A Thesis

MODERN TECHNIQUES IN DIRECT METHODS

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

by

Andrew A. Freer

Chemistry Department

May 1980

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

ACKNOWLEDGEMENTS

I wish to express my sincere thanks to my supervisors Professor G.A. Sim, Dr. C.J. Gilmore and Dr. A.F. Cameron, and in particular to Dr. C.J. Gilmore whose guidance and interest have encouraged and sustained me throughout this research.

Thanks are also due to my other friends in the crystal structure group for their advice, both practical and theoretical.

I am indebted to Dr. D.D. MacNicol for supplying the crystals examined in Chapter 4 and for his helpful discussions on the clathrate structures.

Finally, special thanks are due to the University of Glasgow Computer Service for use of their I.C.L. 2976 computer.

This thesis has been compiled, edited and printed with the aid of a word processor by the kind permission of Dr. D.N.J. White.

SUMMARY

This thesis is concerned with the application of modern direct methods techniques to the solution of difficult structures. The methods employed have been incorporated into the multisolution computer program, MULTAN, thereby producing a comprehensive package which is capable of supplying the crystallographic user with a means of solving the most obstinate of structures.

Chapter 1 contains a review of the techniques used, the first part placing particular emphasis on the application of 4- and 5-phase structure invariants (quartets and quintets) in phase determination. The second part considers the problems of defining a good starting set of reflexions for phase expansion and leads to the introduction of magic-integer / Ψ-map and random phase set / linear equation algorithms. The reliability of these invariant estimates, derived from the various probability formulae introduced in Chapter 1, are investigated in Chapter 2 by analysing the results obtained for two centrosymmetric and three non-centrosymmetric structures.

The use of these probability formulae for structure elucidation is introduced in Chapter 3 by employing quartet invariants in the structure determination of two benzazete dimers which crystallize in space group $P\bar{1}$. The second part of this chapter describes the structure solution of an Ylide : Picric acid complex which involves quartet invariants extended to all space groups.

The use of quartets in an active role (and quintets in a passive role), fully integrated into the MULTAN system, is described in Chapter 4, in which the structures of two hexa-host inclusion compounds are determined. Not only do

both structures comprise fairly large molecules (N=186 and N=188) in space groups $P\bar{1}$ and $P2_1$ respectively and thereby represent a challenge to direct method techniques, but in each analysis interesting chemical features promote the use of such compounds in trapping novel guest species.

Chapters 5 and 6 deal with the structure determination of five natural products, four of which are sesquiterpenoid lactones. Chapter 5 discusses the problems encountered in the structure elucidation of two germacranolides which involved not only the use of quartets and quintets, but also the application of magic-integer / ψ-map and random phase set / linear equation techniques. Chapter 6 involves the structure determination of three molecules using, in two instances, a rescaling of E-magnitudes *via* the temperature coefficient, B. The final structure solution involved a simple modification to the Wilson plot procedure.

The conclusions derived from the techniques employed are discussed briefly at the end of the thesis.

CONTENTS

Chapter 1 Some Aspects of Direct Methods.

I

The Squalene Adduct of Hexakis (p-t-butyl-phenylthiomethyl) benzene.

II

The Acetic Acid Adduct of Hexakis (R-α-phenyl-sulphonylmethyl) benzene.

Tables of observed and calculated structure factors may be found in APPENDIX IV.

Chapter 1

Some Aspects of Direct Methods

1.Ø INTRODUCTION

The essential hurdle remaining in X-ray and neutron-diffraction analyses of crystal structures is the phase problem. The reconstitution of the 'image' which provided the diffraction spectrum - that is, either the electron density or nuclear scattering density in the crystallographic cell - requires a knowledge of both the magnitudes and phases of the Fourier coefficients (reflexion structure factor amplitudes). The relative magnitudes of the structure factors, simply related to the square root of the observed reflexion intensities, may be obtained routinely and with high precision by automatic diffractometry. But determination of reflexion phases is, in general, far from straightforward, and it is the recent developments in this area with which this thesis is concerned.

The intensity of a given X-ray reflexion represents the sum of the coherently scattered wavelets originating from each atom; this sum reflects the phase relationships between the different wavelets - which depend on the positions of the atoms in the unit cell - and the scattering power of each atom. It is this dependence on scattering power which provides the basis for the majority of crystal structure solutions using the heavy atom method. In the absence of a heavy-atom derivative or complex, one is left with two ways of tackling the phase problem.

The 'trial-and-error' method describes itself - molecules of known structure are moved around the unit cell until observed and calculated reflexion intensities are in fair agreement. For a large molecule, or even a medium sized one with a variety of possible conformations, this method is not a practical one although several workers remain interested in the possibility of defining structures in an *ab initio* way through computer calculations restrained by sensible considerations of inter- and intra- molecular potential energies.

The term 'direct' is usually reserved for those methods which attempt to derive the phases of the structure factors directly by mathematical means from the measured X-ray intensities. Ideally, direct methods reduce the phase determining problem to an objective procedure which, once formulated, may be solved by a routine sequence of steps in which any decisions are of a purely mathematical nature.

1.1 DIRECT METHODS

Direct methods of phase solution depend on the physical proposition that the electron density is positive throughout a unit cell. It was realised some thirty years ago that the phases and moduli of reflexion structure factors are related, and this has formed the background to all subsequent work. Harker and Kasper showed that application of Cauchy-Schwartz inequality relationships to the structure factor equation immediately led to relationships between the phases of certain reflexions, which were illustrated in the solution of orthorhombic decaborane[1] .

However, as the complexity of the crystal increases, it becomes necessary to bring in probability methods since inequalities, by themselves, impose restrictions only on phases whose intensities are very large. Hence, a few years after the Harker-Kasper work, a series of papers by Karle and Hauptman[2,3] developed the concept of probable relationships between phases of reflexions.

The basis for these methods lies in the equation derived by Sayre[4] (1952) for the special case of centrosymmetric crystals:

$$F_{\underline{h}} = \frac{1}{V}\frac{f}{g}\sum_{\underline{k}} F_{\underline{k}}\, F_{\underline{h}-\underline{k}} \qquad (1.1)$$

where f is the scattering factor for the real atom, g is the common scattering factor of the squared atom and V is the unit cell volume. However, the derivation of the inequalities makes no assumption about the shape of the atoms in the cell. In order to introduce the concept that atoms are discrete points in reciprocal space, the structure factor(1.1) is expressed as a normalised structure factor, $E_{\underline{h}}$.

1.2 THE NORMALISED STRUCTURE FACTOR

The normalised structure factor, $E_{\underline{h}}$, in its simplest form can be written as:

$$E_{\underline{h}} = \frac{F_{\underline{h}}(\text{observed})}{F_{\underline{h}}(\text{expected})} \qquad (1.2)$$

Expressing the expected structure amplitude as the total electron density in the unit cell, (1.2) becomes:

$$|E_{\underline{h}}|^2 = \frac{K|F_{\underline{h}}|^2}{\zeta \sum_{j=1}^{N} f_j 2} \qquad (1.3)$$

where f_j is the atomic scattering factor of the jth. atom in a unit cell containing N atoms, ζ is a factor which corrects for space group extinctions and $F_{\underline{h}}$ is a relative structure factor which requires the constant scale factor, K, to be applied to bring $F_{\underline{h}}$ to an absolute scale. The fundamental equation (1.3) has two serious drawbacks which prevent it being used in this form for accurate calculation of E-magnitudes.

1.2.1 Temperature Factor, B

The normal scattering factor curves are calculated on the basis of the electron distribution in a stationary atom.

However, the atoms in crystals are always vibrating about their rest points. The effect of thermally induced vibrations was first analysed by Debye in 1913 and later corrected by Waller. The magnitude of the vibration depends on the temperature, the mass of the atom and the firmness with which it is held in place by covalent bonds or other forces.

The effect of such thermal vibration is to spread the electron cloud over a larger volume and thus cause the scattering power of the real atom to fall off more rapidly than in the ideal stationary model. The change in scattering power is given by:

$$f = f_o e^{-B(\sin^2\theta)/\lambda^2} \qquad (1.4)$$

where B is related to the mean-square amplitude ($\bar{U}^2$) of atomic vibration by:

$$B = 8\pi^2 \bar{U}^2$$

f is the ideal and f_o the observed scattering factor for a point atom.

1.2.2 Scale Factor, K

An average observed intensity, corrected for Lorentz and polarisation factors, may be defined such that:

$$I_{rel} = \langle |F_{rel}|^2 \rangle_{average} \qquad (1.6)$$

For a unit cell which contains N atoms, it can be shown that the theoretical average intensity is given by:

$$I_{abs} = \sum_{j=1}^{N} f_j^2$$

Hence the average intensity depends merely on the unit cell contents and not atomic positions. Ideally, the ratio of

I_{abs} to I_{rel} should be a scaling factor required to place individual I_{rel}'s on an absolute scale. However, f's are not constants but are functions of $(\sin\theta)/\lambda$ so that I_{abs} also varies with $\sin\theta/\lambda$. This variation is normally avoided by dividing reciprocal space into concentric shells, each dimensioned such that the change of f with $(\sin\theta)/\lambda$ within the shell can be ignored, and averaging the I_{rel}'s of the reflexions within each shell. In this way I_{rel} can be compared with I_{abs} calculated from the f's appropriate to the shell.

The measured, or observed, intensities, I_{rel} , are related to those calculated on an absolute scale (for atoms that are at rest), I_{abs} , by:

$$I_{rel} = K.I_{abs}\, e^{-2B\sin^2\theta/\lambda^2} \qquad (1.8)$$

where K is a constant scale factor. Rearranging the terms and taking natural logarithms of both sides gives

$$\ln(I_{rel} / \Sigma f^2_j) = \ln K - 2B\sin^2\theta/\lambda^2 \qquad (1.9)$$

A plot of the left side of (1.9) as a function of $(\sin^2\theta)/\lambda^2$ yields a straight line of slope -2B whose intercept on the ordinate at $\theta = \emptyset^{o}$ determines the scale factor, K. This is the procedure devised by A.J.C. Wilson[5]. In practice, however, it is usual to calculate a least-squares line through the Debye points to obtain the Debye-Waller factor, B, and the scale factor, K, used to convert F's to E's. If the Debye points deviate greatly from a straight line, the K-curve method introduced by Karle and Hauptman[6] is used.

It cannot be too strongly emphasised that, since all direct methods calculations commence with the computation of E-magnitudes, it is worthwhile considering the factors which affect their actual values. E-magnitudes calculated using

Wilson plots to obtain values for scale and temperature coefficients are subject to alignment of inter- and intra-atomic vectors, which illustrates that molecules are not random atoms in reciprocal space.

1.3 ORIGIN AND ENANTIOMORPH DEFINITION

1.3.1 Origin Definition

In order to define a crystal structure completely, we must be able to specify the atomic positions with respect to a frame of reference. This entails setting up a system of axes with a fixed origin. In general, there will be a choice of origin locations since crystal symmetry usually determines the direction of the axes but not their absolute positions. A change in origin position will affect the phases of the structure factors (but not their magnitudes), so the selection of a particular origin imposes a corresponding pattern of relative phases on the diffraction pattern.

The normal structure factor expression is:

$$F_{\underline{h}} = \sum_j f_j \exp(2\pi i \underline{h}.\underline{x}_j) \qquad (3.1)$$

If the origin is moved a distance $\Delta\underline{x}$, the structure factor becomes:

$$F'_{\underline{h}} = \sum_j f_j \exp 2\pi i \underline{h}.(\underline{x}_j - \Delta\underline{x})$$

therefore

$$F'_{\underline{h}} = F_{\underline{h}} \exp(-2\pi i \underline{h}.\Delta\underline{x}) \qquad (3.2)$$

Thus an origin shift of $\Delta\underline{x}$ changes the phase of $F_{\underline{h}}$ by $-2\pi i \underline{h}.\Delta\underline{x}$. In terms of its components, the phase shift is

given by:

$$\Delta\phi_{hkl} = -2\pi(h\Delta\underline{x} + k\Delta\underline{y} + l\Delta\underline{z}) \qquad (3.3)$$

Thus in space group $P\bar{1}$ if the origin is moved from (Ø,Ø,Ø) to the point (Ø.5,Ø,Ø) all reflexions with h odd will change by an odd multiple of π (a change of sign), while those with h even will change by an even multiple of π (retain the same sign). Hence it is possible to put reflexions into categories according to the parity of h,k or l, i.e. whether or not their signs depend upon origin position in the x,y or z direction.

1.3.2 Structure Invariants and Seminvariants

A. Structure Invariants

The cataloguing of reflexions during origin definition leads to the concept of structure invariants and derived seminvariants[7-9]. From equation (3.2) it has been shown that the phases of the structure factors may change with a change in origin. Although the signs of individual phases depend on the structure and choice of origin, there exist certain linear combinations of phases, the structure invariants, whose values are independent of the choice of origin. Because of this dependence on structure alone, it is not surprising that the development of the theoretical basis of direct methods of phase determination involves the structure invariant in the central role.

Clearly ,from (3.2)

$$\phi'_{\underline{h}} = \phi_{\underline{h}} - 2\pi\underline{h}.\underline{r} \qquad (3.4)$$

where $\phi'_{\underline{h}}$ is the phase of the structure factor $F'_{\underline{h}}$, with respect to the new origin shifted $\underline{r}$ from its previous position. Considering the linear combinations of both sides

of (3.4) to have integer coefficients $A_{\underline{h}}$ which depend on $\underline{n}$ leads to:

$$\sum_{\underline{h}} A_{\underline{h}} \phi'_{\underline{h}} = \sum_{\underline{h}} A_{\underline{h}} \phi_{\underline{h}} - 2\pi\left(\sum_{\underline{h}} A_{\underline{h}} \underline{h}\right).\underline{r} \qquad (3.5)$$

If

$$\sum_{\underline{h}} A_{\underline{h}} \underline{h} = \emptyset \qquad (3.6)$$

then

$$\sum_{\underline{h}} A_{\underline{h}} \phi'_{\underline{h}} = \sum_{\underline{h}} A_{\underline{h}} \phi_{\underline{h}} \qquad (3.7)$$

no matter what the vector $\underline{r}$ may be, and the linear combination of the phases (3.7) is a structure invariant since it has the same value for every choice of origin.

As a corollary of (3.6); when the summation (3.7) is restricted to three vectors with integer coefficients

$$\phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}}$$

is a structure invariant if $\underline{h}+\underline{k}+\underline{l} = \emptyset$. The use of structure invariants can readily be expanded to include a set of four[2,10,11] or five[12] phases whose linear combination of phases satisfies the integer combination, such that $\underline{h}+\underline{k}+\underline{l}+\underline{m} = \emptyset$, and $\underline{h}+\underline{k}+\underline{l}+\underline{m}+\underline{n} = \emptyset$ respectively. The occurrence of structure invariants and seminvariants lends itself to the rule that reflexions used to define the origin cannot, themselves, combine to give a structure invariant. Therefore, by definition, reflexions used in origin definition should be linearly independent.

B. Structure Seminvariants.

Seminvariants arise from using the restrictions imposed by space group symmetry when defining the origin. Because of

this restriction many linear combinations of phases, in addition to the universal structure invariants previously discussed, will remain unchanged in sign when the origin is shifted only in the restricted ways allowed by space group symmetry. As an example, in the space group $P2_1$, the linear combination

$$\phi_{h_1 \emptyset\ l_1} + \phi_{h_2 \emptyset\ l_2}$$

is a structure seminvariant if:

$$\underline{h}_1 + \underline{h}_2 = \emptyset (\text{mod } 2)$$

$$\underline{l}_1 + \underline{l}_2 = \emptyset (\text{mod } 2)$$

In general, the structure seminvariants are the linear combinations

$$\sum_{\underline{h}} A_{\underline{h}}\ \phi_{\underline{h}}$$

where $A_{\underline{h}}$ are integers satisfying

$$\sum_{\underline{h}} A_{\underline{h}} . \underline{h}_s = \emptyset (\text{mod. } \underline{\omega}_s)$$

$\underline{h}_s$ is the vector seminvariantly associated with the vector $\underline{h}$ and $\underline{\omega}_s$ is the seminvariant modulus. The usefulness of seminvariants is best illustrated in enantiomorph definition.

1.3.3 Enantiomorph Definition

In the non-centrosymmetric case we also need to define the enantiomorph. The effect of changing the enantiomorph is to reverse the signs of all phases, so that ϕ_{hkl} becomes $-\phi_{hkl}$. This, in turn, will reverse the signs of all the structure invariants of the type

$$S = \phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}}$$

This means that the enantiomorph can be defined by specifying the value of a particular invariant. In general, one of the possible values for a structure invariant or seminvariant corresponds to one of the enantiomorphous structures permitted by the structure factor magnitudes and the second value corresponds to the other enantiomorph. When the two enantiomorphs are distinct, one may be selected by specifying arbitrarily the sign of a structure seminvariant or invariant whose value differs from $\emptyset$ or π . For efficient enantiomorph definition this seminvariant must involve large E-magnitudes.

Having determined that

$$\phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}}$$

is a structure invariant and therefore does not depend on origin but only on the position of the atoms we now wish to predict its value. This is done <u>via</u> probability theory.

1.4 PROBABILITY THEORY

1.4.1. Introduction

Consider an experiment yielding a set of possible outcomes. It is assumed, either because the initial conditions are so numerous or poorly defined, or the laws governing the relationships between the initial conditions and the actual outcome are so complex or poorly understood, that it is not possible to predict with certainty the actual result of the given experiment. All that is known is that one of a well defined set of possible outcomes must occur. If the experiment is performed N times, where N is a large number, and the outcome X_i occurs M times, then the probability of X_i

may be taken as $P(X_i) = M/N$.

In an X-ray diffraction experiment we have a large number of invariants and associated E-magnitudes, and wish to know the probable value of the invariant.

Algebraically, the calculated normalised structure factor, $E_{\underline{h}}$, is defined as:

$$E_{\underline{h}} = \frac{\sigma_3}{\sigma_3 \sigma_2^{3/2}} \sum_{j=1} \exp\ (2\pi i \underline{h}.\underline{r}_j) \qquad (4.1)$$

where

$$\sigma_n = \sum_{j=1} Z_j$$

and Z_j is the atomic number of the jth. atom in the unit cell and $\underline{r}_j$ is the position of the atom labelled j. The reciprocal vector $\underline{h}$ may be fixed and the position vectors $\underline{r}_j$ regarded as random variables or, alternatively, the crystal structure may be fixed and the reciprocal vector $\underline{h}$ regarded as the random variable. A random variable is a function which takes on a definite value at every point in a sample space. In either case $E_{\underline{h}}$ is a function of one or more random variables and is itself a random variable and on this basis a probable value of its phase$(2\pi i \underline{h}.\underline{r}_j)$ may be found.

The distributions to be considered here are of the second kind, where one is ordinarily confronted with a fixed but unknown crystal structure and a large number of structure factor magnitudes sampled from reciprocal space.

1.4.2 Triplets

Triplets are three-phase structure invariants of the form:

$$\Phi_3 = \phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}} \qquad (4.2)$$

In the centrosymmetric case the value of Φ_3 is restricted to either $\emptyset$ or π which leads to the use of 'signs' in phase determination; the signs + or - correspond to phase angles of $\emptyset$ or π respectively. Sayre pointed out, however, that for the case where the E-magnitudes involved are large the sign of Φ_3 is determined by the agreement in sign among the products of the component phases,i.e.

$$S(E_{\underline{h}}) \simeq S(E_{\underline{k}}.E_{\underline{l}}) \qquad (4.3)$$

where S means 'sign of' and may be considered as $\pm$ 1. For the invariant, Φ_3, the product of the three signs is most probably +, assuming large E-magnitudes are involved. This probability increases with an increase in the product of the magnitudes of the E's involved.

The first practical probability formula which expressed this quantitatively was put forward by Cochran and Woolfson[13] (1955) for the centrosymmetric case. The probability that:

$$S(E_{\underline{h}}).S(E_{\underline{k}}).S(E_{\underline{l}}) = +1$$

is given by,

$$P^{+} = 1/2 + 1/2 \tanh A/2 \qquad (4.4)$$

where

$$A = 2\sigma_3\sigma_2^{-3/2}|E_{\underline{h}}E_{\underline{k}}E_{\underline{l}}| \qquad (4.5)$$

If there are several indications as to the sign of $S(E_{\underline{h}})$ then (4.3) becomes:

$$S(E_{\underline{h}}) \simeq S(\sum_{\underline{k}} E_{\underline{k}}.E_{\underline{l}}) \qquad (4.6)$$

and the probability relationship (4.4) is generally expressed as:

$$P_3^+ = 1/2 + 1/2 \tanh \Sigma A/2 \qquad (4.7)$$

The corresponding formulae for non-centrosymmetric structures, where the phase of $|E_{\underline{h}}|$ may take on any value from $\emptyset$ to 2π, were first introduced by Cochran[14] in 1955. For coherence with other sections the nomenclature derived by Hauptman is used. The fundamental relationship,

$$\Phi_3 \simeq \phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}} \qquad (4.8)$$

is expressed as a joint conditional probability distribution[15],

$$P_{1/3} = P(\Phi_3|R_1,R_2,R_3) \simeq \frac{1}{2\ I_o(A)} \exp(A \cos\Phi_3) \qquad (4.9)$$

where I_o is a modified Bessel function of the first kind.

$$R_1=|E_{\underline{h}}|;\ R_2=|E_{\underline{k}}|;\ R_3=|E_{\underline{l}}|$$

are, as expected, parameters of the distribution. Graphs of (4.9) are shown in Figures 1(a) and 1(b) for A values of 2.316 and $\emptyset$.731 respectively. Note, that owing to symmetry, it is sufficient to plot these curves in the intervals $\Phi_3 = \emptyset - 180^o$.

It is clear that these distributions always have a unique maximum at $\Phi_3 = \emptyset$ and a unique minimum at $\Phi_3 = 180^o$. Furthermore, the larger the value of A, the smaller the variance of the distribution.

1.4.3 Statistical Parameters from Joint Conditional Probability Distributions

In the non-centric case, the following statistical parameters are derived from the relevant joint conditional probability distributions, $P(\Phi)$.

Modal value of Φ, $(|\Phi_m|)$, is that value which gives the maximum value of $P(\Phi)$.

Mean value of Φ, $(<|\Phi|>)$, where

$$<|\Phi|> = \int_0^{2\pi} \Phi P(\Phi) d\Phi \qquad (4.10)$$

Of immediate interest when dealing with any phase relationship is its variance, V, which is a measure of its reliability.

$$\text{Variance, } V = \int_0^{2\pi} (\Phi - <(\Phi)>)^2 . P(\Phi) d\Phi \qquad (4.11)$$

and standard deviation,

$$\sigma = V^{1/2} \qquad (4.12)$$

In the centrosymmetric case, where the phase estimate will only have one of two possible values (0 or π), probability formulae give an estimate of its reliability directly, i.e. $P^+ > 0.99$ etc.

1.5 THE USE OF TRIPLE PHASE INVARIANTS

1.5.1 Phase Expansion - Σ_2 Relationships - Tangent Formula

Apart from the complexity of the structure, perhaps the most important element influencing success or failure of a direct solution of the phase problem is the choice of starting reflexions. Once the normalised structure factors have been

calculated a subset is chosen for carrying out sign determination.

In the case of centrosymmetric crystals, the most useful sign determining formula has been that termed the Σ_2 relationship[16] and is given in equation (4.6) with its associated probability (4.7). The summation, which is carried out over high E-magnitudes, can involve simply one or many terms, depending on the stage of the analysis. Hence by the use of these triple-phase relationships it is possible, by starting with a limited number of phases, to pyramid to a number large enough to give a recognisable Fourier representation of the structure.

With non-centrosymmetric structures the procedure is similar. Employing relationship (4.8) with the joint conditional probability distribution (4.9), the probable value for the invariant,

$$\phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}}$$

will be zero if large E-magnitudes are used. It is possible to combine all the indications for $\phi_{\underline{h}}$:

$$\begin{aligned}
\phi_{\underline{h}} &+ \phi_{\underline{k}}1 + \phi_{\underline{l}}1 \simeq 0 \\
\phi_{\underline{h}} &+ \phi_{\underline{k}}2 + \phi_{\underline{l}}2 \simeq 0 \\
" \quad &" \quad " \\
" \quad &" \quad " \\
\phi_{\underline{h}} &+ \phi_{\underline{k}}n + \phi_{\underline{l}}n \simeq 0
\end{aligned} \qquad (5.1)$$

When all the associated phases have been calculated, the combined probable value of $\phi_{\underline{h}}$ is given by the tangent formula[17]:

$$\tan \phi_{\underline{h}} = \frac{\sum_{\underline{k}} |E_{\underline{k}}.E_{\underline{l}}| \sin(\phi_{\underline{k}} + \phi_{\underline{l}})}{\sum_{\underline{k}} |E_{\underline{k}}.E_{\underline{l}}| \cos(\phi_{\underline{k}} + \phi_{\underline{l}})} = \frac{T_{\underline{h}}}{B_{\underline{h}}} \qquad (5.2)$$

Refinement of a set of phases can be achieved by a recycling process, employing the tangent formula, which can be continued until negligible shifts in the derived phase angles are observed from one cycle to the next.

1.6 THE MULTISOLUTION TECHNIQUE

The last fifteen years has seen the increasing use of multisolution techniques. The crystallographic user has a wide and varied range of packages from which to choose; X-RAY 78 (J.M.Stewart), N.R.C.(Ahmed), SHELX (G.M.Sheldrick). For mini-computers several systems now exist, e.g. Syntex XTL (R.A.Sparks), CRYSTAN (H.Burzoff et. al.,). However, perhaps the most powerful and efficient is the MULTAN package of Woolfson, Main and Germain[18].

1.6.1 Selecting the Starting Set

The most difficult part of any multisolution technique is the procedure used to select a starting set of reflexions. To make use of the derived Σ_2 relationships (4.6) a small starting set of reflexions, which will efficiently expand to give new phase information, is required. In normal application of the multisolution process a number of phases may be expressed explicitly (origin and enantiomorph), and other phases are represented by reflexions which take on different values for each set of tangent refinements. These reflexions usually take on values of $\pm \pi/4$ and $\pm 3\pi/4$.

In this way the phases of these 'variable' reflexions are permuted until the correct sequence is found. If, for

example, there are n variable reflexions in the starting set, then 4^n sets of phases would be developed. Four variable phases for a non-centrosymmetric structure require 256 permutations. For this reason, the number of starting set reflexions is usually kept to a minimum.

One computer algorithm used to choose the origin-defining reflexions, which form a good starting point for phase determination, is the CONVERGENCE procedure[19]. This algorithm lends itself to automation. The first step is the calculation of $\alpha^2_{\underline{h},est}$ for each reflexion. This is a measure of how reliably each phase can be determined in terms of all others remaining in the data set at any one time and can be calculated without knowledge of individual phase angles.

$$\langle \alpha_{\underline{h}}{}^2 \rangle_{est} = \sum_{\underline{h}'} A_{\underline{h}\underline{h}'}{}^2 + 2 \sum_{\underline{h}'} \sum_{\underline{h}''} A_{\underline{h},\underline{h}'} \, A_{\underline{h},\underline{h}''} \, \frac{I_1(A_{\underline{h},\underline{h}'}) I_1(A_{\underline{h},\underline{h}''})}{I_o(A_{\underline{h},\underline{h}'}) I_o(A_{\underline{h},\underline{h}''})} \qquad (6.1)$$

where $\underline{h}' = \underline{h}'' + \underline{h}$

The summations over $\underline{h}'$ and $\underline{h}''$ are taken over all available Σ_2 interactions involving $\phi_{\underline{h}}$. By using $\alpha^2_{\underline{h},est}$ the least reliably determined phase is eliminated at each stage, leaving at the end those reflexions which are strongly linked together and which lead to reliable phase determination by giving multiple indications using strong relationships. One of the reasons the algorithm is successful is that it looks at the way reflexions are linked together through phase relationships, and does not look at each reflexion or relationship in isolation. It will be shown later that the CONVERGENCE procedure can be used to take account of higher neighbourhoods of structure invariants.

$\alpha_{\underline{h}}{}^2{}_{,est}$ can be usefully compared with another term, $\alpha_{\underline{h},expt}$, which is derived from the tangent formula (5.2) as;

$$\langle a_{\underline{h}} \rangle_{expt} = A_{\underline{h}} |E_{\underline{h}}| (T_{\underline{h}}^2 + B_{\underline{h}}^2)^{1/2} \qquad (6.2)$$

For a random set of phases the expectation value of $a_{\underline{h}}^2$ is given by:

$$\langle a_{\underline{h}}^2 \rangle_{random} = \sum_{\underline{h}'} A_{\underline{h},\underline{h}'}^2 \qquad (6.3)$$

1.6.2 The Tangent Formula in Multisolution Methods

Application of Σ_2 relationships to non-centrosymmetric structures usually yields several estimates for the value of $\phi_{\underline{h}}$. Some, particularly those involving large E-magnitudes, will contribute more strongly to this estimate than others. Therefore reflexions at the bottom of the convergence map will participate significantly during the early stages of phase expansion. In order to use this information in the tangent formula a weighting scheme is applied so that reflexions at the bottom of the convergence map have a weight of approximately unity , while those contributors further up the mapping procedure are downweighted. The term $a_{\underline{h},expt}$, already defined in (6.2), is used as the weighting function.

The weighted tangent formula may be written as:

$$\tan \phi_{\underline{h}} = \frac{\sum_{\underline{k}} \omega_{\underline{k}} \omega_{\underline{l}} |E_{\underline{k}} E_{\underline{l}}| \sin(\phi_{\underline{k}} + \phi_{\underline{l}})}{\sum_{\underline{k}} \omega_{\underline{k}} \omega_{\underline{l}} |E_{\underline{k}} E_{\underline{l}}| \cos(\phi_{\underline{k}} + \phi_{\underline{l}})} \qquad (6.4)$$

where $\omega_{\underline{h}} = \min(\emptyset.2\ a_{\underline{h},expt}, 1.\emptyset)$

which takes into account, by employing $a_{\underline{h},expt}$, the 'strength' of the relationships being used at the time.

This weighting scheme plays an important role during the early stages of phase determination, but all the weights

quickly become unity and so have little effect during the latter stages.

1.6.3 Figures of Merit

In the standard MULTAN package three figures of merit are computed for each set of phases determined and are output with the tangent formula results.

a) The absolute figure of merit , ABSFOM[20], is a measure of internal consistency among the Σ_2 relationships and is calculated from:

$$\text{ABSFOM} = \frac{a_{\underline{h},\text{est}} - a_{\underline{h},\text{random}}}{a_{\underline{h},\text{expt}} - a_{\underline{h},\text{random}}} \qquad (6.5)$$

For a set of phases with almost no self-consistency the value of ABSFOM will be zero. On the other hand, for the correct set of phases one should expect ABSFOM to be close to unity.

b) PSI-ZERO[21] Another figure of merit which is effective is the value of

$$\Psi_o = \sum_{\underline{h}} \left| \sum_{\underline{k}} E_{\underline{k}} . E_{\underline{h}-\underline{k}} \right| \qquad (6.9)$$

where the inner summation is over all the terms available from the set of phases being determined and the outer summation is over a number of reflexions for which $E_{\underline{h}}$ is zero or small in magnitude. For a good set, Ψ_o is expected to be small since the inner summation is essentially a selection of contributors to Sayre's equation.

c) Finally, the R_{Karle} figure of merit described by Karle and Karle[22] can also be used. For each $E_{\underline{h}}$ in the set of phases being determined one calculates

$$\sum_{\underline{h}} \text{calc} = K \sum_{\underline{h}'} E_{\underline{k}} . E_{\underline{h}-\underline{k}} \qquad (6.10)$$

where K is a constant chosen to give

$$\Sigma |E_{\underline{h}}|^2_{obs} = \Sigma |E_{\underline{h}}|^2_{calc} \qquad (6.11)$$

The residual is then defined by

$$R = \frac{\sum_{\underline{h}} | |E_{\underline{h}}|_{obs} - |E_{\underline{h}}|_{calc} |}{\sum_{\underline{h}} |E_{\underline{h}}|_{obs}} \qquad (6.12)$$

and is a measure of the extent to which the 'squared structure' resembles the 'structure' for the set of phases under consideration.

It is also useful to combine these figures of merit to give:

$$C = \omega_1 \frac{\text{ABSFOM} - \text{ABSFOM}_{min}}{\text{ABSFOM}_{max} - \text{ABSFOM}_{min}}$$

$$+ \omega_2 \frac{\Psi_{o\ max} - \Psi_o}{\Psi_{o\ max} - \Psi_{o\ min}}$$

$$+ \omega_3 \frac{R_{max} - R}{R_{max} - R_{min}}$$

where ω_1, ω_2 and ω_3 are weights which are usually chosen as unity. They may be changed to give more weighting to Ψ_o and less to ABSFOM for space groups not involving translational symmetry. C should be as large as possible and has a maximum value of $\omega_1 + \omega_2 + \omega_3$.

1.6.4 Failures of the Multisolution Technique

Sometimes multisolution programs fail and no E-maps can be found from which a recognisable fragment can be recycled to derive the complete structure. The pattern of failure is rather difficult to interpret, as shown in the analysis of a number of MULTAN failures reported by Lessinger[23]. There are, however, two inherent limitations in the multisolution technique.

The first limitation concerns the size of the starting set. The basic aim is to use a starting set of reflexions small enough to give a manageable number of phase permutations yet large enough to lead to a strong development of new phase information. Hence it is possible to start with a small number of reflexions and, by use of Σ_2 relationships and the tangent formula , to pyramid to a number large enough to give a recognisable Fourier representation of the structure. The obvious danger in this approach is that the pyramid is balanced on its point and instability can ensue.

If the number of starting reflexions are to be kept at a minimal level then, as the thesis will endeavour to illustrate, the use of higher invariants with their associated probabilities may be used to ensure better phase expansion during the convergence mapping procedure of the analysis. However, there are alternative methods of initial phase assignment which exist; magic-integer phase representation and random phase sets coupled with linear equations.

A second limitation is the tangent formula itself. For some structures even the correct phases are unstable under the process of tangent formula refinement. This instability in the formula arises from the assumption that all Σ_2 relationships are independent - this is not so. Since the formula itself is derived *via* probability theory it is hoped that by using higher invariants (quartets, quintets etc.) with their better probability relationships, a weighted

tangent formula using these invariants may prove useful.

1.7 NESTED NEIGHBOURHOODS

In general, structure factor magnitudes determine the values of the cosine invariants, and therefore, except for a two-fold ambiguity, the values of the structure invariants themselves. As already discussed in section 1.4, the probability of the sign of a triple-phase invariant, Φ_3 ,being zero depends on the size of the E-magnitudes involved in the construction of the invariant. It is therefore reasonable to assume that the invariant will be more sensitive to some E's than others. In an attempt to define the criteria which bring about this dependence on E-magnitudes the concept of nested neighbourhoods was introduced[24].

A Quartets

The four-phase structure invariant

$$\Phi_4 = \phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}} + \phi_{\underline{m}} \qquad (7.1)$$

where $\underline{h} + \underline{k} + \underline{l} + \underline{m} = \emptyset$

depends, in the first instance, on the values of the principal terms

$$|E_{\underline{h}}|, |E_{\underline{k}}|, |E_{\underline{l}}| \text{ and } |E_{\underline{m}}|.$$

$$(R_1, R_2, R_3 \text{ and } R_4 \text{ respectively})$$

However, by considering that a four-phase structure invariant can be derived from the sum of two triple-phase structure invariants with one vector in common, a number of new (cross) terms can be derived:

$$\Phi_1 = \phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{-\underline{h}-\underline{k}} \qquad (7.2)$$

if $|E_{\underline{h}}|$, $|E_{\underline{k}}|$ and $|E_{-\underline{h}-\underline{k}}|$ are large, then

$$\Phi_1 \simeq \emptyset$$

and

$$\Phi_2 = \phi_{\underline{l}} + \phi_{\underline{m}} + \phi_{\underline{h}+\underline{k}} \qquad (7.3)$$

Again, if $|E_{\underline{l}}|$, $|E_{\underline{m}}|$ and $|E_{\underline{h}+\underline{k}}|$ are large, then

$$\Phi_2 \simeq \emptyset$$

Combining (7.2) and (7.3)

$$\Phi_4 = \Phi_1 + \Phi_2 = \phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}} + \phi_{\underline{m}} \simeq \emptyset$$

By taking various combinations of the four principal terms, viz.

$$|E_{\underline{h}+\underline{k}}|,\ |E_{\underline{k}+\underline{l}}|,\ |E_{\underline{l}+\underline{m}}|,\ |E_{\underline{h}+\underline{l}}|,\ |E_{\underline{k}+\underline{m}}|,\ |E_{\underline{h}+\underline{m}}|$$

it is found that only three of these are unique,

$$|E_{\underline{h}+\underline{k}}|,\ |E_{\underline{k}+\underline{l}}| \text{ and } |E_{\underline{l}+\underline{h}}|$$

(R_{12} , R_{23} and R_{31} respectively)

By addition of the three new cross-terms, derived from the principal $|E|$s, one obtains a different estimate for Φ_4 , dependent on these seven magnitudes. Since the first estimate, dependent on only the four magnitudes, uses less information than the second estimate, dependent on the seven magnitudes it is natural to expect that the latter will, at

least in favourable cases, be more reliable than the former, and this in fact is the case.

With more magnitudes the potential for obtaining a small variance is increased. It is feasible to conjecture that there exists an additional small set of magnitudes which, when added to the previous two sets, will combine to yield a still better estimate for the invariant Φ_4. In so doing the concept of nested neighbourhoods is evolved which is depicted schematically in Figure 2.

The first neighbourhood consists of four principal magnitudes R_1, R_2, R_3, R_4 ; the second is the set-theoretic union of the first neighbourhood and the three additional magnitudes are shown in the second shell. The addition of the third neighbourhood is accomplished by making $\underline{p}$ and $\underline{q}$ arbitrary reciprocal vectors which satisfy

$$\underline{h} + \underline{k} + \underline{p} + \underline{q} = \emptyset \qquad (7.4)$$

such that

$$\Phi_{pq} = \phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{p}} + \phi_{\underline{q}} \qquad (7.5)$$

is a structure invariant.

If,

$$\underline{l} + \underline{m} - \underline{p} - \underline{q} = \emptyset \qquad (7.6)$$

then

$$\Psi_{pq} = \phi_{\underline{l}} + \phi_{\underline{m}} - \phi_{\underline{p}} - \phi_{\underline{q}} \qquad (7.7)$$

is also a structure invariant.

Previously, Φ_4 had been estimated by the 7-magnitudes in its second neighbourhood,

$$|E_{\underline{h}}|,\ |E_{\underline{k}}|,\ |E_{\underline{l}}|,\ |E_{\underline{m}}|,\ |E_{\underline{h}+\underline{k}}|,\ |E_{\underline{k}+\underline{l}}|,\ |E_{\underline{l}+\underline{h}}| \quad (7.8).$$

Φ_{pq} can be estimated by means of the 7-magnitudes in its second neighbourhood,

$$|E_{\underline{h}}|,\ |E_{\underline{k}}|,\ |E_{\underline{p}}|,\ |E_{\underline{q}}|,\ |E_{\underline{h}+\underline{k}}|,\ |E_{\underline{k}+\underline{p}}|,\ |E_{\underline{p}+\underline{h}}| \quad (7.9).$$

and Ψ_{pq} by means of the 7-magnitudes in its second neighbourhood,

$$|E_{\underline{l}}|,\ |E_{\underline{m}}|,\ |E_{\underline{p}}|,|E_{\underline{q}}|,\ |E_{\underline{l}+\underline{m}}|,\ |E_{\underline{m}-\underline{p}}|,\ |E_{-\underline{p}+\underline{q}}| \quad (7.10).$$

However, from (7.1), (7.5) and (7.7) it is clear, that

$$\Phi_4 - \Phi_{pq} - \Psi_{pq} \equiv \emptyset \qquad (7.11)$$

It can therefore be expected that in the favourable case when the 7-magnitude estimates yield values for Φ_4, Φ_{pq} and Ψ_{pq} in accordance with (7.11), then Φ will be well estimated in terms of the 21-magnitudes of which only the following thirteen are unique,

$$|E_{\underline{h}}|,\ |E_{\underline{k}}|,\ |E_{\underline{l}}|,\ |E_{\underline{m}}|,\ |E_{\underline{p}}|,\ |E_{\underline{q}}| \quad (7.12)$$

$$(R_1\ ,\ R_2\ ,\ R_3\ ,\ R_4\ ,\ R_5\ ,\ R_6)$$

$$|E_{\underline{h}+\underline{k}}|,\ |E_{\underline{k}+\underline{l}}|,\ |E_{\underline{l}+\underline{h}}|,\ |E_{\underline{h}+\underline{p}}|,\ |E_{\underline{k}+\underline{p}}|,\ |E_{\underline{l}-\underline{p}}|,\ |E_{\underline{m}-\underline{p}}| \quad (7.13)$$

$$(R_{12}\ ,\ R_{23}\ ,\ R_{31}\ ,\ R_{15}\ ,\ R_{25}\ ,\ R_{3\bar{5}}\ ,\ R_{4\bar{5}})$$

It is expected that the conditional variance of the invariant, Φ_4, given the 13-magnitudes in its third neighbourhood, will be small if the three 7-magnitude subsets of the third neighbourhood, which are the respective second neighbourhoods of the structure invariants Φ_4 , Φ_{pq} and Ψ_{pq} ,

give reliable estimates for the latter in accordance with (7.11). Thus only those third neighbourhoods are useful for which $|E_{\underline{p}}|$ and $|E_{\underline{q}}|$ are both large and where $\underline{p}$ and $\underline{q}$ satisfy (7.4).

When three estimates for the phases of the invariants (7.1), (7.5), and (7.7) combine, as in equation (7.11), the resulting relationship is referred to as a trio. A closer look at the trio shows that if Φ_4 has four large principal terms and three small cross-terms, then, from probability theory its magnitude will be π ; if Ψ_{pq} has four large principal terms and three small cross-terms then the value of Ψ_{pq} is also π and the resultant phase of Φ_{pq} must therefore be zero. This gives a semi-independent estimate of the quartet magnitudes.

B. Quintets

For the quintet structure invariant:

$$\Phi_5 = \phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}} + \phi_{\underline{m}} + \phi_{\underline{n}} \qquad (7.14)$$

where $\underline{h} + \underline{k} + \underline{l} + \underline{m} + \underline{n} = \emptyset$

an analogous argument[25], as applied to quartet structure invariants, can be envisaged whereby the five principal E-magnitudes,

$$|E_{\underline{h}}|, |E_{\underline{k}}|, |E_{\underline{l}}|, |E_{\underline{m}}| \text{ and } |E_{\underline{n}}|$$

may combine to produce a number of cross-terms. A quintet is derived from the addition of a four-phase structure invariant and a three-phase structure invariant, each with one term in common.

$$\Phi_1 = \phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}} + \phi_{-\underline{h}-\underline{k}-\underline{l}} \qquad (7.15)$$

If $E_{\underline{h}}$, $E_{\underline{k}}$, $E_{\underline{l}}$ and $E_{-\underline{h}-\underline{k}-\underline{l}}$ are large, then

$$\Phi_1 \simeq \emptyset$$

and

$$\Phi_2 = \phi_{\underline{m}} + \phi_{\underline{n}} + \phi_{\underline{h}+\underline{k}+\underline{l}} \qquad (7.16)$$

When $E_{\underline{m}}$, $E_{\underline{n}}$ and $E_{\underline{h}+\underline{k}+\underline{l}}$ are large,

$$\Phi_2 \simeq \emptyset$$

Adding (7.15) and (7.16) gives

$$\Phi_5 = \Phi_1 + \Phi_2 = \phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}} + \phi_{\underline{m}} + \phi_{\underline{n}} = \emptyset \qquad (7.17)$$

In this case there are twenty cross-terms associated with the second neighbourhood, of which only ten are unique. The additional magnitudes are:

$$|E_{\underline{h}+\underline{k}}|, |E_{\underline{h}+\underline{l}}|, |E_{\underline{h}+\underline{m}}|, |E_{\underline{h}+\underline{n}}|, |E_{\underline{k}+\underline{l}}|,$$

$$|E_{\underline{k}+\underline{m}}|, |E_{\underline{l}+\underline{n}}|, |E_{\underline{l}+\underline{m}}|, |E_{\underline{m}+\underline{n}}|, |E_{\underline{k}+\underline{n}}|$$

$$(R_{12}, R_{13}, R_{14}, R_{15}, R_{23},$$
$$R_{24}, R_{25}, R_{34}, R_{45}, R_{35})$$

These new cross-terms may then be combined with the five principal terms to give a better estimate for the phase of the invariant, Φ_5, based on the 15-magnitudes of the second neighbourhood. Again, third, fourth and fifth etc. nested neighbourhood sequences can be derived using appropriate reciprocal vectors until 29-, 37- and 51-magnitude neighbourhoods are accumulated respectively. At present only the first and second neighbourhoods are used.

By using the nested neighbourhood concept and involving more E-magnitudes, it is possible to obtain different values for probability estimates and these values should be of greater reliability.

1.8 PROBABILITY USING HIGHER INVARIANTS

A. Quartets

An obvious extension of the probability relationships already derived for the triple-phase structure invariant, Φ_3, is to consider the four-phase structure invariant, Φ_4. This relationship was already derived and published as the Σ_5 relationship by Hauptman[2]. It had also beeen derived, independently, by Simerska (1956). Using Hauptman's notation and applying the assumptions employed for triple-phase probability, the conditional probability distribution, $P_4^{\pm}$, given the four magnitudes, R_1, R_2, R_3, R_4, has been found for centrosymmetric structures[15].

$$P_4^{\pm} \simeq \frac{1}{K_4} . Z_4^{\pm} \qquad (8.1)$$

where

$$Z_4^{\pm} = \exp(\pm B/2) \qquad (8.2)$$

and

$$K_4 = Z_4^{+} + Z_4^{-} \qquad (8.3)$$

with

$$\sigma_4 = 2\sqrt{P_4^{+} . P_4^{-}} \qquad (8.4)$$

and

$$B = \frac{1}{\sigma_2^3}(3\sigma_3^2 - \sigma_2 \sigma_4) R_1 R_2 R_3 R_4 \qquad (8.5)$$

Comparison of the results derived from (8.1) and (4.4) shows little difference, with the most probable value of Φ as zero. However, a zero estimate for Φ with (4.4) is, in general, less reliable due to the smaller value of A compared to B. The theory of the four-phase structure invariant appears to have yielded nothing new.

Consideration of the fact that the structure invariant, Φ_4, must occasionally be in the neighbourhood of 180°, leads to an examination of the influence of the cross-terms involved. By an argument similar to that used to derive $P_4^{\pm}$, Hauptman[15] has calculated the 7-magnitude conditional probability distribution, $P_7^{\pm}$, for centrosymmetric structures as:

$$P_7^{\pm} \simeq \frac{1}{K_7} . Z_7^{\pm} \qquad (8.6)$$

where

$$Z_7^{\pm} = \exp(\mp B)\cosh\left(\frac{\sigma_3}{\sigma_2^{3/2}} R_{12} Y_{12}^{\pm}\right)\cosh\left(\frac{\sigma_3}{\sigma_2^{3/2}} R_{23} Y_{23}^{\pm}\right) X$$

$$\cosh\left(\frac{\sigma_3}{\sigma_2^{3/2}} R_{31} Y_{31}^{\pm}\right) ,$$

$$Y_{12}^{\pm} = R_1 R_2 \pm R_3 R_4$$

$$Y_{23}^{\pm} = R_2 R_3 \pm R_1 R_4$$

$$Y_{31}^{\pm} = R_3 R_1 \pm R_2 R_4$$

and

$$K_7 = Z_7^{+} + Z_7^{-} \qquad (8.7)$$

with

$$\sigma_7 = 2\sqrt{P_7^{+} . P_7^{-}} \qquad (8.8)$$

In the special case when R_{12},R_{23}, R_{31} are all large, then $P_7^{\pm}$ is close to 1 and $\Phi_4 \simeq 0$; if, on the other hand, R_{12}, R_{23} and R_{31} are small, then $P_7^{\pm}$ is close to 0 and $\Phi_4 \simeq \pi$. The invariants derived using small cross-terms which give rise to phase estimates of π are called <u>negative quartets</u>.

Hence the additional information of the cross-terms has been sufficient, under certain cicumstances, to change the most probable value of Φ_4 from 0° when only (8.1) is used , to 180° when (8.6) is used. Thus the comparison between (4.4) and (8.6) illustrates with particular force the great change which may result as one increases the number of magnitudes on which the estimate for the structure invariant depends.

Since the nested neighbourhood concept has developed a third neighbourhood, comprising of 13-magnitudes, Hauptman[11] has derived the conditional probability distribution, $P_{13}^{\pm}$, which is given in Appendix I. By having more magnitudes contributing to the phase estimate, it is expected that phases so determined will be more reliable.

Akin to the centrosymmetric formulae, (8.1) and (8.6), the associated joint conditional probability distributions, $P_{1/4}$ and $P_{1/7}$, for non-centrosymmetric structures in space group P1 have been formulated by Hauptman[15].

$$P_{1/4} \simeq \frac{1}{K_4} \exp(B\cos\Phi) \qquad (8.9)$$

$$P_{1/7} \simeq \frac{1}{K_7} \exp(-2B\cos\Phi) I_o\left(\frac{2\sigma_3}{\sigma_2^{3/2}} R_{12}Y_{12}\right) I_o\left(\frac{2\sigma_3}{\sigma_2^{3/2}} R_{23}Y_{23}\right) X$$

$$I_o\left(\frac{2\sigma_3}{\sigma_2^{3/2}} R_{31}Y_{31}\right) , \qquad (8.10)$$

where

$$Y_{12} = [R_1^2R_2^2 + R_3^2R_4^2 + 2R_1R_2R_3R_4\cos\Phi]^{1/2}$$

$$Y_{23} = [R_2^2R_3^2 + R_1^2R_4^2 + 2R_1R_2R_3R_4\cos\Phi]^{1/2}$$

$$Y_{31} = [R_3^2R_1^2 + R_2^2R_4^2 + 2R_1R_2R_3R_4\cos\Phi]^{1/2}$$

where K_7 is a suitable normalising factor independent of Φ_4. The 13-magnitude third neighbourhood, $P_{1/13}$ formula[10] is given in Appendix II.

