2 / 4F' . |F''_A| \qquad \ldots\ldots\ldots\ldots\ldots \quad 4$$

ΔF_o^2 in this expression is the "Bijvoet difference" $F_H^2 - F_{\bar{H}}^2$. The solution of equation 4 automatically imposes an ambiguity in the sign of θ such that α from equations 2 and 3 becomes :-

$$\alpha = \alpha' + \pi/2 \pm \theta \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 5$$

This problem is inherent in the determination of phases when the data consist of a measured set of Bijvoet pairs and a known heavy atom position. However, in a test calculation, Dale, Hodgkin and Venkatesan[6] showed that the correct phase is more likely to be that which is closer to the heavy atom phase α'. Making this assumption they were able to locate 73 atoms in a structure which contained 78. Subsequent structure determinations making the same assumption have been equally successful.

The "extended FORTRAN" computer program "BIJVOET" was written to calculate phases for ferri-mycobactin using the principles outlined above. The program consists of apreliminary input routine which reads unit-cell, symmetry and scattering factor data. This is followed by the input of the intensity data in any order. The program then sorts the input intensity data into anomalous pairs which are written on a scratch file in a compressed format. The magnitude and phase of the heavy atom

real component (i.e. F'_A and α') are then calculated followed by magnitude of the imaginary component F''_A.

The mean structure factor without anomalous dispersion (F') is then calculated from equation 1 and this is used in equation 4 to determine θ. The two possible values of α are then calculated and that which is nearest to the heavy atom phase accepted as the phase for F'. Finally $\underline{h}$, $\underline{k}$, $\underline{\ell}$, F' and α are written onto an output file which is suitable for input into the X-RAY-67 program system for the calculation of a Fourier map.

In addition to the normal checks which are necessary to ensure that the program executes correctly, BIJVOET contains three user-controlled options. The first of these is a rejection ratio which is used in the expression :-

"If F'_A is less than RRATIO.$(F_H + F_{\overline{H}})/2$" the reflection is rejected from phase calculation. In practice a low value for this ratio is recommended since the phase α' is known accurately even if the heavy atom scattered amplitude is very small. The second option controls the type of anomalous pairs which are included in the phase calculation and makes it possible to include all paired reflections, to include pairs where one reflection is classified as unobserved or to include only those pairs where both reflection are observed. Thirdly it is possible to include unpaired zonal reflections. These are assigned the phase of the heavy atom component.

References

1. D. Coster, K.S. Knol and J.A. Prins, Z. Phys., 1930, 63, 345.

2. J.M. Bijvoet, Nature, 1954, 173, 888.

3. A.F. Peerdeman, A.J. Van Bommel and J.M. Bijvoet, Proc. Koninkl. Ned. Akad. Wetnschap., 1951, B54, 16.

4. Y. Okaya and R. Pepinsky, Phys. Rev., 1956, 103, 1645.

5. G.N. Ramachandran and S. Raman, Curr. Sci., 1956, 25, 348.

6. D. Dale, D.C. Hodgkin and K. Venkatesan, in Crystallography and Crystal Perfection, ed. G.N. Ramachandran, 1963, Acad. Press.