In sharp contrast to the distributions (4.9) and (8.9), the maximum of (8.10) may lie anywhere in the interval 0-180°. Figures (3 a,b and c) show typical examples of the distribution (8.10) for the parameters given. In general, the estimate is good when B is large and when the estimate is near zero or 180°. If B is small, or if the estimate is around 90°, the probability is least reliable. The effect of B as a guide to the probability estimate is discussed in greater detail in Chapter 2.

The change in value of Φ_4 using small cross-terms, immediately suggests an escape from the 'all-positive' problem which has hindered triple-phase sign relationships in symmorphic space groups.

A significant point which emerges is the fact that the above distributions, although initially derived for P1 and $P\bar{1}$, may be applied to any space group. This is accomplished by 'converting' the other space groups to triclinic symmetry. Ideally, each space group could have its individual joint conditional probability formula[26], (Hauptman[27] has already proposed a formula using seminvariants for space group $P2_1$) but the task of originating all 230 distributions is formidable.

B. Quintets

The success of probability formulae as applied to quartet structure invariants, Φ_4, has led to the derivation of several distributions for the five-phase structure invariant, Φ_5. As with quartet invariant estimates using only the principal terms, the quintet has an estimated probability of zero when large principal terms, $|E_{\underline{h}}|$, $|E_{\underline{k}}|$, $|E_{\underline{l}}|$, $|E_{\underline{m}}|$ and $|E_{\underline{n}}|$ are used on their own.

Again, by taking the ten second neighbourhood magnitudes along with the five principal terms, Hauptman[28] has derived the conditional probability formula, $P_{15}^{\pm}$, for centrosymmetric structures.

$$P_{15}^{\pm} \simeq \frac{1}{K_{15}} . Z_{15}^{\pm} \qquad (8.11)$$

where $Z_{15}^{\pm}$ is defined in Appendix(III).

Schenk[29] has proposed a joint conditional probability distribution for quintets based on the purely exponential expression derived by Hauptman and Fortier[12]. For non-centrosymmetric structures the distribution is:

$$P_{1/15} \simeq \frac{1}{L} \exp[(6 - \Sigma R_{12}^2) 2E_5 \cos\Phi_5] \; X \; \prod_{10 \text{ terms}} I_o(2R_{12}Y_{12}) \qquad (8.12)$$

The pure exponential form can be expressed as:

$$P_{1/15} \simeq \frac{1}{L} \exp(\Delta R_1 R_2 R_3 R_4 R_5 \cos\Phi_5) \qquad (8.13)$$

Although not as accurate as the exponential-Bessel function form, it has a distinct advantage in giving the

parameter, Δ, the discriminant of Φ_5. The discriminant uses the cross-terms in a different manner. Instead of estimating the phase of Φ_5 as 0° or 180° the value of Δ is used as a guide to the reliability of Φ_5. For values of $\Delta<<0$ a negative quintet has an increased probability of being correct; and likewise with positive quintets where $\Delta>>0$. For values about zero the phase indication should be downweighted. The use of the discriminant has also been extended to centrosymmetric formulae.

We now have at our disposal several powerful formulae, $P_{1/7}$, $P_{1/13}$, $P_{1/15}$ and related centrosymmetric formulae, which are capable of giving multiple indications of the phase of the structure invariant, Φ. Triple-phase invariants, on the other hand, are limited to principal terms. A second neighbourhood is available to triplets by assigning arbitrary vectors, but at present the probabililty distributions available are unstable.

1.8.1 Figures of Merit Using Higher Invariants

The traditional figures of merit, ABSFOM and R_{Karle}, can be unreliable for symmorphic space groups. By using quartet and quintet invariants it is possible to have two additional figures of merit which are of a more discerning nature for these space groups.

NQEST[30], modified by Gilmore[31].

$$NQEST = \frac{\Sigma_{\underline{hklm}} \ \omega_{\underline{hklm}} \cos(\phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}} + \phi_{\underline{m}})}{\Sigma_{\underline{hklm}} \ \omega_{\underline{hklm}}}$$

where the summation is over all values of quartets predicted to have a modal value of π, and

$$\omega_{\underline{hklm}} = 1/\sigma^2_{\underline{hklm}}$$

whilst for the centrosymmetric case

$$\omega_{\underline{hklm}} = |1-2P^{+}|$$

NQINT is an analogous function to NQEST, but the summation is over all the available quintet invariants:

$$NQINT = \frac{\Sigma_{\underline{hklmn}} \ \omega_{\underline{hklmn}} \cos(\phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}} + \phi_{\underline{m}} + \phi_{\underline{n}})}{\Sigma_{\underline{hklmn}} \ \omega_{\underline{hklmn}}}$$

and

$$\omega_{\underline{hklmn}} = -\Delta/\sigma^{2}$$

In the centrosymmetric situation

$$\omega_{\underline{hklmn}} = -\Delta|1-2P^{+}|$$

Only those quintets having $\Delta \leqslant -1.0$ are included.

1.9 MAGIC INTEGERS

Magic integers[32] are mathematical devices by which a single symbol may be made to represent several phases.

1.9.1 The Basic Concept of Magic Integers

If three phases ϕ_1, ϕ_2, ϕ_3 are expressed in cycles so that

$$0 \leqslant \phi < 1$$

the following equations may be set up:

$$\phi_1 = 3x \bmod(1)$$

$$\phi_2 = 4x \bmod(1)$$

$$\phi_3 = 5x \bmod(1)$$

where, in this case, 3, 4 and 5 are called 'magic integers'. The proposition is that, no matter what the value of ϕ, for some value of the coefficient x in the range $0 \leqslant x < 1$ the three equations can be approximately satisfied. Thus three phases can be expressed in terms of one undefined variable. Similarly another three phases can be defined via magic-integers in terms of the variable y or z.

At the beginning of the phase expansion procedure there exist a number of reflexions that have had phase estimates assigned to them, i.e. origin and enantiomorph reflexions. There also exist a number of unknown phases that can be expressed in magic-integer form[33]. These reflexions are called primary (P) reflexions. From single triple phase relationships containing a pair of P reflexions a second set

of reflexions is obtained, also in symbolic magic integer form. These are secondary (S) reflexions. Having now obtained a larger starting set of reflexions comprising (P) and (S) reflexions expressed in magic-integer form, the problem is how to evaluate the probable values of the variables {x, y, ...}. This is accomplished by setting up a Fourier transform, the terms of which are derived from the triple-phase relationships involving the coefficients x, y,.... etc.

1.9.2 Ψ-maps

Consider a situation where a number of phases are represented in the following manner:

$\phi_1 = 3x \bmod(1)$ $\quad \phi_4 = 3y \bmod(1)$ $\quad \phi_7 = 3z \bmod(1)$

$\phi_2 = 4x \bmod(1)$ $\quad \phi_5 = 4y \bmod(1)$ $\quad \phi_8 = 4z \bmod(1)$

$\phi_3 = 5x \bmod(1)$ $\quad \phi_6 = 5y \bmod(1)$ $\quad \phi_9 = 5z \bmod(1)$

If there is a phase relationship

$$\phi_1 + \phi_3 + \phi_6 + \pi \simeq \emptyset \bmod(1)$$

this can be represented by

$$3x + 5x + 5y + \pi \simeq \emptyset \bmod(1)$$

or

$$8x + 5y + \pi \simeq \emptyset \bmod(1)$$

In general, any relationship involving the phases represented by magic-integers and fixed phases can be expressed as an equation of the type:

$$Hx + Ky + Lz + b \simeq \emptyset \text{mod}(1) \qquad (9.1)$$

The 'b' in equation (9.1) arises because of translational symmetry in certain space groups, since ø's are considered in one asymmetric unit in reciprocal space. Since there must now exist a number of Σ_2 relationships between (P) and (S) phases, these may be written in cosine form:

$$\cos\{2\pi(Hx + Ky + Lz + b)\} \ddagger 1 \qquad (9.2)$$

where $\ddagger$ means 'tends to be close to, but must be less than'. Relationships of this type which link the combined P and S sets give rise to the terms of a Fourier map, the peaks of which take into account the strength of the relationships by looking for maxima of the function

$$\Psi(x,y,z) = \sum_r |E_{1r}E_{2r}E_{3r}|\cos\{2\pi(H_r x + K_r y + L_r z + br)\} \qquad (9.3)$$

The maxima may be sought by evaluating $\Psi(x,y,z)$ over the range ø to 1 for each of the variables x, y and z. Adequate resolution can be obtained by evaluating the function at about four times the maximum index points along each axis; hence there is a requirement to keep the magic-integers as low as possible. A high peak in the Ψ-map can be translated into phases and these should satisfy the phase relationships linking the phases in the initial set.

However, White and Woolfson[32] have found it necessary to refine the initial individual phase angles obtained using the parameter shift technique. This method consists of taking the phases one at a time, changing their values in steps over a small range and shifting them to that value within the range which gives the maximum value of $\Psi(x,y,z)$. At this stage all available Σ_2 relationships are employed.

The origin reflexions are not allowed to change. By varying each phase independently the coupling of phases through magic integers is broken at this stage. The sets of phases so derived are used as starting points for tangent formula expansion and refinement.

It should be pointed out that there are two distinct uses of magic integer phase representation. Firstly, they can be used in the mode which has just been described, viz., in the setting up of equations which are solved in terms of a Fourier summation. Alternatively, they can be used solely in a phase permutation process.

A simple illustration of magic integer phase permutation is given in Figure 4. The 16 phase combinations produced by quadrant permutation on two unknown phases, ϕ_1 and ϕ_2, are plotted in Figure 4(a). A convenient magic integer sequence (2,3) is used to generate a set of lines as shown in Figure 4(b). ϕ_1 now takes only the four values $\pm\pi/4$ and $\pm 3\pi/4$, the corresponding values of ϕ_2 are shown plotted in the diagram. There are now only 12 phase combinations compared to the 16 derived by conventional means. There are two reasons for this reduction in number. The first is that the r.m.s. error of magic integer phases is slightly higher than the error of the phases produced by quadrant permutation. The second reason for a saving in the number of phase sets can be seen from Figure 4(b) where there is produced a much more efficiently packed lattice in phase space, and this will be true of any magic-integer sequence if defined as described by Main[34].

The reduction in the number of phase sets becomes very large as the number of variable reflexions, n, increases. Table 1 compares the number of sets produced by quadrant permutation with those produced by magic-integer sequences.

1.9.3 The Accuracy of Magic Integer Phase Representation

Main[35] has shown that magic-integer phase representation introduces errors in the derived phases. These errors are best determined as r.m.s. errors with a lower bound limit.

The r.m.s. error, $\Delta\phi_{rms}$, can be calculated as:

$$\Delta\phi_{rms} = 2\pi\sqrt{\frac{\bar{d}^2}{n}} \text{ radians} \qquad (9.4)$$

where n is the length of the sequence and $\bar{d}^2$ is the mean square distance of any point in (n-1) dimensional space. Since approximations have to be made for $\bar{d}^2$ the introduction of a lower bound limit of the r.m.s. error as calculated from (9.4) is found to be:

$$\Delta\phi_{lb} = 2\sqrt{\frac{\pi(n-1)}{n(n+1)}}\left[\frac{\Gamma(n+1)/2}{(\sum_i mi^2)^{1/2}}\right]^{1/(n-1)} \text{ radians}$$

1.9.4 Magic Integer Sequences

For an efficient magic integer sequence[35] the following criteria should be met:

a) The integers should be small.

b) The overall r.m.s. error should be small.

c) The r.m.s error should be divided evenly among the phases represented.

Since criteria a) and b) are mutually opposed a compromise must be made. In a practical sense there are four main rules which should be followed for a good magic integer sequence to result.

Firstly, when the largest integer is fixed, the remaining integers should be as large as possible (thus minimising $\Delta\phi_{rms}$). If the integers are also nearly equal, then the phase errors will be more evenly distributed. A second rule states that if any integer in the sequence is complemented, i.e. if m_i is changed to $m_n - m_i$, the effect is merely to reverse the direction of the ith. axis in n-dimensional space. Finally, if the sum or difference of two integers is also a member of the sequence, the r.m.s. error of that sequence will be higher than it would be otherwise.

A combination of these rules leads to the most efficient magic integer sequences found so far.

1.10 RANDOM PHASE SETS AND LINEAR EQUATIONS

A triple-phase relationship can be expressed as[37]:

$$\phi_h \pm \phi_k \pm \phi_l + b \simeq 0 \bmod(2\pi) \qquad (10.1)$$

Expressing phases in cycles and using the appropriate value of K as a weight, this may be transformed to:

$$K\phi_h \pm K\phi_k \pm K\phi_l \simeq K(n-b) \qquad (10.2)$$

where n is some, generally unknown, integer. If the integers are known then the whole system of such equations, normally far more numerous than the constituent phases, may be written in matrix-algebra notation as:

$$\underline{A}\phi = \underline{C} \qquad (10.3)$$

where $\underline{C} = K(\underline{n}-\underline{b})$

to give a least-squares solution:

$$\phi = (\underline{A}^T\underline{A})^{-1}\ \underline{A}^T\underline{C} \qquad (10.4)$$

Hence, if an approximate set of phases is available[37] then nearest integers may be found for the right-hand sides of the equations and a process of cyclic refinement carried out until the refinement reaches a natural termination when the integers do not change. Equation (10.4) has a larger radius of convergence than the tangent formula. This radius of convergence is such that it can be possible to start with a random set of phases which can converge to the correct value.

However, the use of integer values for phase estimates (in cycles) leads to a serious restriction on the use of (10.3) and (10.4). In a situation where phase relationships yield values in the range, say 1.48 - 1.52. For example, it is difficult to establish what integer values should be assigned to them. Apart from straight rejection of such values the sensible course is to introduce a weighting scheme whereby the equation is retained at its nearest integer value but given a rather low weight. The device for achieving this has been shown by Woolfson[36] to be a function of the term α, which is the departure from the nearest integer and satisfies:

$$-0.5 < \alpha < 0.5$$

such that

$$f(\alpha) = 2^{m-1}\alpha^m$$

where $m > 1$. The advantage of this type of weighting scheme is that, with K reintroduced, equation (10.2) is easily altered to:

$$K\phi \pm K\phi_k \pm K\phi_l = K[n-b+f(\alpha)] \qquad (10.6)$$

where only the vector $\underline{C}$ (10.3) is modified at each cycle of refinement. As a result of many experiments, Baggio *et. al.*

conclude that a good weighting function to use is $f(\alpha) = 4\,\alpha^3$. However, the introduction of a weighting scheme tends to obscure the point of completion of the least-squares refinement.

The initial starting sets which are used as input to the linear equations are obtained via a random number generating program which is available on most computers.

By far the most serious problem in using random phase sets and linear equations is how best to recognise the correct solution. This has been partially overcome with the use of negative quartet and negative quintet figures of merit.

REFERENCES

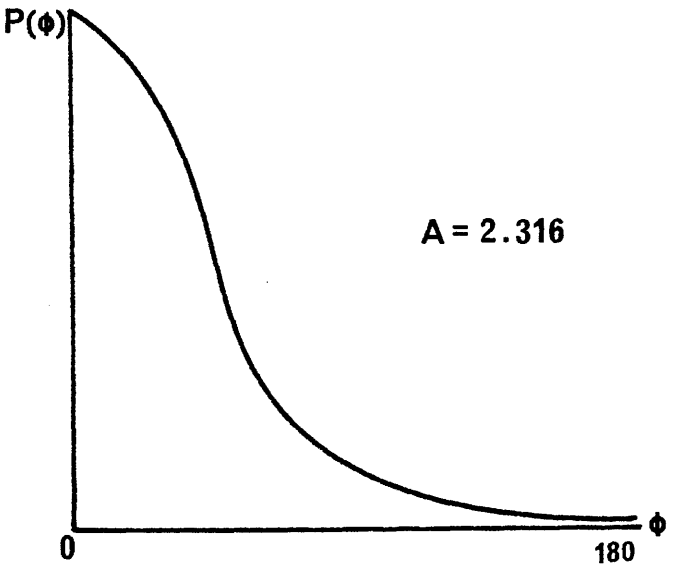
1. J. S. Kasper, C. M. Lucht and D. Harker, Acta Cryst. (1951) 3, 436.
2. H. Hauptman and J. Karle, (1953). Solution of the Phase Problem. I. The Centrosymmetric Crystal. Pittsburgh: Polycrystal Book Service.
3. H. Hauptman and J. Karle, Acta Cryst. (1956) 6, 131.
4. D. Sayre, Acta Cryst. (1952) 5, 60.
5. A. J. C. Wilson, Nature, (1942) pp. 151.
6. J. Karle and H. Hauptman, Acta Crys. (1953a) 6, 131.
7. H. Hauptman and J. Karle, Acta Cryst. (1956) 9, 45.
8. H. Hauptman and J. Karle, Acta Cryst. (1959) 12, 93.
9. H. Hauptman and J. Karle, Acta Cryst. (1961) 14, 217.
10. H. Hauptman, Acta Cryst. (1977) A33, 556.
11. H. Hauptman, Acta Cryst. (1977) A33, 565.
12. H. Hauptman and S. Fortier, Acta Cryst. (1977) A33, 575.
13. W. Cochran and M. M. Woolfson, Acta Cryst. (1955) 8, 1.
14. W. Cochran, Acta Cryst. (1955) 8, 473.
15. H. Hauptman, Acta Cryst. (1976) A32, 87.
16. H. Hauptman and J. Karle, Acta Cryst. (1954) 7, 369.
17. H. Hauptman and J. Karle , Acta Cryst. (1956) 9, 635.
18. P. Main, M. M. Woolfson, L. Lessinger, G. Germain and J. P. Declercq, MULTAN - A System of Computer Programs for Automatic Solution of Crystal Structures from X-ray Difraction Data. Universities of York , England and Louvain, Belgium.
19. G. Germain, P. Main and M. M. Woolfson, Acta Cryst. (1970) B26, 274.
20. G.Germain, P. Main and M. M. Woolfson, Acta Cryst. (1971) A27, 368.
21. W. Cochran and A. S. Douglas, (1957) Proc. R. Soc. London Ser. A, 227, 486.
22. J. Karle and I. Karle, Acta Cryst. (1966) 21, 849.
23. L. Lessinger, Acta Cryst. (1976) A32, 538.
24. H. Hauptman, Acta Cryst. (1977) A33, 553.
25. H. Hauptman, Acta Cryst. (1977) A33, 568.

28. S. Fortier and H. Hauptman, Acta Cryst. (1977) A33, 829.
29. N. Van Der Putten and H. Schenk, Acta Cryst. (1977) A33, 856.
30. G. T. DeTitta, J. W. Edmonds, D. A. Langs and H. Hauptman, Acta Cryst. (1975) A31, 472.
31. C. J. Gilmore, Acta Cryst. (1977) A33, 712.
32. P. S. White and M. M. Woolfson, Acta Cryst. (1975) A31, 53.
33. J. P. Declercq, G. Germain and M. M. Woolfson, Acta Cryst. (1975) A31, 367.
34. P. Main, Acta Cryst. (1978) A34, 750.
35. P. Main, Acta Cryst. (1977) A33, 31.
36. M. M. Woolfson, Acta Cryst. (1977) A33, 219.
37. R. Baggio, M. M. Woolfson, J. P. Declercq and G. Germain, Acta Cryst. (1978) A34, 883.

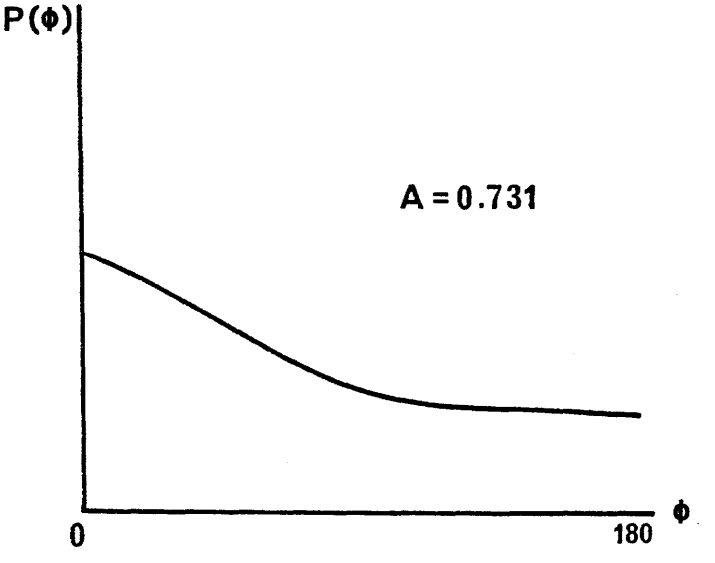
Table 1

n	Quadrant	$\Delta\phi_{rms}$	r=2	$\Delta\phi_{lb}$
1	4	26.0	4	26.0
2	16	26.0	12	29.3
3	64	26.0	28	32.5
4	256	26.0	60	35.2
5	1024	26.0	124	37.0
6	4096	26.0	252	38.4
7	16384	26.0	508	39.4
9	262144	26.0	2040	40.7
10	1048576	26.0	4092	41.1

Probability distribution (4.9)
for three-phase invariants



(a)



(b)

Figure 1

Nested neighbourhoods

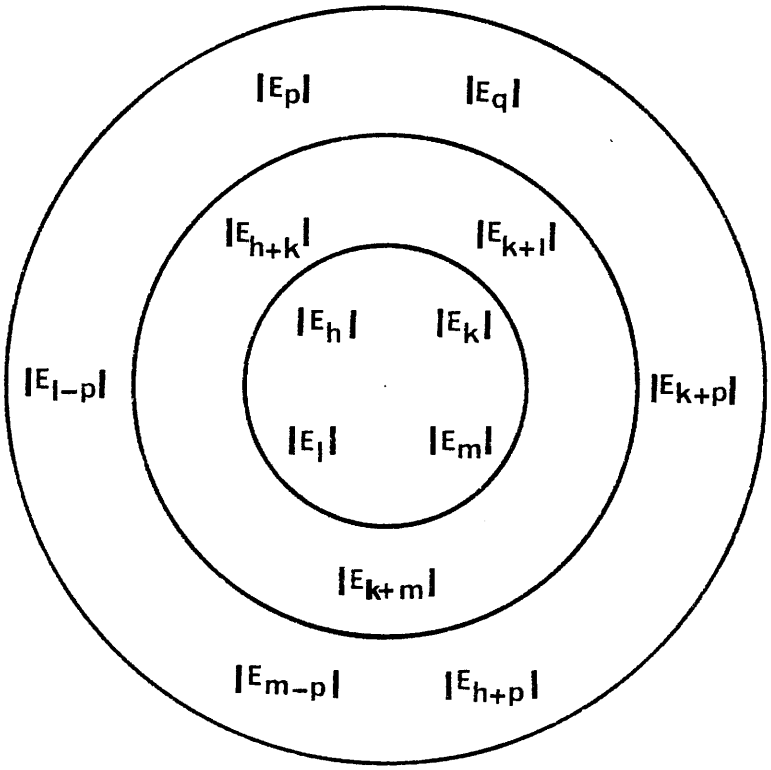
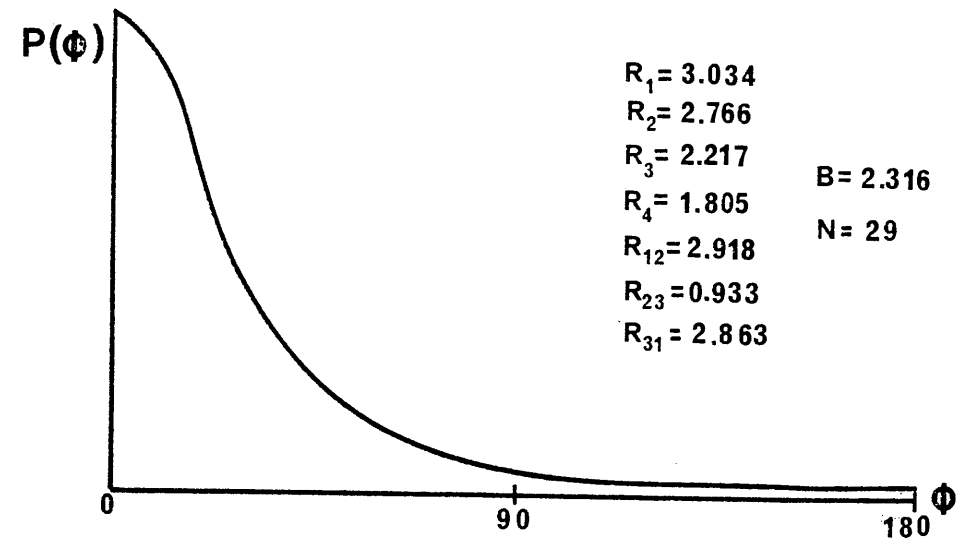
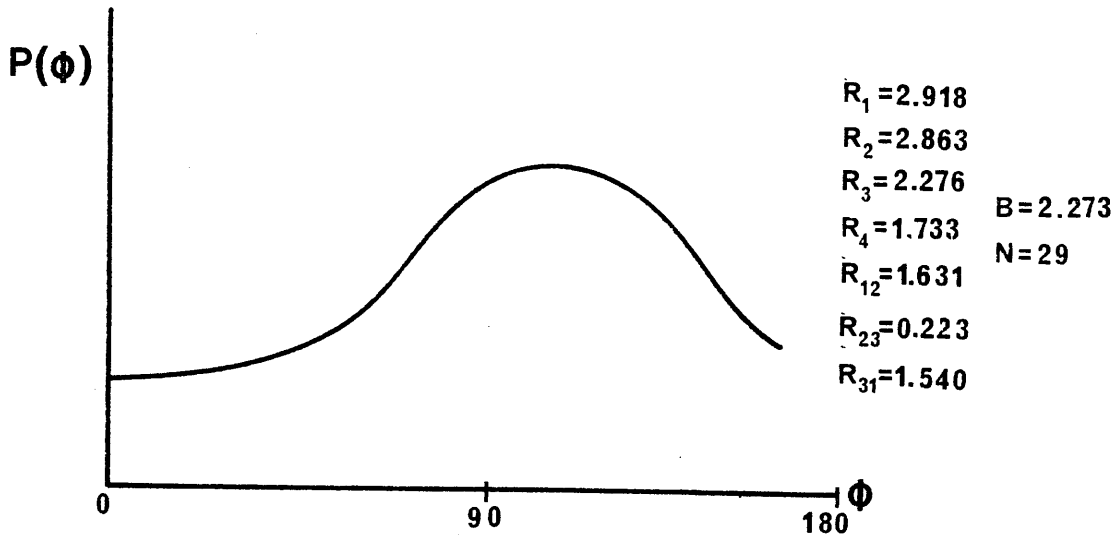


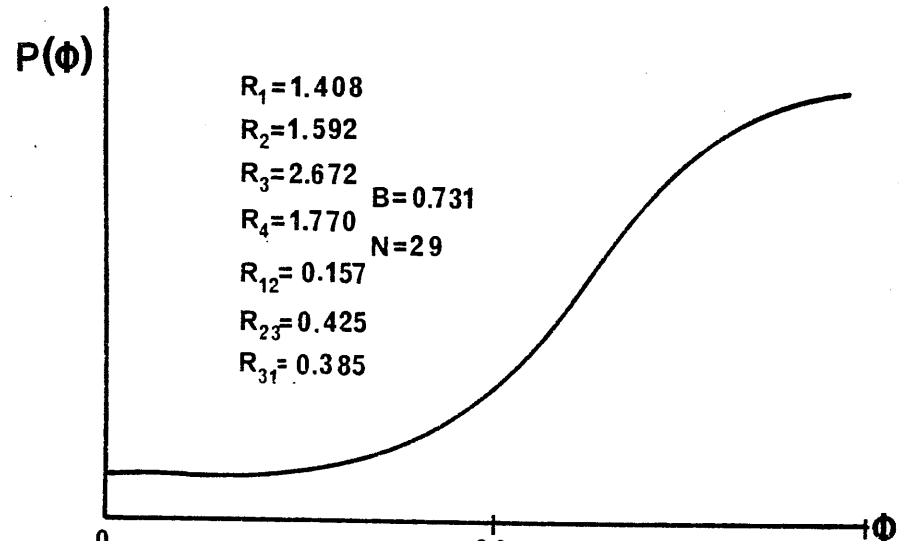
Figure 2


P(Φ)
R1= 3.034
R2= 2.766
R3= 2.217
R4= 1.805
R12= 2.918
R23 =0.933
R31 = 2.863
B= 2.316
N= 29
0
90
180
Φ

(a)


P(Φ)
R1 =2.918
R2 =2.863
R3 =2.276
R4 =1.733
R12 =1.631
R23 =0.223
R31 =1.540
B=2.273
N=29
0
90
180
Φ

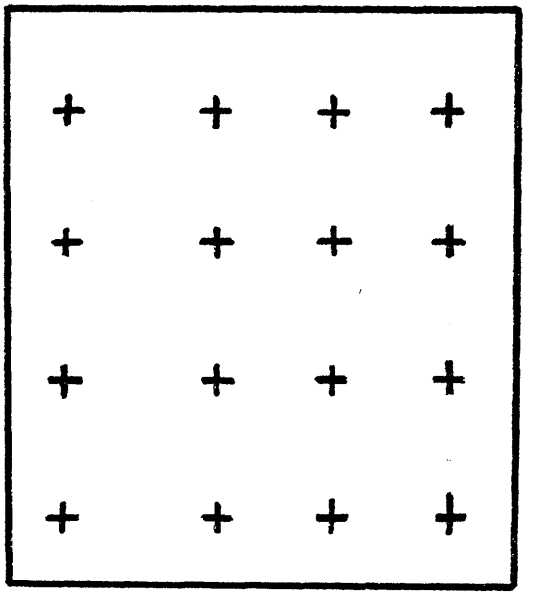
(b)


P(Φ)
R1 =1.408
R2 =1.592
R3 =2.672
R4 =1.770
R12 = 0.157
R23 =0.425
R31 = 0.385
B=0.731
N=29
0
90
180
Φ

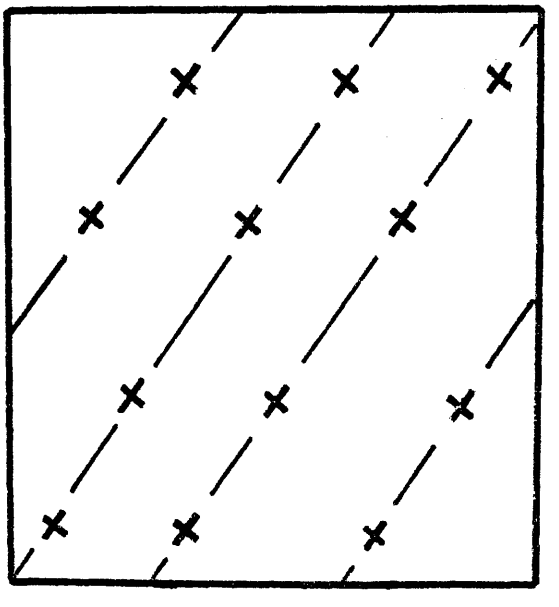
(c)

Figure 3

a) Quadrant and b) magic-integer representation in 2-dimensional phase space



(a)



(b)

Figure 4

Chapter 2

An Investigation of the P_7 and P_{13} Quartet Formulae

2.Ø INTRODUCTION

From Chapter 1, Section 1.8, it has been shown, via probability theory, that the more E-magnitudes available for quartet phase estimation, then the more reliable that estimate should be. The two simplest formulae[1] for estimating the sign of a quartet of reflexions and its associated probability in space group $P\bar{1}$ are the 7-magnitude two neighbourhood $P_7^{\pm}$, and the 13-magnitude, three neighbourhood $P_{13}^{\pm}$. In this chapter the reliability of these sign estimates, and derived probabilities, are investigated for both centrosymmetric formulae. An analysis of the invariant phase estimates gleaned from the non-centrosymmetric $P_{1/7}$ and the $P_{1/13}$ formulae is also carried out.

2.1 THE CENTROSYMMETRIC CASE:- $P_7^{\pm}$ and $P_{13}^{\pm}$

$P_7^{\pm}$ and $P_{13}^{\pm}$ were studied with a view to resolving the following questions:

(i) The relative reliabilities of the two formulae and which is most applicable to a given situation.

(ii) The limits of structural complexity each is capable of attacking.

(iii) The special problems associated with the third neighbourhood and with negative quartets.

To clarify these, and related problems, we consider first how $P_7^{\pm}$ and $P_{13}^{\pm}$ vary in reliability with structural complexity. Using this information, a separate discussion of both formulae follows, with reference to the special problems posed by

negative quartets and their reliable estimation.

2.1.1 The Test Data

Two data sets were generated in space group $P\bar{1}$ to a $\sin\theta$ limit corresponding to the complete copper sphere for two idealised structures containing 90 and 208 identical point atoms in the unit cell respectively. From these, two sets of quartets were generated using the programme QGEN under the following conditions:

(a) ANS N=90

13493 quartets were generated for which:

$$|E_{\underline{h}}|, |E_{\underline{k}}|, E_{\underline{l}}|, |E_{\underline{m}}| > 2.0$$

For the third neighbourhood:

$$|E_{\underline{p}}| \text{ and } |E_{\underline{q}}| > 2.0$$

(b) TETRA N=208

49500 quartets were generated for which:

$$|E_{\underline{h}}| > 3.01; \; |E_{\underline{k}}|, |E_{\underline{l}}|, |E_{\underline{m}}| > 2.0$$

For the third neighbourhood:

$$|E_{\underline{p}}| \text{ and } |E_{\underline{q}}| > 2.0$$

In both cases only those quartets satisfying the simultaneous conditions:

$$P_7^{\pm} \geqslant 0.7 \text{ or } \leqslant 0.3$$

and

$$P_{13}^{\pm} \geqslant \emptyset.7 \text{ or } \leqslant \emptyset.3$$

were accepted.

2.1.2 Reliability of $P_7^{\pm}$ and $P_{13}^{\pm}$ as a Function of B

Equation(8.5), Chapter 1, defines B in terms of the number of atoms in the unit cell, N, as well as the magnitudes of the principal E's involved in the particular invariant. Since B (and hence the probability distribution) is a function of the complexity of the structure under consideration, analyses of $P_7^{\pm}$ and $P_{13}^{\pm}$ as a function of B were undertaken. The reliability of the 7- and 13-magnitude formulae with respect to B is best demonstrated by a series of histograms in which the percentage failure of a quartet indication is plotted against B (A1-A4). In all the histograms a theoretical line is drawn representing the expected distribution of quartet failures as a function of B, assuming that the formulae hold exactly under the conditions specified. The lowest accessible B value under these conditions was ca. $\emptyset.2$.

In all cases, there is a tendency for quartets to be more reliably estimated than either formula would suggest and this bias is more pronounced for B values greater than unity. Of great importance, however, is the lack of any significant decrease in the reliability of quartet sign estimates at B values down to $\emptyset.2$, in any histogram. This is important for two reasons:-

(a) As structural complexity increases, the average B value for the quartet relationship will fall, and, if quartets are to make accessible structure analyses inaccessible via triple-phase invariants, it is important to be able to work confidently with quartets at low values of B.

(b) Table 1.1 summarises the availability of quartets having $P_7^{\pm} \geqslant \emptyset.7$ or $\leqslant \emptyset.3$ and $P_{13}^{\pm} \geqslant \emptyset.7$ or $\leqslant \emptyset.3$ as a function of B. The

largest number of such relationships is to be found in the range $\emptyset.6 \geqslant B \geqslant \emptyset.3$. If we are to make full use of the overdeterminacy of the phase problem, it is advantageous to be able to work with these quartets, even when it is not strictly necessary. An increase in the number of independent relationships can often lead to the correct solution for some difficult structures.

Having established that no detrimental bias exists with respect to B (for values down to $\emptyset.2$) in both formulae, we can now continue the consideration of $P_7^{\pm}$ and $P_{13}^{\pm}$ without specific reference to B limits.

2.1.3 The $P_7^{\pm}$ Formula

Tables adjacent to the histograms (A1-A4) summarise the failure rates in a form independent of B. It can be seen that there is a considerable tendency to underestimate the associated probability and it is more marked for ANS than for the larger TETRA. The negative quartets show this bias to a greater extent than their positive counterparts. These are shown in Histograms and Tables B1-B3.

2.1.4 The $P_{13}^{\pm}$ Formula

(a) The calculation of $P_{13}^{\pm}$

It is of obvious importance that the probability estimate associated with each quartet be as reliable as possible. $P_{13}^{\pm}$ employs a third neighbourhood containing six E-magnitudes;

$$|E_{\underline{p}}|, |E_{\underline{q}}|, |E_{\underline{h}+\underline{p}}|, |E_{\underline{k}+\underline{p}}|, |E_{\underline{l}+\underline{p}}|, |E_{\underline{m}+\underline{p}}|$$

where $\underline{p} + \underline{q} + \underline{l} + \underline{m} = \emptyset$

coupled with the 7-magnitudes from the 1st. and 2nd.

neighbourhoods.

Numerous neighbourhoods can be found that satisfy these criteria. For ANS, restricting $|E_q| \geq 2.0$ gave an average of 10 possible third neighbourhoods for each quartet, whereas, for TETRA, a similar constraint resulted in a 27 neighbour average with some quartets having over 100 possible third neighbourhoods. Each of these neighbourhoods gives a separate probability indication, and the problem arises how best to use this multiplicity. Three possibilities were investigated:

(i) Combine the n individual probability estimates, ${}^{j}P^{\pm}_{13}$, assuming that these estimates are independent, and hence derive a total estimate, P_t, via:

$$\frac{P_t}{(1-P_t)} = \frac{\left[\prod_{j=1}^{n}\right] {}^{j}P^{\pm}_{13}}{\left[\prod_{j=1}^{n}\right] (1-{}^{j}P^{\pm}_{13})}$$

Since each estimate shares the first and second neighbourhood, the assumption of independence is incorrect, and this manifests itself in a gross overestimation of the total probability. Indeed, for both ANS and TETRA ca. 30% of the quartets whose associated probabilities were calculated in this manner had net probabilities of 1.0 or 0.0.

(ii) Take that third neighbourhood which gave the highest associated probability, $P^{\pm}_{13}$, and use only this, at the same time combining it with the 7-magnitude probability estimate, again assuming independence;

$$\frac{P_t}{(1-P_t)} = \frac{P^{\pm}_{7} \cdot P^{\pm}_{13}}{(1-P^{\pm}_{7}) \cdot (1-P^{\pm}_{13})}$$

The results of these calculations showed the expected independence of B, but there was still a severe tendency to over estimate the associated probabilities. Again, this is a consequence of an incorrect assumption of independence.

(iii) Use the best third neighbourhood indication only, without the $P_7^{\pm}$ estimate. This appears to give the most reliable probability estimates, and it is this form of $P_{13}^{\pm}$ that is used in the remainder of the discussion.

(b) The Results of $P_{13}^{\pm}$

The $P_{13}^{\pm}$ formula represents a considerable advance over the 7-magnitude estimates for two reasons. As the tables and histograms (A1-A4) make clear the tendency to underestimate the probability is only slight with $P_{13}^{\pm}$ and hence gives a closer agreement with the theoretical failure rate.

Secondly, this better estimate is accompanied by a greater number of invariants being made available in each of the B ranges. Not only is the total number of quartets increased in each B range, but, as the probability range tends to unity, $P_{13}^{\pm}$ produces many more quartets than its $P_7^{\pm}$ variant. For example, in TETRA, $P^+ \geqslant 0.99$, the $^nP_{13}:^nP_7$ ratio is 522:3217, which represents a six-fold increase in available quartets. A similar effect is also exhibited by ANS. The ratio $^nP_7:^nP_{13}$ is defined as the ratio of the number of quartets in a given probability range estimated by $P_7^{\pm}$ to the number estimated via $P_{13}^{\pm}$. The $^nP_7:^nP_{13}$ ratios are summarised in Table 1.2.

2.1.5 The Special Problems of Negative Quartets

The essence of quartet invariants is their ability to give invariant magnitudes other than zero. If full use is to be made of this property, then it is important to obtain as

many negative quartets as possible with high associated estimates. As indicated in the previous sections, negative quartets have special problems associated with them. In view of their critical importance these difficulties are considered separately:

(i) It is more difficult to obtain a negative quartet with high associated probability than it is for a positive quartet, particularly with more complex structures. In consequence, it is hard to obtain reliable negative quartets at low B values. This is apparent in histograms B1-B3, where the relevant features are tabulated in a manner similar to that used in discussing positive quartets. In particular the complete absence of useful negative quartets at $B \geqslant 0.4$ should be noted.

(ii) The bias towards underestimating the associated probability is worse for negative quartets than for positive quartets: this can be seen from the graphs and adjacent tables (B1-B3), where the average failure is well below the theoretical line in both $P_7^{\pm}$ and $P_{13}^{\pm}$. In neither structure could negative quartets be found having $P^{+} \geqslant 0.99$. However, the ${}^{n}P_7 : {}^{n}P_{13}$ ratios in Table 1.3 show that $P_{13}^{\pm}$ increases the associated probabilities quite dramatically from their $P_7^{\pm}$ values, making more quartets accessible for use in phase determining procedures.

From the practical veiwpoint it is possible to solve $P\bar{1}$ structures using only the positive quartets; but such treatment will, in general, require a multisolution approach and hence remove one of the advantages that quartets have over triplets. In subsequent chapters it will be shown that by including only a small number of negative quartets (<100) the course of phase refinement is influenced significantly, with the result that the structure is obtained at the first attempt.

2.2.1 THE NON-CENTROSYMMETRIC CASE:- $P_{1/7}$ and $P_{1/13}$

With $P_7^{\pm}$ and $P_{13}^{\pm}$ the calculated invariants may be assigned only one of two values, $\emptyset$ or π , and therefore appropriate constraints may be applied. In $P_{1/7}$ and $P_{1/13}$, however, there can be no such constraints, and consequently a spread of errors is expected in the final invariant estimates. How these differences in phase errors between calculated and observed phases ($|\Phi_{obs}-\Phi_{calc}|$) are distributed over different standard deviation intervals was investigated by considering three known structures in space groups P1, $P2_1$ and $P2_12_12_1$.

As with the previous $P_7^{\pm}$ and $P_{13}^{\pm}$ formulae, an indication was required as to the reliability of the quartets generated, and subsequently several topics are discussed:

(i) A comparison of the phase estimate errors employing triplet and quartet invariants.

(ii) The relative reliabilities of the two formulae, $P_{1/7}$ and $P_{1/13}$.

(iii) The effect on the reliability if 1,2 or 3 of the second neighbourhoods are missing in the $P_{1/7}$ formula.

2.2.2 Experimental

(I) FOUTRA, N=62

The first structure was a phthalic anhydride, $C_{26}H_{16}O_5$, space group P1 with Z=1. It was one of the first molecules to be elucidated with the aid of quartets[2,3] and seemed an ideal structure with which to compare calculated and known phases. The known phases were supplied to QGEN and the following conditions used to generate quartets.

For $P_{1/3}$, 851 triplets were generated for which:

$$|E_{\underline{h}}|,|E_{\underline{k}}|,|E_{\underline{l}}| > 1.5$$

Using $P_{1/7}$ and $P_{1/13}$ 3291 quartets were found for which:

$$|E_{\underline{h}}|,|E_{\underline{k}}|,|E_{\underline{l}}|,|E_{\underline{m}}| > 1.50$$

For the third neighbourhood:

$$|E_{\underline{p}}| \text{ and } |E_{\underline{q}}| > 1.50$$

For $P_{1/7}$ with missing neighbourhoods 4700 quartets were derived which satisfied the above conditions.

(II) ACCAGE, N=188

The second structure considered, a novel chiral inclusion compound, is discussed more fully in Chapter 4. The molecule crystallises in space group $P2_1$ with Z=2.

For $P_{1/3}$ 108 triplets were generated for which:

$$|E_{\underline{h}}|,|E_{\underline{k}}|,|E_{\underline{l}}| > 2.2$$

$P_{1/7}$ and $P_{1/13}$ found 1281 quartets for which:

$$|E_{\underline{h}}|,|E_{\underline{k}}|,|E_{\underline{l}}|,|E_{\underline{m}}| > 2.2$$

For the third neighbourhood;

$$|E_{\underline{p}}| \text{ and} |E_{\underline{q}}| > 2.3$$

2097 quartets were generated <u>via</u> $P_{1/7}$ for up to three missing neighbourhoods using the conditions specified above.

(III) DITERP, N=50

The final structure examined was a diterpene, $C_{20}H_{26}O_3$, which crystallised in space group $P2_12_12_1$, Z=4. The structure was solved by A.Maltz using standard MULTAN techniques.

For $P_{1/3}$ 501 triplets were found for:

$$|E_{\underline{h}}|,|E_{\underline{k}}|,|E_{\underline{l}}| > 1.70$$

and $P_{1/7}$ and $P_{1/13}$ found 1353 quartets for which:

$$|E_{\underline{h}}|,|E_{\underline{k}}|,|E_{\underline{l}}|,|E_{\underline{m}}| > 1.70$$

For the third neighbourhood:

$$|E_{\underline{p}}| \text{ and } |E_{\underline{q}}| > 1.80$$

Using $P_{1/7}$ with missing neighbourhoods 391 quartets were generated employing the same criteria as above.

2.2.3 The Relative Reliability of Quartets via $P_{1/7}$ Compared to Triplets via $P_{1/3}$

Many of the factors affecting the reliabilities of quartet invariants, already discussed for the centrosymmetric case, are also relevant in the case of the three non-centrosymmetric structures. In particular, the approximate independence of B as a function of the quartet failure rate, as found for the centrosymmetric examples, is also manifest in the $P_{1/7}$ and $P_{1/13}$ results. In this event the quartet statistics were analysed by comparing the error between calculated and final phase values at different intervals of σ (equation 4.12, Chapter 1). Tables 2.1, 2.2 and 2.3 summarise these results.

For all structures studied, and for all formulae used,

there was a general increase in invariant error as σ increased. Within the range of E-magnitudes used , there was a marked increase of available invariants when going from $P_{1/3}$ to $P_{1/7}$. In FOUTRA, for example, only 851 triple-phase invariants were found, whereas 3284 quartets were generated for the same lower bound value of σ . ACCAGE produced a 10-fold increase, whilst DITERP shows a less dramatic 2-fold increase in available invariants.

2.2.4 The $P_{1/13}$ Formula

As with $P_{13}^{\pm}$, the $P_{1/13}$ estimate gives rise to several third neighbourhoods which will contribute to the final invariant magnitude. To determine the usefulness of the 3rd. neighbourhood formula a detailed listing of all third neighbourhood information was generated for two of the structures under discussion, viz. FOUTRA and DITERP. The results showed that where several 3rd. neighbourhood contributors were employed, occasionally one or more discrepant neighbourhoods were found for the $P_{1/13}$ estimate. A discrepant neighbourhood can be defined as follows. As previously described for the centrosymmetric case (section 2.1.4(a)) several (often as many as 20) 3rd. neighbourhoods were available as contributors to the final phase estimate for a quartet. In many instances there were one or more estimates (discrepancies) having different values from the majority of calculated phase estimates. The differences can range from 30° to 180°. When a discrepant relationship occurred the quartet estimate was likely to be poor. The greater the number of discrepancies between 3rd. neighbourhoods, the greater the possibility of the quartet estimate being incorrect.

Tables 2.4, 2.5, 2.6 and 2.7 list the number of quartets calculated <u>via</u> $P_{1/13}$ with values of 0 or π for <u>all</u> third neighbourhood contributors and display them with a percentage error to their final refined values at error intervals of 20,

40 and 60 degrees. Also tabulated are the statistics for invariants where one of the 3rd. neighbourhood estimates is wrong.

2.2.5 $P_{1/7}$ - Missing neighbourhoods

Tables 2.8, 2.9 and 2.10 show the effect on the quartet phase estimates when one or more of the second neighbourhoods are missing from the $P_{1/7}$ calculation. The results conclude that there is a worthwhile increase in the availability of higher invariants and this could prove useful in data sets which are poorly resolved and subsequently few reliable Σ_2 relationships are available. The distributions derived by Heinerman[4] were used for the calculation of $P_{1/7}$ with missing neigbourhoods.

2.3 Advantages and Disadvantages of the 13-magnitude Formula over the 7-magnitude Formula.

(a) Advantages

It is now possible to consider the relative merits of the two formulae. The P_{13} formula has the following advantages over its P_7 counterpart.

(i) It gives a more reliable estimate for both positive and negative quartets, although the underestimation still exists.

(ii) There is a considerable increase in the availability of quartets $({}^{n}P_7 : {}^{n}P_{13})$ having a high associated probability. This increase is most dramatic in the range $1.0 \geq P^{+} \geq 0.99$.

(iii) There is also a pronounced increase in the availability of negative quartets. Without $P_{13}^{\pm}$ no quartets having $P^{+} \leq 0.05$ are available for either structure.

(iv) In cases of limited data, the multiplicity of third neighbourhoods can generate strong indications in situations where P_7 alone is of limited use.

(v) Multiplicity of individual estimates also enhances the phase estimation procedure.

(b) Disadvantages

The principal disadvantage of P_{13} is the computing overhead. The determination of all possible third neighbourhoods is equivalent to the computation of all accessible trios involving the quartet under examination. This is a very time consuming process. By allowing constraints to be levied on $|E_p|$ and $|E_q|$ this increase in time can be partially alleviated without reducing, significantly, the accuracy of the formula.

2.4 The Limits of the Formulae

It is a difficult and unreliable process to extrapolate the results of the two analyses to more complex situations, but several general conclusions may be drawn.

It is clear that either of the formulae is capable of solving both structures. $P_7^{\pm}$ is adequate for ANS, but less so for TETRA. In this case the greater number of reliable relationships made available by the third neighbourhood make it more readily accessible to the $P_{13}^{\pm}$ formula.

If, as a general guide, we require 40 phase relationships per atom, then for ANS (N=90) we need ca. 3600 relationships and for TETRA (N=208) ca. 8000. Let us take as 'useful' those quartets having $P^+ \geqslant 0.90$. For the $P_7^{\pm}$ formula this requires us to use quartets with lower bound limits of B set at 1.0 and 0.4 for ANS and TETRA respectively. For $P_{13}^{\pm}$ these limits are

raised by Ø.1 to 1.1 and Ø.5. (A difference of Ø.1 in B may seem small, but it should be remembered that the number of available quartets increases exponentially with decrease in B). Simple extrapolation of these results to a B limit of Ø.2, using the histograms, indicates that structures having N=5ØØ should be accessible to $P^{\pm}_{13}$ and structures having N=3ØØ possibly accessible to $P^{\pm}_{7}$. The B limit is somewhat arbitrarily chosen - if the formulae hold up at B=Ø.1 then structures of twice this complexity may well prove amenable to treatment <u>via</u> quartet invariants, assuming that high quality, high resolution data are available. In the context of this thesis the largest $P\bar{1}$ structure solved <u>via</u> quartet invariants was a squalene inclusion compound where N=186.

REFERENCES

1. See references 11 and 15 Chapter 1.
2. C. J. Gilmore, Acta Cryst. (1977) A33, 712.
3. A. A. Freer, C. J. Gilmore, D. M. Mant and J. McCormick, J. C. S. Chem. Comm. (1977) 296.
4. J. J. Heinerman. In Direct Methods in Crystallography, edited by H. Hauptman, pp. 174-194. Pittsburgh:Polycrystal.

Table 1.1

Distribution of quartets as a function of B

B-limits	Number of quartets ANS (N=90)	TETRA (N=208)	TOTAL
>1.6	796	35	831
1.6-1.2	1802	217	2019
1.2-1.0	2119	527	2646
1.0-0.8	3272	1546	4818
0.8-0.6	3918	5358	9276
0.6-0.4	1586	21933	23519
0.4-0.2	2	19884	19886

Table 1.2

Ratio of P^{+}_{13} to P^{+}_{7}

Probability range(P^{+})	P^{+}_{13}/P^{+}_{7} ANS	TETRA
>0.80	1.34	1.03
>0.90	1.70	2.54
>0.95	2.02	3.10
>0.99	3.10	6.16

Table 1.3

Ratio of P^{-}_{13} to P^{-}_{7} for negative quartets only.

Probability range(P^{+})	P^{-}_{13}/P^{-}_{7} ANS	TETRA
<0.20	3.00	11.56
<0.10	11.82	
<0.05		-

Table 2.1

FOUTRA:- Triplets and quartets for top 173 |E|'s, no missing neighbourhoods.

	$P_{1/3}$			$P_{1/7}$			$P_{1/13}$		
	$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$			$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$			$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$		
Sigma range	<MODE>	<MEAN>	N	<MODE>	<MEAN>	N	<MODE>	<MEAN>	N
21.2-28.3	24.2	20.9	10	-	-	-	-	-	-
28.3-35.3	32.3	19.3	125	32.9	24.3	14	41.9	24.9	124
35.3-42.4	40.7	26.7	230	41.6	25.1	184	47.7	30.3	504
42.4-49.5	44.5	27.3	279	45.6	28.8	457	51.3	31.6	752
49.5-56.6	53.8	33.9	179	50.4	31.8	667	60.6	36.6	770
56.6-63.6	71.4	44.9	28	59.1	37.1	894	65.2	39.8	728
63.6-70.7	-	-	-	70.5	39.2	1068	71.4	39.9	395
Number of invariants			851			3284			3273

Table 2.2

ACCAGE:- Triplets and quartets for top 100 |E|'s, no missing neighbourhoods.

	$P_{1/3}$			$P_{1/7}$			$P_{1/13}$		
	$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$			$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$			$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$		
Sigma range	<MODE>	<MEAN>	N	<MODE>	<MEAN>	N	<MODE>	<MEAN>	N
28.3-35.3	26.7	30.1	32	26.2	22.7	61	25.3	20.9	122
35.3-42.2	37.2	34.5	66	36.2	31.7	185	36.2	31.2	229
42.2-49.5	53.2	38.8	10	41.8	36.0	193	38.1	29.7	186
49.5-56.6	-	-	-	45.4	38.8	220	55.4	42.8	177
56.6-63.6	-	-	-	62.2	45.6	235	55.3	42.1	168
63.6-70.7	-	-	-	66.8	47.7	387	65.4	48.6	314
Number of invariants			108			1281			1196

Table 2.3

DITERP:- Triplets and quartets for top 115 |E|'s, no missing neighbourhoods.

	$P_{1/3}$			$P_{1/7}$			$P_{1/13}$		
	$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$			$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$			$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$		
Sigma range	<MODE>	<MEAN>	N	<MODE>	<MEAN>	N	<MODE>	<MEAN>	N
28.3-35.3	47.2	46.4	5	-	-	-	18.Ø	1Ø.Ø	2
35.3-42.4	29.4	23.8	29	26.6	25.3	5	48.5	32.6	81
42.4-49.5	37.3	24.1	85	49.9	29.9	84	5Ø.9	32.2	192
49.5-56.6	44.7	29.9	2Ø5	43.7	28.5	2Ø6	53.6	32.5	3Ø9
56.6-63.6	48.4	32.2	285	55.Ø	34.7	4Ø4	59.4	37.2	395
63.6-7Ø.7	48.8	3Ø.9	47	62.8	4Ø.4	654	62.6	4Ø.5	348
Number of invariants			656			1353			1327

Table 2.4

FOUTRA:- Quartets using $P_{1/7}$ with up to two missing neighbourhoods.

	NO MISSING NEIGHBOURHOODS			ONE MISSING NEIGHBOURHOOD			TWO MISSING NEIGHBOURHOODS		
	$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$			$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$			$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$		
Sigma range	<MODE>	<MEAN>	N	<MODE>	<MEAN>	N	<MODE>	<MEAN>	N
28.3-35.3	32.0	22.6	4	53.5	32.1	6	20.5	13.3	4
35.3-42.4	33.4	17.5	25	31.8	22.1	82	33.4	18.4	51
42.4-49.5	39.7	30.1	92	45.6	29.5	252	40.3	27.4	119
49.5-56.6	46.7	31.9	143	49.3	30.4	377	48.8	34.1	208
56.6-63.6	50.4	35.2	202	57.4	35.5	540	56.8	35.0	230
63.6-70.7	57.8	40.6	324	69.6	41.6	960	72.9	44.1	399
Number of invariants			770			2217			1011

Table 2.5

ACCAGE:- Quartets using $P_{1/7}$ with up to two missing neighbourhoods.

	NO MISSING NEIGHBOURHOODS			ONE MISSING NEIGHBOURHOOD			TWO MISSING NEIGHBOURHOODS		
	$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$			$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$			$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$		
Sigma range	<MODE>	<MEAN>	N	<MODE>	<MEAN>	N	<MODE>	<MEAN>	N
28.3-35.3	26.4	22.5	61	43.7	42.3	20	-	-	-
35.3-42.4	36.8	32.0	188	44.0	42.0	100	51.1	50.7	8
42.4-49.5	41.2	35.7	190	57.3	50.8	107	14.0	30.0	2
49.5-56.6	45.9	38.6	222	64.4	52.6	132	67.6	62.8	16
56.6-63.6	61.3	45.2	236	61.6	47.0	174	77.4	54.2	19
63.6-70.7	67.2	48.1	391	81.1	53.9	206	73.8	51.6	22
Number of invariants			1288			739			67

Table 2.6

DITERP:- Quartets using $P_{1/7}$ with up to two missing neighbourhoods.

	NO MISSING NEIGHBOURHOODS			ONE MISSING NEIGHBOURHOOD			TWO MISSING NEIGHBOURHOODS		
	$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$			$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$			$\frac{\Sigma\lvert\Phi_{obs}-\Phi_{calc}\rvert}{N}$		
Sigma range	<MODE>	<MEAN>	N	<MODE>	<MEAN>	N	<MODE>	<MEAN>	N
35.3-42.4	34.0	23.4	2	57.5	49.5	4	-	-	-
42.4-49.5	95.3	63.0	14	35.6	25.0	10	-	-	-
49.5-56.6	63.6	40.2	16	74.9	51.1	31	32.9	14.4	15
56.6-63.6	54.8	30.0	39	70.1	50.9	32	31.4	27.5	39
63.6-70.7	79.0	46.8	88	79.7	51.7	59	59.4	39.7	42
Number of invariants			159			136			96

Table 2.7

3rd. neighbourhood Discrepancies.

FOUTRA:- Positive quartets only. Derived from $P^{\pm}_{1/13}$

(a) Where all indications are 0°.

Deviation($^{\circ}$) from zero	20	40	60	>60
Number correct	167	134	93	181
% correct	29	52	69	31

(b) Where all but one indication are 0°.

Deviation($^{\circ}$) from zero	20	40	60	>60
Number correct	44	30	27	67
% correct	26	30	58	42

Table 2.8

FOUTRA:- Negative quartets only. Derived from $P^{\pm}_{1/13}$

(a) Where all indications are 180°.

Deviation($^{\circ}$) from 180°	20	40	60	>60
Number correct	30	31	25	64
% correct	20	41	57	42

(b) Where all but one of the indications are 180°.

Deviation($^{\circ}$) from 180°	20	40	60	>60
Number correct	5	7	10	34
% correct	9	13	18	60

Table 2.9

DITERP:- Positive quartets only. Derived from $P^{\pm}_{1/13}$

(a) Where all the indications are zero.

Deviation(°) from zero	20	40	60	>60
Number correct	226	211	155	354
% correct	24	46	63	37

(b) Where all but one indication is zero.

Deviation(°) from zero	20	40	60	>60
Number correct	30	28	25	70
% correct	20	38	54	46

Table 2.10

DITERP:- Negative quartets only. Derived from $P^{\pm}_{1/13}$

(a) Where all the indications are 180°.

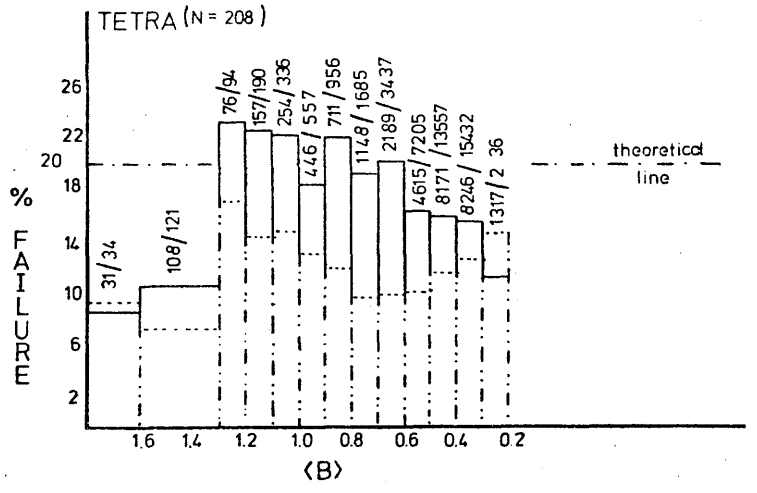
Deviation(°) from 180°	20	40	60	>60
Number correct	6	1	4	9
% correct	30	1	55	45

(b) Where all but one of the indications are 180°.

There were insufficient invariants available for analyses.

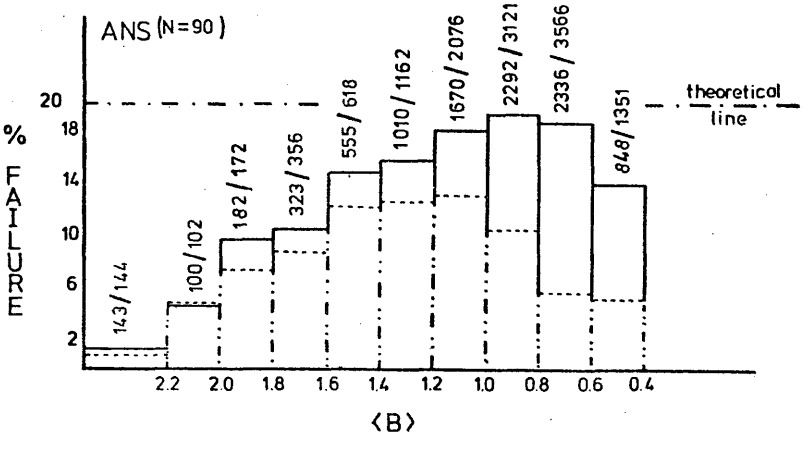
A1 :- P_7^{1+} and P_{13}^{1+} $P^+ > 0.8$ or $P^+ \leq 0.2$

$P_{13}^{\pm}$ CONDITIONS :- Best 3rd. neighbourhood only



	$P_7^{\pm}$	$P_{13}^{\pm}$
Total number of quartets	27469	28195
" " " failures	3253	3009
Average (%)	11.8	10.8

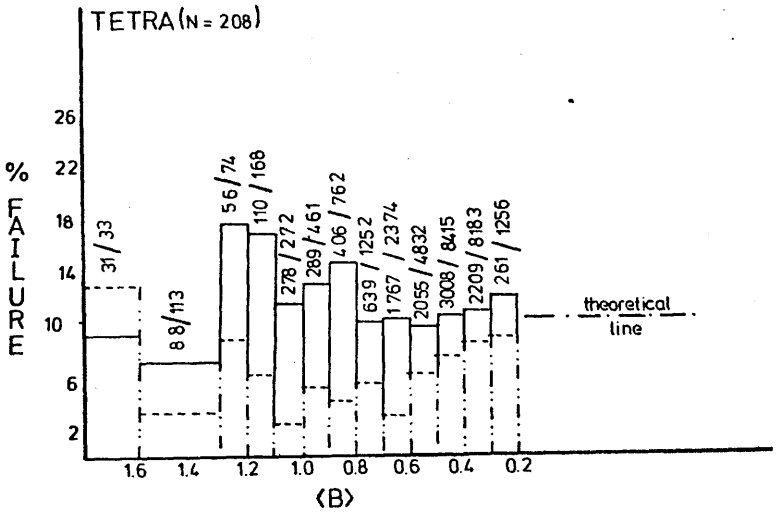
$P_7^{\pm}$; $P_{13}^{\pm}$



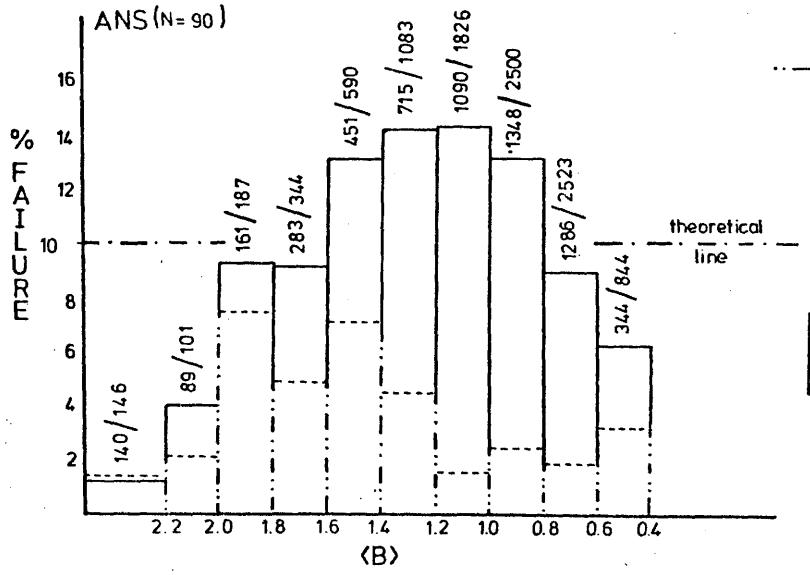
	$P_7^{\pm}$	$P_{13}^{\pm}$
Total number of quartets	9458	12688
" " " failures	910	2192
Average (%)	9.6	17.3

A2 :- P_7^+ and P_{13}^+ $\quad P^+ \geq 0.9$ or $P^+ \leq 0.1$

P_{13}^+ CONDITIONS :- Best 3rd. neighbourhood only



	P_7^+	P_{13}^+
Total number of quartets	11192	28195
" " " failures	705	3009
Average (%)	6.3	10.7

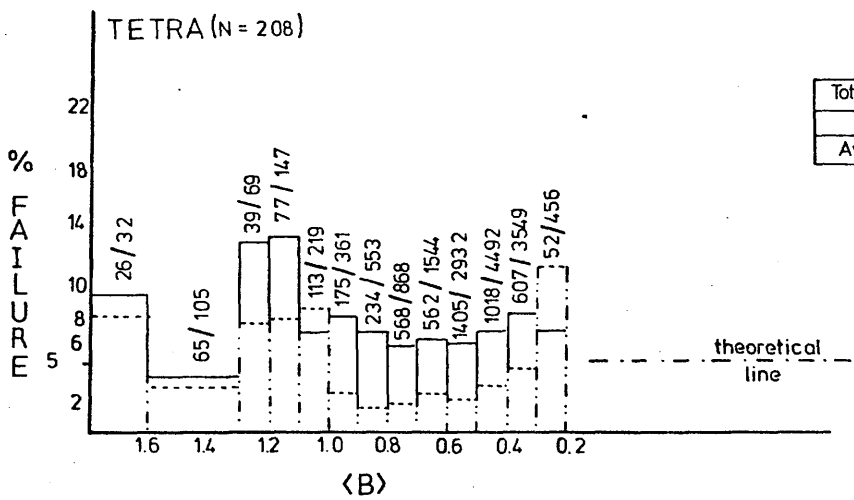


$\cdots\cdots P_7^+$; $\text{——} P_{13}^+$

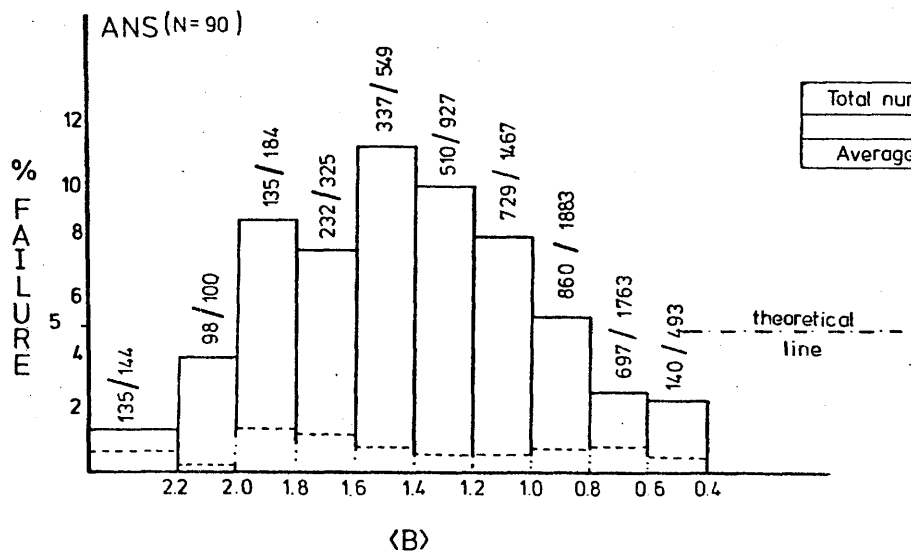
	P_7^+	P_{13}^+
Total number of quartets	5957	10142
" " " failures	180	1179
Average (%)	3.0	11.6

A3 :- P_7^+ and P_{13}^+ $P^+ \geq 0.95$ or $P^+ \leq 0.05$

P_{13}^+ CONDITIONS :- Best 3rd. neighbourhood only



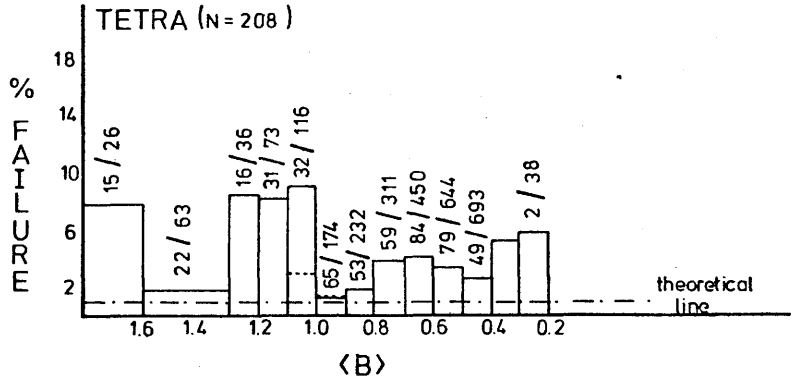
	P_7^+	P_{13}^+
Total number of quartets	4941	15322
failures	167	1091
Average (%)	3.4	7.1



	P_7^+	P_{13}^+
Total number of quartets	3873	7840
failures	35	491
Average (%)	0.9	6.3

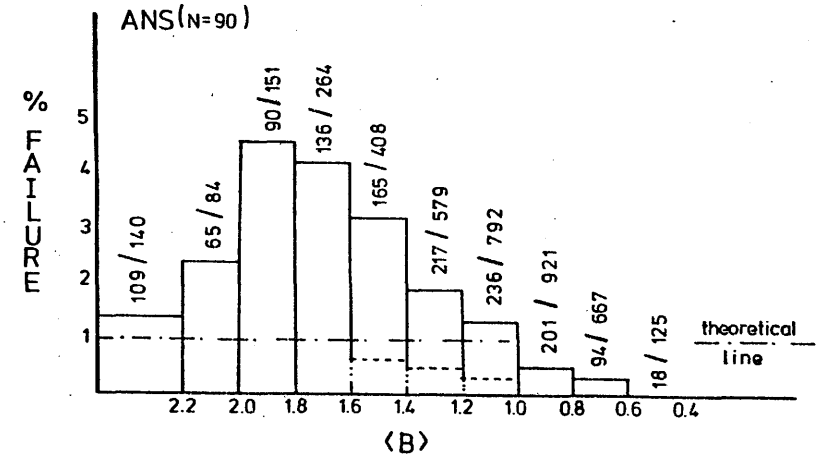
A4 :- $P_7^{\pm}$ and $P_{13}^{\pm}$ $P^{+} \geq 0.99$ or $P^{+} \leq 0.01$

P_{13}^{+} CONDITIONS :- Best 3rd. neighbourhood only



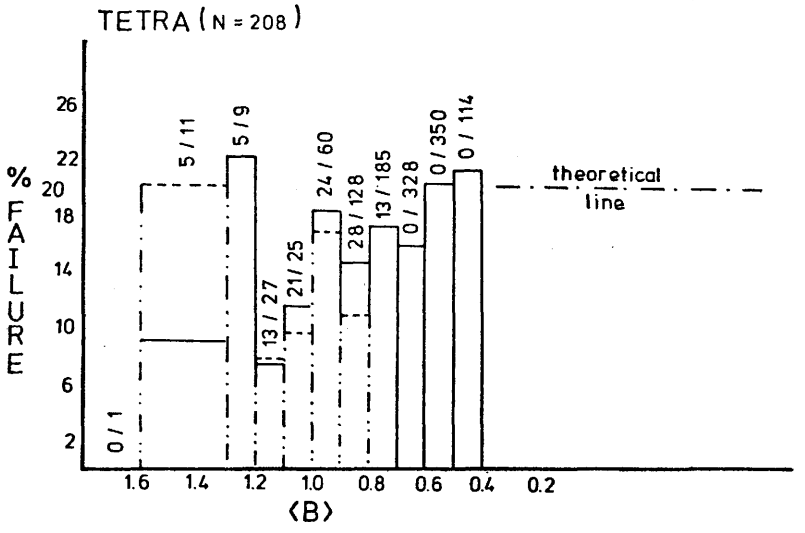
	$P_7^{\pm}$	$P_{13}^{\pm}$
Total number of quartets	522	3217
" " " failures	6	103
Average (%)	1.1	3.2

$\cdots\cdot\cdots\cdot P_7^{\pm}$; —— $P_{13}^{\pm}$



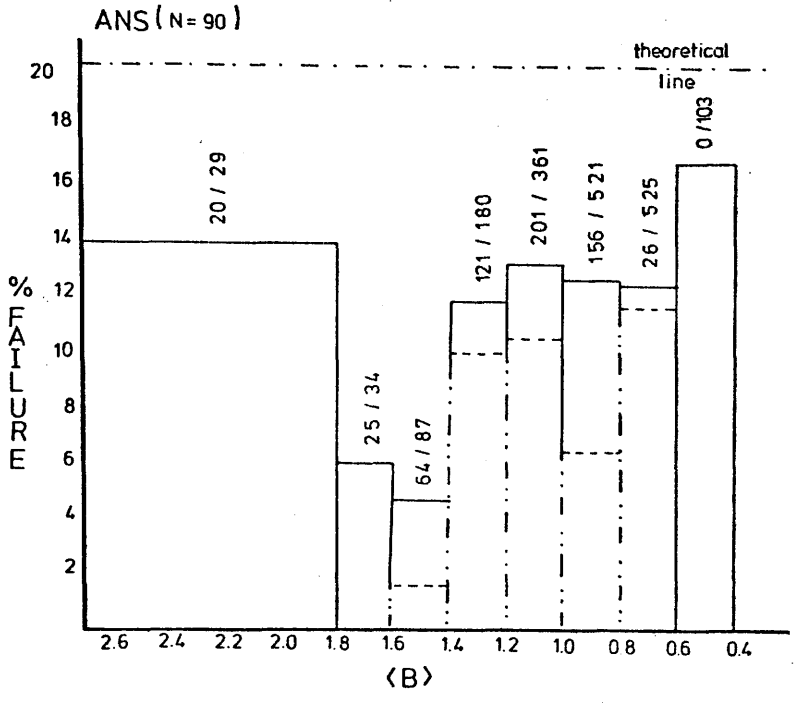
	$P_7^{\pm}$	$P_{13}^{\pm}$
Total number of quartets	1331	4131
" " " failures	4	63
Average (%)	0.3	1.5

B1 :- P_7^- and P_{13}^- $\quad P^+ \leq 0.2$



	P_7^-	P_{13}^-
Number of quartets	109	1260
" " failures	11	210
Average (%)	10.1	17.4

P_7^- ; P_{13}^-

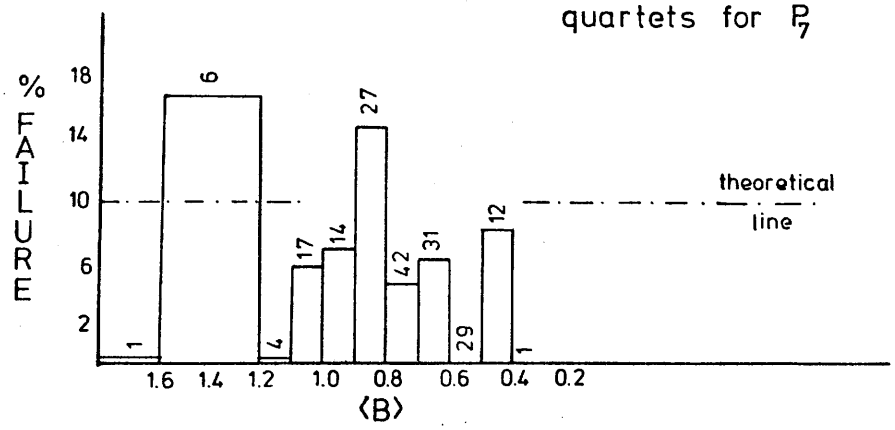


	P_7^-	P_{13}^-
Number of quartets	613	1840
" " failures	47	222
Average (%)	7.7	12.2

B 2 :- P_7^- and P_{13}^- $P^+ \leq 0.1$

TETRA (N = 208)

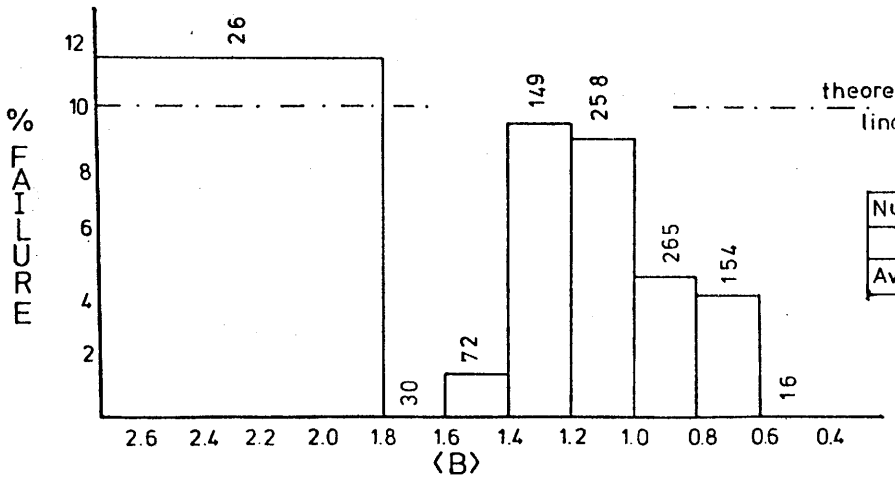
There were no negative quartets for P_7



	P_{13}^-
Number of quartets	184
" " failures	12
Average (%)	6.5

·· — ·· — P ——— P

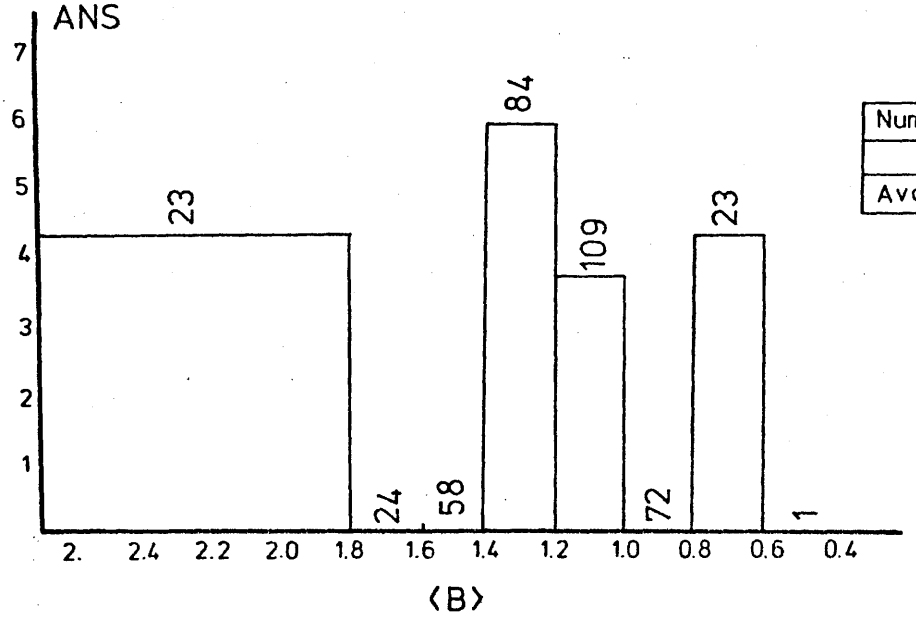
ANS (N = 90)



	P_7^-	P_{13}^-
Number of quartets	82	970
" " failures	1	57
Average (%)	-	6.1

B 3 :- P_{13}^{-} only. $P^{+} \leq 0.05$

TETRA There was only one negative quartet in this range (it was correct).



	P_{13}^{-}
Number of quartets	394
" " failures	11
Average (%)	2.8

For $P^{+} \leq 0.01$ there were 49 quartets, all of which were correct.

Chapter 3

The Use of Quartets in Small Structures:

A: Structure Determination of Three Benzazete Dimers.

B: Structure Determination of an Ylide:Picric Acid Complex.

A: CRYSTAL STRUCTURE SOLUTION OF THREE ANGULAR DIMERS

3.1.Ø INTRODUCTION

Crystal structures in symmorphic space groups are traditionally the most difficult problems to solve *via* direct methods, since, as discussed in Chapter 1, the lack of translational symmetry tends to make the process ill-conditioned, and the lack of equivalent reflexions can give rise to a paucity of sign relationships of high associated probability. Such factors force a need for a relatively large number of possible solutions from which the correct set may be difficult to extract. The strategy outlined below was therefore employed in the solution of two benzazete dimers, (IIIa) and(IIIc), which crystallize in the space group $P\bar{1}$. In part B the crystal structure solution of a picric acid : ylide complex is described, where the use of quartet invariants has been expanded to structures in any space group.

3.1.1 AUTOMATION OF QUARTET INVARIANT ANALYSIS

As the first development of a standard program package whereby quartet invariants could be used routinely in structure determination, a modification to the program PHASE[1] in the X-ray 72 System[2] of Stewart *et al*. (1972) was carried out, allowing these higher invariants to be used in an automatic procedure. Used conventionally, PHASE generally produces a single solution for space groups having translational symmetry, where it employs quartets (calling them 'relationships of the second kind') which are derived *via* an overlap of two triplets, as demonstrated in Section 1.7 of Chapter 1. Since all the associated E-magnitudes are large, all such derived quartets in symmorphic space groups will have

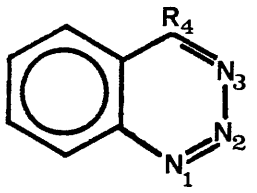
a value of 0° for Φ_4, with the result that the problems discussed previously still remain.

It is, however, a simple task to adapt PHASE to read quartets calculated by the quartet generation program QGEN, utilising $P_7^{\pm}$ and $P_{13}^{\pm}$, instead of computing them *via* triplet overlap. The presence of negative quartets ($\Phi_4 = \pi$) alleviates the lack of translational symmetry and, as seen in Chapter 2, there is a large increase in the number of available phase relationships of high reliability (albeit these are interdependent).

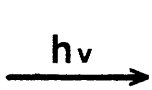
For structures (IIIa) and (IIIc) this procedure was used to phase the top 100 E-magnitudes and, in both cases, a single phase set was produced which revealed the complete crystal structure.

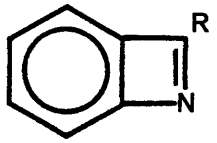
3.1.2 THREE BENZAZETE DIMERS

Photolysis of the 4-aryl- and 4-alkyl-1,2,3-benzotriazines (I,a-c) yielded the products (IIIa-c), which result from the dimerisation of the corresponding benzazetes (IIa-c). However, the structures of the dimers (IIIa-c) were in doubt because of the difficulty of differentiating clearly between the possible regioisomeric angular dimers, IV-VI[3,4]. The analysis of the three compounds (IIIa-c) was undertaken to confirm the structures shown, and to reveal that, at least in those cases studied, different 4-substituents in the 1,2,3-benzotriazenes (Ia-c) have not resulted in the production of different isomers.



Ia-c



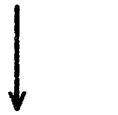


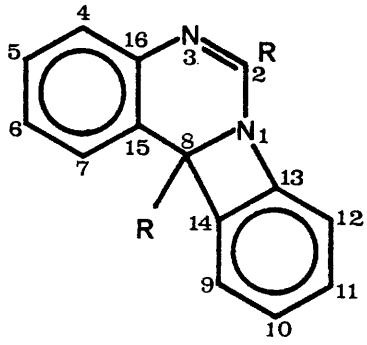
IIa-c

a) R= Ph

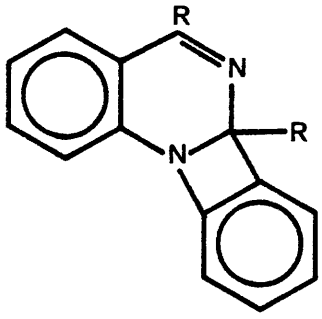
b) R= p-MeOC$_6$H$_4$

c) R= t-butyl

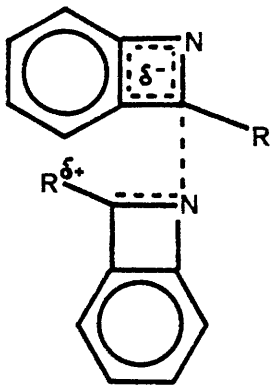




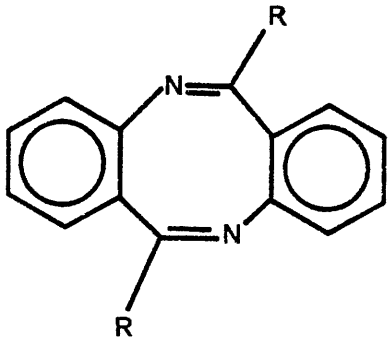
IIIa-c



IV



V



VI

3.1.3 EXPERIMENTAL

(IIIa)

Crystal data

Angular dimer of 2-phenylbenzazete, $C_{26}H_{18}N_2$, M_r=358.6, triclinic, a=6.711(1), b=12.456(2), c=11.796(2) Å, α=87.0(1), β=78.2(1), γ=74.4(1)°, U=929.8 Å^3, D_c=1.27, D_m=1.28 Mgm^{-3}, Z=2, F(000)=376, space group $P\bar{1}$, μ(Mo-K_α)=0.80 cm^{-1}.

Data collection

Instrument used: Hilger-Watts Y290
Radiation used: Mo-K_α λ=0.71069 Å
Filter: Graphite monochromator, $\cos^2 2\theta$ =0.965
Upper limit for data collection: $2\theta_{max}$=60°
Number of independent reflexions: m=2114
Unobserved cut-off: $3\sigma_I$
Number of parameters refined: n=324
Number of reflexions per parameter: m/n=6.53

(IIIb)

Crystal data

Angular dimer of 2-(p-methoxyphenyl) benzazete, $C_{28}H_{22}O_2N_2$, M_r=418.5, monoclinic, a=8.526(1), b=23.003(3), c=12.415(1) Å, β=112.5(1)°, U=2250.2 Å^3, D_c=1.24, D_m=1.24 Mgm^{-3}, Z=4, F(000)=880, space group $P2_1/c$, μ(Mo-K_α)=0.88 cm^{-1}

Data collection

Instrument used: Hilger-Watts Y290
Radiation used: Mo-K_α , λ=0.71069 Å
Filter: Graphite monochromator, $\cos^2 2\theta$ =0.965
Upper limit for data collection: $2\theta_{max}$=54°

Number of independent reflexions: m=3184
Unobserved cut-off: $3\sigma_I$
Number of parameters refined: n=376
Number of reflexions per parameter: m/n=8.47

(IIIc)

Crystal data

Angular dimer of 2-(p-tert-butylphenyl) benzazete, $C_{24}H_{26}N_2$, M_r=342.5, triclinic, a=10.185(1), b=9.307(1), c=10.170(1) Å, α=91.8(1), β=94.9(1), γ=103.2(1)°, U=933.9 Å^3, Z=2, D_c=1.21, D_m=1.22 Mgm^{-3}, F(000)=368, space group $P\bar{1}$, μ(Mo-K_α)=0.77 cm^{-1}.

Data collection

Instrument used: Hilger-Watts Y290
Radiation used: Mo-K_α λ=0.71069 Å
Filter: Graphite monochromator, $\cos^2 2\theta$=0.965
Upper limit for data collection: $2\theta_{max}$=60°
Number of independent reflexions: m=3752
Unobserved cut-off: $3\sigma_I$
Number of parameters refined: n=312
Number of reflexions per parameter: m/n=12.03

3.1.4 STRUCTURE DETERMINATION

For the triclinic structures, (IIIa) and (IIIc), application of QGEN provided a set of 1926 and 1652 quartet invariants respectively; generated for the top 100 E-magnitudes ($|E| \geq 1.4$, for (IIIa); $|E| \geq 2.2$ for (IIIb)), using the 7-magnitude, 2nd. neighbourhood formula, $P_7^{\pm}$, derived by Hauptman and Green, where $P^+ \geq 0.7$ and 0.8 for (IIIa) and (IIIc) respectively. The probabilities for acceptance of a triplet,

a positive quartet and a negative quartet were Ø.85, Ø.7Ø and Ø.32 for (IIIa) and Ø.75, Ø.95 and Ø.Ø7 for (IIIc). The phases thus derived were expanded, as described previously, via routine application of triplet phase relationships to phase 242 |E|'s> 1.8 for (IIIa) and 236 |E|'s> 1.8 for (IIIc). The resulting E-maps revealed the positions of all non-hydrogen atoms.

Structure (IIIb) crystallizes in space group $P2_1/c$, and was solved by routine application of PHASE, with triplets alone. An E-map derived from 246 |E|'s> 1.85 produced the complete structure.

It is noteworthy that application of the unmodified version of PHASE to (IIIa) and (IIIc) failed to produce a unique solution in either case.

3.1.5 STRUCTURE REFINEMENT

In all three cases the hydrogen atoms were located from a difference electron density synthesis. The non-hydrogen atoms were refined with anisoptropic thermal parameters using CRYLSQ. The hydrogen atom coordinates and thermal parameters were refined, except for (IIIc), where the temperature factors were fixed at B = 4.Ø$Å^2$. A polynomial weighting scheme was employed. At convergence the R factors were Ø.Ø45, Ø.Ø43 and Ø.Ø6Ø for (IIIa-c) respectively (R_w=Ø.Ø5, Ø.Ø6l, Ø.Ø57). Tables 1, 2 and 3 give the atomic coordinates and thermal parameters for structures (IIIa-c) respectively, while Table 4 summarises, comparatively, bond distances, interbond angles and torsion angles for the three dimers. Figures 1, 2 and 3 show ORTEP drawings for each molecule.

3.1.6 DISCUSSION

Confirmation of the structures (IIIa-c) indicates that formation of these angular dimers may be the result of a Diels-Alder reaction, since the regioselectivity observed is that anticipated if the addition proceeds through a zwitterion or involves a transition state with appreciable polar character[3]. The structure analysis of (IIIa-c) has also resolved the ambiguities surrounding the structures of the angular dimers formed by 2-alkyl and by 2-arylbenzazetes, and confirms that the latter compounds are produced by photolysis of the corresponding 4-substituted 1,2,3-benzotriazines.

Comparable bond lengths for all three molecules are experimentally identical, with the C(2)-N(3) bond virtually localised. Bond angles are also virtually identical, with the exception of C(13)-N(1)-C(2) in (IIIc), which has a value of 134.1(2)^{o} in contrast to the values of 128.6(3) and 125.8(1)^{o} in (IIIa) and (IIIb) respectively. This seems to reflect the different degrees of pyramidal nature of N(1) in the three molecules, rather than distortion of other interbond angles. Thus, N(1) in (IIIc) is the most pyramidal (sum of angles at N(1) 346.0^{o}), while that in (IIIb) is least pyramidal (sum of angles 337.1^{o}). In all three instances the four-membered rings are planar and subtend identical dimensions.

The conformation of the six-membered heterocyclic rings differ from compound to compound, although all adopt twisted and distorted boat conformations in which N(3) forms a shallow prow, and C(8) forms a more pronounced prow. Interestingly, the conformation in (IIIa) lies between that of (IIIb) and (IIIc), which could be inferred from the valency angles around the pyramidal nitrogen atoms where (IIIb)<(IIIa)<(IIIc).

Moreover, it is difficult to rationalise these conformational differences in terms of steric interactions. However, the flattening of N(1) in (IIIc) and associated

expansion of the C(13)-N(1)-C(2) interbond angle has the effect of relieving interactions between C(22) of the t-butyl group and the aromatic ring C(9)-C(14). In this context the interbond angles at C(2) in (IIIc) compensate for this effect by showing slight expansions and contractions.

On the other hand the same effect associated with C(81) of the t-butyl group as opposed to the C(81) phenyl groups of (IIIa) and (IIIb) is absent. However, the overall conformation of the molecule is such that this t-butyl group projects well away from the less congested face of the molecule, whereas the planar configuration of C(2) constrains the C(21) substituents within the sphere of serious interaction with the rest of the molecule.

REFERENCES

1. J. M. Stewart, (1970) Crystallographic Computing, pp. 71-74. Edited by F. R. Ahmed, Copenhagen:Munskgaard.
2. J. M. Stewart, G. J. Kruger, H. L. Ammon, C. Dickinson and S. R. Hall, (1972). The X-ray System - version June 1972. Technical Report TR-192 of the Computer Science Center, University of Maryland.
3. C. A. Rees, R. Storr and D. Whittle, J. C. S. Chem. Comm. (1976) pp. 411-412.
4. C. A. Rees, J. C. S. Perkin I (1976) pp. 2166-2169.

Table 1

Atomic coordinates and thermal parameters for (IIIa).

(a) Atomic coordinates ($x10^4$)

ATOM	x/a	y/b	z/c
N(1)	2381(4)	-7467(2)	7003(2)
C(2)	2061(5)	-6684(3)	7864(3)
N(3)	2731(4)	-6628(2)	8819(2)
C(4)	5276(6)	-6972(3)	9824(3)
C(5)	6726(6)	-7531(4)	9967(4)
C(6)	7116(6)	-8363(4)	9212(4)
C(7)	6054(6)	-8647(3)	8300(3)
C(8)	3324(5)	-8417(2)	7209(3)
C(9)	6058(6)	-8436(3)	5083(3)
c(10)	6643(7)	-7745(4)	4093(3)
C(11)	5757(6)	-6869(3)	4026(3)
C(12)	4227(6)	-6609(3)	4922(3)
C(13)	3719(5)	-7292(3)	5877(3)
C(14)	4594(5)	-8169(3)	5961(3)
C(15)	4565(5)	-8105(3)	8516(3)
C(16)	4164(5)	-7254(3)	8918(3)
C(21)	0787(5)	-5942(2)	7655(3)
C(22)	0596(6)	-5054(3)	8410(3)
C(23)	-0667(6)	-4383(3)	8256(4)
C(24)	-1702(7)	-4560(3)	7334(4)
C(25)	-1501(7)	-5421(3)	6574(4)
C(26)	-0285(6)	-6122(3)	6740(3)
C(81)	1609(5)	-9511(2)	7394(3)
C(82)	0326(6)	-9852(3)	6578(3)
C(83)	-1331(6)	-10815(3)	6739(4)
C(84)	-1729(7)	-11441(3)	7729(4)
C(85)	-0472(7)	-11125(3)	5537(4)
C(86)	1204(6)	-10160(3)	8378(3)

Table 1 (continued)

(b) Anisotropic temperature factors ($Å^2 \times 10^4$)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
N(1)	413	271	299	124	-059	011
C(2)	330	287	333	069	-021	015
N(3)	417	374	336	152	-100	-039
C(4)	455	485	413	104	-141	-020
C(5)	446	601	539	108	-198	080
C(6)	430	606	662	219	-150	146
C(7)	510	445	468	205	-077	022
C(8)	404	279	317	136	-040	009
C(9)	447	442	461	114	-027	-048
C(10)	489	617	395	048	045	-037
C(11)	532	537	342	-047	-043	092
C(12)	533	339	377	046	-128	036
C(13)	369	327	309	035	-067	-010
C(14)	377	315	331	060	-039	009
C(15)	363	339	327	082	-051	055
C(16)	341	375	349	067	-086	042
C(21)	336	278	349	085	-066	-001
C(22)	444	378	373	143	-097	-023
C(23)	585	407	554	222	-108	-079
C(24)	566	514	728	293	-234	-042
C(25)	573	537	659	207	-337	-037
C(26)	475	375	472	119	-182	-069
C(81)	387	293	322	106	-013	-014
C(82)	504	390	386	072	-065	050
C(83)	509	457	478	026	-122	-021
C(84)	518	352	561	-016	-020	011
C(85)	683	391	459	035	-045	116
C(86)	539	366	391	060	-095	051
Average e.s.d.'s						
C,N	21	19	21	16	17	15

Table 1 (continued)

(c) Hydrogen atom fractional coordinates ($x10^4$)

ATOM	x/a	y/b	z/c	U_{iso}
H(4)	5030(56)	-6325(31)	10349(32)	0.058
H(5)	7645(67)	-7295(35)	10559(37)	0.081
H(6)	8165(59)	-8687(30)	9267(31)	0.054
H(7)	6349(58)	-9231(31)	7724(32)	0.059
H(9)	6630(54)	-9034(29)	5141(30)	0.049
H(10)	7871(65)	-7843(33)	3428(36)	0.076
H(11)	6354(63)	-6348(33)	3286(36)	0.075
H(12)	3591(51)	-6012(28)	4934(28)	0.041
H(22)	1465(52)	-4904(27)	8985(30)	0.045
H(23)	-0817(58)	-3767(31)	8843(33)	0.064
H(24)	-2633(66)	-4118(34)	7258(36)	0.077
H(25)	-2250(64)	-5573(33)	5960(35)	0.070
H(26)	-0218(47)	-6743(26)	6222(27)	0.033
H(82)	0567(60)	-9405(32)	5922(34)	0.067
H(83)	-2230(55)	-11041(28)	6158(31)	0.052
H(84)	-2857(59)	-12040(32)	7826(32)	0.059
H(85)	-0806(62)	-11602(33)	9246(35)	0.073
H(86)	2206(57)	-9947(26)	8926(29)	0.041

Table 2

Atomic coordinates and thermal parameters for (IIIb)

(a) Fractional coordinates ($x10^4$)

ATOM	x/a	y/b	z/c
N(1)	5205(3)	1407(1)	2024(1)
C(2)	4996(2)	1054(1)	2870(2)
N(3)	3928(2)	1162(1)	3354(1)
C(4)	1276(4)	1591(1)	3176(2)
C(5)	-0082(4)	1962(1)	2636(2)
C(6)	-0070(4)	2330(1)	1762(2)
C(7)	1285(4)	2332(1)	1415(2)
C(8)	4180(2)	1984(1)	1603(2)
C(9)	3188(4)	1845(1)	-0812(2)
C(10)	3445(4)	1407(1)	-1504(2)
C(11)	4358(3)	0907(1)	-1035(2)
C(12)	5095(4)	0808(1)	0170(2)
C(13)	4825(2)	1246(1)	0828(2)
C(14)	3911(2)	1746(1)	0369(2)
C(15)	2654(2)	1962(1)	1936(2)
C(16)	2649(2)	1582(1)	2821(2)
C(21)	6113(2)	0541(1)	2225(1)
C(22)	7603(2)	0493(1)	3015(2)
C(23)	8639(2)	0009(1)	3376(2)
C(24)	8203(2)	-0434(1)	3967(1)
O(24)	9110(3)	-0930(1)	4363(1)
C(241)	10587(4)	-1018(1)	4170(2)
C(25)	6725(2)	-0389(1)	4191(2)
C(26)	5684(2)	0086(1)	3812(2)
C(81)	5267(2)	2400(1)	2027(2)
C(82)	6285(3)	2560(1)	3203(2)
C(83)	7280(4)	3049(1)	3621(2)
C(84)	7249(4)	3491(1)	2859(2)
O(84)	8151(2)	3999(1)	3180(2)
C(841)	9299(7)	4059(2)	4362(4)
C(85)	6239(4)	3440(1)	1686(2)
C(86)	5261(2)	2946(1)	1272(2)

Table 2 (continued)

(b) Anisotropic temperature parameters ($Å^2 \times 10^4$)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
N(1)	376	354	329	052	174	039
C(2)	333	379	320	016	135	038
N(3)	426	479	453	114	247	120
C(4)	474	514	646	061	353	089
C(5)	436	545	786	044	355	011
C(6)	349	512	649	099	164	004
C(7)	400	455	449	062	141	037
C(8)	344	352	338	067	147	057
C(9)	429	475	365	-030	133	053
C(10)	560	603	329	-142	161	-016
C(11)	656	501	431	-110	266	-115
C(12)	543	387	459	-020	245	-032
C(13)	379	402	338	-024	167	008
C(14)	366	403	359	-007	159	022
C(15)	339	363	367	012	137	-015
C(16)	358	406	424	035	197	030
C(21)	334	393	323	022	129	029
C(22)	361	435	392	003	167	049
C(23)	331	514	424	034	166	049
C(24)	368	385	342	056	089	-013
O(24)	475	445	539	143	189	041
C(241)	455	618	637	176	183	-046
C(25)	402	418	456	020	174	073
C(26)	366	454	436	037	194	070
C(81)	316	378	424	051	175	004
C(82)	433	459	443	045	155	009
C(83)	446	580	538	007	150	-101
C(84)	382	469	787	-035	258	-118
O(84)	669	590	1084	-226	337	-205
C(841)	843	984	1140	-363	255	-425
C(85)	445	443	750	019	273	100
C(86)	391	479	487	017	181	060
Average esd's						
C,N	10	10	10	8	8	8

Table 2 (continued)

(c) Hydrogen atom fractional coordinates ($x10^4$)

ATOM	x/a	y/b	z/c	U_{iso}
H(4)	1308(34)	1332(11)	3791(23)	0.065
H(5)	-0977(47)	1948(14)	2961(29)	0.089
H(6)	-1025(32)	2590(10)	1353(20)	0.063
H(7)	1256(29)	2573(10)	0796(20)	0.059
H(9)	2633(31)	2164(11)	-1143(21)	0.046
H(10)	3019(34)	1459(11)	-2271(24)	0.059
H(11)	4490(36)	0633(13)	-1580(23)	0.059
H(12)	5708(31)	0469(11)	0512(20)	0.054
H(22)	7891(26)	0811(10)	2607(18)	0.048
H(23)	9648(28)	-0013(10)	3208(17)	0.053
H(25)	6515(30)	-0704(11)	4644(17)	0.063
H(26)	4566(27)	0120(9)	3947(17)	0.050
H(241A)	10258(34)	-1035(11)	3311(25)	0.068
H(241B)	11092(39)	-1372(13)	4519(26)	0.090
H(241C)	11430(41)	-0713(10)	4385(27)	0.101
H(82)	6321(30)	2236(10)	3745(20)	0.058
H(83)	8033(44)	3090(13)	4469(29)	0.107
H(85)	6180(31)	3753(11)	1159(21)	0.065
H(86)	4645(29)	2915(9)	0449(20)	0.054
H(841A)	8806(48)	4061(17)	4871(35)	0.116
H(841B)	10286(75)	3734(23)	4597(47)	0.173
H(841C)	10000(75)	4437(24)	4191(51)	0.298

Table 3

Atomic coordinates and thermal parameters for (IIIc)

(a) Fractional coordinates (x10^4)

ATOM	x/a	y/b	z/c
N(1)	2248(3)	3677(2)	8889(2)
C(2)	2860(2)	2818(2)	9707(2)
N(3)	3511(2)	1897(2)	9256(2)
C(4)	4902(2)	1369(3)	7601(3)
C(5)	5367(3)	1545(4)	6372(3)
C(6)	4847(3)	2385(4)	5486(3)
C(7)	3824(3)	3057(3)	5800(2)
C(8)	2121(2)	3450(2)	7397(2)
C(9)	2353(2)	6333(3)	6637(2)
C(10)	2712(3)	7800(3)	8703(2)
C(11)	2580(3)	7699(3)	7346(3)
C(12)	2630(3)	6565(3)	9474(3)
C(13)	2430(2)	5248(2)	8760(2)
C(14)	2294(2)	5135(2)	7388(2)
C(15)	3297(2)	2846(2)	7014(2)
C(16)	3863(2)	2039(2)	7942(2)
C(21)	2710(2)	2922(3)	1182(2)
C(211)	1319(2)	3173(4)	1418(3)
C(212)	2883(4)	1505(4)	1805(4)
C(213)	3847(2)	4200(4)	1806(3)
C(81)	0701(2)	2509(2)	6846(2)
C(811)	0554(3)	2587(4)	5343(3)
C(812)	0544(3)	0921(3)	7226(4)
C(813)	-0404(3)	3135(4)	7411(4)

Table 3(continued)

(b) Anisotropic thermal parameters($Å^2 \times 10^4$)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
N(1)	437	340	407	112	064	-032
C(2)	390	350	465	065	063	009
N(3)	511	387	547	152	075	003
C(4)	495	515	798	197	059	-163
C(5)	510	823	854	208	156	-309
C(6)	608	932	630	129	232	-215
C(7)	597	704	472	125	124	-082
C(8)	422	364	398	105	048	-041
C(9)	550	504	623	150	054	111
C(10)	699	354	933	153	045	-104
C(11)	671	403	867	149	051	138
C(12)	650	434	568	163	054	-103
C(13)	415	337	518	107	057	-029
C(14)	434	375	501	116	050	003
C(15)	410	401	465	071	072	-101
C(16)	413	369	548	083	084	-098
C(21)	544	504	440	145	092	037
C(211)	746	991	606	322	225	021
C(212)	918	723	623	263	163	207
C(213)	924	795	529	025	-049	-091
C(81)	433	469	583	106	-015	-132
C(811)	618	983	643	171	-137	-231
C(812)	555	481	1085	-022	-044	-092
C(813)	420	781	941	135	012	-238
Average e.s.d's						
C,O,N	14	14	14	11	11	11

Table 3(continued)

(c) Hydrogen atom fractional coordinates ($x10^4$)

ATOM	x/a	y/b	z/c
H(4)	5328(24)	0886(27)	8283(24)
H(5)	6115(24)	1050(26)	6157(23)
H(6)	5011(24)	2403(26)	4585(25)
H(7)	3414(24)	3674(26)	5118(24)
H(9)	2178(23)	6205(26)	5662(24)
H(10)	2925(24)	8726(28)	9096(24)
H(11)	2601(23)	8605(27)	6926(23)
H(12)	2703(24)	6611(27)	10322(24)
H(211A)	0609(25)	2304(28)	11004(24)
H(211B)	1342(23)	4381(28)	11041(23)
H(211C)	1729(25)	3711(28)	12206(25)
H(212A)	2245(25)	0701(28)	11336(24)
H(212B)	3833(25)	1463(27)	11722(25)
H(212C)	2771(23)	1594(26)	12785(25)
H(213A)	4746(26)	4056(28)	11370(25)
H(213B)	3739(23)	5262(28)	11370(24)
H(213C)	3742(24)	4359(26)	12737(25)
H(811A)	-0261(26)	2081(27)	5031(25)
H(811B)	1144(26)	2220(28)	4970(25)
H(811C)	0798(23)	3799(29)	5040(23)
H(812A)	-0282(25)	0421(27)	6849(24)
H(812B)	0639(25)	0830(27)	8190(25)
H(812C)	1105(26)	0555(29)	6955(26)
H(813A)	-0487(24)	4134(29)	7108(24)
H(813B)	-0317(24)	3125(27)	8407(25)
H(813C)	-1223(26)	2550(27)	7171(24)

Table 4

Interatomic distances(Å) and angles(°) for (IIIa-c)

(a) Bonded distances

	DIMER(IIIa)	DIMER(IIIb)	DIMER(IIIc)
N(1) -C(2)	1.395(4)	1.390(2)	1.380(3)
N(1) -C(8)	1.535(4)	1.525(2)	1.516(3)
N(1) -C(13)	1.434(4)	1.444(2)	1.444(2)
C(2) -N(3)	1.288(4)	1.292(2)	1.292(3)
C(2) -C(21)	1.476(4)	1.474(2)	1.522(3)
N(3) -C(16)	1.416(4)	1.417(2)	1.415(3)
C(4) -C(5)	1.380(5)	1.386(3)	1.376(4)
C(4) -C(16)	1.403(5)	1.401(3)	1.407(3)
C(5) -C(6)	1.375(6)	1.379(3)	1.362(4)
C(6) -C(7)	1.388(6)	1.386(3)	1.386(4)
C(7) -C(15)	1.389(5)	1.389(3)	1.389(3)
C(8) -C(14)	1.544(4)	1.543(2)	1.538(3)
C(8) -C(15)	1.501(4)	1.510(2)	1.509(3)
C(8) -C(81)	1.522(4)	1.511(2)	1.555(3)
C(9) -C(11)	1.414(5)	1.396(3)	1.403(4)
C(9) -C(14)	1.368(5)	1.374(2)	1.364(3)
C(10)-C(11)	1.387(6)	1.374(2)	1.373(5)
C(10)-C(12)	1.406(5)	1.403(3)	1.403(4)
C(12)-C(13)	1.378(5)	1.370(3)	1.370(3)
C(13)-C(14)	1.389(4)	1.385(3)	1.389(3)
C(15)-C(16)	1.398(5)	1.401(2)	1.391(3)
C(21)-C(22)	1.400(5)	1.395(2)	
C(21)-C(26)	1.392(5)	1.401(2)	
C(22)-C(23)	1.382(5)	1.385(3)	
C(23)-C(24)	1.383(3)	1.386(2)	
C(24)-C(25)	1.379(6)	1.394(2)	
C(25)-C(26)	1.390(5)	!.385(3)	
C(81)-C(82)	1.395(5)	1.392(2)	
C(81)-C(86)	1.388(5)	1.388(2)	

Table 4(continued)

(b) Interbond angles

			DIMER(IIIa)	DIMER(IIIb)	DIMER(IIIc)
C(8)	-N(1)	-C(2)	121.1(2)	121.9(1)	122.2(2)
C(13)	-N(1)	-C(2)	128.6(3)	125.8(1)	134.1(2)
C(13)	-N(1)	-C(8)	89.4(2)	89.4(1)	89.7(1)
N(3)	-C(2)	-N(1)	123.2(3)	123.6(2)	122.0(2)
C(21)	-C(2)	-N(1)	116.1(3)	116.0(2)	118.5(2)
C(21)	-C(2)	-N(3)	120.6(3)	120.3(2)	119.4(2)
C(16)	-N(3)	-C(2)	118.3(3)	117.9(2)	117.5(2)
C(16)	-C(4)	-C(5)	120.5(4)	120.0(2)	119.8(2)
C(6)	-C(5)	-C(4)	120.2(4)	120.5(2)	120.6(3)
C(7)	-C(8)	-C(5)	120.3(4)	120.0(2)	120.5(3)
C(15)	-C(7)	-C(6)	120.2(4)	120.5(2)	120.0(2)
C(14)	-C(8)	-N(1)	84.9(2)	85.3(1)	85.4(1)
C(15)	-C(8)	-N(1)	110.0(2)	110.1(1)	108.3(2)
C(15)	-C(8)	-C(14)	116.0(3)	117.2(1)	116.8(2)
C(81)	-C(8)	-N(1)	110.1(3)	111.8(1)	112.1(2)
C(81)	-C(8)	-C(14)	116.3(3)	116.6(1)	115.7(2)
C(81)	-C(8)	-C(15)	115.1(3)	112.6(1)	114.6(2)
C(14)	-C(9)	-C(10)	115.3(3)	115.2(3)	115.4(2)
C(11)	-C(10)	-C(9)	121.6(4)	122.4(2)	121.6(3)
C(12)	-C(11)	-C(10)	123.0(4)	122.3(2)	123.0(2)
C(13)	-C(12)	-C(11)	113.7(3)	114.0(2)	114.3(2)
C(12)	-C(13)	-N(1)	141.1(3)	141.2(2)	142.9(2)
C(14)	-C(13)	-C(12)	124.1(3)	124.3(2)	123.1(2)
C(14)	-C(13)	-N(1)	94.8(3)	94.5(1)	94.0(2)
C(9)	-C(14)	-C(8)	146.9(3)	147.2(2)	146.5(2)
C(13)	-C(14)	-C(8)	90.7(3)	90.8(1)	90.9(2)
C(13)	-C(14)	-C(9)	122.4(3)	121.9(2)	122.6(2)
C(15)	-C(16)	-N(3)	124.1(3)	123.9(2)	124.3(2)
C(4)	-C(16)	-N(3)	117.0(3)	116.7(2)	116.4(2)
C(15)	-C(16)	-C(4)	118.9(3)	119.3(2)	119.3(2)
C(8)	-C(15)	-C(7)	121.8(2)	121.8(2)	123.4(3)
C(16)	-C(15)	-C(7)	120.0(3)	119.7(2)	119.6(2)
C(16)	-C(15)	-C(8)	118.2(3)	118.5(2)	117.9(2)

Table 4(b) continued

	DIMER(IIIa)	DIMER(IIIb)	DIMER(IIIc)
C(22) -C(21) -C(2)	119.6(3)	122.6(2)	
C(26) -C(21) -C(2)	121.6(3)	119.1(2)	
C(26) -C(21) -C(22)	118.8(3)	118.3(2)	
C(23) -C(22) -C(21)	120.3(3)	121.2(2)	
C(24) -C(23) -C(22)	120.4(4)	119.7(2)	
C(25) -C(24) -C(23)	119.9(4)	119.7(2)	
C(26) -C(25) -C(24)	120.2(4)	120.4(2)	
C(21) -C(26) -C(25)	120.4(3)	120.7(2)	
C(82) -C(81) -C(8)	119.2(3)	120.0(2)	
C(86) -C(81) -C(8)	122.2(3)	121.7(2)	
C(86) -C(81) -C(82)	118.5(3)	118.3(2)	
C(84) -C(83) -C(82)	119.4(4)	119.5(2)	
C(83) -C(82) -C(81)	121.3(3)	121.4(2)	
C(85) -C(84) -C(83)	120.2(4)	120.0(2)	
C(86) -C(85) -C(84)	120.7(4)	120.2(2)	
C(85) -C(86) -C(81)	119.9(3)	120.7(2)	
C(23) -C(24) -O(24)		125.3(2)	
C(25) -C(24) -O(24)		115.0(2)	
C(24) -O(24) -C(241)		117.9(2)	
C(83) -C(84) -O(84)		124.5(2)	
C(85) -C(84) -O(84)		115.6(2)	
C(84) -O(84) -C(841)		118.0(2)	
C(211)-C(21) -C(2)			110.4(2)
C(212)-C(21) -C(2)			110.0(2)
C(213)-C(21) -C(2)			107.6(2)
C(211)-C(21) -C(212)			109.2(2)
C(211)-C(21) -C(213)			115.3(2)
C(212)-C(21) -C(213)			108.3(2)
C(811)-C(81) -C(8)			108.9(2)
C(812)-C(81) -C(8)			110.3(2)
C(831)-C(81) -C(8)			109.9(2)
C(811)-C(81) -C(812)			110.5(3)
C(811)-C(81) -C(813)			108.4(2)
C(812)-C(81) -C(813)			108.9(2)

Table 4(continued)

(c) Torsion angles (selected)

	(IIIa)	(IIIb)	(IIIc)
C(8) -N(1) -C(2) -N(3)	7.8	Ø.7	8.9
N(1) -C(2) -N(3) -C(16)	11.4	14.7	14.9
C(2) -N(3) -C(16)-C(15)	-14.1	-12.4	-15.9
N(3) -C(16)-C(15)-C(8)	-3.2	-5.5	-7.5
C(16)-C(15)-C(8) -N(1)	19.7	18.4	27.1
C(15)-C(8) -N(1) -C(2)	-22.8	-17.1	-29.4
C(13)-N(1) -C(8) -C(14)	-2.3	-1.3	-1.1
N(1) -C(8) -C(14)-C(13)	2.4	1.4	1.2
C(8) -C(14)-C(13)-N(1)	-2.6	-1.4	-1.2
C(14)-C(13)-N(1) -C(8)	2.6	1.5	1.3
C(13)-N(1) -C(2) -C(21)	73.3	66.6	64.4
C(16)-N(3) -C(2) -C(21)	-172.1	-166.9	-167.6
C(8) -N(1) -C(2) -C(21)	-168.9	-177.8	-168.6
C(13)-C(14)-C(8) -C(81)	112.4	113.4	113.6
C(7) -C(15)-C(8) -C(81)	73.5	71.8	81.1
C(16)-C(15)-C(8) -C(81)	-1Ø5.5	-1Ø7.1	-98.9
C(2) -N(1) -C(8) -C(81)	1Ø5.2	1Ø8.9	98.Ø
C(13)-N(1) -C(8) -C(81)	-118.5	-118.1	-117.1

Dimer IIIa

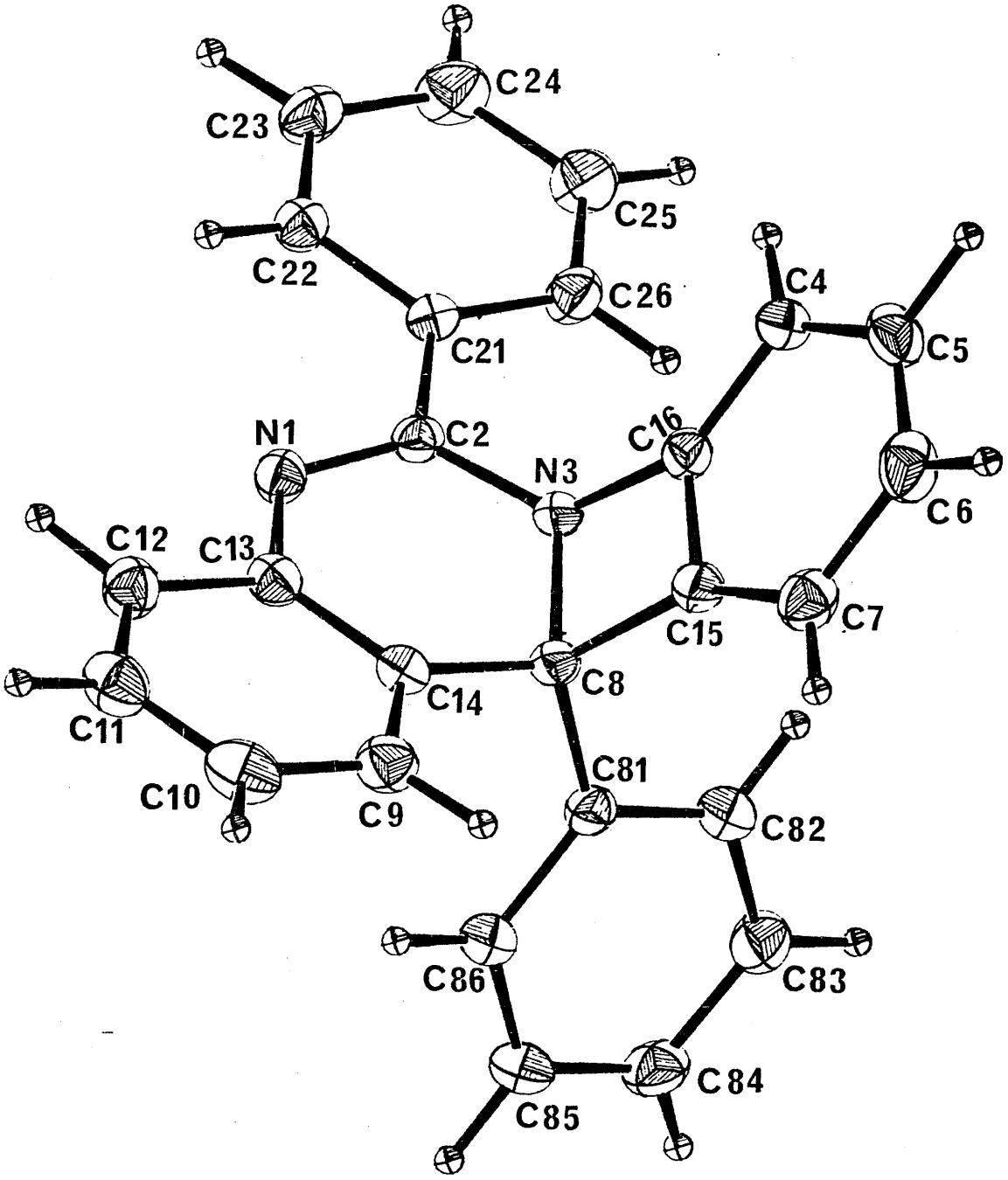


Figure 1

Dimer IIIb

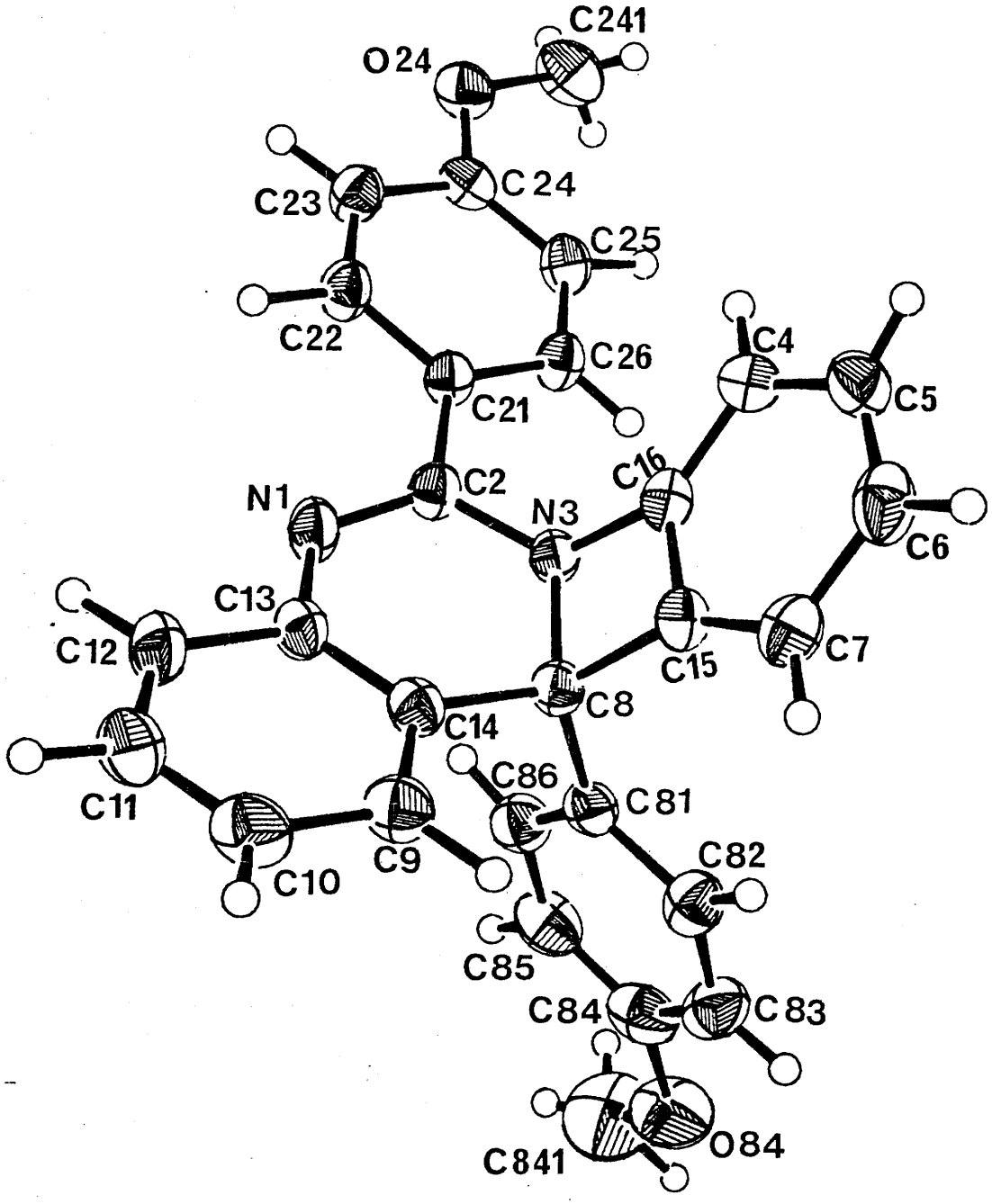


Figure 2

Dimer IIIc

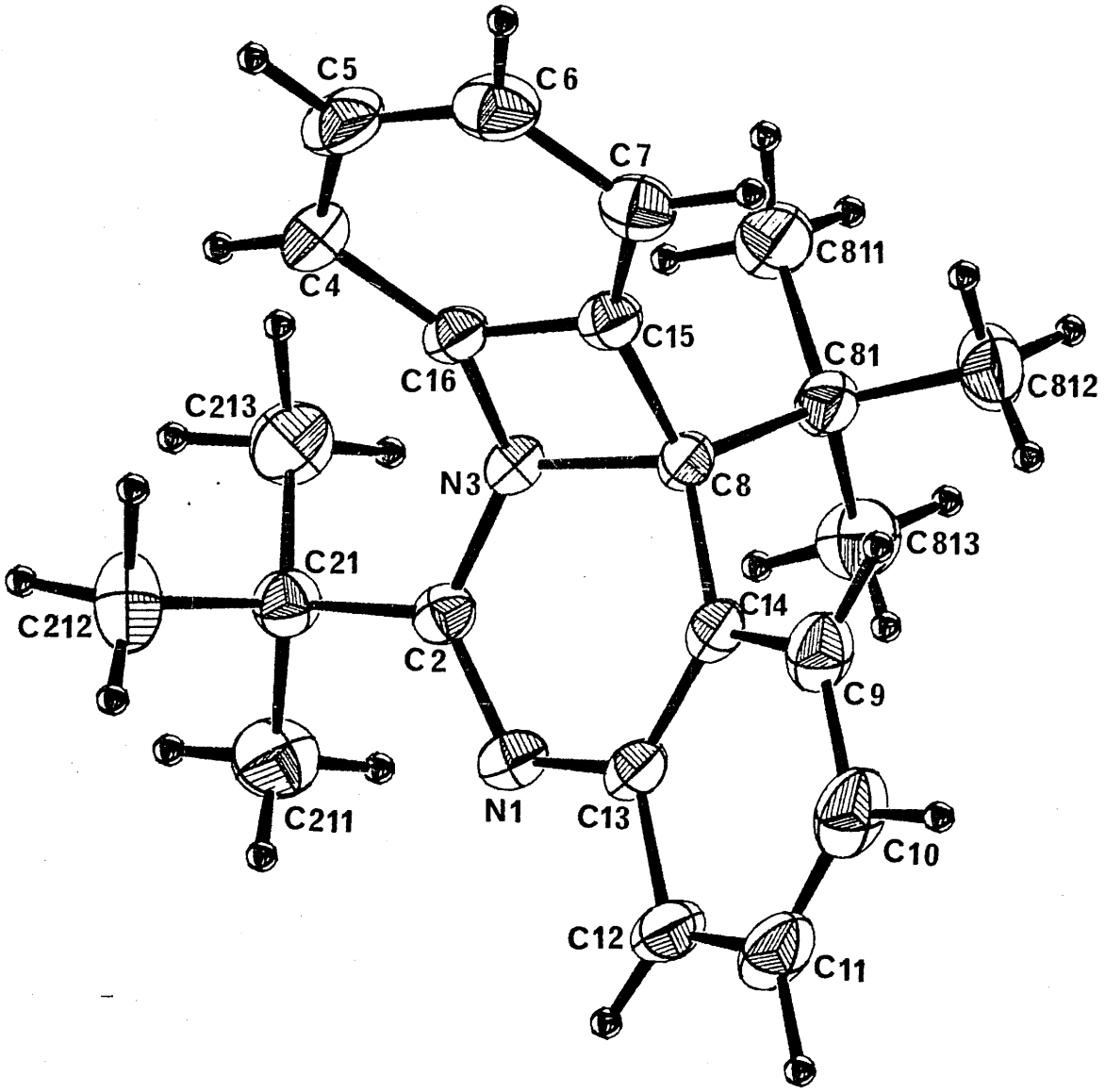


Figure 3

B: CRYSTAL STRUCTURE SOLUTION OF A PICRIC ACID : YLIDE COMPLEX

3.2.1 INTRODUCTION

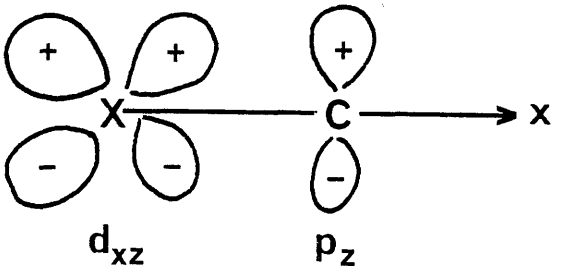
Ylides have been described as stabilized anions[1] which have a vicinal zwitterionic electronic structure represented generally as $>X^{+}-Y^{-}<$, where $X^{+}=(P,N,As,Sb,S,Se)$ and $Y^{-}=(C,N,O)$. Compounds which are included in this category are carbon-sulphur ylides ($>S^{+}-C^{-}<$), carbon-phosphorous ylides ($>P^{+}-C^{-}<$), and 'onium-imine ylides ($>X^{+}-N^{-}-$).

That ylides are stable chemical compounds, whereas anions are reactive, must result from some unique stabilization afforded the negative atom, Y, by the presence of the adjacent X^{+} grouping, otherwise known as the 'onium residue. However, although such stabilization probably occurs, almost all stable compounds of this type have an electron-withdrawing group attached to the Y^{-} portion of the molecule. This electron-withdrawing group will have a stabilizing effect on the molecule by virtue of its ability to delocalise part of the negative charge residing on the Y^{-} atom.

It is well known that ylides formed from second and higher row elements (X=P,As,Sb,S,Se,Br,I) possess greater stability than their first row analogues (X=N,O,F)[2-5]. This enhanced stability of ylides containing second row 'onium species (X^{+}) has been attributed to the possibility of a π-interaction of vacant 3d orbitals on the onium species with the lone pairs of electrons on the anionic atom (Y^{-}). Such valence shell expansion is not possible for first row elements[6] since the energy gap to the next vacant orbital is sufficiently large to preclude their involvement in any appreciable bonding.

$(d-p)_{\pi}$ interactions are perhaps best illustrated in carbanionic ylides of the second row elements where the

carbanionic moiety exhibits trigonal (sp^2) geometry. The lone pair of electrons may be assumed to occupy a p-orbital perpendicular to the plane defined by the σ-bonds in the $>X^+-C^-<$ moiety. Such an arrangement is ideally suited to $d_\pi-p_\pi$ overlap with a suitable d-orbital of the 'onium species.

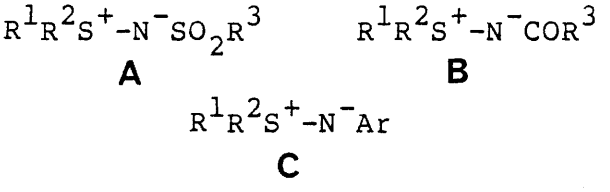


For 'omium imines (Y=N-), however, a similar general inference is difficult to justify, since two lone-pairs are available on the nitrogen for bonding. Possibly it would be more correct to envisage the two lone pairs occupying equivalent orbitals, both of which can interact equally with suitable combinations of the d-orbitals of the 'onium group. The problem can then be simplified by resolving the d-overlap into two mutually perpendicular components, one in the plane of the σ-bond framework and the other perpendicular to this plane.

Thus a possible description of the π-bonding arrangement in the second row 'onium imines is that two equivalent lone pairs on the negatively charged nitrogen atom can interact with suitable combinations of the d-orbitals on the 'onium residue to form bonding overlaps which may theoretically be resolved into two mutually perpendicular components: (i) σ', in the plane of the σ-bond framework, and (ii) π, perpendicular to the σ-bond framework.

Moreover, it has been found[7,8] that stability in such cases may be enhanced by the competition of 'onium and stabilizing groups to delocalise the negative charge on the anionic atom. Thus, by attaching different groups, Z , to the anionic atom centre (in sulphur-nitrogen ylides) additional

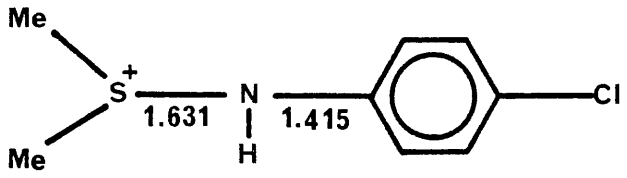
stabilizing - destabilizing effects can be compared. If Z is an electron withdrawing group (eg. carbonyl, cyano carbonyl etc.), the ylides are relatively stable, as seen with N-sulphonyliminosulphuranes (A) and N-acyliminosulphuranes (B) where they are hygroscopic and stable solids at room temperature, whereas the N-aryliminosulphuranes (C), with electron withdrawing groups (CNOR, NO_2) on the ring, are only moderately stable (1-12 months at room temperature) and N-aryliminosulphuranes with electron donating substituents on the ring are hygroscopic and decompose within a few days or weeks at room temperature.



A **B** **C**

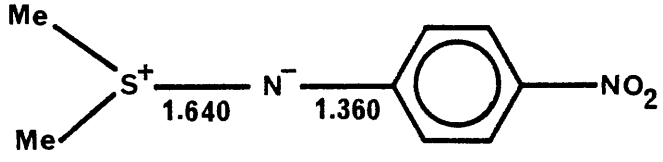
Recent reports[9] of chemical bonds in ylides have claimed to demonstrate the importance of ionic interactions as opposed to $(d\text{-}p)_\pi$ interactions. However, the nature of the stabilization remains complex and debatable, and might be attributed to several factors which are not necessarily independent.

As part of an examination of the bonding, charge distributions and conformation of second row ylides, the crystal structure of N-(p-chlorophenyl) iminodimethylsulphur (IV) : picrate, which is representative of the class of ylides (X=S, $Y=N^-$), has been determined. The choice of this compound for study was influenced by the recent analyses of **A** and **B** [7] with which comparisons in geometry can be made in respect to the protonation and subsequent elimination of a lone-pair on the former anionic nitrogen.

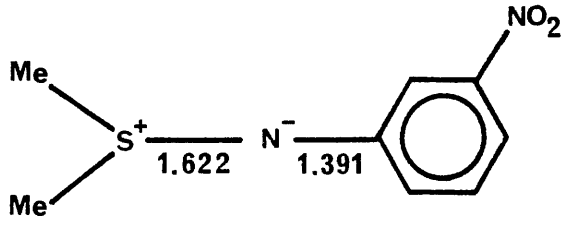


I

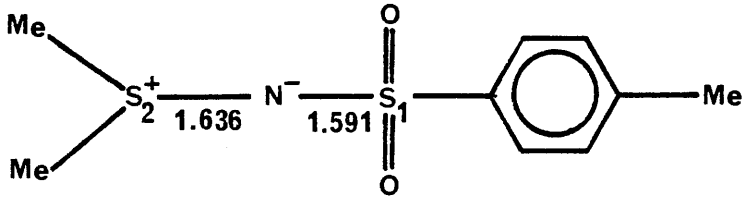
(picrate)$^-$



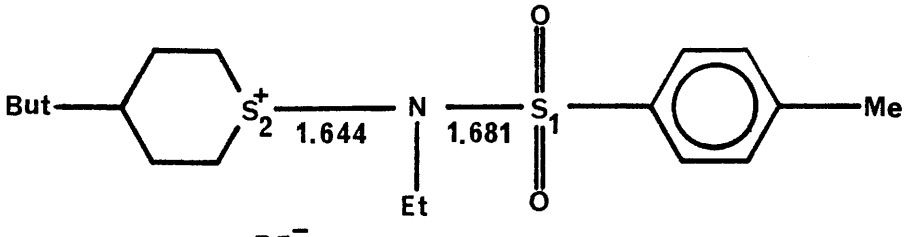
II



III



IV



V

BF_4^-

3.2.2 EXPERIMENTAL

(i) Preparation of Crystals

The sulphonium ylide was prepared by the method described by Claus and Vilsmaier[10] (1975). Picric acid in ether solution was prepared by standard techniques[11].

1.2g of the sulphamide was dissolved in the minimum of dry ether. The picric acid solution (62ml.) was added dropwise (to excess) and stirred for an hour. The mixture was filtered and the filtrate (yellow) recrystallised from an acetone-ether solution. On standing for a few days the solution produced needle-shaped crystals which were used in the analysis.

(ii) Crystal data

N-(p-chlorophenyl) iminodimethylsulphur (IV) picrate, $C_{14}H_{13}O_7N_4SCl$, M_r=416.8, monoclinic, a=7.924(1), b=9.406(1), c=23.078(2) Å, β=92.19(6)°, U=1718.8 Å^3, D_c=1.61, D_m=1.61Mgm^{-3}, Z=4, F(000)=856, space group $P2_1/n$, $\mu(MoK_\alpha)$=3.94 cm^{-1}.

Data colection

Instrument used: Hilger Watts Y290
Radiation used: Mo-K_α λ=0.71069 Å
Filter: Graphite monochromator, $\cos^2 2\theta$=0.965
Upper limit for data collection: $2\theta_{max}$=60°
Number of independent reflexions: m=3687
Unobserved cut-off: $3\sigma_I$
Number of parameters refined: n=153
Number of reflexions per parameter: m/n=24.1

3.2.3 STRUCTURE DETERMINATION

The structure was solved using quartets in a modified version of PHASE which is incorporated in the X-RAY 72 suite of programs. 1031 quartets were produced using QGEN for E-magnitudes having $|E| \geqslant 2.10$ with $P^{+} \geqslant 0.90$ ($P^{-} \leqslant 0.10$), employing both $P_7^{\pm}$ and $P_{13}^{\pm}$ formulae. Since the quartets generated were no longer in a symmorphic space group there appeared groups of invariants which had all four principal terms in common. These 'families' of quartets[12] are eliminated by the 'TIDY' feature in QGEN. The number of quartets remaining after elimination of families was 955.

Triplet phase relationships were generated for E-magnitudes greater than 1.6. The probabilities for the acceptance of a triplet, a positive quartet, and a negative quartet were 0.8, 0.95, and 0.07 respectively. The program arbitrarily selected the three reflexions for origin definition, which were expanded to phase 341 reflexions (156+, 185-) for the E-map calculation from which the complete structure was revealed.

3.2.4 STRUCTURE REFINEMENT

The structure was refined using CRYLSQ. Atomic and isotropic thermal parameters were adjusted by full matrix least-squares to an R-value of 0.098. At this point a difference map located all the hydrogen atom positions which were then introduced into the refinement procedure. Anisotropic refinement of all non-hydrogen atoms and subsequent isotropic refinement of hydrogen parameters, with the introduction of a weighting scheme of the type:

$$w = (A/F_{obs})^2 \qquad \text{where } (A=14.0)$$

reduced R to a final value of Ø.Ø33 (R_w=Ø.Ø44). A drawing of the molecule is included in Figure I. Final fractional coordinates and thermal parameters appear in Table 1, while Table 2 contains all bond lengths and interbond angles.

3.2.5 DISCUSSION

The structure analysis of N-(p-chlorophenyl) iminodimethylsulphur picrate has allowed a comparison to be made with the bonding and charge distribution found in other S(IV) ylides and picrate ion moieties.

a) The picrate ion

Table 3 gives relevant bond lengths and angles for the title compound with four other picrate ions. As a comparison, bond lengths and interbond angles (the latter applying only to the benzene ring) for two picrate acid molecules are also given. A common feature of all four ions is the shortness of the O(1)-C(1) bond, 1.24Å, compared to the longer distance, 1.32Å, exhibited by the molecular structures. This distance is comparable to that found in many ketones, aldehydes and carboxylic acids and their salts[13]. The C(1)-C(2) and C(1)-C(6) bond lengths, 1.453 and 1.457Å, correspond to those found in conjugated unsaturated systems such as acraldehyde and acrylic acid, where the π-electrons of the benzene ring have reduced delocalisation to a degree similar to that for 'single bonds' in short conjugated unsaturated systems.

The picrate molecule, on the other hand, has the shorter (benzene type) bond of 1.40Å. The remaining bonds in the ring do not deviate significantly from values normally found in benzene and its derivatives. The small bond angle, C(2)-C(1)-C(6), of 111.7° is characteristic of picrate ions, which, together with the increased bond lengths of adjacent atoms, leads the benzene ring to deviate from planarity.

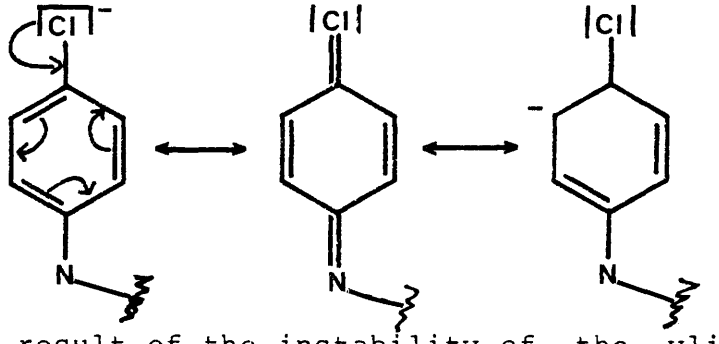
Intra-molecular distances, O(1)-O(21) and O(1)-O(61), of 2.646 and 2.697Å respectively, highlight the dissimilar twisting of the nitro- groups to the benzene ring; the dihedral angles between the planes bounded by N(2)-O(21)-O(22) and N(6)-O(61)-O(62) and the benzene ring plane being 13.8 and 30.2° respectively. The para-nitro- group, as may be expected, shows a lesser degree of twisting, with a dihedral angle of 9° with the ring plane.

b) Iminodimethylsulphur(IV) cation

Protonation of the former anionic nitrogen atom does not appear to have altered the d-orbital involvement associated with the S-N bonding system of the ylide moiety. The S-N bond (1.631Å) is in good experimental agreement with (II)[14] (1.640Å), (III)[14] (1.622Å), (IV)[9] (S(2)-N(1) 1.636Å) and (V)[15] (S(2)-N(1) 1.644Å). Delocalisation of the single lone pair of electrons on the nitrogen atom would seem to suggest that $(d\text{-}p)_{\pi}$ overlap requires only the involvement of one pair of electrons to completely fill the vacant molecular orbitals of the sulphur atom.

The significant lengthening of the N-C(phenyl) bond, 1.415Å, compared to (II) and (III), may be attributed to two factors. Firstly, the exclusion of a lone pair of electrons from taking part in σ' bonding must contribute to the increased dimension. Lengthening of the N-Z bond is also seen when comparing (IV) and (V), where N-alkylation gives an increase in bond length S(1)-N(1) from 1.591Å in (IV) to 1.681Å in (V), while leaving the S(2)-N(1) length relatively unchanged. Furthermore, an increase in N-Z bond distance associated with protonation of the anionic nitrogen can be found in the comparison of dibenzenesulphonamide[16] and its sodium salt cited by Cotton and Stokely., where the only major change attributed directly to deprotonation of $(C_6H_5SO_2)_2NH$ is the decrease of 0.070(16)Å in the mean S-N bond length.

The decrease in the N-C bond length in (I), however, is not as large as may have been expected from consideration of the previous examples. This is almost certainly due to the inductive effect of the p-chlorine which, via canonical representations (see below), will impart a small amount of double bond character to the N-C bond. Evidence for this indirect contribution of the chlorine on the N-C bond is seen when comparing the nitro- deactivated ylides (II) and (III). Average values for the bonds adjacent to the carbon adjoining the nitrogen are longer, 1.410Å in (II) and 1.423Å in (III), compared to the benzene type 1.391Å for (I).



As a result of the instability of the ylide (I), when removed from solution, it was not possible to make a direct comparison of the protonated and deprotonated forms of (I). Initially it was hoped that a study could be made of the picrates formed with (II) and (III). However, experimentation using different solvents proved unsuccessful, due to the greater stability of the latter ylides wrought by the deactivating effect of the nitro- groups.

Trigonal, sp^2 ,geometry of the nitrogen atom is suggested by the valence angle of 120.4°. Moreover, the torsion angles along the N(44)-C(44) bond show the sulphur to be twisted out of the benzene ring plane by 25.4°, which is greater than the -9.3 and -9.0° experienced in (II) and (III). This is in accordance[15] with the single lone pair of electrons of the nitrogen atom being located in a p-type orbital approximately perpendicular to the S-N-C plane , such that valence lone-pair repulsions are minimised.

REFERENCES

1. R. F. Hudson, Chem. in Brit. (1971) 7, 287.
2. A. F. Cameron in Molecular Structure by Diffraction Methods, Vol I, specialist periodical reports to the Chemical Society, and subsequent volumes.
3. A. W. Johnson in Organic Compounds of Sulphur, Selenium and Tellurium, Volume I, specialist reports to the Chemical Society, and subsequent volumes.
4. A. W. Johnson, Ylid Chemistry, Academic Press, London, 1966.
5. J.L. Atwood and W. A. Sheppard, Acta Cryst. (1975) B31, 2638.
6. B. C. Webster and R. F. Stewart, J. Chem. Soc. (A), 1971, 2987.
7. J. McEllhatton, Ph. D. Thesis, (1977) Glasgow University.
8. F. D. Duncanson, Ph. D. Thesis, (1976) Glasgow University.
9. N. J. Hair, Ph. D. Thesis, (1975) Glasgow University.
10. D. Claus and H. Vilsmaier, Tetrahedron (1975) 31, 505.
11. Organic Vogal, pp. 666.
12. L. Lessinger, Intercongress Symposium, Direct Methods of Crystallography, Buffalo, 1976.
13. Tables of Interatomic Distances and Configuration in Molecules and Ions, 1965, p. 1958.
14. A. Maltz, Ph. D. Thesis, (1978) Glasgow University.
15. R. E. Cook, M. D. Glick, J. J. Rigau and C. R. Johnson, J. Amer. Chem. Soc. (1971) 93, 924.
16. F.A. Cotton and P. F. Stokely, J. Amer. Chem. Soc. (1970) 92, 294.
17. K. Maartmann-Koe, Acta Cryst. (1969) B25, 1452.
18. J. Palenick, Acta Cryst. (1972) B28, 1633.
19. C. E. Bugg and U. Thewalt, Acta Cryst. (1975) B31, 121.
20. B. Jensen, Acta Chem. Scand. (1975) B29, 891.
21. F. H. Bernstein and M.Kaftory, Acta Cryst. (1976) B32, 387.

Table 1

Atomic coordinates and thermal parameters for Ylide:Picric Acid Complex

(a) Atomic coordinates ($x10^4$)

ATOM	x/a	y/b	z/c
Cl	7707(1)	5009(1)	3161(1)
S	11200(1)	11683(1)	3567(1)
O(1)	6164(1)	8733(1)	5256(1)
C(1)	5689(1)	8387(1)	4756(1)
C(2)	6130(1)	9135(1)	4231(1)
N(2)	7200(1)	10382(1)	4268(1)
O(21)	7999(2)	10644(1)	4722(1)
O(22)	7300(1)	11144(1)	3838(1)
C(3)	5572(2)	8751(2)	3682(1)
C(4)	4533(2)	7577(2)	3612(1)
N(4)	3975(2)	7172(2)	3032(1)
O(41)	4268(2)	7993(2)	2633(1)
O(42)	3199(2)	6049(2)	2963(1)
C(5)	4006(2)	6807(1)	4080(1)
C(6)	4547(2)	7204(1)	4629(1)
N(6)	3894(2)	6375(1)	5101(1)
O(61)	4762(2)	6232(1)	5546(1)
O(62)	2492(2)	5827(2)	5028(1)
C(11)	8786(2)	6542(2)	3383(1)
C(22)	9252(2)	6710(2)	3963(1)
C(33)	10100(2)	7937(2)	4140(1)
C(44)	10489(2)	8972(1)	3738(1)
N(44)	11369(2)	10202(1)	3932(1)
C(441)	13308(2)	12081(3)	3368(1)
C(442)	10882(3)	13016(2)	4099(1)
C(55)	10019(2)	8781(2)	3154(1)
C(66)	9150(2)	7564(2)	2983(1)

Table 1 (continued)

(b) Thermal parameters ($A^2 \times 10^4$)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{23}	U_{23}
Cl	704(3)	507(2)	925(4)	-062(2)	-208(3)	-225(2)
S	373(2)	442(2)	418(2)	-044(1)	-123(1)	133(1)
O(1)	453	459	304	026	-103	-038
C(1)	294	358	321	078	-046	-061
C(2)	274	368	330	040	-017	-082
N(2)	324	428	393	006	008	-087
O(21)	545	684	425	-205	-058	-134
O(22)	524	524	545	-080	-024	064
C(3)	313	460	308	050	-006	-072
C(4)	335	484	328	046	-052	-137
N(4)	460	739	389	-045	-056	-120
O(41)	1051	1243	316	-438	-067	-088
O(42)	661	714	575	-077	-156	-280
C(5)	330	371	427	030	-049	-103
C(6)	358	341	364	052	-022	-031
N(6)	527	355	445	060	020	002
O(61)	846	545	464	-024	-115	113
O(62)	547	688	713	-123	060	126
C(11)	375	400	561	057	-070	-131
C(22)	411	351	494	049	-036	005
C(33)	381	371	364	051	-061	008
C(44)	314	379	344	052	-043	-014
N(44)	487	400	373	-055	-168	081
C(441)	467	780	430	-127	-018	134
C(442)	463	407	806	052	-039	026
C(55)	489	531	328	002	-033	-004
C(66)	517	560	393	049	-081	-119
Average e.s.d.'s						
O,N,C	6	8	7	6	5	5

Table 1 (continued)

(c) Hydrogen atom fractional coordinates ($\times 10^4$)

ATOM	x/a	y/b	z/c	U_{iso}
H(1)	11914(29)	10188(24)	4229(11)	0.061(6)
H(3)	5891(24)	9273(21)	3355(9)	0.050(5)
H(5)	3265(24)	6016(20)	4047(8)	0.047(5)
H(22)	8957(27)	5982(23)	4232(10)	0.052(6)
H(33)	10403(26)	8092(23)	4542(9)	0.054(6)
H(55)	10342(27)	9498(23)	2885(10)	0.064(6)
H(66)	8864(35)	7461(29)	2550(13)	0.073(8)
H(441a)	14015(26)	12051(23)	3697(10)	0.064(5)
H(441b)	13309(31)	13002(30)	3215(11)	0.090(7)
H(441c)	13581(32)	11382(28)	1062(12)	0.081(7)
H(442a)	9880(35)	12866(26)	4239(11)	0.068(7)
H(442b)	11800(32)	13005(26)	4383(11)	0.063(7)
H(442c)	10966(34)	13897(29)	3903(12)	0.079(8)

Table 2

Interatomic distances(Å) and angles(°) for Ylide:Picric Acid Complex.

(a) Bonded distances

O(1) -C(1)	1.244(2)	C(1) -C(2)	1.453(2)
C(1) -C(6)	1.457(2)	C(2) -N(2)	1.448(2)
C(2) -C(3)	1.375(2)	N(2) -O(21)	1.229(2)
N(2) -O(22)	1.229(2)	C(3) -C(4)	1.383(2)
C(4) -N(4)	1.443(2)	C(4) -C(5)	1.378(2)
N(4) -O(41)	1.231(2)	N(4) -O(42)	1.230(2)
C(5) -C(6)	1.373(2)	C(6) -N(6)	1.452(2)
N(6) -O(61)	1.222(2)	N(6) -O(62)	1.231(2)
Cl -C(11)	1.743(2)	S -N(4)	1.631(1)
S -C(441)	1.788(2)	S -C(442)	1.780(2)
C(11)-C(22)	1.383(2)	C(11)-C(66)	1.374(2)
C(22)-C(33)	1.390(2)	C(33)-C(44)	1.386(2)
C(44)-N(44)	1.415(2)	C(44)-C(55)	1.396(2)
C(55)-C(66)	1.386(3)		

Average C-H bond distance: 0.950(22); N(4)-H(1) 0.796(24)Å

Table 2 (continued)

(b) Interbond angles

C(2)	-C(1)	-O(1)	124.9(1)	C(6)	-C(1)	-O(1)	123.3(1)
C(6)	-C(1)	-C(2)	111.7(1)	N(2)	-C(2)	-C(1)	120.2(1)
C(3)	-C(2)	-C(1)	124.2(2)	C(5)	-C(6)	-C(1)	124.3(1)
N(6)	-C(6)	-C(1)	119.7(1)	O(21)	-N(2)	-C(2)	119.6(1)
O(22)	-N(2)	-C(2)	118.8(1)	C(4)	-C(3)	-C(2)	119.1(1)
O(22)	-N(2)	-O(21)	121.7(1)	N(4)	-C(4)	-C(3)	118.6(1)
C(5)	-C(4)	-C(3)	121.7(1)	N(4)	-C(4)	-C(5)	119.8(1)
O(41)	-N(4)	-C(4)	117.9(1)	O(42)	-N(4)	-C(4)	118.7(1)
C(6)	-C(5)	-C(4)	119.1(1)	O(42)	-N(4)	-O(41)	123.3(2)
N(6)	-C(6)	-C(5)	116.0(1)	O(61)	-N(6)	-C(6)	119.0(1)
O(62)	-N(6)	-C(6)	117.9(1)	O(62)	-N(6)	-O(61)	123.1(1)
C(22)	-C(11)	-Cl	119.2(1)	C(66)	-C(11)	-Cl	119.6(1)
C(442)	-S	-N(44)	104.6(1)	C(44)	-S	-N(44)	104.8(1)
C(44)	-N(44)	-S	120.4(1)	C(442)	-S	-C(441)	100.8(1)
C(66)	-C(11)	-C(22)	121.2(2)	C(33)	-C(22)	-C(11)	119.1(2)
C(55)	-C(66)	-C(11)	119.2(2)	C(44)	-C(33)	-C(22)	120.3(2)
N(44)	-C(44)	-C(33)	118.9(1)	C(55)	-C(44)	-C(33)	119.8(1)
C(55)	-C(44)	-N(44)	121.4(1)	C(66)	-C(55)	-C(44)	119.7(2)

Table 3

Comparison of bond lengths and angles for the picrate moiety

(a) Bond lengths

		PICRATE ION				PICRATE MOLECULE	
		(I)[16]	(II)[17]	(III)[18]	(IV)[19]	(V)[20]	(VI)[19]
O(1)-C(1)	1.244	1.243	1.239	1.239	1.237	1.327	1.321
C(1)-C(2)	1.453	1.453	1.450	1.462	1.450	1.403	1.406
C(2)-C(3)	1.375	1.372	1.372	1.388	1.368	1.370	1.378
N(2)-C(2)	1.448	1.457	1.461	1.452	1.463	1.467	1.468
N(2)-O(21)	1.229	1.229	1.237	1.242	1.214	1.218	1.237
N(2)-O(22)	1.229	1.232	1.206	1.211	1.225	1.221	1.205
C(3)-C(4)	1.383	1.382	1.368	1.374	1.390	1.393	1.377
C(4)-C(5)	1.378	1.382	1.362	1.370	1.379	1.373	1.364
N(4)-C(4)	1.443	1.436	1.457	1.454	1.443	1.469	1.481
N(4)-O(41)	1.231	1.228	1.212	1.215	1.240	1.204	1.228
N(4)-O(42)	1.230	1.228	1.212	1.218	1.214	1.224	1.222
C(5)-C(6)	1.373	1.382	1.362	1.363	1.374	1.379	1.379
N(6)-C(6)	1.452	1.448	1.461	1.454	1.449	1.457	1.471
N(6)-O(61)	1.231	1.229	1.237	1.217	1.195	1.235	1.210
N(6)-O(62)	1.222	1.229	1.206	1.205	1.244	1.215	1.200
C(6)-C(1)	1.457	1.453	1.450	1.459	1.449	1.406	1.401

Table 3 (continued)

(b) Angles

		PICRATE ION				PICRATE MOLECULE	
		(I)	(II)	(III)	(IV)	(V)	(VI)
C(2)-C(1)-C(6)	111.7	111.1	111.5	111.6	112.2	115.3	115.6
C(1)-C(2)-C(3)	124.2	124.9	124.2	124.Ø	124.5	123.9	123.2
C(2)-C(3)-C(4)	119.1	118.5	119.1	119.5	119.1	117.5	117.7
C(3)-C(4)-C(5)	121.7	122.Ø	121.9	121.4	121.3	122.Ø	122.5
C(4)-C(5)-C(6)	124.3	125.9	124.2	124.1	125.Ø	122.7	122.4
C(5)-C(6)-C(1)	124.3	125.9	124.2	124.1	125.Ø	122.7	122.4

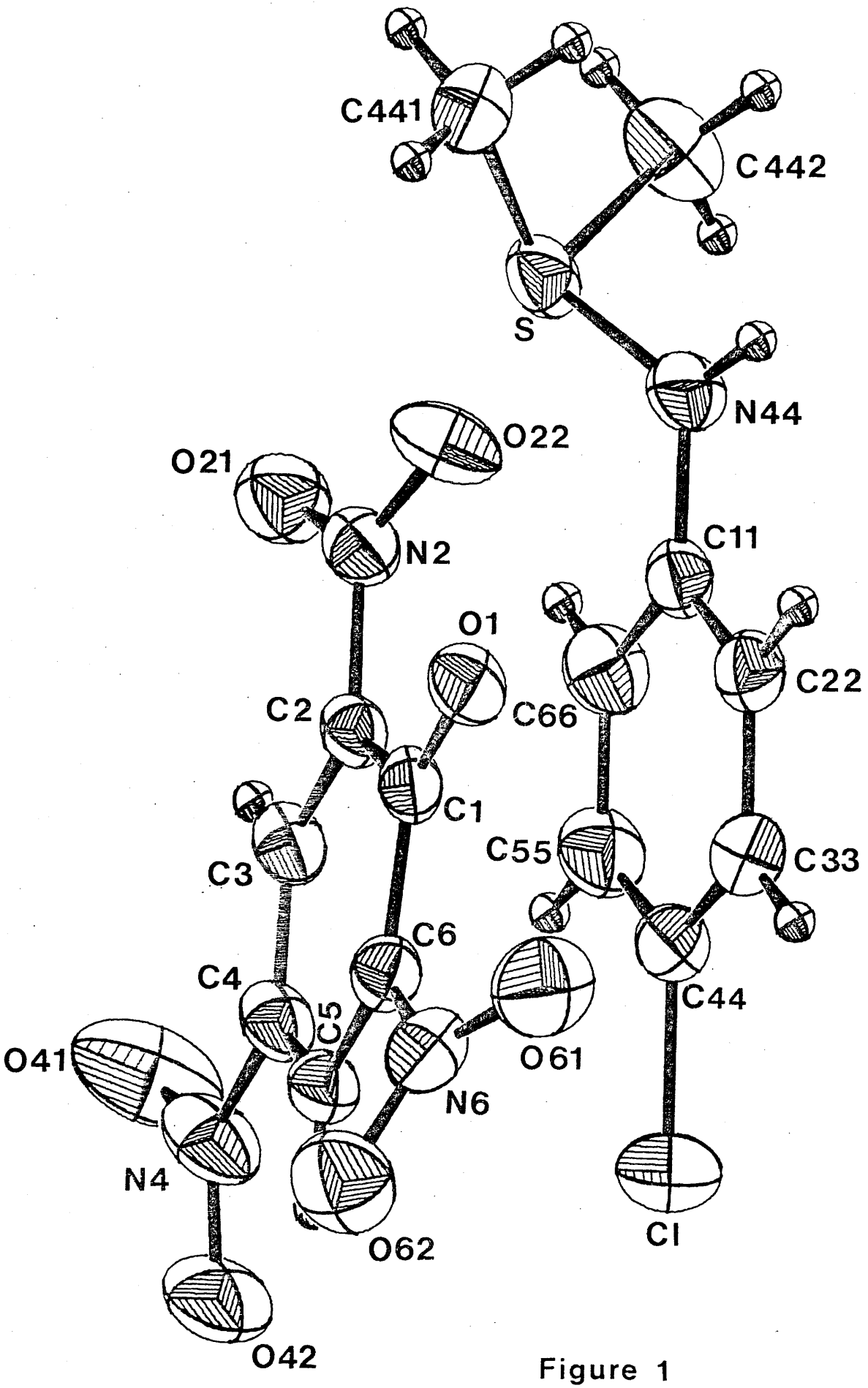

C441
C442
S
N44
O22
O21
N2
C11
O1
C66
C22
C2
C1
C55
C33
C3
C6
C4
C44
C5
O61
O41
N6
N4
O62
Cl
O42

Figure 1

Chapter 4

Quartets in Larger Structures:

Two Hexa-host Inclusion Compounds.

4.Ø INTRODUCTION

The structures of two hexa-host inclusion compounds have been determined *via* an enhanced version of MULTAN 78 which included quartet and quintet invariants as an integral part of the program. The first molecule discussed has squalene ($C_{3Ø}H_{5Ø}$) as the guest, which is, to date (March 198Ø), the largest molecule to be trapped in such compounds, and together with the host moiety, comprises 93 atoms in the crystallographic asymmetric unit.

As a step towards the concept of asymmetric synthesis, in which the crystalline lattice may determine the course of a reaction, the structure of a chiral hexa-host clathrate has been elucidated in which acetic acid dimers form the guest species to give a total asymmetric unit cell content of 86 atoms.

4.1 HIGHER INVARIANTS IN MULTAN

Several extensions to the MULTAN system have been described recently in which magic integer / Ψ-map and random phase set - linear equation algorithms have been employed[1]. However, they are confined to the use of three-phase invariants. The considerable activity in deriving formulae for estimating the magnitudes of four- and five-phase structure invariants[2-6] has already been stressed in previous chapters. Since these relationships contain new phase information it is logical to incorporate them as an extension of the MULTAN procedure.

Two modes of useage of higher invariants must be distinguished:

a) The active mode in which the invariants are used to generate new phase information.

b) The passive mode where the invariants are used only for figures of merit for selecting the most probable phase set.

I Quartets

The MULTAN 78 program was modified to use the 7-magnitude, 2nd. neighbourhood $P_{1/7}$ and the 13-magnitude, 3rd. neighbourhood $P_{1/13}$ joint conditional probability distributions of Hauptman[2] for the non-centrosymmetric case and the corresponding $P_7^{\pm}$ and $P_{13}^{\pm}$ formulae for centrosymmetric space groups[3]. These formulae, as described in Chapter 1, give reliable estimates for all combinations of the principal and cross-terms that comprise a quartet. The 7-magnitude formulae are straightforward in their estimation, but as pointed out in Chapter 2, $P_{1/13}$ and $P_{13}^{\pm}$ are used in such a way that any discrepancies occurring between the individual 3rd. neighbourhood estimates for Φ_4 are excluded from active use in phasing procedures. In a similar way quartets which exhibit discrepancies between 2nd. and 3rd. neighbourhood Φ_4 estimates are also excluded.

The use of the 13-magnitude formulae is thus an option in the program. When it is applied, a single estimate of V_{13} or $P_{13}^{\pm}$ is required for later use; the best value encountered in the 3rd. neighbourhood search is selected. Missing cross-terms in the 2nd. neighbourhood (but not the 3rd.) are also permitted as an option. From Chapter 2 it can be seen that this option can be useful for poor quality data sets.

II Quintets

As with quartets, there are several quintet distributions available (Chapter 1) employing the 1st. and 2nd. neighbourhoods. Of the formulae we have tested, the $P_{1/15}$

distribution of van der Putten and Schenk[5] (1977) gives reliable quintet estimates for the non-centrosymmetric case whilst the $P_{15}^{\pm}$ formula of Fortier and Hauptman[6] (1977) is used for centrosymmetric space groups. These formulae are available, in the passive mode only, as a user option in the enhanced MULTAN 78 system. Quintets are expensive to generate. In this program only the negative quintet subset is calculated. As a further check on reliability, the discriminant, Δ , as defined by Fortier et al. is applied. As an empirical observation, quintets having $\Delta \leqslant -1.0$ and which also have an estimated magnitude of 180° are the most reliably negative, and only these invariants are accepted.

Quintets are utilised only in the passive mode since their information content for one individual phase is quite low and can readily give rise to large accumulated errors when used to generate new phase angles, especially in conjunction with magic integer phase representation.

III Active Use of Quartets

In its original form, MULTAN 78 employs only triplets. As a measure of the reliability of each three-phase invariant an associated variable $A_{\underline{hkl}}$ is used, where:

$$A_{\underline{hkl}} = 2\sigma_3\sigma_2^{-3/2} |E_{\underline{h}}E_{\underline{k}}E_{\underline{l}}| \qquad (1)$$

and $E_{\underline{h}}$, $E_{\underline{k}}$ and $E_{\underline{l}}$ are the three E-magnitudes involved in the triplet. This variable is employed throughout the convergence mapping and tangent procedures. In order to mix the quartets with these relationships it is necessary to apply the same scale of reliability. For the non-centrosymmetric case, this is carried out as follows:

a) For each quartet the relevant joint conditional probability distribution $P(\Phi)$ (where $P(\Phi)$ is either $P_{1/7}$ or $P_{1/13}$) is calculated in 45° intervals. The mode, $|\Phi|$, is found.

b) The distribution is normalised *via* numerical integration using Simpson's rule, such that:

$$\int_0^{2\pi} P(\Phi)d\Phi = 1 \qquad (2)$$

c) The associated variance, V,is also found *via* numerical integration of the normalised distribution:

$$V = \int_0^{2\pi} (\Phi - |\Phi_m|)^2 P(\Phi)d\Phi \qquad (3)$$

d) Each quartet is assigned an equivalent A value, $A^{eq}_{\underline{hklm}}$, related to V (in degrees^2) by an empirically derived equation.

$$A^{eq}_{\underline{hklm}} \simeq 5583/(V+255) \qquad (4)$$

From experimentation it has been found that only quartets for which $A^{eq}_{\underline{hklm}} > 0.6$ can be accepted.

A similar procedure is used in the centrosymmetric case, but here the probability, P^+, is converted to $A^{eq}_{\underline{hklm}}$ by the relationship:

$$A^{eq}_{\underline{hklm}} = 0.5\log_{10}(\max(P^+,1-P^+)/(1.0-\max(P^+,1-P^+)) \qquad (5)$$

In practice this procedure is readily automated and does not require large amounts of computer time. The 3- and 4-phase invariants can now be freely mixed together throughout the convergence mapping and tangent refinement routines.

For the latter a version of the formula of van der Putten and Schenk[7] (1979) is used:

$$\tan \phi_{\underline{h}} = \frac{\sin_{tri} + \sin_{qua}}{\cos_{tri} + \cos_{qua}} = \frac{T_{\underline{h}}}{B_{\underline{h}}} \qquad (6)$$

where:

$$\sin_{tri} = \sum_{\underline{k}} \omega_{\underline{k}} \omega_{\underline{l}} A_{\underline{hkl}} \sin(\phi_{\underline{k}} + \phi_{\underline{l}})$$

and

$$\sin_{qua} = \sum_{\underline{k}} \sum_{\underline{l}} \omega_{\underline{k}} \omega_{\underline{l}} \omega_{\underline{m}} A^{eq}_{\underline{hklm}} (\sin\phi_{\underline{k}} + \phi_{\underline{l}} + \phi_{\underline{m}} + S|\Phi_4|)$$

with corresponding cosine expressions for $\cos_{tri}$ and $\cos_{qua}$;

$$\omega_{\underline{h}} = \tanh\{\sigma_3 \sigma_2^{-3/2} (T_{\underline{h}}^2 + B_{\underline{h}}^2)^{1/2}\}$$

The variable $S = \pm 1$ and is chosen such that:

$$\phi_{\underline{h}} + \phi_{\underline{k}} + \phi_{\underline{l}} + \phi_{\underline{m}} + S|\Phi_4|$$

is closest to zero. This is only relevant for enantiomorph sensitive quartets, which can thus only be employed when a value, albeit approximate, can be assigned to Φ_4.

It must be emphasised that equation (6) assumes that quartets and triplets are independent. In the case of triplets and negative quartets this is to a large extent true, since the latter utilises small E-magnitudes in the cross-terms. This mixture of quartets and triplets is one option of the program.

For the strongly positive quartets (those with a zero mode and low variance), one or more of the cross-terms will involve large E-magnitudes. If these E's are also used in the triplets then the independence of the 3- and 4-phase invariants is lost. As the number of such cross-terms in the quartet increases, this correlation will also increase until a point is reached where all three cross-terms involve large E-magnitudes. Under these circumstances, the quartet can be considered as an overlap of triplets with common phase angles. Thus the triplet and quartet contain similar information

although they are employed in a different way[8]

In the absence of a theoretical estimate of the covariances of these relationships, a simple linear weighting scheme is employed in which $A^{eq}_{\underline{hklm}}$ for the quartet is modified to give $A'_{\underline{hklm}}$:

$$A'_{\underline{hklm}} = A^{eq}_{\underline{hklm}} \cdot (1-n/3) \qquad (8)$$

where n is the number of cross-terms in the 2nd. neighbourhood for which the corresponding phase angle has been determined. The triplets keep the same weight. Thus at the beginning of phase determination, where very few phases are known, most quartets have their full weight in the tangent formula, but as the phasing procedure continues this weight is progressively reduced to zero for the strongly positive quartets, whilst the best negative invariants maintain their full weight.

This dynamic use of A poses problems during convergence mapping, since it is no longer possible to predict an a-priori value of A. In this case, the quartets are still included but they are given the minimum value of A likely to be achieved during phase expansion and refinement. This procedure seems to be satisfactory.

The use of quartet and quintet information in the passive mode in MULTAN is evoked in the calculation of the figures of merit, NQEST and NQINT, already described in Chapter 1. Figure 1 displays a flow diagram outlining the procedures and options of the program.

4.2 CLATHRATES AND MOLECULAR INCLUSION CHEMISTRY

The first organic clathrate to be discovered was the hydrogen sulphide included form of quinol by Wohler[9] in 1849. Subsequent investigations showed that quinol formed a series

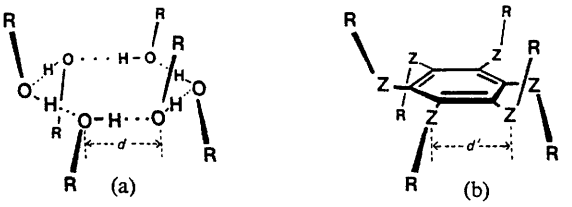
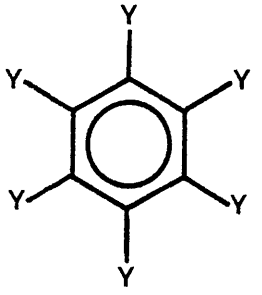


Figure 2

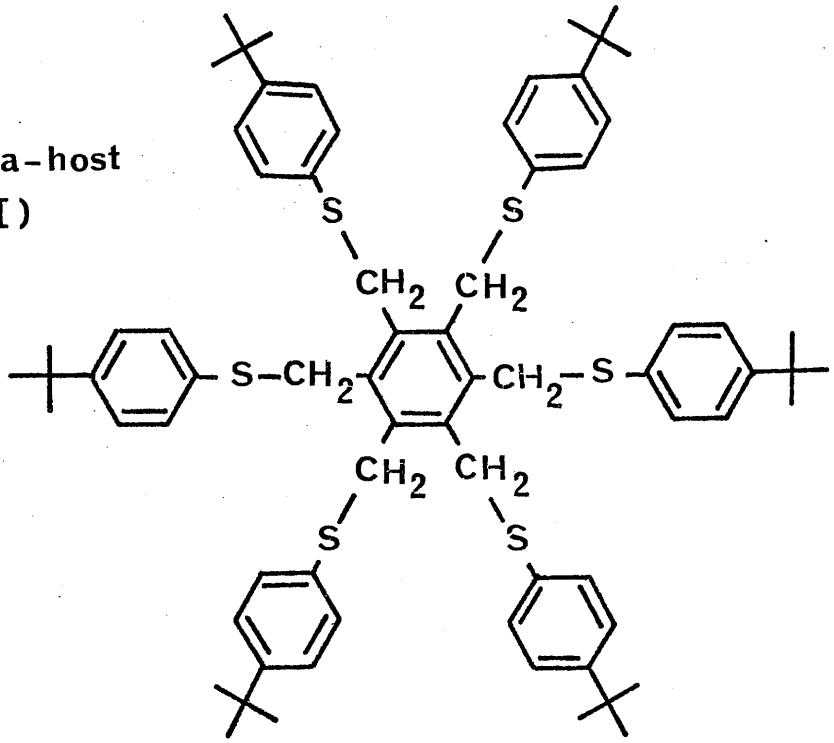
Following the idea that suitable hexa-substituted benzenes might have an increased chance of crystallising to form non-close packed structures, compounds with general formula (I) have been synthesised[21,22,23].

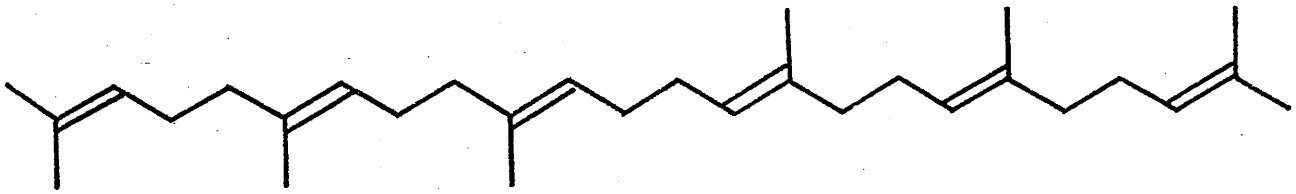


(I) a) Y = SPh
b) Y = CH_2OPh
c) Y = CH_2SPh
d) Y = CH_2SCH_2Ph
e) Y = $CH_2SC_6H_4But$-p
f) Y = $CH_2SeC_6H_4but$-p
g) Y = $CH_2SC_6H_4$(1-adamantyl)-p
h) Y = CH_2S-(2-naphthyl)

All of the compounds (Ia-h) exhibit inclusion ability, and (Ie), for example, forms adducts with toluene, cycloheptane, phenyl acetylene, bromoform, and iodo-benzene, with a host to guest ratio of 1:2 in each case. In some cases remarkable guest selectivity is found, 95% o-xylene and 5% p-xylene being included by host (Ie) when it is crystallized from an equimolar mixture of these solvents. Using this versatile hexa-host two compounds were crystallized containing unique guests which are reported herein.

Hexa-host
(I)





Squalene guest

I - THE SQUALENE ADDUCT OF HEXAKIS (p-t-BUTYLPHENYLTHIOMETHYL) BENZENE

4.1.1 INTRODUCTION

The advent of hexa-host inclusion chemistry, with its ability to trap more complex and interesting guests, leads to a convenient method whereby molecules that are liquids under normal conditions can be studied. In this manner the large triterpene, squalene ($C_{30}H_{50}$), was made available to X-ray diffraction techniques.

Squalene, found in large quantities in shark liver oil, is of biogenetic importance as a precursor to cholesterol *via* lanosterol. The precise course of cyclisation and concerted rearrangement of this reaction is determined by the conformation in which the flexible all-trans-squalene molecule is folded. The conformation adopted when squalene is 'frozen' in the crystalline state, as in I, can be directly compared to the conformation found in squalene (-110°C) by Sheldrick *et al.*[24].

4.1.2 EXPERIMENTAL

Crystal data

Squalene adduct of hexakis(*p*-*t*-butylphenylthiomethyl) benzene, $C_{72}H_{90}S_6 \cdot 1/2(C_{30}H_{50})$, M_r=1353.3, triclinic, a=14.710(5), *b*=15.773(6), *c*=20.417(5) Å, α=107.40(2), β=113.93(3), γ=81.93(3)°, U=4131.8 Å^3, D_c=1.09, D_m=1.10 Mgm^{-3}, *Z*=2, F(000) = 1466, space group $P\bar{1}$, μ(Mo-K_α) = 2.04 cm^{-1}.

Data collection:

Instrument used: Enraf Nonius CAD-4
Radiation used: Mo-K_α λ= 0.7107Å
Upper limit for data collection: $2\theta_{max} = 46^\circ$
Number of independent reflexions: m = 4230
Unobserved cut-off*: $2.5\sigma_I$
Number of parameters refined: n =374
Number of reflexions per parameter: m/n =11.3

* This cut-off was applied only in least-squares refinement; during structure solution, where it is important to have all available data for quartet analyses, some 12297 independent reflexions were used.

4.1.3 STRUCTURE DETERMINATION

The structure was solved by application of quartet invariants to the enhanced version of MULTAN described in the preceding section. 582 quartets and 2299 quintets were generated via QGEN for the top 120 E-magnitudes $\geq$ 2.5, employing the $P_7^{\pm}$, $P_{13}^{\pm}$ and the $P_{15}^{\pm}$ formulae respectively. For a quartet to be accepted P_7^+ was $\geq$ 0.60 and P_{13}^+ was $\geq$ 0.70, whilst for a quintet the associated probability was $\geq$ 0.70. This quartet information was added to the triple phase relationships derived from 450 E-magnitudes ($|E| \geq 2.1$) to give a total of 2881 invariants for the subsequent CONVERGENCE procedure. 49 negative quartets and 351 negative quintets ($\Delta \geq -1.0$) were used as contributors to their respective figures of merit.

Three origin defining reflexions and three variable reflexions resulted in eight solutions being computed which gave two E-maps capable of yielding a solution. Table 1.0 lists the figures of merit derived for these eight solutions.

The negative quartet-quintet figures of merit, NQEST and NQINT, without doubt gave the best indications as to the correct E-maps, having values of -Ø.83, -Ø.83 and -Ø.83, -Ø.87 respectively for the two correct solutions. As seen from Table 1.Ø, the three standard figures of merit were not quite so decisive, although an indication as to the correct solution is still possible.

From the best E-map (No.2), 65 out of a possible 93 atoms were located and recycled *via* the Sim-weighted Fourier option in MULTAN. At this stage no squalene molecule had been found. The consequent map completed all but two carbon atoms of the *t*-butyl groups and revealed eleven out of fifteen atoms of the squalene moiety. A further weighted Fourier calculation found no more atoms, and at this stage least-squares refinement was initiated.

The method by which the crystal structure of (I) was elucidated has shown quartet invariants as a worthwhile inclusion into an integrated multisolution technique. As an experiment, the data were processed *via* the standard MULTAN package, i.e. using triplet information only. Employing the same number of variables as before, with the default values set by the program, no solution was found; probably accountable to the appearance of 'holes' in the early stages of the convergence map.

Persevering with the standard MULTAN and changing the default values by expanding the number of variables used, one could perhaps, obtain the correct solution. However, it is worthwhile noting that with 93 atoms in the asymmetric unit, the number of Σ_2 relationships attained falls short of the required quota, if the conventional ratio of relationships per atom is to be observed. As illustrated in Chapter 2, quartet invariants provide the necessary increase to satisfy this ratio for the number of atoms involved, without having to increase the number of E-magnitudes to more than 5ØØ.

Moreover, by successful application of the enhanced version of MULTAN, it took only 1Ø3 seconds to generate the necessary quartet invariant information, and this compares favourably with the more time consuming process (2ØØsecs.) for re-running the program, even once, if the standard MULTAN package had been employed.

4.1.4 STRUCTURE REFINEMENT

The structure was refined using SHELX. Two cycles of isotropic refinement produced a difference map which indicated two features of the host:guest molecule from which problems were to arise. Firstly, on subsequent location of all the atoms of the t-butyl groups there remained spurious peaks adjacent to the refined terminal t-butyl coordinates. This was diagnosed as disorder of the t-butyl groups. Bond length and angle calculations led to the conclusion that the residual peaks were atoms occupying proportional population with the existing terminal groups. The disordered peaks were incorporated in the isotropic least-squares calculations by assignment, and refinement, of the site occupation parameters via a free variable for each t-butyl group, such that the sum of their values was fixed at 1.Ø.

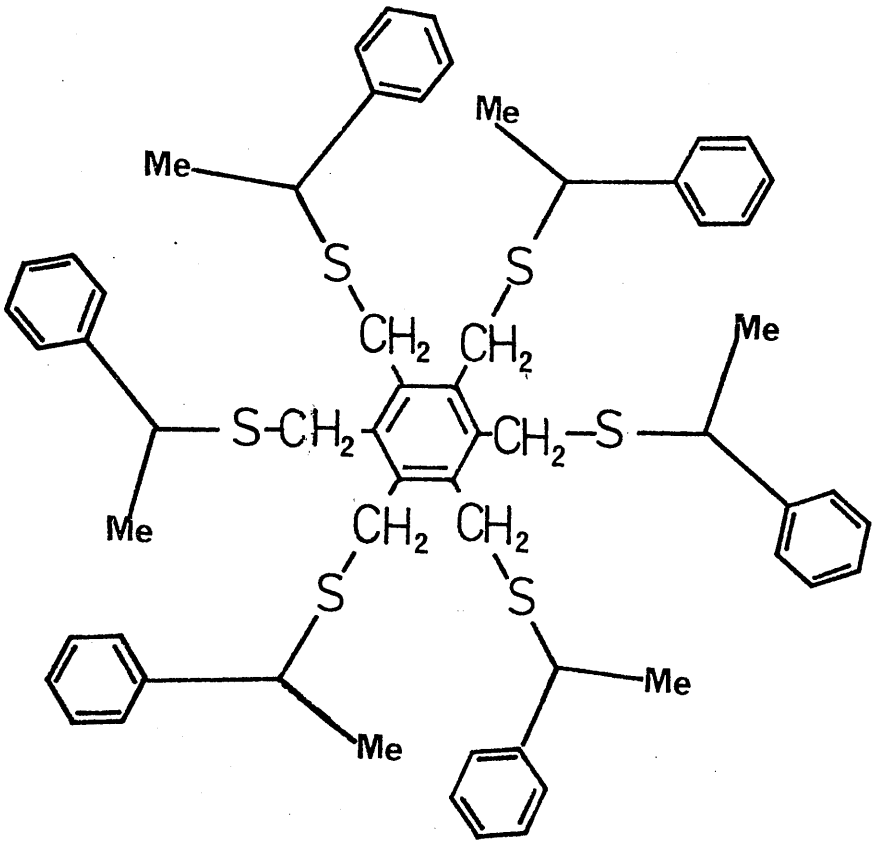
Secondly, the squalene molecule also appeared to be disordered at each end of the chain. However, it was possible to resolve this disorder by consideration of an overlap of two squalene molecules. From a difference map it was observed that the squalene molecule was lying diagonally across the unit cell, i.e. from (Ø,Ø,Ø), through the centre of symmetry at (Ø,1/2,1/2), to (Ø,1,1) as a continuous chain. To make sense of this continuous chain disorder it is necessary to envisage two squalene molecules overlapping, with one of the squalene molecules moved by one isoprene unit with respect to the other molecule. Figure 3 shows this diagramatically where the second molecule is overlayed on the first, but with the

required shift. The effect of carrying out this manoeuvre is to impart a centre of symmetry at either side of the original molecular centre of of symmetry, depending on which way the single isoprene shift is applied (in this case moving to the left of the molecular centre). In this way the crystal or space group centre of symmetry falls across the double bond which has had this additional symmetry element conferred on it by the isoprene shift manoeuvre, and thus effects the continuity of the chain. As a result the methyl groups at either 'end' of the chain must assume half population since at each end it has been reflected through the centre of symmetry.

The single isoprene displacement is the only shift possible, since on moving the second molecule two or more isoprene units would result in other methyl groups appearing as atoms on the other side of the double bonds - a careful study of the difference maps gave no indication of this.

At this stage isotropic refinement converged to an R value of Ø.142. The six benzene rings on the hexa-host were then refined as rigid bodies, while the sulphur atoms and the fourteen atoms of the squalene moiety that had unitary population parameters were refined anisotropically. This procedure was adopted due to the restrictions, imposed by SHELX, on the number of atoms that are allowed to be refined anisotropically. The final least-squares calculation only included 12 out of a possible 115 hydrogen atom positions, a fact that is reflected in the final weighted R-value, R_w = Ø.1Ø4. The 12 hydrogens included were the methylene hydrogens associated with the carbon atoms adjacent to the sulphurs in the hexa-host.

Tables 1.1 record the atomic coordinates and temperature parameters for both host and guest molecules, with Table 1.2 showing bond lengths and angle calculations for the host and Table 1.3 giving bond lengths and angles for the guest molecule. Figure 8 gives relevant torsion angles for the host molecule.



Chiral hexa-host

(II)

II - THE ACETIC ACID ADDUCT OF HEXAKIS (R-α-PHENYLETHYLSULPHONYL METHYL) BENZENE

4.2.1 INTRODUCTION

The possibility of utilising the chirality of crystals to achieve asymmetric synthesis was considered as early as 1908[25]. Asymmetric synthesis through reactions in chiral crystals involves two aspects: generating chiral crystals and performing topochemically controlled, solid-state reactions which yield chiral products[26].

Since many clathrates (e.g. urea, tri-o-thymotide) form chiral crystals, the use of these materials may represent a widely applicable method for 'engineering' chiral crystalline matrices with achiral guest molecules. However, by design of a chiral hexa-substituted host, where the intermolecular features are conducive to chiral guest molecules, the prospect of asymmetric synthesis is enhanced. Leiserowitz[27] has considered the packing modes of carboxylic acids and has shown that the commonly observed centrosymmetric, hydrogen bonded dimer generally results in centrosymmetric, achiral crystals.

The crystal structure of the acetic acid adduct of hexakis(R-α-phenylethylsulphonylmethyl)benzene (II) was therefore undertaken to resolve the chirality of the guest species, and the conformation adopted by the novel hexa-host.

4.2.2 EXPERIMENTAL

Crystal data

Acetic acid adduct of hexakis(R-α-phenylethylsulphonylmethyl) benzene, $C_{60}H_{66}S_6O_{12}.4(CH_3CO_2H)$, M_r=1411.6, monoclinic, a=16.319(5), b=13.869(4), c=16.731(5)Å, β=106.47(3)°, U=3631.7 Å^3, D_c=1.11, D_m=1.12 Mgm^{-3}, Z=2, F(000) = 1492, space group $P2_1$, μ(Mo-K_α)=2.6 cm^{-1}.

Data collection

Instrument used: Enraf Nonius CAD-4
Radiation used: Mo-K_α λ= 0.71069 Å
Upper limit of data collection: $2\theta_{max}$ = 54°
Number of independent reflexions: m = 3378
Unobserved cut-off*: $2.5\sigma_I$
Number of parameters refined: n = 337
Number of parameters per reflexion: m/n= 10.0

* For structure solution the complete data set of 9408 reflexions was used, which was then reduced for least-squares analyses.

4.2.3 STRUCTURE DETERMINATION

The structure was solved by direct methods using MULTAN which incorporated quartet invariants in an active mode. From QGEN, 1279 unique quartets and 6094 quintets were obtained for the top 100 E-magnitudes using $P_{1/7}$ and $P_{1/13}$ (quartets) and $P_{1/15}$ (quintets) with a σ_7/σ_{13} and σ_{15} limit of 50°.

A subset of 110 negative quartets was integrated with 4848 triple phase relationships derived from 470 E-magnitudes $\geq$ 1.76 to yield a starting set of reflexions in which four variable reflexions were permuted via magic integer

phase representation. 266 negative quintets contributed to NQINT.

Subsequent tangent refinement resulted in 40 phase solutions being calculated with the derived figures of merit in close agreement. From the best indication (NQINT = -0.394, NQEST =-0.127) an E-map revealed 32 peaks of the host molecule which were recycled via Sim-weighted Fourier techniques to produce a complete structure in which two acetic acid dimers were located.

4.2.4 STRUCTURE REFINEMENT

These approximate atomic coordinates were adjusted by several cycles of full-matrix least-squares calculations employing the SHELX system, whereby isotropic refinement converged at R = 0.126. Subsequently, the six phenyl rings were refined as isotropic rigid bodies, and the temperature factors of all other atoms assigned to the anisotropic mode. Introduction of the weighting scheme of the type:

$$w = k/(\sigma^2(F) + abs(g).F.F)$$

where k and g were redetermined after each structure factor calculation by fitting $(F_o-F_c)^2$ to $(\sigma^2(F) + abs(g).F.F)/k$, converged the least-squares calculations after a further four cycles of refinement to a final R-value (R_w) of 0.078. As with the squalene adduct it was not possible to include the 80 hydrogen atoms remaining in the structure. A difference map located the positions of the two protons associated with one of the dimer guest moieties - these hydrogen atoms were included in the final least-squares cycle. Atomic fractional coordinates and temperature parameters for (II) are given in Table 2.1, while Table 2.2 summarises bond length and angle calculations along with relevant mean plane geometry. Figure 9 displays the torsion angles for the host and

Table 2.3 gives the relevant geometry for both dimer guests.

4.2.5 DISCUSSION

The crystal structure determination of the title compounds show that in (I) an inclusion compound of the channel type prevails, with the squalene moiety existing as a continuous chain running throughout the crystal lattice sandwiched between hexa-host molecules in a host to guest ratio of 2:1. In (II) the acetic acid dimers are found as discrete entities dispersed in the chiral hexa-host lattice in such a manner as to form a true clathrate with a host to guest ratio of 1:2, where the latter figure refers to a dimer molecule. Figures 6 and 7 illustrate the packing arrangements for both structures.

a) The Host Molecules.

A view looking directly onto the plane of the central benzene ring for each host molecule, (I) and(II), is shown in Figures 4 and 5 respectively. In (I) the centroid of the central benzene ring is located, within experimental error, at Ø,1/4,1/4 (and Ø,3/4,3/4) in the unit cell (cf.squalene), whereas a general position is encountered for (II). A comparison may be sought between the two hosts, (I) and (II), and several other hexa-host molecules, described by MacNicol et al., in which the molecule (III), hexakis(benzylthiomethyl) benzene[23], is indicative of the series. The four structures referenced therein crystallize in centrosymmetric space groups with the centroid of the central benzene ring positioned at a centre of symmetry.

It would therefore have been reasonable to expect the achiral hexa-host (I), which crystallized in the centrosymmetric space group $P\bar{1}$, to adopt a similar position in the unit cell whereby the centroid of the central ring was

placed on a centre of symmetry. The fact that this is not so suggests competition between the centrosymmetric guest, squalene, and the host for the centre of symmetry: a competition which the guest wins, and the host is relegated to the non-crystallographic special position at (0,1/4,1/4). This unique preference afforded the squalene guest is possibly due to the difficulty of imposing a centre of inversion on the multi-legged host with its encumbent phenyl rings and disordered t-butyl groups; whilst it is simpler to satisfy the conditions of an intramolecular inversion centre on the rationally disordered squalene.

On the other hand, the chiral host, (II), and, for that matter, the acentric acetic acid dimer guests, occupy general positions in the cell and, as the geometry of both host and guest moieties will show, this does not preclude the occurrence of approximate pseudo-symmetrical arrangements within each species.

This possible correlation between molecular structure and crystal symmetry has been shown by Jacques[28] et al., who reported that molecules having a two-fold symmetry axis, C_2, tend to crystallize in chiral structures. Furthermore, the postulate by Green[26], that the probability of adopting a chiral structure is enhanced with molecules having three-fold symmetry, gives added interest to the conformation of (II).

The sulphur atoms in the 'legs' of (I) and (II), as with (III), are situated alternately above and below the plane of the central benzene ring. In each case the legs are staggered with the methylene carbon going in the opposite direction, with respect to the ring plane, from its sulphur neighbour. Tables 1.2c and 2.2c give the deviations of the relevant atoms and average displacement of the terminal phenyl rings from the mean plane of the central ring for (I) and (II) respectively.

It is noteworthy that in (I) and (III) the deviations for the methylene carbons are in the same direction as the

deviations of the adjacent atoms that comprise the central ring; whereas in (II) this sympathetic deviation is absent. Although only small deviations are experienced, it is thought to be related to the steric hinderance encountered by the introduction of the bulkier chiral substituents. In both cases the phenyl groups alternate above and below the central ring plane.

Bond lengths and angles for both hexa-host moieties are in good agreement having average central ring values of 1.399(20)Å, 1.397(12)Å and 120.2°(12),120.0°(9) for (I) and (II) respectively, while the longer methylene-sulphur bonds experienced in (I), compared to (II), are a result of delocalization of the sulphone group.

Consideration of the six torsion angles about the ring-methylene plane gives values close to 90° in both (I) and (II), which, coupled with the proximity of the >C-CH_2-S-C* torsion angles to 180°, reveals approximate three-fold core symmetry in each case. This is also observed for (III). An interesting feature of the chiral host is the all positive signs for the torsion angles about the methylene-sulphur bonds, and is due to the terminal phenyl rings adopting a preferred orientation to avoid steric overcrowding associated with the additional methyl on the chiral carbon.

The approximate three-fold symmetry is extended in (II) beyond the core atoms to the periphery of the host molecule, where -CH_2-S-C*-Me torsion angles have values of 180, 68, -177, 65, -176 and 66° while -CH_2-S-C*-Ph torsion angles have values of 56, -62, 60, -60, 61 and -63°. Since this three-fold symmetry is only observed for the chiral species it could be argued that this is the corollary of the earlier proposal by Green.

b) Squalene

Table 1.3 gives relevant bond distances and interbond angles for the guest of (I). The longer bond lengths which occur at each end of the molecule are undoubtedly due to the disorder already discussed. However, thermal parameters and e.s.d's for the central portion of the molecule, C(71)-C(81), do not deviate significantly from values normally associated with such molecules, and therefore the salient geometry of the squalene guest is worthy of comment.

Double bonds, C(72)-C(74) 1.278(37)Å and C(77)-C(79) 1.305(33)Å, though slightly shorter than those found in squalene at -110°C, display a characteristic lack of delocalization also found in β-carotene and its derivatives. Figure 10 shows an ORTEP drawing of the squalene guest whilst Figure 11 compares the torsion angles of squalene in (I) with those of squalene at -110°C. Distinct conformational differences, as could be expected, exist between the two conformers. The conformation adopted in (I) can be likened to a'square-wave' form, comprised of alternating planar isoprene units. The dihedral angle between these alternating isoprene units is, on average, ca. 12°. At -110°C a more twisted squalene structure is observed.

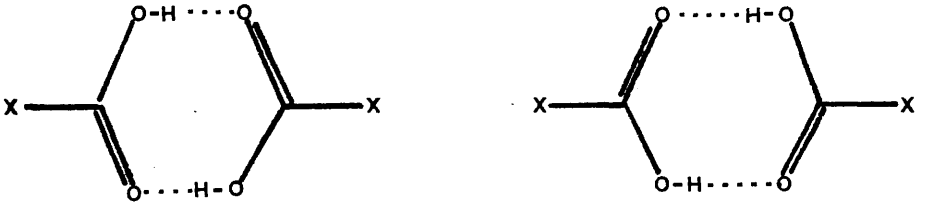
c) The Acetic Acid Dimers

The guest moiety of (II) consists of two discrete pairs of non-coplanar acetic acid dimers occupying different positions within the host environment. The first dimer considered is sandwiched between the central rings of two host molecules in such a manner that each terminal methyl group points directly at the central host ring. The average distance from the central ring to the methyl group is 3.59Å.

Mean plane calculations involving the ring atoms C(1)-C(6); O(100), O(101), C(100), C(101), and O(200), O(201), C(200), C(201) show the mean dimer plane at an angle of 89.8°

to that of the central ring plane, whilst a dihedral angle of 17.4^{o} is subtended between the two acid groups.

The bond lengths, Table 2.3, show almost complete orientational disorder of the carbonyl groups. The average C-O distance 1.242(31)Å can be compared to the distinct C=O and C-O distances, 1.206 and 1.321 respectively, found for acetic acid (-140^{o}C) by Jönsson[29] (1971) using neutron diffraction data. The disorder is due to the existence of two mutually indistinguishable equivalent orientations, (a) and (b) for the carbonyl dimer.



(a) (b)

Currie, Speakman and Curry[30] have shown the differences between C=O (carbonyl) and C-OH (hydroxyl) distances vary widely (0.04-0.12Å), accountable to the existence of orientational disorder between (a) and (b). Leiserowitz, in his comprehensive study of carboxylic acids, has described two modes of disorder. These modes are termed static or dynamic, depending on whether there is (the latter case) or is not proton transfer across the O-H.....O bond.

In the case of dimer A both protons were located. However, one hydroxyl proton is bonded to O(100) 0.93 Å and the other hydroxyl proton is approximately centrally situated in a symmetric O-H-O bond 1.60 Å. It is therefore difficult to define exactly the disorder present in this dimer, although it would appear that in the C(100)-O(101) and C(200)-O(201) bond systems dynamic disorder, where the C-O bonds are in a state of resonance as the proton undergoes rapid oscillations across the O(101)-H-O(201) bond, is the major contributor.

In trimesic acid[31] where similar bond lengths (1.255A(C-OH) and 1.244Å(C=O)) are found, there is introduced the interesting correlation between disorder and thermal motion of the oxygen atoms involved, a factor which undoubtedly contributes to the equivalence of the C-O bonds found in dimer A.

The second dimer trapped by the host species, Dimer B, is to be found adjacent to the legs carrying atoms S(1) and S(6). The high temperature parameters associated with this dimer makes interpretation of bond lengths and angles meaningless, even on the basis of orientational disorder. From several I.R. spectra, recorded prior to data collection, it was evident that the host:guest ratio was dependent on the sample medium used. Since in our X-ray analysis we had enclosed the crystal in a capillary with mother liquor it would appear that this second dimer is the more volatile component.

REFERENCES

1. J. P. Declercq, G. Germain and M. M. Woolfson, Acta Cryst. (1979) A35, 622.
2. H. Hauptman, Acta Cryst. (1977a) A33, 556.
3. H. Hauptman, Acta Cryst. (1977b) A35, 565.
4. C. Giacovazzo, Acta Cryst. (1976) 32, 91.
5. N. van der Putten and H. Schenk, Acta Cryst. (1977) A33, 856.
6. S. Fortier and H. Hauptman, Acta Cryst. (1977) A33, 829
7. N. van der Putten and H. Schenk, Acta Cryst. (1979) A35, 381.
8. C. Giacovazzo, Acta Cryst. (1980) A36, 74.
9. F. Wohler, Annalen, (1849) 69, 294.
10. F. Mylius, Ber. (1886) 19, 999.
11. D. E. Palin and H. M. Powell, J. Chem. Soc. (1947) 208.
12. D. Lawton and H. M. Powell, J. Chem. Soc. (1958) 2339.
13. M. von Stackelberg, A. Hoverath and C. H. Scheringer, Z. Elektrochem (1958) 62, 123.
14. A. P. Dianin, J. Soc. Phys. Chim. Russe. (1914) 46, 1310.
15. D. D. MacNicol, H. H. Mills and F. B. Wilson, Chem. Comm. (1969) 1332.
16. J. L. Flippen, J. Karle and I. L. Karle, J. Amer. Chem. Soc. (1970) 92, 3749.
17. H. M. Powell, 'Non-stoichiometric Compounds', Academic Press, New York, (1964) pp. 438.
18. L. C. Fetterly, Ibid. pp.491.
19. D. D. MacNicol and D. R. Wilson, Chem. Comm. (1976) 494.
20. J. L. Flippen and J. Karle, J. Phys. Chem. (1971) 75, 3556.
21. H. H. Mills, D. D. MacNicol and F. B. Wilson, unpublished results.
22. D. D. MacNicol and D. R. Wilson, Chem. and Ind. (1977) 84.
23. D. D. MacNicol, A. D. U. Hardy and D. R. Wilson, Nature (1977) 266, 611
24. J. Ernst, W. S. Sheldrick and J. H. Fuhrhop, Agnew. Chem. Int. Ed. Engl. (1976) 15, 778.
25. T. Ostromusslensky, Chem. Ber. 41, 3035.
26. B. S. Green, M. Lahav and D. Rabinovich, Acc. of Chem. Res. (1979) 12, 6.

27. L. Leisserowitz, Act Cryst. (1976) B32, 775.
28. A. Collet, M. Brienne and J. Jacques, Bull. Soc. Chim. Fr. (1972) 127.
29. P. G. Jönsson, Acta Cryst. (1971) B27, 893.
30. M. Currie, J. C. Speakman and N. A. Curry, J. Chem. Soc. (A), (1967) pp. 1862.
31. D. J. Duchamp and R. E. Marsh, Acta Cryst. (1969) B25, 5.

Table 1.0

Figures of Merit (Squalene adduct)

NUMSET	ABSFOM	PSI ZERO	RESID	NQEST	NQINT	COMB. FOM
1	1.0195	1.559	24.25	-0.88	-0.88	4.4464
2	1.0375	1.991	22.95	-0.83	-0.83	4.4565
3	1.0166	2.331	24.42	-0.10	-0.47	2.1619
4	0.9961	1.807	26.02	-0.10	-0.47	2.1517
5	1.0053	2.197	25.90	-0.33	-0.41	2.0229
6	0.9862	1.677	27.30	-0.33	-0.41	2.0849
7	0.9720	1.873	27.61	0.09	0.04	0.6078
8	0.9912	2.357	26.37	0.09	0.04	0.5578

Table 1.1A

Atomic coordinates and thermal parameters for hexa-host(I)

(i) Atomic coordinates ($x10^4$)

ATOM	x/a	y/b	z/c
C(1)	8962(9)	7495(7)	7330(6)
C(01)	7878(8)	7405(7)	7101(6)
S(1)	7375(3)	6502(2)	6254(2)
C(11)	6328(5)	6191(5)	6340(4)
C(12)	6049	5308	5996
C(13)	5212	5011	6015
C(14)	4656	5597	6378
C(15)	4935	6481	6722
C(16)	5772	6778	6704
C(17)	3727(10)	5287(8)	6404(7)
C(171)	3504(29)	4280(25)	5964(22)
C(172)	3928(29)	5365(26)	7232(21)
C(173)	2870(30)	5984(25)	6199(23)
C(171')	2969(24)	4950(21)	5623(17)
C(172')	3982(23)	4484(21)	6772(18)
C(173')	3177(25)	5988(21)	6830(19)
C(2)	9639(10)	6955(7)	7759(7)
C(02)	9256(9)	6229(7)	7924(6)
S(2)	8983(3)	6641(2)	8764(2)
C(21)	8443(7)	5660(4)	8691(5)
C(22)	8993	5077	9113
C(23)	8579	4284	9018
C(24)	7615	4074	8500
C(25)	7064	4658	8078
C(26)	7478	5451	8173
C(27)	7171(12)	3158(10)	8393(8)
C(271)	6196(21)	3367(17)	8481(16)
C(272)	7548(21)	2449(17)	7895(16)
C(273)	7634(19)	2891(17)	9152(15)
C(271')	6038(39)	3142(33)	7881(31)
C(272')	8058(39)	2451(32)	8580(30)
C(273')	6739(40)	2744(32)	7470(28)

Table 1.1A (continued)

ATOM	x/a	y/b	z/c
C(3)	10656(10)	7059(7)	7995(6)
C(03)	11373(9)	6525(7)	8501(6)
S(3)	11645(3)	5466(2)	7926(2)
C(31)	12410(6)	4883(5)	8589(4)
C(32)	13169	4342	8428
C(33)	13777	3837	8906
C(34)	13626	3872	9546
C(35)	12867	4413	9707
C(36)	12259	4918	9229
C(37)	14275(11)	3327(9)	10073(7)
C(371)	15340(23)	3053(20)	9967(17)
C(372)	14554(25)	3865(21)	10921(17)
C(373)	13753(24)	2463(20)	9905(17)
C(371')	15284(29)	3724(26)	10529(22)
C(372')	14318(31)	2326(26)	9650(22)
C(373')	13922(32)	3389(27)	10732(22)
C(4)	11008(9)	7658(7)	7784(6)
C(04)	12104(9)	7707(7)	7983(7)
S(4)	12636(3)	8595(2)	8833(2)
C(41)	13686(5)	8903(5)	8733(4)
C(42)	13992	9782	9062
C(43)	14814	10051	9004
C(44)	15329	9443	8617
C(45)	15022	8565	8288
C(46)	14200	8295	8346
C(47)	16243(11)	9744(9)	8558(8)
C(471)	16384(23)	10747(19)	8861(17)
C(472)	17172(24)	9114(20)	8848(18)
C(473)	16013(22)	9518(22)	7696(16)
C(471')	16922(36)	10397(30)	9281(26)
C(472')	16983(36)	8965(31)	8371(28)
C(473')	15864(34)	10185(34)	7900(25)

Table 1.1A (continued)

ATOM	x/a	y/b	z/c
C(5)	10340(10)	8201(7)	7359(6)
C(05)	10724(9)	8918(7)	7190(6)
S(5)	10953(3)	8461(2)	6327(2)
C(51)	11542(6)	9388(4)	6340(5)
C(52)	11146	9703	5707
C(53)	11591	10408	5677
C(54)	12430	10798	6278
C(55)	12826	10484	6911
C(56)	12381	9779	6942
C(57)	12951(10)	11556(8)	6245(7)
C(571)	12566(27)	12455(21)	6632(18)
C(572)	14111(25)	11357(21)	6546(20)
C(573)	12621(25)	11610(22)	5393(18)
C(571')	12205(32)	12052(28)	5605(22)
C(572')	13308(34)	12260(26)	6992(23)
C(573')	13857(31)	11134(27)	6034(25)
C(6)	9321(10)	8094(7)	7107(7)
C(06)	8604(9)	8622(7)	6607(6)
S(6)	8297(3)	9669(2)	7182(2)
C(61)	7642(6)	10273(5)	6500(4)
C(62)	6609	10371	6423
C(63)	6112	10884	5745
C(64)	6649	11299	5504
C(65)	7683	11201	5761
C(66)	8179	10688	6259
C(67)	6107(10)	11881(9)	4985(7)
C(671)	5990(2)	11372(17)	4199(15)
C(672)	5031(21)	12170(18)	4942(15)
C(673)	6611(22)	12844(17)	5338(14)
C(671')	5221(39)	11374(31)	4315(26)
C(672')	6848(37)	12188(32)	4668(27)
C(673')	5772(40)	12722(32)	5349(27)

Table 1.1A (continued)

(ii) Thermal parameters ($Å^2 x 10^3$)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
C(1)	50	35	51	23	19	00
C(01)	30	59	61	18	14	-06
S(1)	50	84	62	10	25	-16
C(11)	49					
C(12)	69					
C(13)	73					
C(14)	51					
C(15)	71					
C(16)	61					
C(17)	62					
C(171)	110					
C(172)	110					
C(173)	110					
C(171')	110					
C(172')	110					
C(173')	110					
C(2)	51	38	63	18	28	-09
C(02)	67	48	52	20	31	-06
S(2)	99	56	70	19	48	-15
C(21)	57					
C(22)	78					
C(23)	75					
C(24)	72					
C(25)	89					
C(26)	76					
C(27)	87					
C(271)	110					
C(272)	110					
C(273)	110					
C(271')	110					
C(272')	110					
C(273')	110					

Table 1.1A (continued)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
C(3)	48	41	47	14	19	-09
C(03)	50	53	56	18	18	08
S(3)	73	63	60	26	25	20
C(31)	53					
C(32)	75					
C(33)	86					
C(34)	64					
C(35)	89					
C(36)	83					
C(37)	77					
C(371)	110					
C(372)	110					
C(373)	110					
C(371')	110					
C(372')	110					
C(373')	110					
C(4)	36	51	48	14	15	-08
C(04)	52	58	70	20	28	-12
S(4)	52	79	65	14	22	-11
C(41)	50					
C(42)	78					
C(43)	81					
C(44)	56					
C(45)	63					
C(46)	67					
C(47)	78					
C(471)	110					
C(472)	110					
C(473)	110					
C(471')	110					
C(472')	110					
C(473')	110					

Table 1.1B

Atomic coordinates and thermal parameters for squalene guest

(i) Atomic coordinates ($x10^4$)

ATOM	x/a	y/b	z/c
C(68)	9761(27)	0821(21)	10254(23)
C(69)	10250(23)	0100(23)	9750(21)
C(70)	10350(14)	0950(12)	9300(16)
C(71)	11345(11)	1099(13)	8998(15)
C(72)	11059(12)	1950(11)	8701(10)
C(73)	10920(15)	2781(14)	9260(11)
C(74)	11165(16)	1925(12)	8093(11)
C(75)	10842(14)	2696(13)	7723(9)
C(76)	9972(13)	2384(10)	6984(9)
C(77)	9572(11)	3120(11)	6564(9)
C(78)	9096(16)	3904(11)	6917(12)
C(79)	9748(14)	2978(13)	5953(11)
C(80)	9407(13)	3687(12)	5489(11)
C(81)	10356(12)	4127(12)	5579(10)
C(82)	10084(21)	4688(11)	4936(25)
C(83)	10095(33)	4069(22)	4164(21)

Table 1.1B

Atomic coordinates and thermal parameters for squalene guest

(i) Atomic coordinates ($x10^4$)

ATOM	x/a	y/b	z/c
C(68)	9761(27)	0821(21)	10254(23)
C(69)	10250(23)	0100(23)	9750(21)
C(70)	10350(14)	0950(12)	9300(16)
C(71)	11345(11)	1099(13)	8998(15)
C(72)	11059(12)	1950(11)	8701(10)
C(73)	10920(15)	2781(14)	9260(11)
C(74)	11165(16)	1925(12)	8093(11)
C(75)	10842(14)	2696(13)	7723(9)
C(76)	9972(13)	2384(10)	6984(9)
C(77)	9572(11)	3120(11)	6564(9)
C(78)	9096(16)	3904(11)	6917(12)
C(79)	9748(14)	2978(13)	5953(11)
C(80)	9407(13)	3687(12)	5489(11)
C(81)	10356(12)	4127(12)	5579(10)
C(82)	10084(21)	4688(11)	4936(25)
C(83)	10095(33)	4069(22)	4164(21)

Table 1.1B (continued)

(ii) Thermal parameters ($Å^2 x 10^3$)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
C(68)	127	120	237	-012	114	-019
C(69)	092	157	758	142	-129	-082
C(70)	268	261	768	-049	129	000
C(71)	147	165	141	105	-003	025
C(72)	088	113	093	037	014	-005
C(73)	161	156	143	021	077	024
C(74)	186	117	112	031	049	000
C(75)	134	167	092	066	002	-047
C(76)	127	103	097	047	028	-024
C(77)	071	104	082	023	-005	-023
C(78)	169	086	206	-003	098	018
C(79)	129	168	107	087	014	-013
C(80)	111	169	180	128	053	000
C(81)	095	159	158	127	015	-007
C(82)	069	099	313	018	028	-013
C(83)	208	074	121	013	110	015
Average e.s.d.'s						
C	24	19	30	16	18	15

Table 1.1C

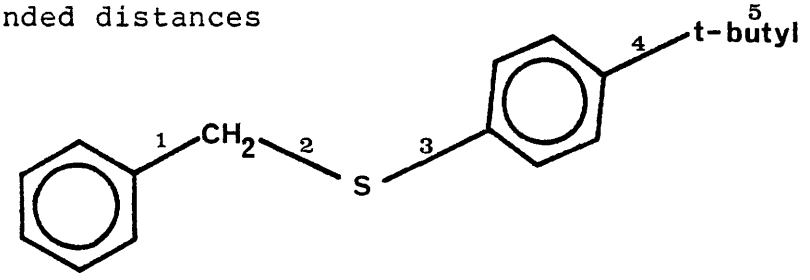
Atomic coordinates and isotropic thermal parameters for methylene hydrogen atoms.

ATOM	x/a	y/b	z/c	U_{iso}
H(1)	7748(8)	7268(7)	7543(6)	0.11
H(2)	7502(8)	8021(7)	7000(6)	0.11
H(3)	8584(9)	5969(7)	7459(6)	0.11
H(4)	9814(9)	5704(7)	7998(6)	0.11
H(5)	12052(9)	6885(7)	8845(6)	0.11
H(6)	11042(9)	6404(7)	8850(6)	0.11
H(7)	12460(9)	7081(7)	8072(7)	0.11
H(8)	12224(9)	7839(7)	7534(7)	0.11
H(9)	10180(9)	9455(7)	7125(6)	0.11
H(10)	11410(9)	9165(7)	7644(6)	0.11
H(11)	8939(9)	8762(7)	6269(6)	0.11
H(12)	7936(9)	8250(7)	6253(6)	0.11

Table 1.2

Interatomic distances(Å) and angles(°) for hexa-host(I)

(a) Bonded distances



(i) Endocyclic bonds

C(1)-C(2)	1.411(16)	C(2)-C(3)	1.396(19)
C(3)-C(4)	1.402(22)	C(4)-C(5)	1.386(17)
C(5)-C(6)	1.382(20)	C(6)-C(1)	1.419(22)

(ii) Exocyclic bonds

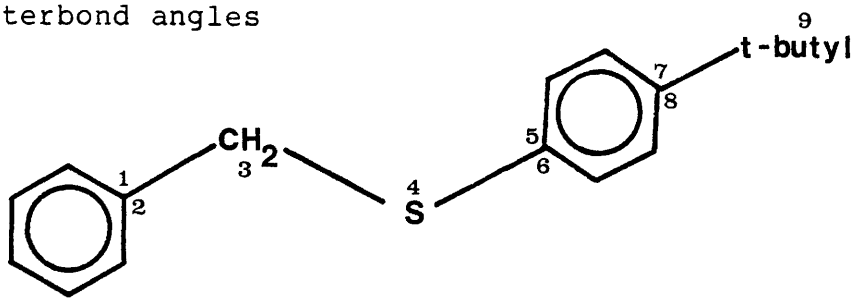
Bond number	Side-chain number 1	2	3	4	5	6
1	1.478(18)	1.518(21)	1.520(17)	1.473(19)	1.497(21)	1.504(17)
2	1.832(10)	1.821(14)	1.835(12)	1.838(11)	1.844(15)	1.830(12)
3	1.767(10)	1.766(10)	1.772(9)	1.779(10)	1.789(10)	1.798(9)
4	1.546(19)	1.594(19)	1.538(17)	1.550(21)	1.563(22)	1.506(17)
5*	1.518(41)	1.566(48)	1.564(43)	1.560(40)	1.563(45)	1.564(47)

* The value given is the average value of both contributors to the disordered t-butyl groups.

(iii) Rigid body refinement constrained all bond lengths within the terminal phenyl groups to 1.395(12)Å

Table 1.2 (continued)

(b) Interbond angles



(i) Endocyclic angles

C(6)-C(1)-C(2)	118.8(13)	C(1)-C(2)-C(3)	119.4(12)
C(2)-C(3)-C(4)	120.8(11)	C(3)-C(4)-C(5)	119.7(13)
C(4)-C(5)-C(6)	120.4(12)	C(5)-C(6)-C(1)	120.6(11)

(ii) Exocyclic angles

Angle number	Side-chain number 1	2	3	4	5	6
1	122.1(11)	120.3(13)	118.6(12)	119.9(11)	120.1(13)	121.3(12)
2	119.1(11)	120.1(11)	120.6(13)	120.4(12)	119.5(11)	118.1(13)
3	111.4(9)	112.3(8)	108.3(8)	110.3(9)	109.8(8)	108.5(8)
4	102.2(5)	97.5(5)	103.4(5)	102.7(6)	99.9(8)	101.0(5)
5	116.1(6)	120.6(8)	116.6(6)	117.8(6)	122.6(6)	120.4(7)
6	123.9(6)	119.4(6)	123.4(7)	122.2(8)	117.4(8)	119.6(9)
7	120.7(7)	118.5(8)	121.2(9)	119.9(9)	121.0(9)	119.0(10)
8	119.3(8)	121.5(10)	118.8(10)	120.1(10)	119.0(8)	121.0(11)
9*	110.4(21)	108.5(22)	110.7(18)	111.0(23)	110.5(20)	112.0(19)

* Average value over both t-butyl contributors.

(iii) Rigid body constraints fix all bond angles within the terminal phenyl groups to 120.0(10)°

Table 1.2 (continued)

(c) Least-squares Planes, given in the form:

$lX' + mY' + nZ' = d$, where X', Y', and Z' are coordinates in Å.

a) Plane equation (I)

$$-0.26605X' + 0.49670Y' + 0.82614Z' = 12.54031$$

b) Deviations of atoms (Å) from plane (I)

C(1) 0.009, C(2) -0.007, C(3) 0.011, C(4) -0.018,
C(5) 0.022, C(6) -0.018

c) Deviations of exocyclic atoms from plane (I)

CH_2	0.137	-0.136	0.088	-0.118	0.142	-0.089
S	-1.650	1.481	-1.623	1.553	-1.536	1.616
<Ph>	1.137	-1.194	0.341	-1.280	1.012	-0.798

Table 1.3

Interatomic distances(Å) and angles(°) for squalene guest

(a) Bonded distances

C(68)-C(69)	1.59(4)	C(69)-C(70)	1.88(3)
C(70)-C(71)	1.87(4)	C(71)-C(72)	1.57(2)
C(72)-C(73)	1.51(3)	C(74)-C(72)	1.30(3)
C(75)-C(74)	1.54(3)	C(75)-C(76)	1.52(2)
C(77)-C(76)	1.56(3)	C(77)-C(78)	1.47(3)
C(77)-C(79)	1.32(3)	C(79)-C(80)	1.59(3)
C(80)-C(81)	1.57(3)	C(81)-C(82)	1.69(5)
C(82)-C(83)	1.59(6)		

(b) Interbond angles

C(68)-C(69)-C(70)	91(2)	C(69)-C(70)-C(71)	130(2)
C(70)-C(71)-C(72)	107(1)	C(71)-C(72)-C(73)	114(2)
C(71)-C(72)-C(74)	119(1)	C(73)-C(72)-C(74)	126(2)
C(72)-C(74)-C(75)	123(2)	C(74)-C(75)-C(76)	108(2)
C(75)-C(76)-C(77)	112(1)	C(76)-C(77)-C(78)	116(2)
C(76)-C(77)-C(79)	115(2)	C(78)-C(77)-C(79)	128(2)
C(77)-C(79)-C(80)	120(2)	C(79)-C(80)-C(81)	108(1)
C(80)-C(81)-C(82)	111(2)	C(81)-C(82)-C(83)	113(2)

Table 2.1A

Atomic coordinates and thermal parameters for hexa-host(II)

(i) Fractional coordinates ($x10^4$)

ATOM	x/a	y/b	z/c
C(1)	8177(5)	8437	2894(5)
C(Ø1)	9087(6)	8310(8)	2834(6)
S(1)	9642(2)	9438(2)	2907(2)
O(1)	9527(5)	9981(5)	3600(5)
O(Ø1)	9402(4)	9853(6)	2095(4)
C(1Ø)	10755(6)	9072(9)	3184(7)
C(11)	11259(7)	10007(12)	3253(8)
C(12)	10896(7)	8386(6)	2556(6)
C(13)	10918	8701	1770
C(14)	11116	8053	1214
C(15)	11290	7091	1443
C(16)	11268	6777	2229
C(17)	11070	7424	2785
C(2)	8003(5)	8405	3670(5)
C(Ø2)	8744(6)	8380(7)	4456(6)
S(2)	9023(2)	7132(2)	4772(2)
O(2)	8338(5)	6700(6)	5013(6)
O(Ø2)	9304(5)	6658(6)	4137(5)
C(2Ø)	9931(7)	7236(1Ø)	5687(7)
C(21)	9620(9)	7703(15)	6406(7)
C(22)	10798(5)	8703(5)	5594(5)
C(23)	11517	9126	5448
C(24)	12106	8558	5198
C(25)	11975	7567	5094
C(26)	11256	7143	5240
C(27)	10667	7711	5490

Table 2.1A (continued)

ATOM	x/a	y/b	z/c
C(3)	7176(6)	8534	3732(5)
C(Ø3)	7ØØØ(7)	8456(7)	4581(6)
S(3)	7Ø52(2)	9626(2)	5Ø69(2)
O(3)	6483(5)	1Ø285(6)	4491(5)
O(Ø3)	7914(5)	99Ø6(6)	5389(5)
C(3Ø)	6577(7)	9381(9)	5899(7)
C(31)	6531(11)	1Ø388(12)	6342(9)
C(32)	7Ø72(5)	8617(6)	6497(5)
C(33)	7856	88Ø9	7Ø77
C(34)	8266	8Ø94	7634
C(35)	7891	7187	7611
C(36)	7Ø7Ø	6994	7Ø31
C(37)	6698	77Ø9	6474
C(4)	65ØØ(5)	8595	299Ø(6)
C(Ø4)	5593(6)	8788(7)	3Ø56(7)
S(4)	5Ø22(2)	7666(2)	3Ø92(2)
O(4)	4851(5)	7215(7)	23Ø3(6)
O(Ø4)	5468(5)	7Ø7Ø(6)	38Ø6(6)
C(4Ø)	4Ø28(7)	8Ø25(13)	3241(8)
C(41)	3494(7)	86Ø3(16)	2481(1Ø)
C(42)	4162(6)	8593(7)	4Ø7Ø(4)
C(43)	4249	8Ø57	4795
C(44)	4325	8527	5550
C(45)	4313	9532	5579
C(46)	4225	1Ø067	4857
C(47)	4149	9597	4Ø99

Table 2.1A (continued)

ATOM	x/a	y/b	z/c
C(5)	6641(5)	8602	2230(6)
C(05)	5920(7)	8654(8)	1432(6)
S(5)	5648(2)	9877(3)	1057(2)
O(5)	6409(5)	10328(7)	0988(5)
O(05)	5181(5)	10308(6)	1555(5)
C(50)	4986(8)	9666(11)	0006(8)
C(51)	4785(11)	10711(13)	-0375(8)
C(52)	4217(5)	9105(7)	0009(6)
C(53)	3566	9515	0287
C(54)	2847	8971	0287
C(55)	2779	8018	0010
C(56)	3430	7608	-0267
C(57)	4148	8152	-0268
C(6)	7494(5)	8558	2179(6)
C(06)	7686(6)	8628(8)	1338(5)
S(6)	7694(2)	7495(3)	0844(2)
O(6)	8395(6)	6943(6)	1326(5)
O(06)	6849(5)	7081(7)	0660(6)
C(60)	7906(7)	7788(12)	-0151(7)
C(61)	8802(8)	8185(15)	0048(8)
C(62)	7181(5)	8407(7)	-0688(6)
C(63)	6459	7944	-1192
C(64)	5799	8485	-1712
C(65)	5862	9487	-1728
C(66)	6584	9949	-1224
C(67)	7243	9409	-0705

Table 2.1A (continued)

(ii) Thermal parameters ($Å^2 \times 10^3$)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
C(1)	30	30	41	01	08	-03
C(01)	37	57	45	03	17	07
S(1)	38	52	56	03	15	-05
O(1)	52	56	76	-18	22	-12
O(01)	50	68	57	25	12	-05
C(10)	26	70	66	-02	07	01
C(11)	44	109	89	-10	16	-23
C(12)	62					
C(13)	85					
C(14)	115					
C(15)	122					
C(16)	138					
C(17)	92					
C(2)	39	28	37	03	15	01
C(02)	34	51	47	11	05	-01
S(2)	42	52	52	16	00	04
O(2)	54	64	101	23	27	-04
O(02)	61	57	60	01	07	06
C(20)	61	76	52	12	-09	16
C(21)	89	185	26	04	25	13
C(22)	71					
C(23)	73					
C(24)	97					
C(25)	87					
C(26)	72					
C(27)	55					

Table 2.1A (continued)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
C(3)	38	40	35	-05	08	-03
C(03)	62	41	43	03	26	-05
S(3)	59	58	49	-06	20	-06
O(3)	95	54	55	03	23	04
O(03)	65	84	72	-21	25	-32
C(30)	70	72	54	00	28	17
C(31)	137	94	80	-19	55	92
C(32)	59					
C(33)	75					
C(34)	98					
C(35)	98					
C(36)	109					
C(37)	83					
C(4)	33	38	45	07	07	03
C(04)	35	40	67	-04	11	-06
S(4)	38	62	71	-09	13	-11
O(4)	73	72	94	-34	19	-19
O(04)	56	62	106	11	24	-03
C(40)	30	135	80	01	07	01
C(41)	33	188	110	-11	-04	19
C(42)	68					
C(43)	85					
C(44)	105					
C(45)	104					
C(46)	103					
C(47)	88					

Table 2.1A (continued)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
C(5)	22	49	42	-Ø1	-Ø3	Ø2
C(Ø5)	63	53	45	ØØ	-Ø1	14
S(5)	57	71	47	Ø5	-Ø9	15
O(5)	72	74	73	2Ø	-Ø2	-Ø8
O(Ø5)	89	71	59	-18	Ø7	18
C(5Ø)	68	86	51	11	-Ø6	24
C(51)	115	112	59	21	-Ø7	35
C(52)	64					
C(53)	1Ø2					
C(54)	148					
C(55)	142					
C(56)	12Ø					
C(57)	1ØØ					
C(6)	26	4Ø	43	ØØ	Ø5	Ø1
C(Ø6)	41	68	19	-Ø7	-Ø5	ØØ
S(6)	63	75	56	-25	19	-11
O(6)	99	6Ø	81	-Ø7	27	11
O(Ø6)	8Ø	97	99	-28	46	-27
C(6Ø)	58	13Ø	53	-29	24	-16
C(61)	65	19Ø	54	-25	2Ø	-35
C(62)	8Ø					
C(63)	1Ø4					
C(64)	121					
C(65)	139					
C(66)	131					
C(67)	92					
Average e.s.d.'s						
S	1	2	2	1	1	1
O,C	5	6	5	5	4	4

Table 2.1B

Atomic coordinates and thermal parameters for dimer guests.

(i) Fractional coordinates ($x10^4$)

ATOM	x/a	y/b	z/c
O(100)	3397(7)	7558	-2746(8)
O(101)	2173(8)	7655(8)	-2495(7)
C(100)	2722(13)	7166(11)	-2692(10)
C(101)	2560(12)	6105(11)	-2891(11)
O(200)	3565(7)	9434(8)	-2681(7)
O(201)	2346(9)	9567(10)	-2427(9)
C(200)	2970(10)	9939(13)	-2595(8)
C(201)	2993(14)	10973(13)	-2674(13)
O(300)	10856(10)	5740(13)	9774(9)
O(301)	9617(16)	6197(12)	9097(10)
C(300)	10449(18)	6291(19)	9190(14)
C(301)	10803(17)	6858(14)	8656(13)
O(400)	9807(21)	9905(15)	9092(13)
O(401)	11041(14)	10395(22)	9873(17)
C(400)	10572(23)	9985(23)	9168(42)
C(401)	10903(28)	9672(20)	8447(20)
H(1)	3297(182)	8300(261)	-2171(200)
H(2)	2327(62)	8903(90)	-2404(61)

Table 2.1B (continued)

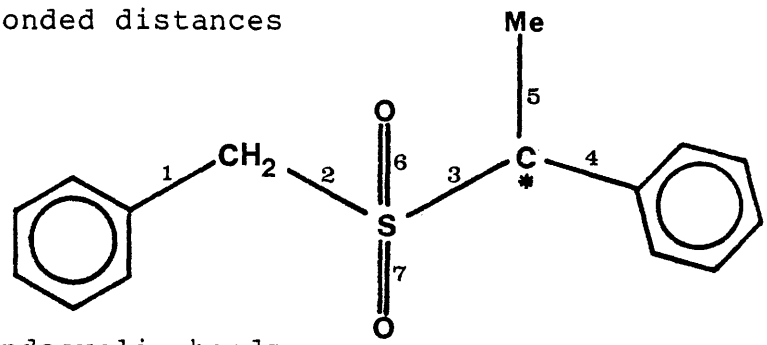
(ii) Thermal parameters ($Å^2 \times 10^3$)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
O(100)	95	83	116	04	27	12
O(101)	109	83	99	04	31	-18
C(100)	101	57	91	08	19	13
C(101)	126	58	104	19	-26	-20
O(200)	79	79	146	12	44	-01
O(201)	125	77	171	-18	87	07
C(200)	68	103	64	-10	03	-02
C(201)	155	59	126	11	00	-07
O(300)	141	167	94	-04	-08	09
O(301)	227	146	113	03	-20	54
C(300)	146	137	90	-53	-42	20
C(301)	232	103	123	57	00	-56
O(400)	328	155	166	-41	101	-59
O(401)	131	245	270	73	91	-21
C(400)	118	101	588	29	168	-05
C(401)	360	103	242	-09	208	03
H(1)	245					
H(2)	72					
Average e.s.d.'s						
O,C	20	15	15	12	18	15

Table 2.2

Interatomic distances(Å) and angles(°) for hexa-host(II)

(a) Bonded distances



(i) Endocyclic bonds

C(1)-C(2)	1.401(12)	C(2)-C(3)	1.397(13)
C(3)-C(4)	1.406(13)	C(4)-C(5)	1.359(13)
C(5)-C(6)	1.416(12)	C(6)-C(1)	1.402(12)

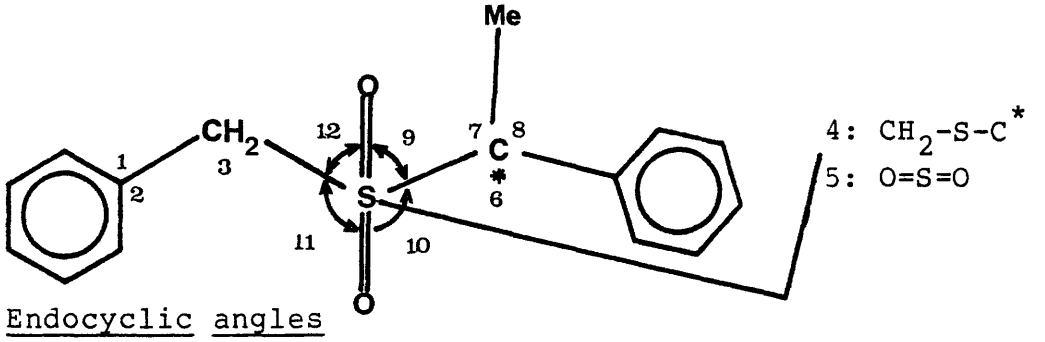
(ii) Exocyclic bonds

Bond number	Side-chain number 1	2	3	4	5	6
1	1.527(12)	1.522(13)	1.536(13)	1.548(13)	1.509(14)	1.526(13)
2	1.800(11)	1.816(11)	1.814(11)	1.819(11)	1.828(12)	1.771(11)
3	1.815(10)	1.814(12)	1.802(11)	1.786(12)	1.810(13)	1.836(12)
4	1.534(19)	1.584(19)	1.594(21)	1.552(23)	1.581(24)	1.506(20)
5	1.475(15)	1.483(15)	1.535(15)	1.547(16)	1.477(16)	1.522(16)
6	1.431(8)	1.423(9)	1.453(9)	1.412(10)	1.430(9)	1.430(10)
7	.427(8)	1.439(8)	1.411(9)	1.436(10)	1.408(9)	1.447(10)

(iii) Rigid body refinement constrained all bond lengths within the terminal phenyl groups to 1.395(11)Å.

Table 2.2 (continued)

(b) Interbond angles



(i) Endocyclic angles

C(6)-C(1)-C(2)	118.1(9)	C(1)-C(2)-C(3)	121.4(9)
C(2)-C(3)-C(4)	118.1(9)	C(3)-C(4)-C(5)	122.3(10)
C(4)-C(5)-C(6)	118.8(9)	C(5)-C(6)-C(1)	121.0(9)

(ii) Exocyclic angles

Angle number	Side-chain number 1	2	3	4	5	6
1	120.7(9)	118.6(9)	120.0(9)	118.5(9)	118.8(9)	120.7(9)
2	121.2(9)	119.3(9)	121.3(10)	118.8(9)	122.4(10)	118.3(9)
3	111.2(8)	109.8(6)	111.0(7)	111.6(6)	113.6(8)	114.0(7)
4	102.6(6)	104.2(7)	101.7(6)	105.2(7)	101.7(7)	105.0(7)
5	119.6(5)	117.4(6)	117.6(6)	117.0(6)	119.5(6)	117.9(6)
6	110.2(9)	112.1(9)	112.9(10)	111.4(9)	110.7(9)	109.8(9)
7	105.3(10)	108.4(10)	106.5(9)	110.2(10)	103.7(10)	107.8(9)
8	114.2(11)	116.5(13)	112.9(10)	112.2(14)	113.8(12)	117.6(14)
9	106.9(6)	108.4(6)	107.3(6)	107.9(7)	106.6(6)	107.6(6)
10	109.8(6)	104.2(7)	101.7(6)	105.2(7)	101.7(7)	105.0(7)
11	106.8(6)	109.1(5)	109.2(6)	110.4(6)	108.1(6)	108.6(6)
12	109.9(6)	109.2(5)	109.0(6)	108.4(6)	107.6(7)	108.8(5)

(iii) Rigid body constraints fix all bond angles within the terminal phenyl groups at 120.0(10)°.

Table 2.2 (continued)

(c) Least-squares planes, given in the form:

$lX' + mY' + nZ' = d$, where X', Y' and Z' are coordinates in Å.

(i) Plane equation (I)

$$0.08317X' + 0.99480Y' + 0.05883Z' = 12.91623$$

(ii) Deviation of atoms (Å) from plane (I)

C(1) -0.006, C(2) -0.032, C(3) 0.037, C(4) -0.011
C(5) -0.023, C(6) 0.029.

(iii) Deviations of exocyclic atoms from plane (I)

CH_2	-0.076	0.085	-0.046	0.130	-0.094	0.101
S	1.573	-1.587	1.602	-1.487	1.538	-1.480
C*	1.232	-1.271	1.247	-1.123	1.098	-1.107
<Ph>	-0.627	0.145	-0.612	0.425	-0.625	0.242

Table 2.3

Bond lengths (Å) and angles (°) for acetic acid guests

DIMER A

(i) Bond lengths

O(1Ø0)-C(1Ø0)	1.25(3)	O(2Ø0)-C(2Ø0)	1.24(2)
O(1Ø1)-C(1Ø0)	1.24(3)	O(2Ø1)-C(2Ø0)	1.24(2)
C(1Ø1)-C(1Ø0)	1.53(2)	C(2Ø1)-C(2Ø0)	1.45(3)
O(1Ø0)-H(1)	1.46(24)	O(2Ø0)-H(2)	1.41(24)
Ø(1Ø1)-H(1)	1.75(21)	Ø(2Ø1)-H(2)	Ø.93(21)
O(1Ø0)-O(2Ø0)	2.61(1)	O(1Ø1)-O(2Ø1)	2.67(2)

(ii) Bond angles

O(1Ø0)-C(1Ø0)-O(1Ø1)	12Ø(1)	O(2Ø0)-C(2Ø0)-O(2Ø1)
O(1Ø0)-C(1Ø0)-C(1Ø1)	12Ø(2)	O(2Ø0)-C(2Ø0)-C(2Ø1)
O(1Ø1)-C(1Ø0)-C(1Ø1)	12Ø(2)	O(2Ø1)-C(2Ø0)-C(2Ø1)
C(1Ø0)-O(1Ø0)-H(1)		C(2Ø0)-O(2Ø1)-H(2)
O(1Ø0)-H(1) -O(2Ø0)		Ø(1Ø1)-H(2) -O(2Ø1)

(iii) Least-squares planes, given in the form:

lX' + mY' + nZ' = d, where X', Y' and Z' are coordinates in Å.

(a) Plane (Ia) : Ø.16782X' - Ø.22ØØ6Y' + Ø.96Ø94Z' = -5.39122
Plane (IIa): Ø.19569X' + Ø.Ø8145Y' + Ø.97728Z' = -1.75403

(b) Deviations of atoms (Å) from planes:

Plane (Ia) : O(1Ø0) -Ø.ØØ2, O(1Ø1) -Ø.ØØ2, C(1Ø0) Ø.Ø14,
C(1Ø1) -Ø.ØØ3
Plane (IIa): O(2Ø0) Ø.ØØ1, O(2Ø1) Ø.ØØ1, C(2Ø0) -Ø.ØØ5,
C(2Ø1) Ø.ØØ2

Table 2.3 (continued)

(c) Dihedral angles between:

Host(II) (Plane (I) Table 2.2c) and Plane (Ia)	98.54
Host(II) (Plane (I) Table 2.2c) and Plane (IIa)	81.09
Plane (Ia) and (IIa)	17.44

DIMER B

(i) Bond lengths

O(300)-C(300)	1.28(3)	O(400)-C(400)	1.26(5)
O(301)-C(300)	1.32(4)	O(401)-C(400)	1.32(6)
C(301)-C(300)	1.42(3)	C(401)-C(400)	1.53(8)
O(300)-O(400)	2.69(3)	O(301)-O(401)	2.53(4)

(ii) Bond angles

O(300)-C(300)-O(301)	108(2)	O(400)-C(400)-O(401)	116(5)
O(300)-C(300)-C(301)	127(3)	O(400)-C(400)-C(401)	116(5)
O(301)-C(300)-C(301)	124(2)	O(401)-C(400)-C(401)	127(4)

(iii) Least-squares planes, given in the form:

$lX' + mY' + nZ' = d$, where X', Y' and Z' are coordinates in Å.

(a) Plane (Ib) : $-0.12368X' + 0.76042Y' + 0.63752Z' = 14.44083$

Plane (IIb): $0.06265X' + 0.91069Y' + 0.40832Z' = 14.05032$

Table 2.3 (continued)

(b) Deviations of atoms (Å) from planes:

Plane (Ib) : O(300) -0.010, O(301) -0.009, C(300) 0.029,
C(301) -0.010.
Plane (IIb): O(400) 0.006, O(401) 0.007, C(400) -0.019,
C(401) 0.006.

(c) Dihedral angles between:

Host(II) and Plane (Ib)	38.4
Host(II) and Plane (IIb)	20.7
Plane (Ib) and Plane (IIb)	19.0

Flow diagram of MULTAN 78

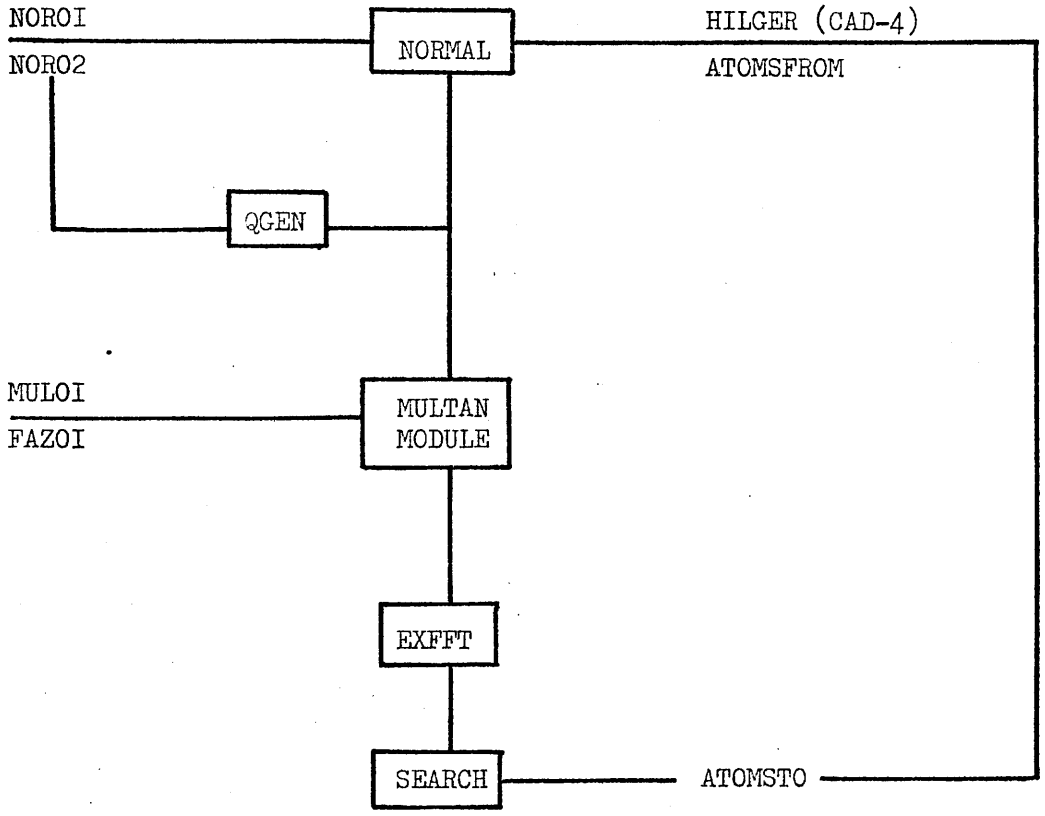


Figure I

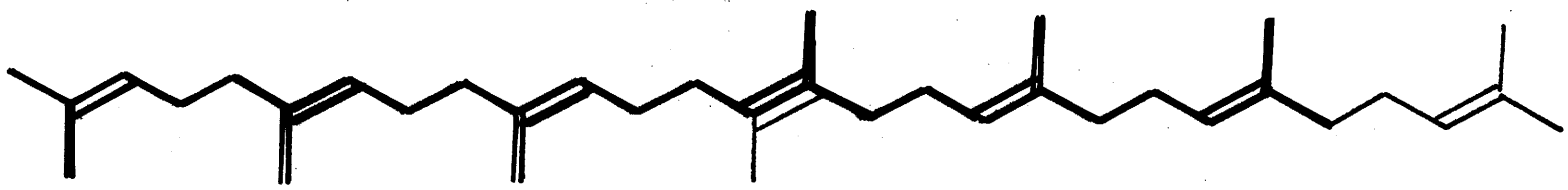


Figure 3

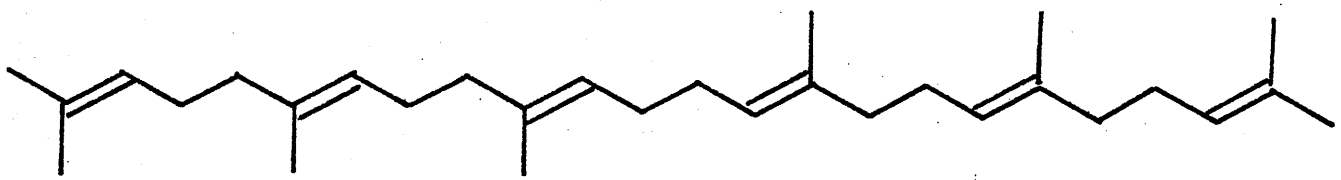


Figure 3

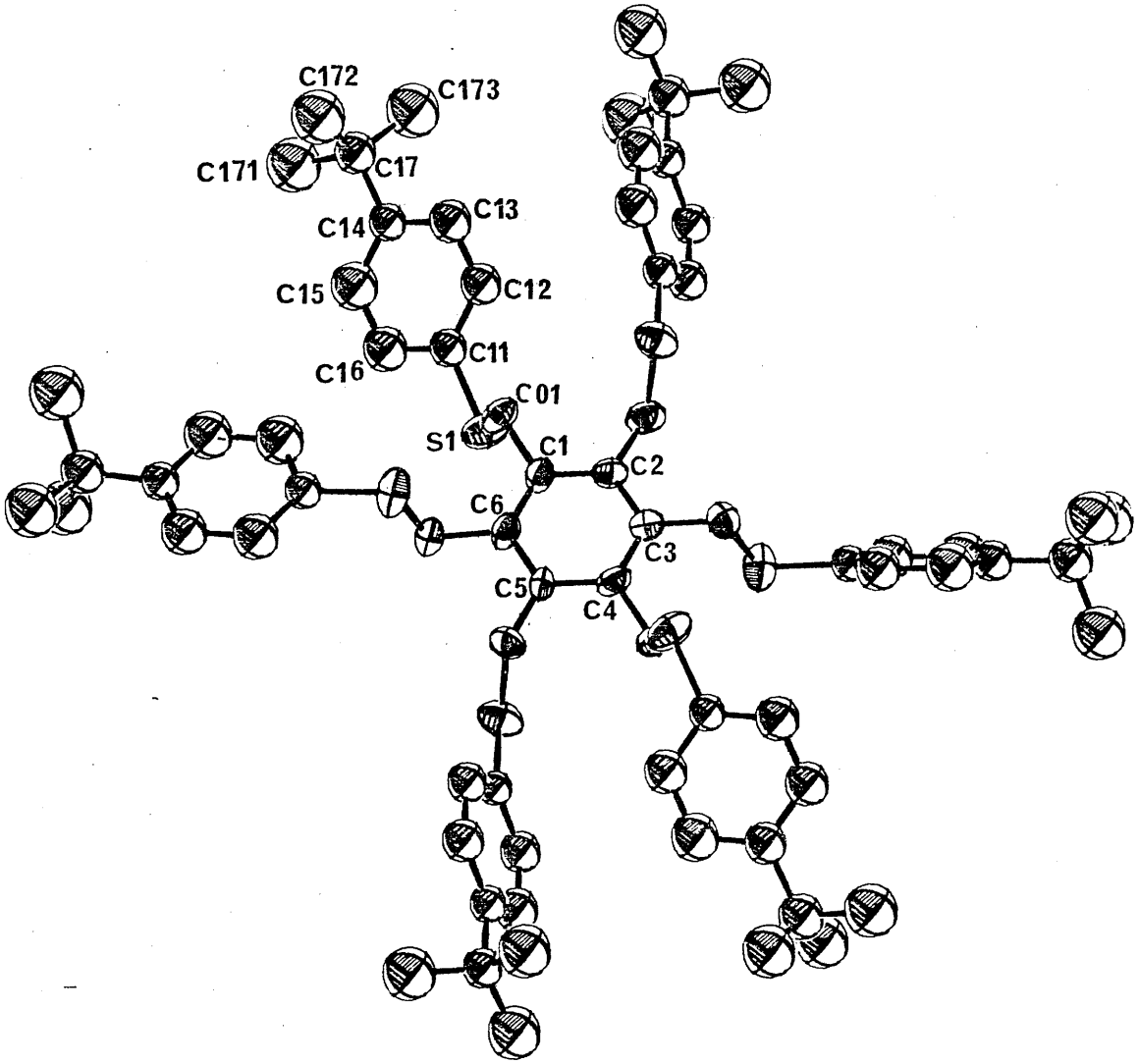


Figure 4

Achiral host

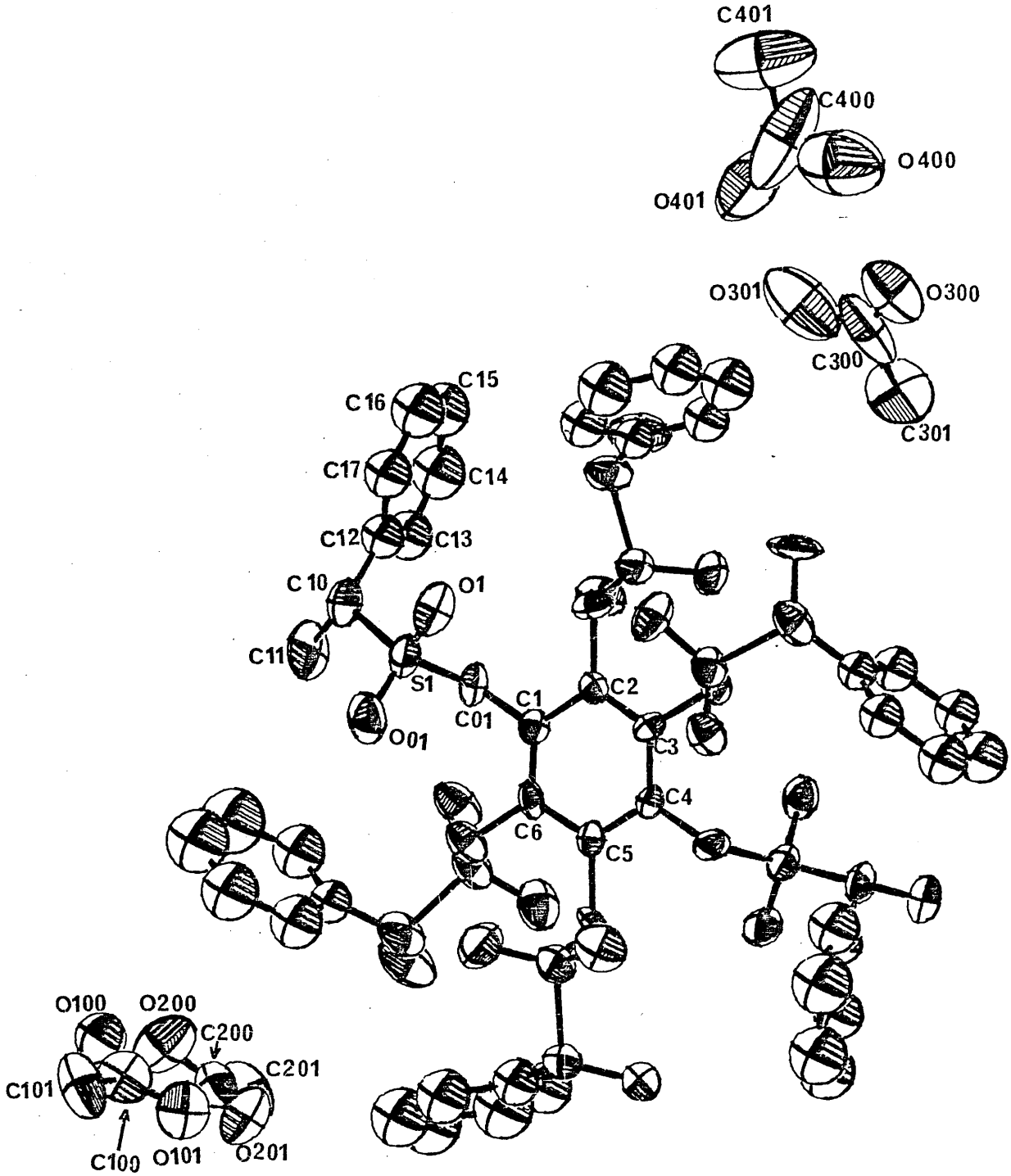
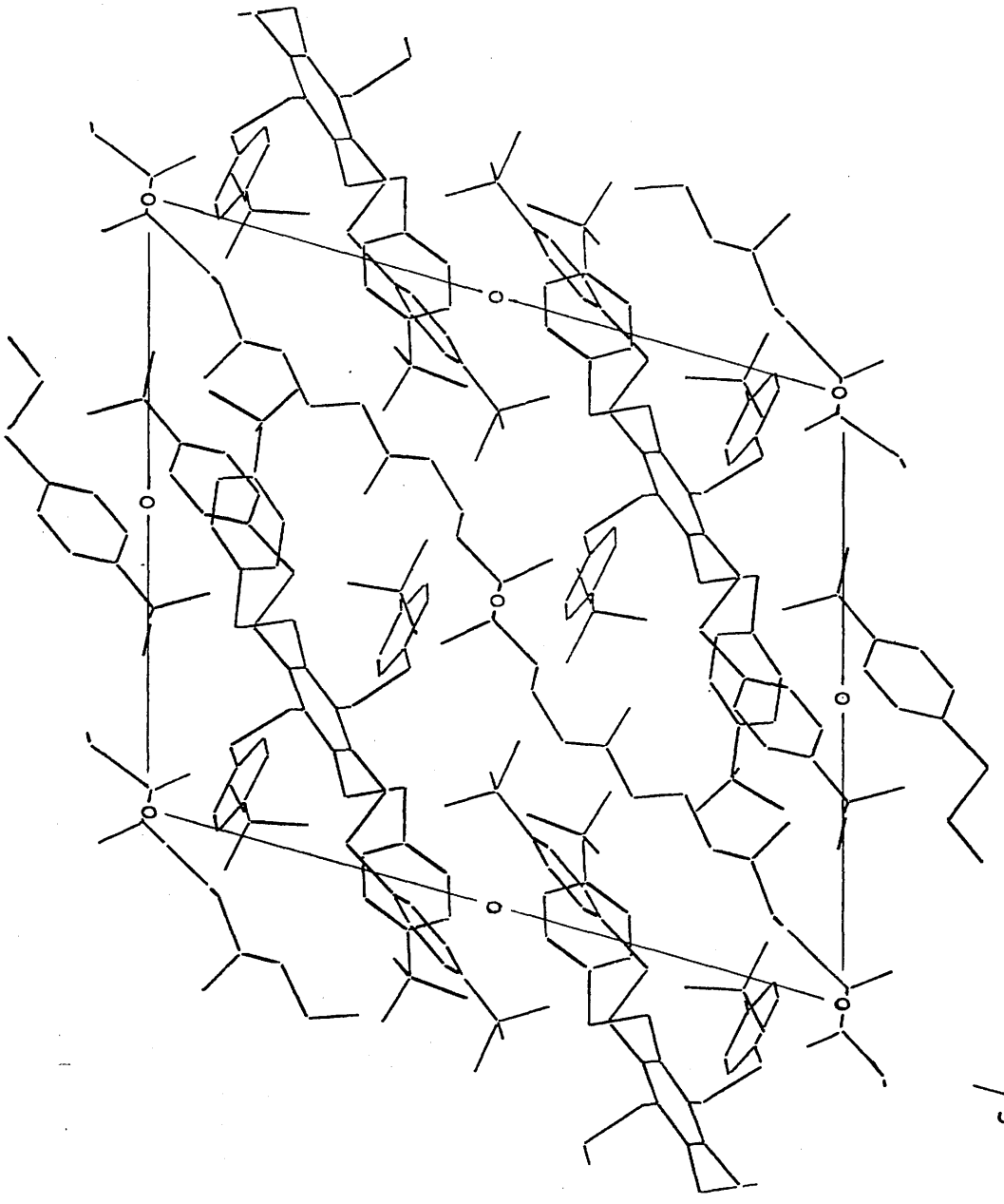


Figure 5

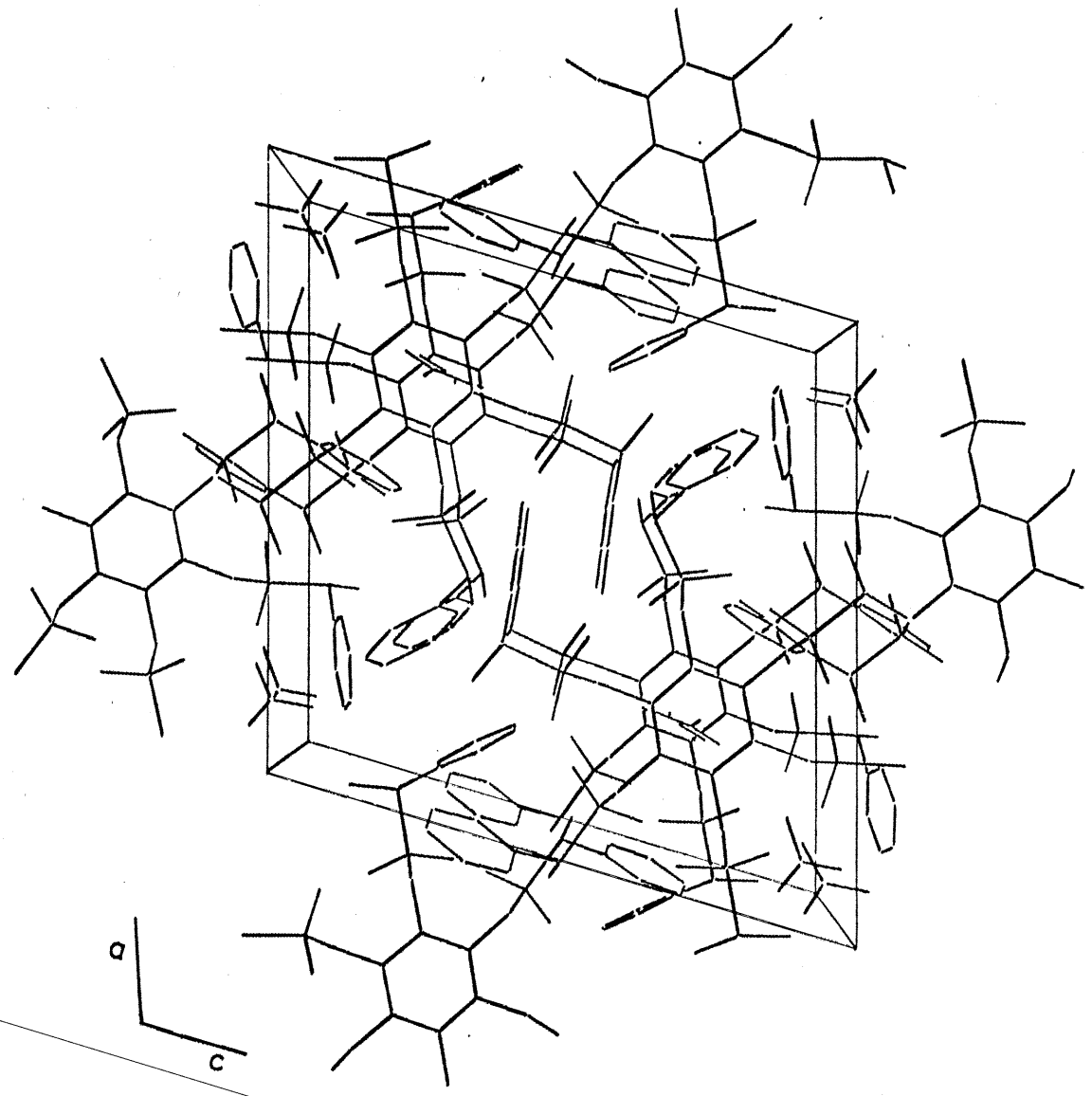
Chiral host and dimer guests



Crystal packing arrangement for I

Figure 6

Figure 7



Crystal packing arrangement for II

Torsion angles for host I

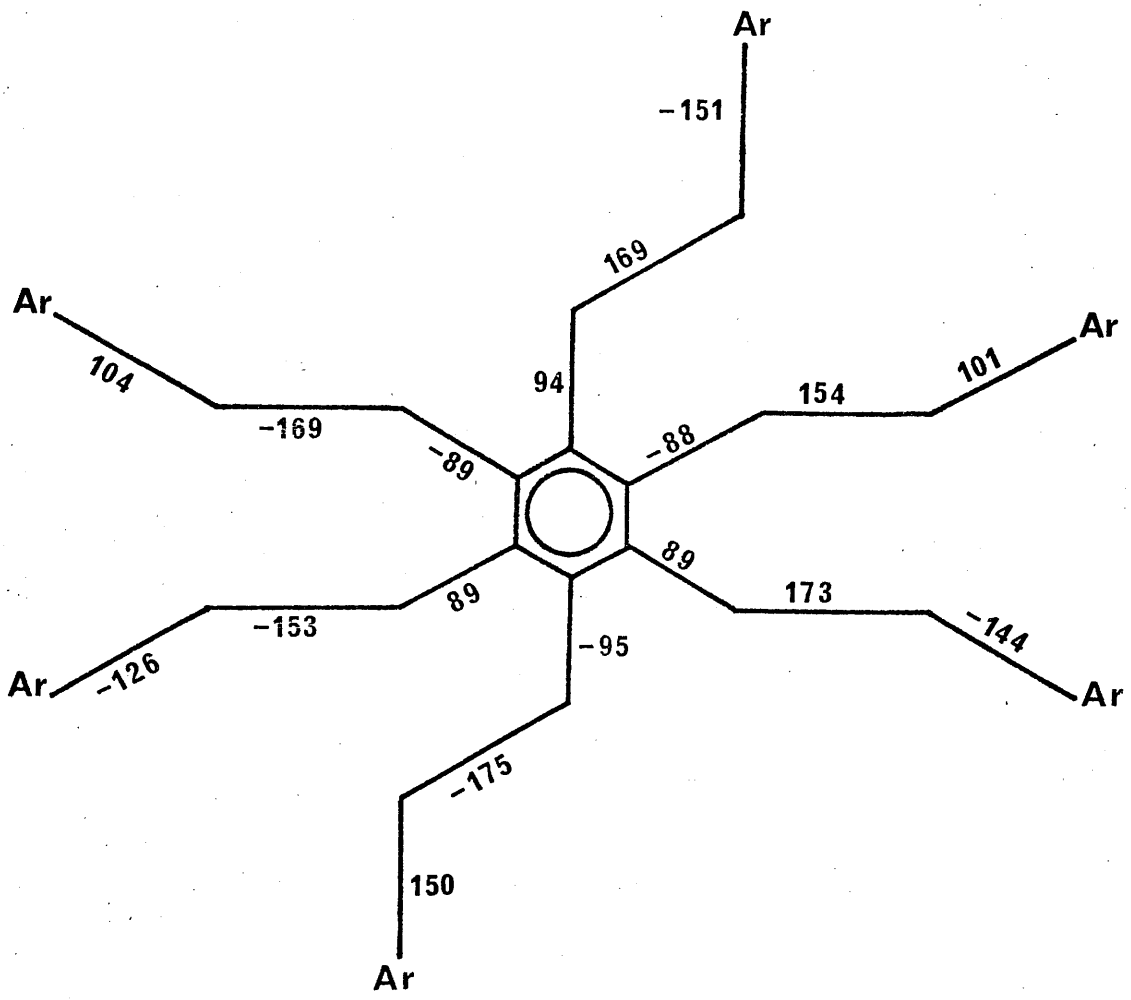
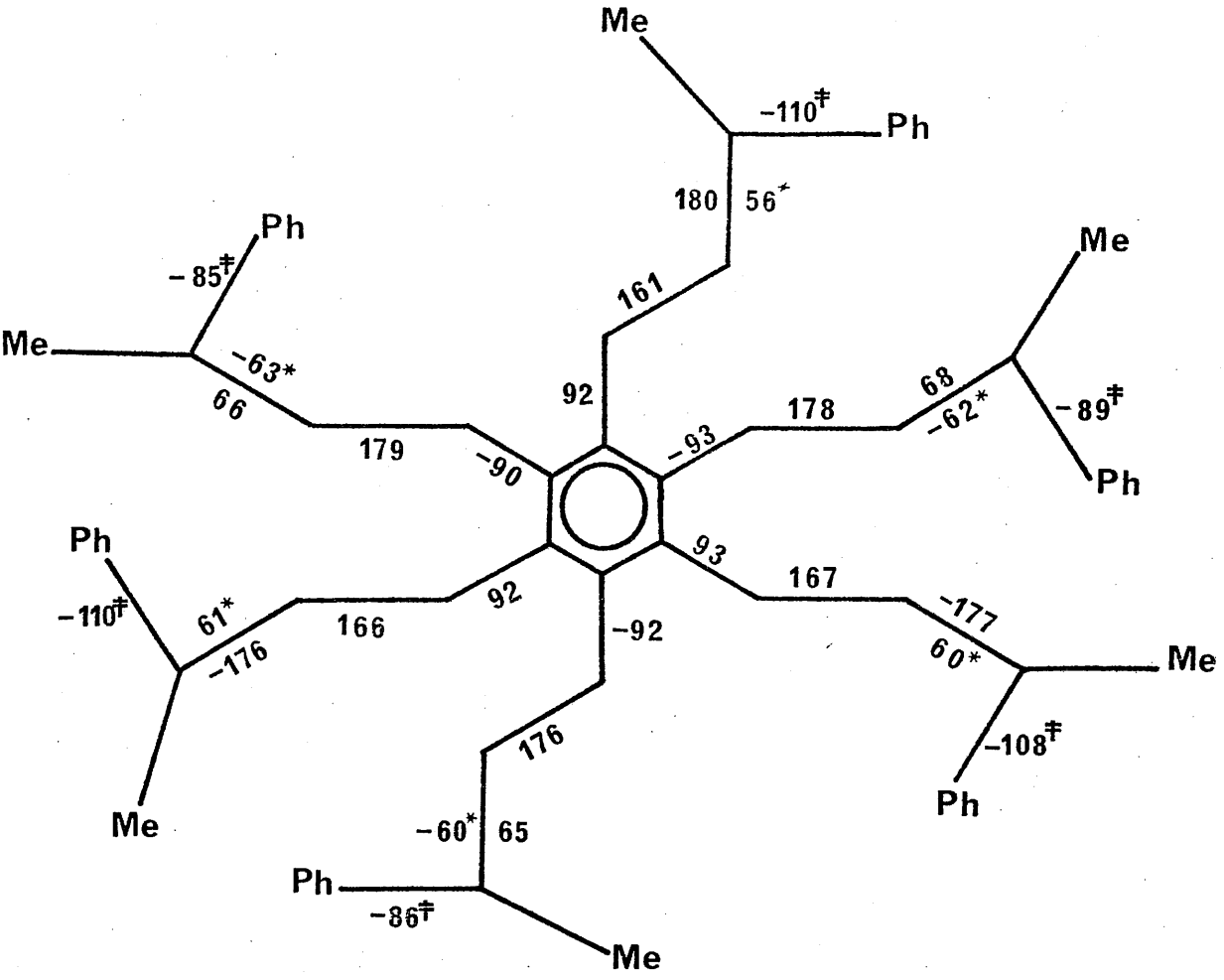


Figure 8

Torsion angles for host II



* Refers to the torsion angle CH_2-S-C^*-Ph

‡ " " " " " $S-C^*-C_{12}-C_{13}$ etc.

Figure 9

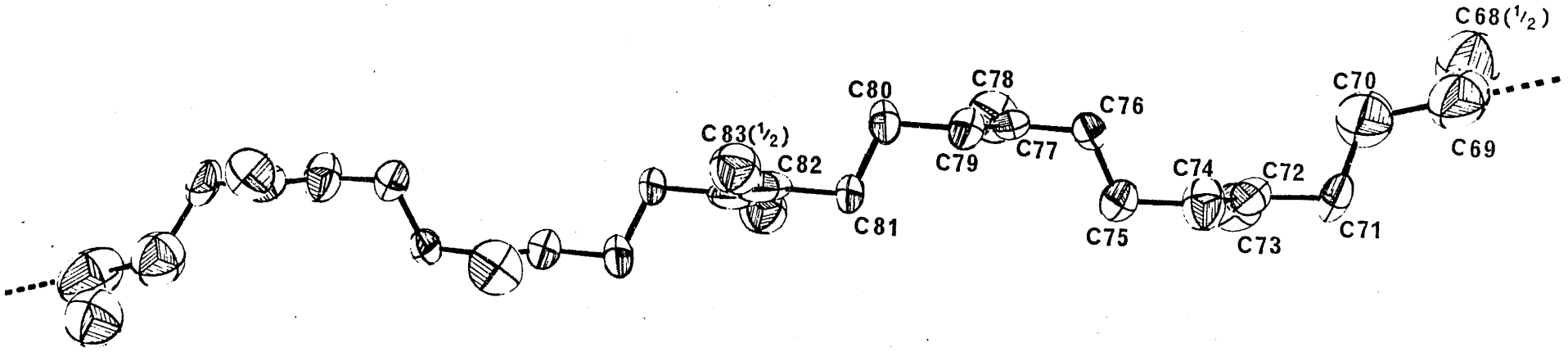


Figure 10

Squalene guest

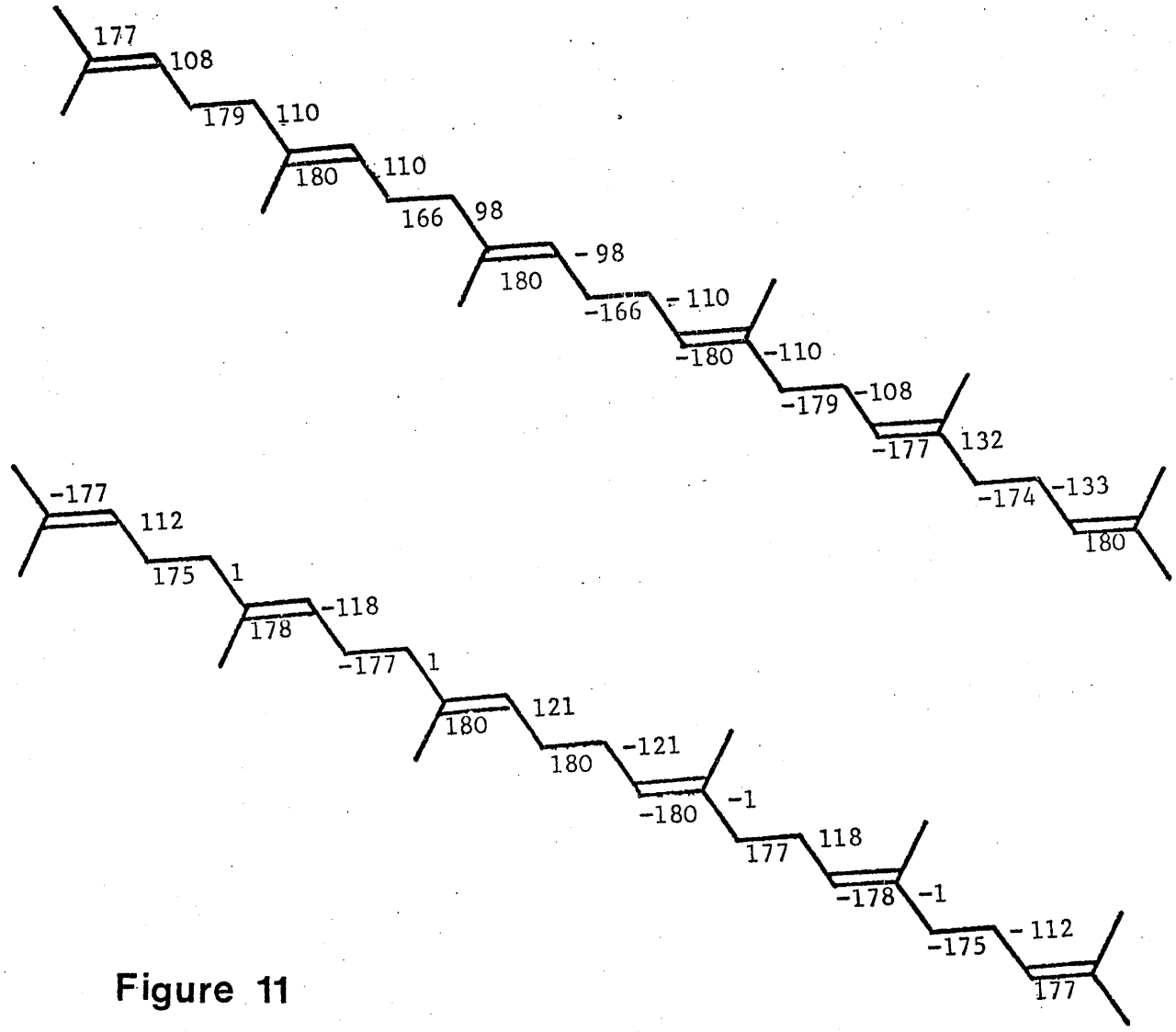

177
108
179
110
180
110
166
98
180
-98
-166
-110
-180
-110
-179
-108
-177
132
-174
-133
180
-177
112
175
1
178
-118
-177
1
180
121
180
-121
-180
-1
177
118
-178
-1
-175
-112
177

Figure 11

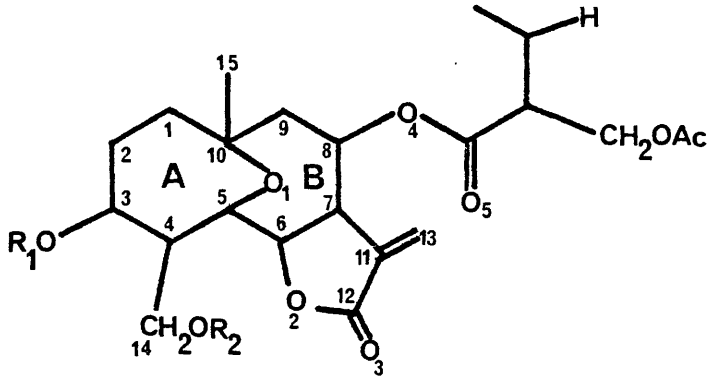
Chapter 5

The Use of Magic Integers and Random Phase Sets:

Structure Determination of Iso- and Acetylchapliatrin.

5.1 INTRODUCTION

The isolation of three closely related germacranolides[1], which have been named chapliatrin (Ia), isochapliatrin (Ib) and acetylchapliatrin(Ic)[2], afforded two crystalline samples, viz. (Ib) and (Ic).



(Ia) R_1=H, R_2=Ac
(Ib) R_1=Ac, R_2=H
(Ic) R_1, R_2=Ac

All three compounds possess the hitherto unreported 5,1Ø-oxygen linkage. Although deduction of the stereochemistry[1] at C(5), C(6), C(7) and C(8) by NMR spectroscopy was relatively straightforward, the stereochemistry assigned to C(3), C(4) and (1Ø) was tentative and therefore called for further examination via X-ray diffraction techniques. However, structure elucidation in both cases was far from routine and many attempts with various direct methods techniques were required before the structures were finally solved.

5.2 THE USE OF MAGIC INTEGER-Ψ MAP AND RANDOM PHASE SETS-LINEAR EQUATIONS IN MULTAN

The previous chapter has already illustrated the usefulness of quartet and quintet invariants as an integral part of the MULTAN package. However, it was found necessary to expand the package even further and to incorporate magic integer phase representation, MAGIC, (see Section

8, Chapter 1) and random phase sets-linear equations, RANDOM, (see Section 9, Chapter 1) as fully integrated options within the mainframe program. This necessity was brought about by the difficulties encountered in the solution of iso- and acetylchapliatrin. These difficulties are discussed later. An outline of these procedures now follows.

Both MAGIC and RANDOM are run between CONVERGE and FASTAN, and provide a large starting set of reflexions for phase expansion at very little extra cost.

5.2.1 MAGIC

MAGIC consists of three parts:

(i) A primary set of reflexions (P) is chosen containing the origin and enantiomorph defining reflexions plus symbolic phases expressed in terms of magic integers. The magic integer sequence is selected by the user. A single symbol is used to represent three phases. A secondary set (S) of reflexions is derived using triplets which involve a pair of P reflexions.

(ii) Relationships which link the P and S sets give rise to terms in a Fourier map (Ψ-map). The peaks derived from this Fourier are then chosen to represent likely values of the symbols and hence each peak represents a likely set of phases for all the P and S reflexions under consideration. The number of peaks thus determined, and translated into crude phases, are selected by the user as an input parameter.

(iii) Finally, all the relationships linking the P and S set reflexions are used to refine the phases in the P and S sets employing the parameter shift procedure involving all available triplets.

5.2.2 RANDOM

After a convergence map has been run, the bottom (<150) reflexions are selected - usually about 100. These reflexions are given random phases from a random number generator and are then refined to convergence using linear equations. A second set of random phases is then selected and the procedure repeated. In this way a specified number of phase sets is produced, among which, hopefully, there is a correct set. The sets usually consist of about 100 trials but the number is under the control of the user. During linear equation refinement pseudo-weighting and enantiomorph discriminating weights are employed as discussed in Chapter 1

When either MAGIC or RANDOM is used, the phase sets obtained are subsequently input to the FASTAN link of MULTAN for phase expansion and refinement. The phases derived from the above procedures are fixed until the final cycle of tangent refinement to prevent any drifting away from their true values.

When using the MAGIC procedure it is not necessary to specify a reflexion for enantiomorph definition. It will be fixed automatically by the Ψ-map peak searching routine, which limits its search to just one half of the Ψ-map.

In both MAGIC and RANDOM, if quartets and/or quintets are available, they will be used to filter out any obviously incorrect solutions before tangent refinement. Since the phasing process is independent of the quartet and quintet figures of merit, this filtering is very efficient. It is only employed if the following requirements are met:

(a) There are at least 25 reliable negative invariants available.

(b) Quartets/quintets were loaded into the system at the Σ_2 stage of the calculation.

A flow diagram of MULTAN in which the options, MAGIC and RANDOM may be used, along with quartet and quintet invariants, is shown in Figure I. Neither MAGIC nor RANDOM will work for centrosymmetric structures.

5.3 EXPERIMENTAL

Isochapliatrin(Ib)

Crystal data

Isochapliatrin, $C_{24}H_{32}O_{10}$, M_r=480.5, orthorhombic, a=7.449(2), b=19.898(1), c=16.367(2) Å, U=2425.9 Å^3, D_m=1.32, D_c=1.31 Mgm^{-3}, Z=4, F(000)=1024, space group $P2_12_12_1$, μ(Mo-K_α)=1.10 cm^{-1}.

Data collection

Instrument used: Enraf-Nonius CAD-4
Radiation used: Mo-K_α , λ=0.71069 Å
Upper limit for data collection: $2\theta_{max}=56^\circ$
Number of independent reflexions: m=1967
Unobserved cut-off*: $2.5\sigma_I$
Number of parameters refined: n=327
Number of reflexions per parameter: m/n=6.0

* The $2.5\sigma_I$ cut-off was only applied after structure solution in order to decrease the time for least-squares analyses.

Acetylchapliatrin (Ic)

Crystal data

Acetylchapliatrin, $C_{25}H_{32}O_{11}$, M_r=508.5, monoclinic, a=34.836(3), b=13.648(2), c=16.970(2) Å, β=138.32(8)°, U=5367.31 Å^3, D_m=1.26, D_c=1.26 Mgm^{-3}, Z=8, F(000)=2224, space goup C2, μ(Mo-K_α)=1.06cm^{-1}.

Data collection

Instrument used: Hilger Watts Y290
Radiation used: Mo-K_α , λ=0.71069 Å
Upper limit of data collection: $2\theta_{max}$=56°
Number of independent reflexions: m=2681
Unobserved cut-off*: $3.0\sigma_I$
Number of parameters refined: n=334
Number of reflexions per parameter: m/n=8.0

* As with isochapliatrin, the $2.5\sigma_I$ cut-off was only applied for least-squares refinement

5.4 STRUCTURE SOLUTION

(i) Isochapliatrin (Ib)

Initial Wiessenberg photographs defined the space group as $P2_12_12$, distinguished from $P2_12_12_1$ since all reflexions were present along the 001 axis. Exhaustive MULTAN runs, with and without quartet invariants, failed to produce a solution. Several MULTAN runs incorporating MAGIC with several different magic integer sequences were also unsuccessful. 800 random phase sets were computed employing negative quartet and quintet figures of merit, all with the same lack of success.

The structure was eventually solved by an intuitive change in the space group to $P2_12_12_1$, whereby application of

standard MULTAN techniques produced an E-map based on 250 E-magnitudes having $|E| \geq 1.61$ which revealed the positions of 21 out of a possible 33 atoms. Sim-weighted Fourier syntheses revealed all but the terminal atoms of the longest side-chain.

The forbidden reflexions (00l, where l=3 to 21 inclusive) along the c axis can only be regarded as genuine reflexions and not artefacts, since they do not disappear when X-rays of different wavelengths are used; thus one of the prerequisites for the Renninger effect is transgressed. A possible explanation can be derived by considering that the crystal symmetry is really monoclinic (say, $P2_1$ with two molecules in the asymmetric unit) whereby, as a result of molecular packing the crystal manifests almost exact orthorhombic symmetry. This problem, however, remains unresolved.

(ii) Acetylchapliatrin (Ic)

Both Weissenberg and precession photographs showed the space group to be triclinic, P1, with four molecules in the asymmetric unit; however, after several cycles of Delauney reduction, varying the parameters of the program, a reduced cell containing two molecules in the asymmetric unit of the monoclinic space group C2 resulted. Structure elucidation via MULTAN with and without quartet invariants was unsuccessful. This situation was due to two predominant failings within the phase determining process.

In the symmorphic space group C2, only two reflexions are required to define the origin - usually an hkl reflexion where k=2n, and another when k=2n+1. Examination of the convergence map revealed the existence of phase islands. In this case the first 120 reflexions at the bottom of the convergence map involved only those relationships between reflexions with k even - this constituted the first, and most serious phase island. After these reflexions, when the second origin reflexion (k=2n+1) was introduced, relationships involving

reflexions with k=2n+1 were predominant.

The inclusion of quartet information, although decreasing the size of these islands, failed to produce a solution - although it should be noted that several fragments were found, but these failed to generate the rest of the structure. It should be noted that 700 random phase sets were also computed without success.

In all cases the figures of merit, in particular Ψ_o, derived from these phase sets gave no clear indication as to the correct solution.

Associated with the problem of phase islands, is the loss of enantiomorph definition, which was possibly a source of difficulty in this structure. The MAGIC procedure can be very useful in such cases by employing a larger starting set of reflexions in which enanatiomorph definition is resolved by computing only half the Ψ-map. MAGIC was therefore run in the manner described previously.

Fifteen primary reflexions were represented by the elements of {8 11 13 14 15} (xyz) and resulted in 22 secondary reflexions. The 37 reflexions of the combined primary and secondary sets were linked by 14 triplets and a Ψ-map was calculated for these. The top 50 maxima in the Ψ-map were refined by a parameter shift technique in which the shifts were restricted to be < 15 degrees. This was necessary to prevent large parameter shifts. These solutions were then refined using the tangent formula. Quartets and quintets, derived for the top 100 E-magnitudes ($|E| \geqslant 2.3$) were used to provide additional figures of merit. The solution with the highest CFOM (3.22 with a theoretical maximum of 5.00) revealed the positions of 32 atoms in the asymmetric unit. The structure was completed with difficulty using Fourier techniques.

The incorporation of quartets and quintets as figures of merit was essential for the success of this technique, since the correct solution was ranked 10th in Ψ_o, 23rd in both ABSFOM and R_{Karle}, and would not normally have been investigated. However, this solution was the only one with both negative NQEST (-0.03) and NQINT (-0.28).

5.5 STRUCTURE REFINEMENT

Both structures were refined using SHELX. Refinement of (Ib) was hindered by the disorder encountered in the terminal acetyl group of the longest side-chain which extends from C(8). Several difference Fourier syntheses finally resolved the disorder as occurring around O(7), the terminal oxygen of the acetyl function, which has two possible orientations, each with a population parameter of ca. 0.5. A final R value of 0.087 (R_w=0.087) was obtained for full-matrix least-squares adjustment of positional and anisotropic thermal parameters to which the calculated hydrogen parameters had been added, but not refined.

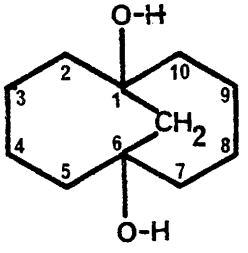
The refinement of (Ic) was straightforward with full-matrix least-squares adjustment of positional and isotropic thermal parameters converging at R=0.121. Anisotropic refinement produced a final weighted R value of 0.087. No hydrogen parameters were included in the final cycles of refinement due to the limits imposed by SHELX on the number of atoms in anisotropic refinement. Contribution of the 64 hydrogen atoms is thought to be significant, since, on convergence of isotropic refinement, addition of these calculated hydrogen positions reduced R from0.121 to 0.104.

Tables 1 and 2 list the final atomic and anisotropic thermal parameters for (Ib) and (Ic) respectively. Bond lengths and angles are tabulated for both molecules in Table 2. Relevant torsion angles have been collated in

Figure 2. ORTEP drawings are shown in Figures 3 and 4 for (Ib) and (Ic) respectively, the former, for clarity, showing only one of the O(7) positions; whilst the latter shows only one of the molecules in the asymmetric unit.

5.6 DISCUSSION

The structure elucidation of iso- and acetylchapliatrin has shown that in both cases the cycloheptane rings adopt the minimum-energy twist-chair (C_2) conformation and the 10-membered ring resembles a low-energy cyclodeca-1,6 diene conformer. In this respect, both structures may be directly compared to bicyclo [4:4:1] undecane-1,6 diol[3] (II) in which the 5,10 oxygen bridge of the chapliatrins has been replaced by a carbon bridge.



(II)

Replacement of C(11) in (II) with an oxygen as in (Ib) and (Ic) appears to have little efect on the stereochemistry of either of the cycloheptane rings; showing , in both cases, minimum energy conformations. The analyses of several sesquiterpene lactones by McPhail and Sim[4] lead to the derivation of two useful parameters, Σ_2 and Σ_s, which can be used as a guide to the relevant stereochemistry of such molecules.

The values of Σ_2 and Σ_s for (Ib) and (Ic) are given below along with the calculated values for the pure twist-chair conformer.

	(Ib)		(Ic)		(Ic')		Twist
	Ring A	Ring B	Ring A	Ring B	Ring A	Ring B	chair
Σ_2	20	15	14	10	16	12	0
Σ_s	117	148	113	142	117	148	141

The C_2 symmetry axis, in all three molecules, passes through C(10) and midway between the C(3)-C(4) bond in ring A, and C(5) and bond C(8)-C(9) in ring B. The conformation of the 10-membered rings in (Ib) and (Ic) also approximates closely to C_2 symmetry, the axis passing through O(1) normal to the plane of least inertia. These results concur with the findings for (II) (the torsion angles are listed in Table 3) with all three molecules resembling a trans-trans-cyclodeca-1,6-diene variant[5] described by White and Bovil (1975).

REFERENCES

1. W. Hertz, I. Wahlberg, C. S. Stevens and P. S. Kalyanarman, Phytochemistry (1975) 14, pp. 1803-1808.
2. H. Wagner.M. A. Iyengar and W. Hertz, Phytochemistry (1973) 12, 2063.
3. D. N. J. White and M. J. Bovill, Acta Cryst. (1977) B33, 3029.
4. A. T. McPhail and G. A. Sim, Tetrahedron (1973) 29, pp. 1751-1758.
5. D. N. J. White and M. J. Bovill, Tetrahedron Lett. (1975) pp. 2239-2240.

Table 1A

Atomic coordinates and thermal parameters for Isochapliatrin(Ib)

(i) Atomic coordinates ($x10^4$)

ATOM	x/a	y/b	z/c
O(1)	12137(8)	7127(3)	7411(4)
O(2)	9575(8)	8689(3)	7210(3)
O(3)	7398(9)	9348(4)	7689(4)
O(4)	12914(9)	8580(3)	9049(4)
O(5)	13154(16)	8461(5)	10414(5)
O(6)	12135(17)	10098(7)	8898(14)
O(7)	12524(31)	11029(11)	8859(23)
O(7')	12254(49)	10219(28)	7871(25)
O(8)	9405(10)	7927(4)	5471(5)
O(9)	16563(9)	8958(3)	5590(4)
O(10)	14108(7)	8762(3)	6369(3)
C(1)	15050(12)	7729(4)	7628(6)
C(2)	15472(12)	7676(5)	6711(7)
C(3)	14222(11)	8043(3)	6110(5)
C(4)	12245(11)	7774(4)	6150(5)
C(5)	11342(11)	7678(4)	6977(6)
C(6)	11069(10)	8271(4)	7548(5)
C(7)	10411(12)	8073(4)	8372(5)
C(8)	11925(13)	7956(5)	9015(6)
C(9)	13102(13)	7347(5)	8810(7)
C(10)	13693(13)	7209(4)	7929(7)
C(11)	9126(12)	8629(5)	8585(6)
C(12)	8546(13)	8937(5)	7844(6)
C(13)	13603(17)	6123(7)	10710(7)
C(14)	11125(12)	8221(5)	5566(6)
C(15)	14568(15)	6501(4)	7889(9)
C(16)	13367(15)	8812(6)	9825(8)
C(17)	14111(15)	9483(6)	9793(12)
C(18)	14841(16)	9778(7)	10402(12)
C(19)	15027(19)	9502((8)	11325(11)

Table 1A (continued)

C(20)	13960(16)	9875(6)	8994(10)
C(21)	11543(22)	10551(13)	8606(11)
C(22)	9718(22)	10720(7)	8547(8)
C(23)	15353(12)	9165(4)	6019(5)
C(24)	14960(13)	9877(4)	6191(6)

(ii) Thermal parameters ($Å^2 \times 10^3$)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
O(1)	39	36	98	-12	-12	09
O(2)	48	58	42	02	-02	-15
O(3)	57	89	68	01	-02	-28
O(4)	48	61	63	-12	-14	05
O(5)	189	138	64	-14	-51	80
O(6)	99	126	454	-161	83	-53
O(7)	126	136	462	-31	-54	48
O(7')	133	437	147	75	-50	-176
O(8)	52	91	95	40	-22	-11
O(9)	51	65	74	01	14	02
O(10)	37	38	58	04	00	-01
C(1)	40	44	78	-03	-18	01
C(2)	33	48	102	01	-06	-05
C(3)	41	43	51	16	11	02
C(4)	41	38	61	24	-03	03
C(5)	21	47	78	06	-12	01
C(6)	29	47	38	00	-11	03
C(7)	43	61	43	-12	-05	11
C(8)	51	64	64	-28	-17	12
C(9)	51	57	85	-27	-30	05
C(10)	46	42	108	-25	-19	-02
C(11)	31	78	70	-27	-05	15
C(12)	44	55	68	03	-09	01
C(13)	70	121	54	05	-23	03
C(14)	48	73	47	05	-21	-01

Table 1A (continued)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
C(15)	67	40	145	-21	-29	-14
C(16)	53	78	90	-06	-13	-03
C(17)	29	66	210	39	07	-06
C(18)	33	89	229	51	16	00
C(19)	53	170	238	102	-43	-19
C(20)	56	69	166	02	20	00
C(21)	88	179	97	-84	-19	17
C(22)	161	87	96	21	-27	-56
C(23)	46	50	39	07	-10	08
C(24)	53	45	90	-17	-10	03
Average e.s.d.'s						
O,C	7	8	9	8	7	8

(iii) Hydrogen atom fractional coordinates ($x10^4$)

ATOM	x/a	y/b	z/c	U_{iso}
HO(8)	16257	11636	10439	0.103
H(1)	16284	7662	7964	0.103
H(2)	14516	8224	7748	0.103
H(3)	15452	7150	6553	0.103
H(4)	16808	7874	6621	0.103
H(5)	14762	7969	5504	0.103
H(6)	12317	7253	5965	0.103
H(7)	9978	7580	6787	0.103
H(8)	12355	8518	7597	0.103
H(9)	9775	7586	8373	0.103
H(10)	11378	7832	9607	0.103
H(11)	14315	7399	9166	0.103

Table 1A (continued)

ATOM	x/a	y/b	z/c	U_{iso}
H(12)	12374	6908	9010	0.103
H(13)	10989	8719	5823	0.103
H(14)	11784	8254	4980	0.103
H(15)	15769	6496	8259	0.103
H(16)	13633	6131	8117	0.103
H(17)	14911	6385	7264	0.103
H(18)	15400	10270	10285	0.103
H(19)	15713	9870	11695	0.103
H(20)	13707	9410	11574	0.103
H(21)	15785	9039	11323	0.103
H(22)	14849	10304	9010	0.103
H(23)	14320	9554	8488	0.103
H(24)	8969	10290	8333	0.103
H(25)	9225	10869	9140	0.103
H(26)	9556	11131	8122	0.103
H(27)	15967	10187	5907	0.103
H(29)	14968	9956	6843	0.103
H(30)	13652	6363	10153	0.103
H(31)	12930	5642	10485	0.103

Hydrogen atoms included in structure factor calculation but not refined.

Table 1B

Atomic coordinates and thermal parameters for Acetylchapliatrin(Ic)

(i) Atomic coordinates ($x10^4$)

ATOM	x/a	y/b	z/c
O(1)	7196(3)	5485	7572(5)
O(2)	5616(3)	5365(4)	5226(5)
O(3)	4794(3)	4567(5)	3704(6)
O(4)	6504(3)	3519(5)	7934(5)
O(5)	6364(4)	1896(6)	7758(7)
O(6)	6814(3)	3899(6)	10106(5)
O(7)	6698(4)	5404(7)	10390(8)
O(8)	6268(3)	6667(5)	7851(6)
O(9)	6362(5)	7857(8)	8866(11)
O(10)	6097(4)	8424(6)	6252(8)
O(11)	5246(5)	8751(7)	4445(9)
C(1)	7298(4)	5404(7)	9189(7)
C(2)	7300(4)	6527(8)	9239(8)
C(3)	6761(4)	7030(7)	8139(10)
C(4)	6603(4)	6921(8)	7024(9)
C(5)	6612(4)	5868(7)	6701(8)
C(6)	6179(4)	5118(6)	6386(8)
C(7)	6290(3)	4073(7)	6277(7)
C(8)	6719(4)	3486(7)	7460(7)
C(9)	7328(4)	3866(7)	8346(8)
C(10)	7459(4)	4984(8)	8623(9)
C(11)	5694(4)	3680(6)	5285(8)
C(12)	5295(4)	4530(8)	4618(9)
C(13)	5504(5)	2773(7)	5007(9)
C(14)	6033(5)	7410(8)	5957(10)
C(15)	8115(4)	5135(10)	9452(10)
C(16)	6331(4)	2682(9)	8000(8)
C(17)	6094(4)	2916(9)	8458(8)
C(18)	5744(5)	2281(12)	8305(10)
C(19)	5547(6)	1284(11)	7756(13)

Table 1B (continued)

ATOM	x/a	y/b	z/c
C(20)	6271(4)	3908(9)	8999(8)
C(21)	7003(5)	4699(11)	10729(10)
C(22)	7643(5)	4674(11)	11911(10)
C(23)	6083(6)	7179(9)	8213(11)
C(24)	5786(6)	10028(10)	5895(17)
C(25)	5658(5)	9026(10)	5414(12)
C(26)	5577(6)	6788(12)	7779(12)
O(1')	7182(3)	3000(5)	2785(5)
O(2')	5614(3)	2763(5)	0300(5)
O(3')	4861(3)	1908(5)	-1255(6)
O(4')	6485(3)	0947(5)	3041(5)
O(5')	6359(4)	-0665(7)	2812(8)
O(6')	6813(3)	1171(6)	5233(7)
O(7')	6712(4)	2672(10)	5556(10)
O(8')	6158(3)	4104(5)	2843(5)
O(9')	6271(4)	5209(8)	3984(8)
O(10')	6042(3)	5861(6)	1261(7)
O(11')	5196(5)	6105(8)	-0674(8)
C(1')	7221(4)	2900(8)	4293(7)
C(2')	7203(4)	4003(9)	4336(7)
C(3')	6652(4)	4495(7)	3184(8)
C(4')	6534(4)	4369(7)	2104(8)
C(5')	6602(4)	3327(7)	1880(8)
C(6')	6178(4)	2555(7)	1533(8)
C(7')	6334(4)	1532(7)	1467(8)
C(8')	6745(4)	0950(7)	2659(8)
C(9)	7354(4)	1392(7)	3629(9)
C(10')	7433(4)	2495(8)	3848(8)
C(11')	5739(4)	1085(8)	0411(7)
C(12')	5349(4)	1907(7)	-0292(8)
C(13')	5569(6)	0169(9)	0109(11)
C(14')	5986(5)	4843(7)	1036(9)

Table 1B (continued)

ATOM	x/a	y/b	z/c
C(15')	8073(4)	2701(10)	4723(11)
C(16')	6305(4)	0098(9)	3069(9)
C(17')	6068(4)	0217(9)	3501(8)
C(18')	5722(5)	-0417(11)	3314(10)
C(19')	5503(7)	-1337(16)	2720(15)
C(20')	6170(5)	1164(11)	4071(10)
C(21')	7002(6)	1959(12)	5883(11)
C(22')	7650(6)	1950(12)	7027(11)
C(23')	6007(4)	4559(9)	3286(9)
C(24')	5778(9)	7513(13)	0714(19)
C(25')	5610(7)	6438(9)	0294(15)
C(26')	5448(5)	4140(10)	2741(12)

(ii) Thermal parameters ($Å^2 \times 10^3$)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
O(1)	43	41	54	17	37	08
O(2)	30	36	33	09	19	06
O(3)	36	56	47	-09	22	-01
O(4)	38	38	32	-02	25	-06
O(5)	111	36	99	-15	88	-26
O(6)	44	66	38	03	30	04
O(7)	71	70	73	-02	46	18
O(8)	68	49	59	-07	48	01
O(9)	121	82	140	-44	100	-10
O(10)	55	37	92	-04	43	-09
O(11)	90	62	91	15	51	18
C(1)	49	42	28	-06	24	-06
C(2)	46	44	33	01	16	-15
C(3)	60	29	71	-19	49	-17
C(4)	48	48	56	01	39	-10
C(5)	34	37	41	-01	26	-09

Table 1B (continued)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
C(6)	43	24	45	-07	35	-14
C(7)	26	41	36	03	21	-05
C(8)	36	34	46	-04	34	-03
C(9)	38	40	41	22	26	12
C(10)	32	49	49	16	27	07
C(11)	45	21	39	-01	32	05
C(12)	41	49	43	-12	35	-09
C(13)	64	37	50	-09	44	-13
C(14)	60	26	52	-09	29	-16
C(15)	23	89	39	08	08	-01
C(16)	39	50	33	16	24	-07
C(17)	35	88	27	14	19	02
C(18)	48	122	60	34	36	-03
C(19)	73	92	100	19	53	-38
C(20)	40	71	26	00	20	01
C(21)	52	98	55	00	45	-06
C(22)	50	95	47	04	23	09
C(23)	105	39	73	-32	66	-10
C(24)	77	40	167	-06	77	05
C(25)	43	61	74	23	35	19
C(26)	74	93	94	-33	66	-10
O(1')	57	34	56	-09	50	-09
O(2')	32	37	41	-01	22	-04
O(3')	51	48	46	-03	28	-02
O(4')	44	48	46	06	36	02
O(5')	113	54	105	01	88	-16
O(6')	58	68	56	09	40	18
O(7')	87	150	104	-22	61	39
O(8')	49	43	49	-03	39	-05
O(9')	78	122	84	-59	61	-22
O(10')	59	37	62	01	33	05
O(11')	90	64	66	18	23	16
C(1')	48	49	34	14	29	18

Table 1B (continued)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
C(2')	41	72	27	-08	18	03
C(3')	33	36	47	-01	26	-06
C(4')	32	34	48	-17	29	-15
C(5')	28	39	45	00	27	02
C(6')	44	35	43	02	37	08
C(7')	47	36	49	-10	40	-13
C(8')	49	34	45	-03	36	03
C(9')	42	39	55	02	33	13
C(10')	24	58	43	02	20	05
C(11')	43	63	23	04	24	-03
C(12')	41	43	37	-09	29	-10
C(13')	83	39	57	-18	41	-29
C(14')	62	28	46	-07	38	-08
C(15')	41	75	74	-13	42	-12
C(16')	43	48	43	01	26	-10
C(17')	43	72	39	14	27	-03
C(18')	46	86	56	18	29	-05
C(19')	67	157	99	61	34	-22
C(20')	44	100	49	15	30	16
C(21')	90	106	81	-08	74	21
C(22')	66	109	56	02	36	05
C(23')	47	73	49	-09	37	02
C(24')	115	48	137	-11	19	-04
C(25')	86	30	104	-04	52	05
C(26')	70	84	98	-17	72	03
Average e.s.d's	7	6	6	5	6	5

Table 2

Interatomic distances(Å) and angles(°) for Iso- and Acetylchapliatrin

(a) Bonded distances

	(Ib)	(Ic)	(Ic')
O(1) -C(5)	1.43(1)	1.45(1)	1.42(1)
O(1) -C(10)	1.44(1)	1.43(1)	1.46(1)
O(2) -C(6)	1.49(1)	1.44(1)	1.48(1)
O(2) -C(12)	1.38(1)	1.37(1)	1.36(1)
O(3) -C(12)	1.21(1)	1.19(1)	1.19(1)
O(4) -C(8)	1.44(1)	1.45(1)	1.46(1)
O(4) -C(16)	1.39(2)	1.33(1)	1.33(1)
O(5) -C(16)	1.20(2)	1.18(1)	1.20(2)
O(6) -C(20)	1.44(2)	1.42(1)	1.52(1)
O(6) -C(21)	1.11(3)	1.31(2)	1.31(2)
O(7) -C(21)	1.27(3)	1.21(2)	1.20(2)
O(8) -C(3)	1.50(1)	1.48(1)	1.45(1)
O(8) -C(23)	1.35(1)	1.34(2)	1.37(2)
O(9) -C(23)	1.22(1)	1.18(2)	1.20(2)
O(10)-C(14)	1.42(1)	1.43(1)	1.42(1)
O(10)-C(25)	-	1.37(2)	1.37(1)
O(11)-C(25)	-	1.19(2)	1.20(2)
C(1) -C(2)	1.54(1)	1.53(1)	1.51(2)
C(1) -C(10)	1.53(1)	1.54(1)	1.50(1)
C(2) -C(3)	1.54(1)	1.50(1)	1.53(1)
C(2) -C(4)	1.57(1)	1.53(2)	1.56(1)
C(4) -C(5)	1.52(1)	1.55(1)	1.54(1)
C(4) -C(5)	1.55(1)	1.52(2)	1.48(1)
C(5) -C(6)	1.52(1)	1.55(1)	1.53(1)
C(6) -C(7)	1.49(1)	1.53(1)	1.53(1)
C(7) -C(8)	1.56(1)	1.56(1)	1.56(1)
C(7) -C(11)	1.50(1)	1.49(1)	1.53(1)
C(8) -C(9)	1.53(1)	1.50(1)	1.53(1)
C(9) -C(10)	1.53(2)	1.56(1)	1.53(1)

Table 2 (continued)

	(Ib)	(Ic)	(Ic')
C(10)-C(15)	1.55(1)	1.56(1)	1.52(1)
C(11)-C(12)	1.43(1)	1.49(1)	1.45(1)
C(11)-C(13)	1.32(1)	1.31(1)	1.31(2)
C(16)-C(17)	1.45(2)	1.54(1)	1.46(1)
C(17)-C(18)	1.28(2)	1.35(2)	1.32(2)
C(17)-C(20)	1.53(2)	1.51(2)	1.49(2)
C(18)-C(19)	1.61(2)	1.50(2)	1.42(3)
C(21)-C(22)	1.40(2)	1.52(2)	1.52(3)
C(23)-C(24)	1.47(1)	1.41(2)	1.51(2)
C(25)-C(26)	-	1.48(2)	1.54(2)

b) Interbond angles

	(Ib)	(Ic)	(Ic')
C(10)-O(1) -C(5)	122(1)	123(1)	123(1)
C(4) -C(5) -O(1)	111(1)	111(1)	112(1)
C(6) -C(5) -O(1)	119(1)	112(1)	113(7)
C(1) -C(10)-O(1)	115(1)	114(1)	113(1)
C(9) -C(10)-O(1)	110(1)	110(1)	109(1)
C(15)-C(10)-O(1)	102(1)	102(1)	102(1)
C(12)-O(2) -C(6)	110(1)	109(1)	110(1)
C(5) -C(6) -O(2)	108(1)	107(1)	107(1)
C(7) -C(6) -O(2)	104(1)	105(1)	104(1)
O(3) -C(12)-O(2)	118(1)	122(1)	120(1)
C(11)-C(12)-O(2)	109(1)	108(8)	111(1)
C(11)-C(12)-O(3)	133(1)	131(1)	129(1)
C(16)-O(4) -C(8)	116(1)	118(1)	119(1)
C(7) -C(8) -O(4)	105(1)	108(1)	108(1)
C(9) -C(8) -O(4)	113(1)	111(1)	110(1)
O(5) -C(16)-O(4)	120(1)	126(1)	122(1)
C(17)-C(16)-O(4)	111(1)	108(1)	112(1)
C(17)-C(16)-O(5)	128(1)	126(1)	125(1)

Table 2 (continued)

	(Ib)	(Ic)	(Ic')
C(21)-O(6) -C(20)	132(2)	115(1)	116(1)
C(17)-C(20)-O(6)	109(1)	106(1)	105(1)
O(7) -C(21)-O(6)	104(2)*	123(1)	124(1)
C(22)-C(21)-O(6)	128(2)*	114(1)	114(1)
C(22)-C(21)-O(7)	114(2)*	123(1)	122(1)
C(23)-O(8) -C(3)	114(1)	120(1)	117(1)
C(2) -C(3) -O(8)	108(1)	109(1)	109(1)
C(4) -C(3) -O(8)	105(1)	108(1)	107(1)
O(9) -C(23)-O(8)	123(1)	119(1)	124(1)
C(24)-C(23)-O(8)	111(1)	113(1)	110(1)
C(24)-C(23)-O(9)	126(1)	128(1)	125(1)
C(25)-O(10)-C(14)	-	118(1)	116(1)
C(4) -C(14)-O(10)	109(1)	107(1)	108(1)
O(11)-C(25)-O(10)	-	122(1)	123(1)
C(26)-C(25)-O(10)	111(1)	110(1)	107(1)
C(26)-C(25)-O(11)	-	128(1)	130(2)
C(10)-C(1) -C(2)	114(1)	115(1)	116(1)
C(3) -C(2) -C(1)	118(1)	116(1)	116(1)
C(9) -C(10)-C(1)	112(1)	113(1)	115(1)
C(15)-C(10)-C(1)	109(1)	111(1)	111(1)
C(4) -C(3) -C(2)	112(1)	117(1)	115(1)
C(5) -C(4) -C(3)	120(1)	116(1)	116(1)
C(14)-C(4) -C(3)	106(1)	112(1)	111(1)
C(14)-C(4) -C(5)	112(1)	110(1)	113(1)
C(6) -C(5) -C(4)	120(1)	118(1)	117(1)
C(7) -C(6) -C(5)	113(1)	113(1)	111(1)
C(8) -C(7) -C(6)	114(1)	114(1)	114(1)
C(11)-C(7) -C(6)	103(1)	102(1)	102(1)
C(11)-C(7) -C(8)	114(1)	117(1)	116(1)
C(9) -C(8) -C(7)	113(1)	113(1)	113(1)

* Disordered side-chain

Table 2 (continued)

	(Ib)	(Ic)	(Ic')
C(12)-C(11)-C(7)	108(1)	107(1)	106(1)
C(13)-C(11)-C(7)	128(1)	130(1)	131(1)
C(10)-C(9) -C(8)	121(1)	121(1)	120(1)
C(15)-C(10)-C(9)	109(1)	108(1)	107(1)
C(13)-C(11)-C(12)	122(1)	122(1)	123(1)
C(18)-C(17)-C(16)	124(1)	121(1)	124(1)
C(20)-C(17)-C(16)	118(1)	121(1)	118(1)
C(20)-C(17)-C(18)	118(1)	120(1)	117(1)
C(19)-C(18)-C(17)	127(1)	129(1)	131(1)

Flow chart for MULTAN modules

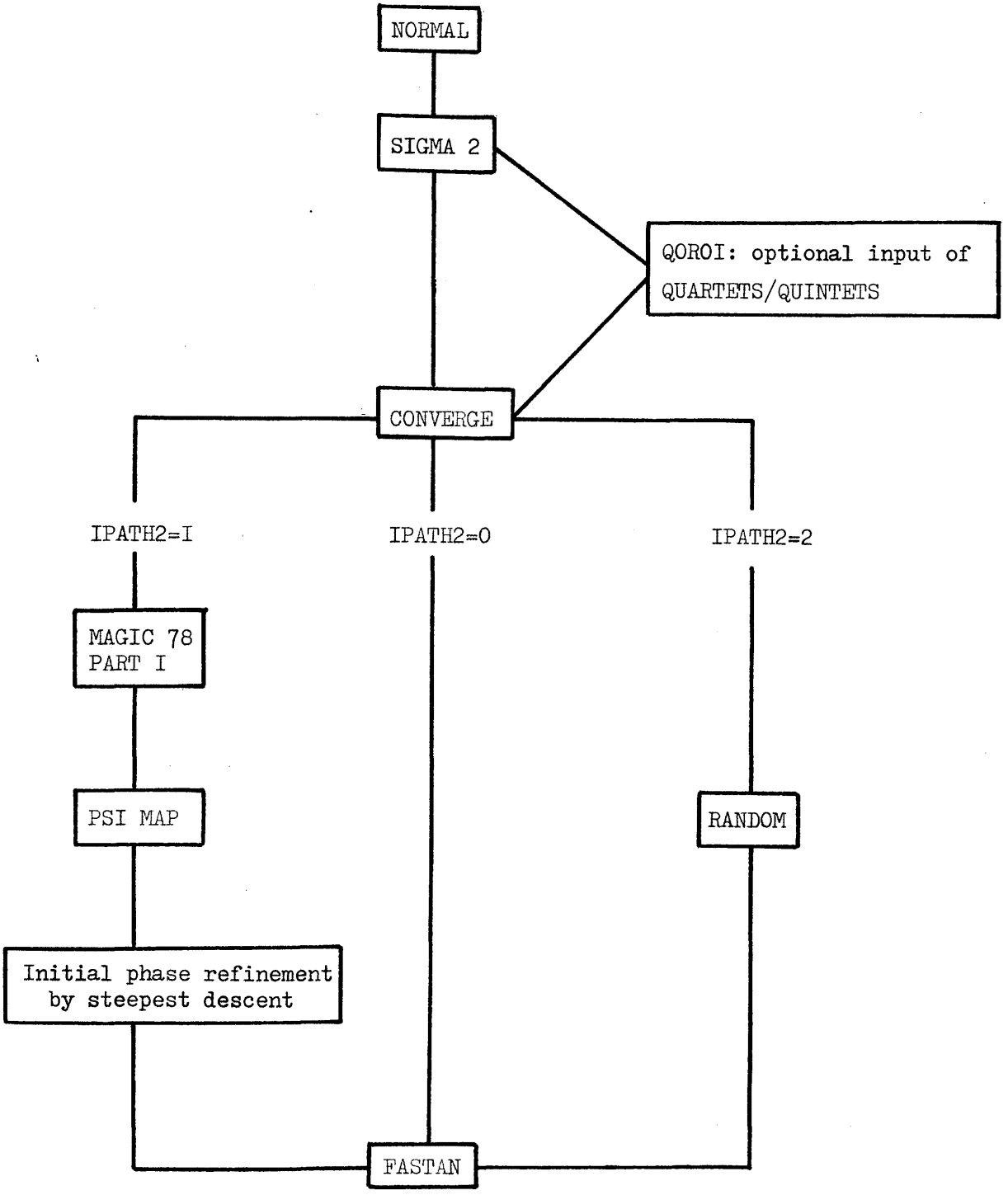
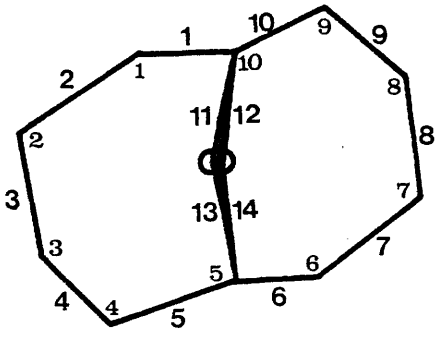


Figure I

Figure 2

Ring torsion angles



11 C(5) -O(1) -C(1Ø)-C(1)
12 C(5) -O(1) -C(1Ø)-C(9)
13 C(4) -C(5) -O(1) -C(1Ø)
14 C(6) -C(5) -O(1) -C(1Ø)

Angle number	Isochapliatrin (Ib)	Acetylchapliatrin (Ic)	(Ic')	(II)
1	-171	-168	-171	-164
2	86	84	86	85
3	-62	-64	-63	-73
4	47	49	47	54
5	61	63	65	63
6	-172	-171	-174	-166
7	93	89	91	85
8	-66	-68	-69	-71
9	41	48	46	52
1Ø	7Ø	64	65	65
11	-35	-37	-38	-43
12	92	9Ø	91	86
13	87	88	9Ø	86
14	-48	-46	-46	-42

Isochapliatrin (Ib)

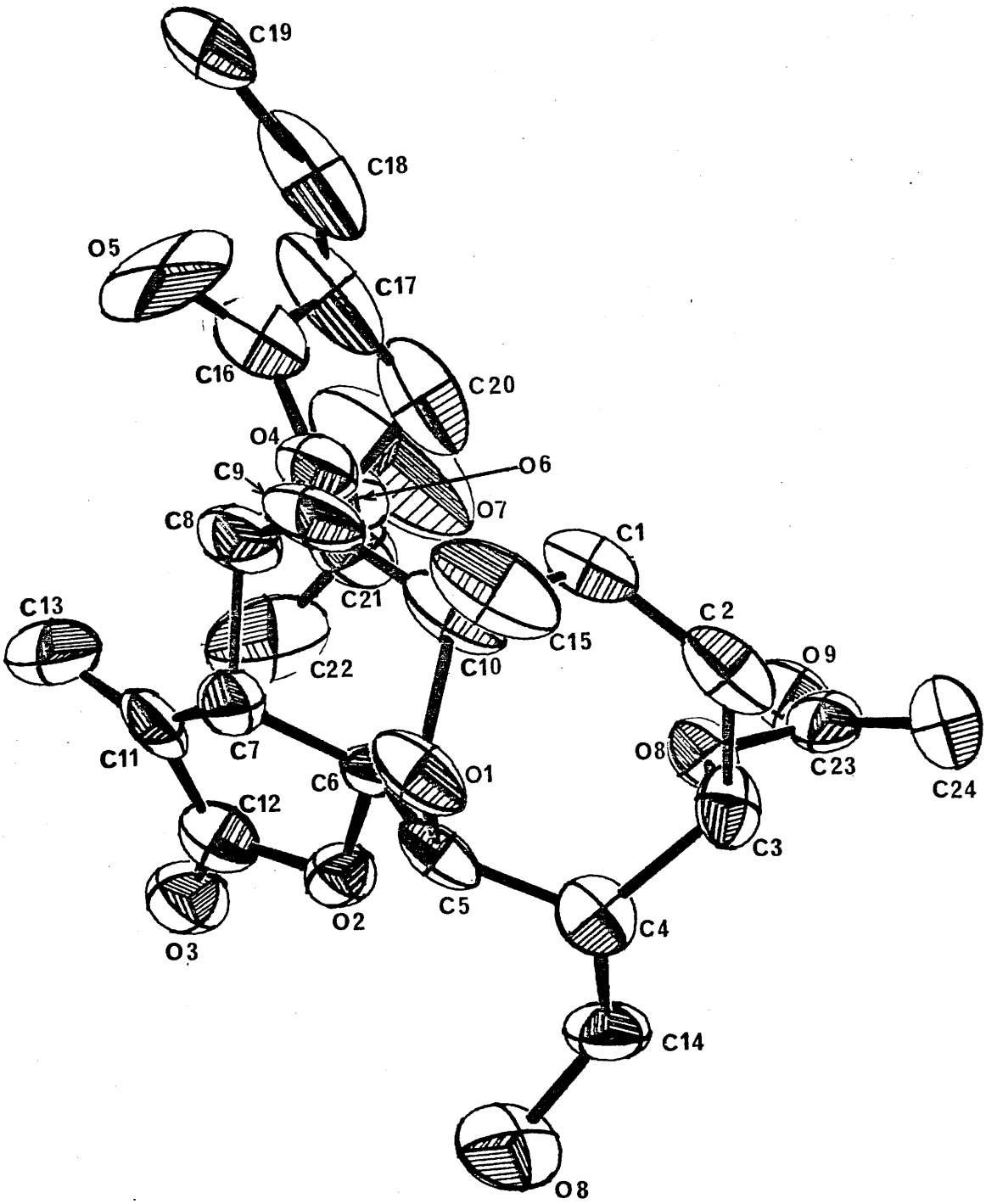


Figure 3

Acetylchapliatrin (I c)

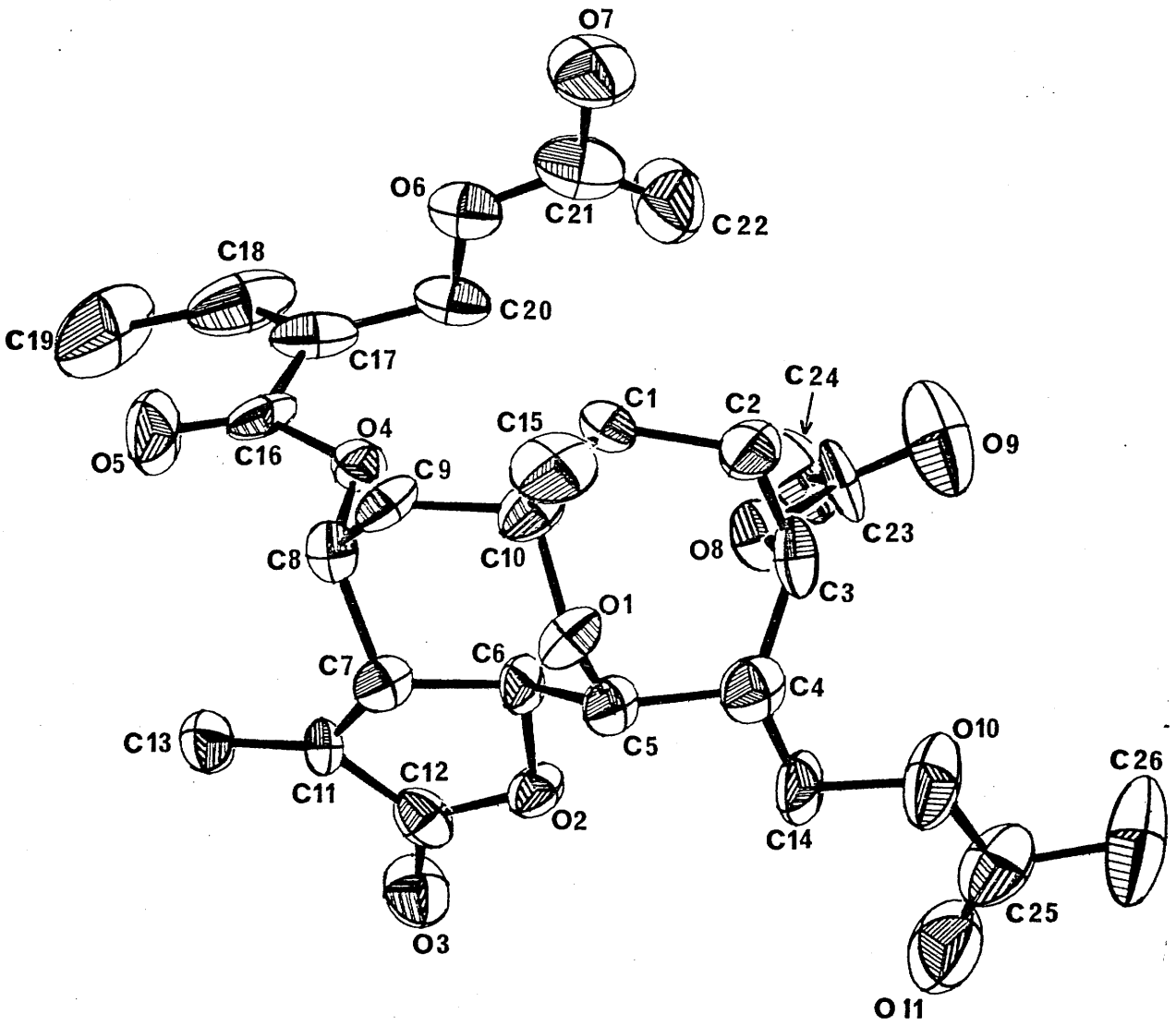


Figure 4

Chapter 6

Controlling the Normalisation Process:

The Structure Elucidation of Three Natural Products.

1.0 GENERAL INTRODUCTION

The three structures which follow were each solved by routine applications of standard MULTAN procedures, but only after recalculation of the E-magnitudes obtained from NORMAL. In each case, the successive use of MULTAN with various input parameters failed to find a solution for any of the three structures. (It should be added that neither MULTAN with quartets nor the optional MAGIC procedure described previously were available at the time).

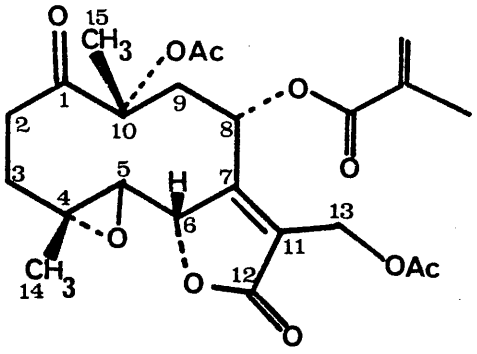
I

THE STRUCTURE ELUCIDATION OF GLAUCOLIDE-F, A SESQUITERPENE LACTONE, BY RESCALING OF E-MAGNITUDES

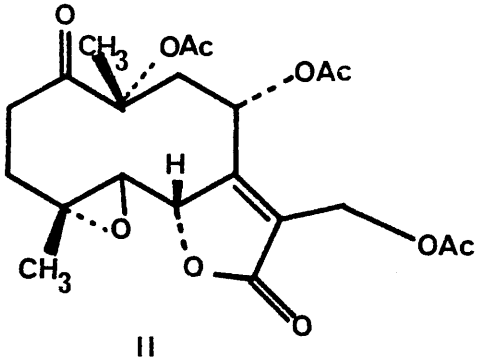
6.1.0 INTRODUCTION

In the first structure to be discussed, a germacranolide-type sesquiterpene (glaucolide-F), quartet and quintet invariants were employed in an active role using an invariants least-squares program described by Gilmore[1]. However, even with the use of higher invariants, it was difficult to expand the initial starting phases to give strong indications for reflexions which could be used as a basis for E-map calculations.

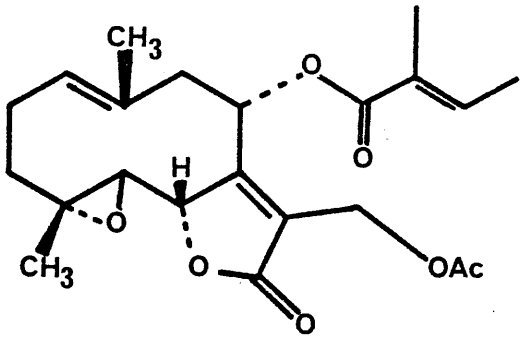
The investigation of several novel, highly oxygenated germacranilide sesquiterpene lactones derived from the species of genus Vernonia, has resulted in the structure elucidation of several of its crystalline forms, viz. glaucolide-A(I)[2], B(II)[3], D(IIIa)[4], E(IIIb)[4] and marginatin(IV)[5]. Spectroscopic and chemical investigation of a further substrate of the genus Vernonia, glaucolide-F, indicated the presence of a saturated ten-membered ring and characteristic C(13)-allylic acetate function[6]. The X-ray crystallographic examination of glaucolide-F was therefore undertaken to establish the relative stereochemistry of the molecule and the conformation of the 10-membered ring.



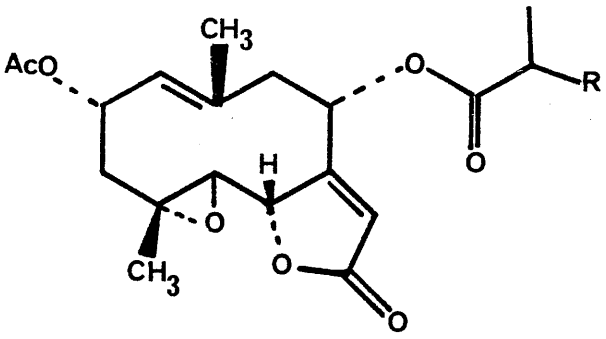
I



II

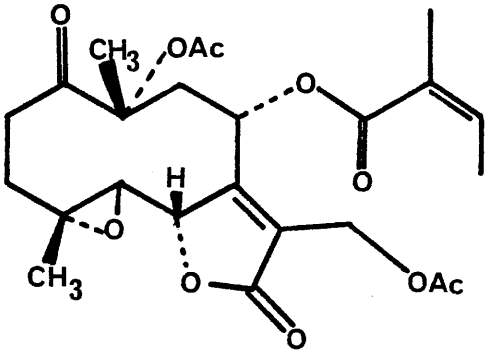


IV

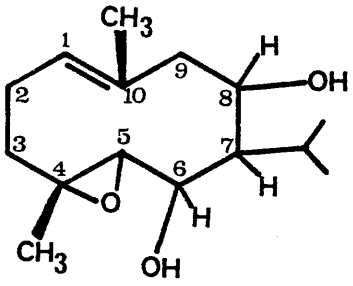


IIIa R= epoxide

IIIb R= ethylene



V



VI

6.1.1 EXPERIMENTAL

Crystal data

Glaucolide-F, $C_{24}H_{30}O_{10}$, M_r=477.9, orthorhombic, a=8.481(1), b=14.448(2), c=19.580(2) Å, U=2399.2 Å^3, D_m=1.31, D_c=1.32Mgm^{-3}, Z=4, F(000)=1016, space group $P2_12_12_1$, μ(Mo-K_α)=1.11 cm^{-1}

Data collection

Instrument used: Hilger Watts Y290
Radiation used: Mo-K_α , λ=0.71069 Å
Filter: Graphite monochromator, $\cos^2 2\theta$=0.960
Upper limit for data collection: $2\theta_{max}=60^o$
Number of independent reflexions: m=1106
Unobserved cut-off: $2.5\sigma_I$
Number of independent parameters: n=308
Number of reflexion per parameter: m/n=3.6

6.1.2 STRUCTURE DETERMINATION AND REFINEMENT

Despite the inclusion of the atomic coordinates for glaucolide-A when calculating E-magnitudes[7] (included as a randomly orientated and randomly positioned fragment) numerous attempts at structure solution using various options in MULTAN 76 were unsuccessful. The generation of quartets for the top 120 E-magnitudes produced only about 180 useful invariants. Symbolic addition performed by hand through the phase expansion procedure encountered difficulty in evaluating reliable phase estimates for sufficient reflexions which could be used in an E-map calculation. RANDOM (as described in Chapter 5) was used to obtain a larger starting set of reflexions by employing the bottom 100 reflexions from the convergence map. However, after the calculation of 1500 phase sets no least-squares phase minimum had been found.

As an experiment, the normalised structure factors were rescaled by assigning specific values to the temperature coefficient, B. Table 1.1(a) shows the results of this process, and, not unexpectedly, as B increases the distribution of E^2 with $\sin\theta/\lambda$ varies markedly from its approximate value of unity. Hence, as B increases the E-magnitudes associated with high ordered reflexions are given increased magnitudes, while reflexions at lower $\sin\theta/\lambda$ values are correspondingly reduced.

Table 1.1(b) shows the changes in the origin reflexions selected by MULTAN as B is increased. Initially, reflexions with low $\sin\theta/\lambda$ values are predominant, but, as B changes high order reflexions become more involved, until, at B = 5.0 all the origin defining reflexions have high $\sin\theta/\lambda$ values. For each increase in B MULTAN was re-run with default options for Σ_2, CONVERGE and FASTAN, until, at B =4.0, the complete structure was revealed from an E-map employing 250 E-magnitudes ($|E| \geqslant 1.22$).

The atomic parameters derived from this E-map were refined by full-matrix isotropic least-squares, using CRYLSQ from the X-RAY suite of programs, to an R of 0.136. A further six cycles of anisotropic refinement reduced R to 0.099. Since the number of reflexions per parameter was now only 3.6, refinement of the hydrogen atom coordinates was not possible; but their inclusion in the structure factor calculation in the final least-squares cycle reduced R to a final value of 0.093 (R_w=0.098). Table 1.2 lists the final fractional coordinates and thermal parameters for the molecule, while Table 1.3 summarises relevant bond lengths and angles.

A probable explanation as to the success of the method employed in the eventual structure solution can be found by considering that by constraining the starting set of reflexions the phase determining path was changed sufficiently to allow a completely new phase expansion and determination process to proceed.

Several different starting sets of reflexions were used to redefine the origin; but, as often occurs during the convergence mapping procedure, these reflexions quickly generated the original origin reflexions and thereby produced solutions which were little different from the first attempts. This is because the E-magnitudes, and therefore the probability relationships, had not changed significantly. By completely changing the E-magnitudes the probability relationships were changed and therefore the phase expansion procedure would take a different route.

6.1.3 DISCUSSION

Figure 1 shows an ORTEP projection of the molecule. The molecular structure of glaucolide-F shows the conformation of the ten-membered ring to be identical (within experimental error) to that found for glaucolide-A and dihydrodesacetoxyglaucolide-A[2], and thereby closely resembles one of the less-favourable conformations derived for cyclodecane by strain-energy minimisation calculations[8]. The conformations are also similar to that of shiromodiol acetate p-bromobenzoate(VI)[9], where the macrocycle contains a double bond at C(4)-C(5). Although these germacrane rings have related conformations, as shown by the endocyclic torsion angles of the rings (Table 4), corresponding bonds have different torsion angles. For example, C(7) in shiromodiol(VI), the atom carrying the exocyclic substituent C(11), plays the same role as C(1) in glaucolide-A(I), B(II) and F(V).

The endocyclic double bond at C(7)-C(11) is typical of the glaucolide series of germacranilides with the five-membered γ-lactone ring almost planar having no internal torsion angles greater than 5^{o}. The shift of the C(11)-C(13) double bond, found in most germacranolides, to C(7)-C(11) (cf.

chapliatrins) in the glaucolide series, although responsible for the lack of significant cytotoxic activity of these molecules[10] does not greatly influence the preferred conformation of the 10-membered ring. The conformation is such that the C(14) and C(15) methyl groups are cis and lie on the α-face of the macrocycle: this is also observed in gaucolide-A and its dihydrodesacetoxy derivative. Conformational similarities also exist among germacrane sesquiterpenoids in which C(14) and C(15) carbon atoms are cis and β [11].

Glaucolides-D and E, however, adopt considerably different conformations from the glaucolides discussed so far. This is due to the presence of the double bond at C(1)-C(10) in glaucolide-D (and E) instead of the keto function at C(1). In glaucolide-F the C(8) ester has a trans configuration about the double bond. The torsion angles and stereochemistry of glaucolide-F conform with the convention designated by Neidle and Rogers[12].

REFERENCES

1. C. J. Gilmore, Acta Cryst. (1977) A33, 712.
2. P. J. Cox and G. A. Sim, J. Chem. Soc. Perkin II, (1974) pp. 1722-1724.
3. W. G. Padolina, R.Davis, P. J. Cox, G. A. Sim, W.H. Watson and I. Beth Wu, Tetrahedron (1974) 30, pp.1161-1170.
4. I. F. Taylor, W. H. Watson, M. Betkouski, W. G. Padolina and T. J. Mabry, Acta Cryst (1976) B32, 107.
5. W. G. Padolina, N. Nakatani, H. Yoshioka, T. J. Mabry and S. A. Monti, Phytochemistry (1974) 12, pp.2225-2229.
6. R. Toubiana, M. J. Toubiana and B. C. Das, Tetrahedron Lett. (1972) 207.
7. P. Main, Acta Cryst. (1975) S31, O2I-1.
8. J. D. Dunitz, Perspectives in Structural Chem. (1968) 2, 1.
9. R. J. McClure, G. A. Sim, P. Coggon and A. T. McPhail, J. C. S. Chem. Comm. (1970) 128.
10. W. H. Watson, M. G. Reinecke and J. C. Hitt, Rev. Latinoamer. Quim. (1975) 6, pp. 1-12.
11. F. H. Allen and D. Rogers, J. Chem. Soc.(B) (1971) 257.
12. D. Rogers, G. P. Moss and S. Neidle, J. C. S. Chem. Comm. 1972, 142.

Table 1.1

(a) Distribution of E^2 with $\sin\theta/\lambda$ as B is increased.

	$\sin\theta/\lambda$									
B	0.0457	0.0081	0.1445	0.2014	0.2585	0.3149	0.3721	0.4306	0.4826	0.5385
3.0	0.0386	0.7081	1.6946	1.0991	0.9808	0.9827	0.8976	0.9786	0.9748	1.0530
3.5	0.2388	1.1038	1.4887	0.8683	0.8259	0.7006	0.7793	1.2091	1.2694	1.1544
4.0	0.2055	0.9538	1.1834	0.7744	0.7569	0.6618	0.7686	1.2477	1.3709	1.3188
4.5	0.1724	0.8154	1.1191	0.6808	0.6825	0.6166	0.7460	1.2683	1.4564	1.4891
5.0	0.1460	0.6965	0.8781	0.5994	0.6171	0.5772	0.7261	1.2920	1.5519	1.6797

(b) The change in starting set reflexions as B is increased.

B						
3.0	1 3 1 (.12),	0 5 8 (.27),	2 0 3 (.14),	2 1 1 (.13),	3 1 3 (.19),	1 1 0 (.07)
3.5	1 1 0 (.07),	2 0 3 (.14),	1 0 6 (.16),	1 3 1 (.12),	2 1 1 (.12)	
4.0	2 0 3 (.14),	1 1 0 (.07),	1 0 6 (.16),	7 1 4 (.43),	7 1 9 (.47),	4 0 14 (.43)
4.5	1 1 0 (.07),	3 0 15 (.42),	4 0 13 (.41),	1 3 1 (.25),	1 0 6 (.16),	4 0 14 (.43)
5.0	3 0 15 (.42),	9 3 0 (.54),	4 0 13 (.41),	7 1 4 (.43),	3 1 13 (.38),	4 0 14 (.43)

Figures in parentheses refer to sin Θ value for that reflexion.

Table 1.2

Atomic coordinates and thermal parameters for Glaucolide-F

(i) Atomic coordinates ($x10^4$)

ATOM	x/a	y/b	z/c
O(1)	9308(14)	5175(11)	3829(7)
O(2)	10555(15)	2609(10)	5125(8)
O(3)	9133(16)	3101(11)	6450(7)
O(4)	7491(19)	2913(10)	7327(6)
O(5)	5247(12)	4720(8)	3826(6)
O(6)	6277(16)	4999(11)	2785(6)
O(7)	7226(14)	6009(9)	5672(6)
O(8)	9609(16)	6120(14)	6134(9)
O(9)*	4516	5101	6464
O(10)*	5136	6152	7143
C(1)	8190(19)	4643(14)	3885(10)
C(2)	8145(25)	3635(13)	3763(10)
C(3)	9778(25)	3185(15)	3956(12)
C(4)	10374(21)	3416(13)	4677(12)
C(5)	9271(24)	3233(15)	5235(10)
C(6)	9343(22)	3757(14)	5881(11)
C(7)	7901(21)	4410(13)	5967(11)
C(8)	7698(21)	5106(12)	5408(9)
C(9)	6277(21)	4792(15)	4926(8)
C(10)	6546(22)	5120(14)	4179(9)
C(11)	7057(23)	4143(13)	6530(9)
C(12)	7885(29)	3311(16)	6853(10)
C(13)	5510(32)	4451(19)	6940(12)
C(14)	11698(25)	4155(17)	4777(14)
C(15)	6661(24)	6194(13)	4092(10)
C(16)	5169(23)	4738(15)	3138(10)
C(17)	3616(29)	4387(20)	2860(11)
C(18)	8347(28)	6503(16)	6006(11)
C(19)	7859(39)	7405(20)	6341(25)

Table 1.2 (continued)

ATOM	x/a	y/b	z/c
C(20)	7141(41)	8021(30)	6117(31)
C(21)	6423(48)	8076(34)	5398(22)
C(22)	8581(37)	7609(23)	7108(13)
C(23)*	4658	5730	6610
C(24)*	3472	6518	6278

* These atoms, belonging to the disordered side-chain, were not refined during the final least-squares cycles.

(ii) Thermal parameters ($Å^2 x 10^3$)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
O(1)	34	90	63	-03	06	-06
O(2)	51	58	99	36	07	07
O(3)	59	114	55	21	00	20
O(4)	106	73	47	16	-02	20
O(5)	28	50	46	-13	-06	-03
O(6)	60	104	56	-12	11	-09
O(7)	45	53	59	03	04	-14
O(8)	41	181	134	-08	-19	-53
O(9)	237	102	103	09	80	32
O(10)	149	190	186	17	45	21
C(1)	15	73	53	08	08	23
C(2)	677	44	46	06	04	-12
C(3)	50	60	102	02	05	-11
C(4)	24	44	107	09	13	09
C(5)	51	63	54	02	02	-20
C(6)	39	68	67	10	03	35
C(7)	35	41	77	-02	-15	04
C(8)	40	19	69	-04	-10	-07
C(9)	42	76	30	09	-09	22
C(10)	53	65	36	14	-16	-24

Table 1.2 (continued)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
C(11)	53	42	44	19	-06	06
C(12)	90	83	48	-32	-32	11
C(13)	108	121	74	19	23	45
C(14)	47	107	129	-44	-17	29
C(15)	73	39	52	17	-17	11
C(16)	52	78	55	-05	-10	-24
C(17)	79	156	49	-39	-05	-07
C(18)	98	77	82	-57	50	-25
C(19)	90	49	295	-27	103	-33
C(20)	78	135	387	36	106	126
C(21)	129	238	230	24	-12	108
C(22)	138	163	64	-78	-40	18
C(23)	173	280	90	-126	80	-119
C(24)	375	213	91	40	10	64
Average e.s.d's						
O,C	11	10	9	12	10	12

Table 1.3

Interatomic distances(Å) and angles(°) for Glaucolide-F

(a) Bonded distances

O(1) -C(1)	1.225(22)	C(3) -C(4)	1.536(32)
O(2) -C(4)	1.468(25)	C(4) -C(5)	1.463(29)
O(2) -C(5)	1.430(24)	C(4) -C(14)	1.561(29)
O(3) -C(6)	1.475(25)	C(5) -C(6)	1.475(29)
O(3) -C(12)	1.355(26)	C(6) -C(7)	1.553(26)
O(4) -C(12)	1.141(24)	C(7) -C(8)	1.496(26)
O(5) -C(10)	1.423(21)	C(7) -C(11)	1.371(27)
O(5) -C(16)	1.349(21)	C(8) -C(9)	1.596(24)
O(6) -C(16)	1.226(23)	C(9) -C(10)	1.555(23)
O(7) -C(8)	1.460(20)	C(10)-C(15)	1.564(27)
O(7) -C(18)	1.356(26)	C(11)-C(12)	1.529(29)
O(8) -C(18)	1.230(28)	C(11)-C(13)	1.600(32)
O(9) -C(13)	1.568	C(16)-C(17)	1.513(31)
O(9) -C(23)	0.960	C(18)-C(19)	1.516(40)
O(10)-C(23)	1.276(49)	C(19)-C(20)	1.165(55)
C(1) -C(2)	1.477(28)	C(19)-C(22)	1.649(53)
C(1) -C(10)	1.658(25)	C(20)-C(21)	1.535(71)
C(2) -C(3)	1.575(29)	C(23)-C(24)	1.652(46)

(b) Interbond angles

C(5) -O(2) -C(4)	60.6(12)	C(12)-O(3) -C(6)	113.0(15)
C(16)-O(5) -C(10)	121.0(13)	C(18)-O(7) -C(8)	116.7(14)
C(14)-O(9) -C(13)	118.9	C(10)-C(1) -O(1)	114.9(17)
C(2) -C(1) -O(1)	128.6(16)	C(10)-C(1) -C(2)	116.4(15)
C(3) -C(2) -C(1)	110.2(16)	C(4) -C(3) -C(2)	114.8(17)
C(5) -C(4) -O(2)	58.4(12)	C(14)-C(4) -O(2)	113.1(16)
C(3) -C(4) -O(2)	114.3(16)	C(5) -C(4) -C(3)	115.9(16)
C(14)-C(4) -C(3)	120.1(18)	C(14)-C(4) -C(5)	119.3(19)
C(4) -C(5) -O(2)	61.0(12)	C(6) -C(5) -O(2)	114.9(16)
C(6) -C(5) -C(4)	121.4(17)	C(5) -C(6) -O(3)	108.2(16)

Table 1.3 (continued)

C(7) -C(6) -O(3)	102.3(14)	C(7) -C(6) -C(5)	111.9(16)
C(8) -C(7) -C(6)	114.8(16)	C(11)-C(7) -C(6)	109.2(16)
C(11)-C(7) -C(8)	135.8(16)	C(7) -C(8) -O(7)	111.9(14)
C(9) -C(8) -O(7)	104.9(13)	C(9) -C(8) -C(7)	109.2(14)
C(10)-C(9) -C(8)	111.0(14)	C(9) -C(10)-O(5)	102.7(14)
C(15)-C(10)-O(5)	113.5(14)	C(9) -C(10)-C(1)	108.8(14)
C(15)-C(10)-C(1)	108.8(14)	C(1) -C(10)-O(5)	108.3(13)
C(15)-C(10)-C(9)	114.5(15)	C(12)-C(11)-C(7)	108.3(16)
C(13)-C(11)-C(7)	138.9(17)	C(13)-C(11)-C(12)	112.8(16)
O(4) -C(12)-O(3)	126.1(21)	C(11)-C(12)-O(3)	107.1(16)
C(11)-C(12)-O(4)	126.7(21)	C(11)-C(13)-O(9)	108.1(20)
O(6) -C(16)-O(5)	122.1(17)	C(17)-C(16)-O(5)	113.3(16)
C(17)-C(16)-O(6)	124.5(17)	O(8) -C(18)-O(7)	118.1(20)
C(19)-C(18)-O(7)	118.0(21)	C(19)-C(18)-O(8)	122.4(24)
C(20)-C(19)-C(18)	129.5(47)	C(22)-C(19)-C(18)	116.6(26)
C(22)-C(19)-C(20)	113.6(39)	C(21)-C(20)-C(19)	126.3(48)
C(24)-C(23)-O(9)	117.4(37)	C(24)-C(23)-O(10)	100.7(29)
O(10)-C(23)-O(9)	137.3(44)		

(c) Ring torsion angles

C(10)-C(1) -C(2) -C(3)	-143.3(16)
C(2) -C(1) -C(10)-C(9)	69.5(19)
C(1) -C(2) -C(3) -C(4)	52.7(22)
C(2) -C(3) -C(4) -C(5)	53.0(24)
C(3) -C(4) -C(5) -C(6)	-153.1(18)
C(4) -C(5) -C(6) -C(7)	110.7(20)
C(5) -C(6) -C(7) -C(8)	-58.3(21)
C(6) -C(7) -C(8) -C(9)	104.9(17)
C(7) -C(8) -C(9) -C(10)	-149.4(15)
C(8) -C(9) -C(10)-C(1)	61.1(17)

Table 1.4

Comparison of torsion angles for glaucolide-F.

Bond	Glaucolide-A	Dihydrodesacetoxy-glaucolide-A	Glaucolide-F	Bond	Shiromodiol
C(1) -C(2)	-143	-142	-141	C(7) -C(8)	-129
C(2) -C(3)	55	54	53	C(8) -C(9)	53
C(3) -C(4)	55	60	54	C(9) -C(10)	64
C(4) -C(5)	-149	-152	-152	C(10)-C(1)	-166
C(5) -C(6)	102	99	109	C(1) -C(2)	112
C(6) -C(7)	-57	-53	-59	C(2) -C(3)	-49
C(7) -C(8)	103	103	105	C(3) -C(4)	86
C(8) -C(9)	-148	-152	-149	C(4) -C(5)	-151
C(9) -C(10)	61	59	62	C(5) -C(6)	79
C(10)-C(1)	68	68	69	C(6) -C(7)	57

Glaucolide-F

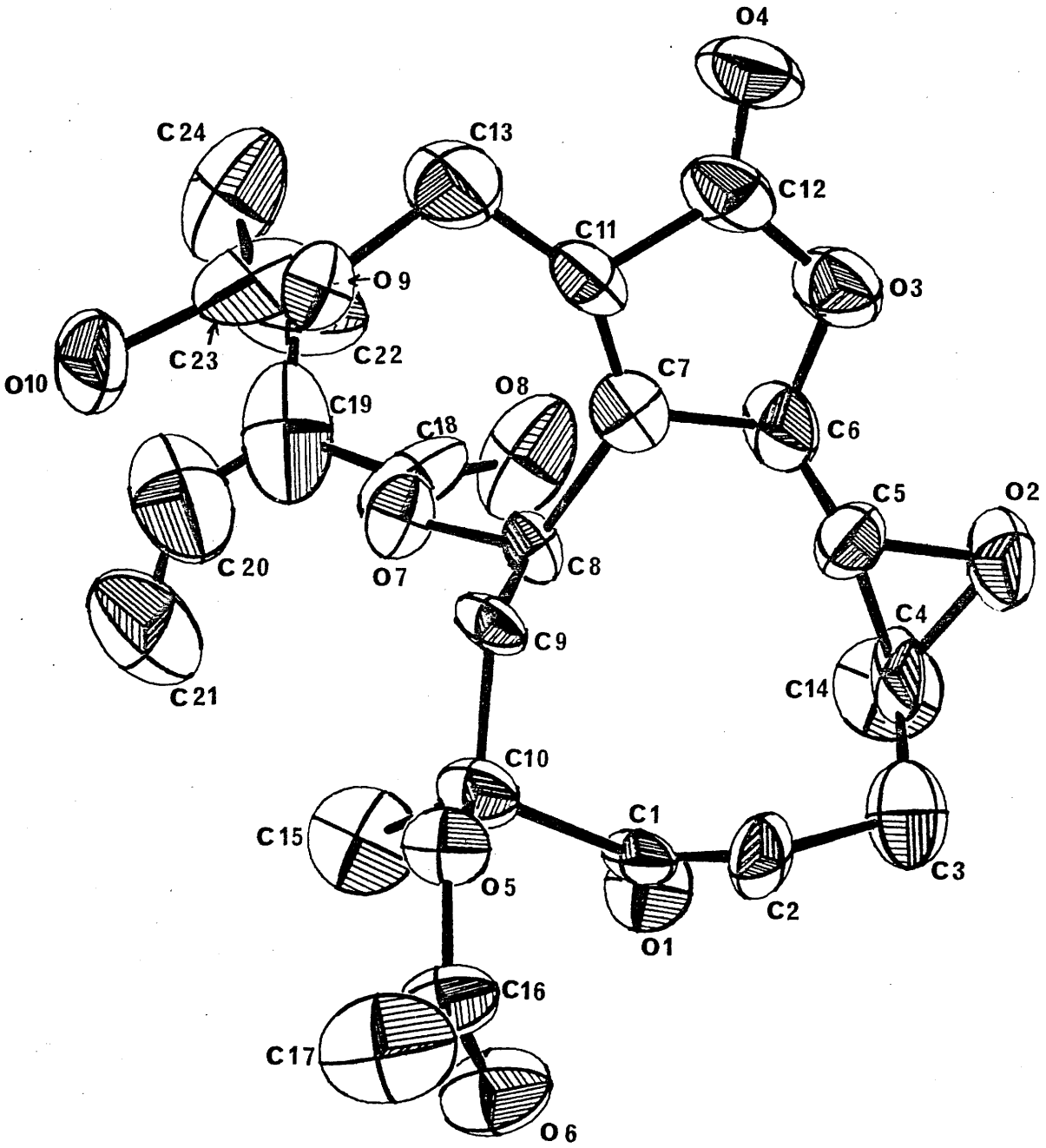


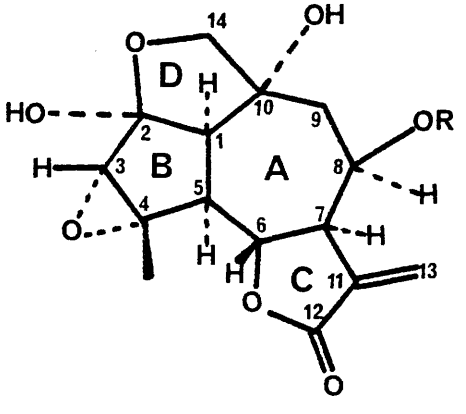
Figure 1.1

II

THE STRUCTURE ELUCIDATION OF A MIXTURE OF TWO NOVEL ISOMERIC SESQUITERPENOIDS

6.2.0 INTRODUCTION

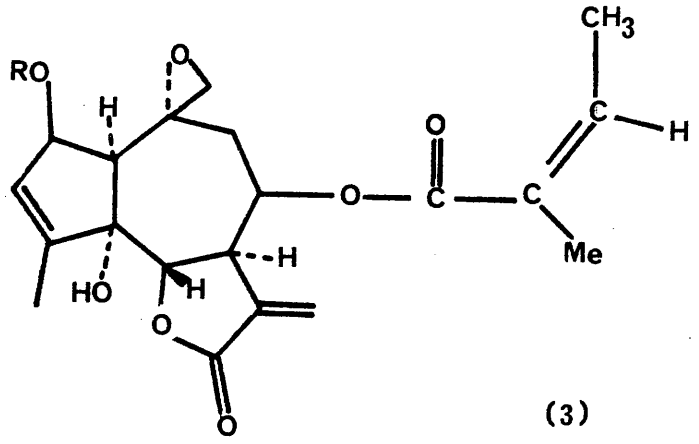
In a search for biologically active natural products, several new metabolites have been isolated from Eupatorium anomalum Nash. Spectroscopic investigations failed to define with certainty the skeletal features of one of the extracts from which a crystalline material of composition $C_{20}H_{26}O_8$ (M.Pt. 152-3°) was derived. X-ray diffraction investigations have established that this extract is a mixture of two novel sesquiterpenoid lactones with isomeric ester side-chains, 2-methylbutanoate (major component,1) and isovalerate (minor component,2), co-crystallising in ca. 2:1 ratio, in the orthorhombic space group $P2_12_12_1$.



1 R=$COCHMeCH_2Me$

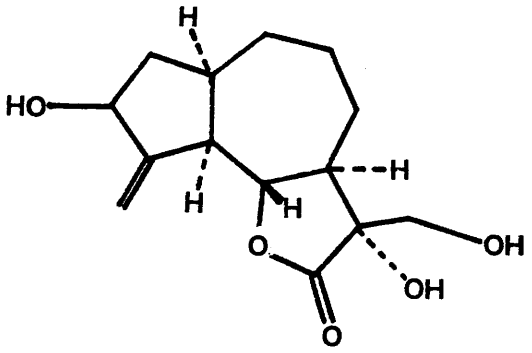
2 R=$COCH_2CHMe_2$

The structure so determined is a highly oxygenated quaianolide and the pattern of cis-1,5 and trans-6,7 ring junctions places it in a category of sesquiterpenoid lactones three other examples of which, Euparotin[1](3), solstitalian[2](4) and bromodihydroisophotosantoni lactone acetate[3](5) have been characterised by X-ray studies.

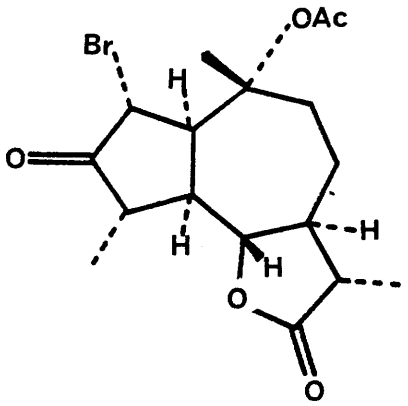


(3)

R = $COCH_2Br$



(4)



(5)

6.2.1 EXPERIMENTAL

Crystal data

Eupatorium, $C_{20}H_{26}O_8.1/2H_2O$, M_r=403.4, orthorhombic, a=6.849(3), b=28.479(7), c=21.659(5) Å, U=4224.6 Å^3, D_c=1.27, D_m=1.26 Mgm^{-3}, Z=8, F(000)=1720, space group $P2_12_12_1$, μ(Mo-K_α)=1.05 cm^{-1}.

Data collection

Instrument used: Hilger Watts Y290
Radiation used: Mo-K_α , λ=0.71069 Å
Upper limit for data collection: $2\theta_{max}=56^\circ$
Number of independent reflexions: m=2785
Unobserved cut-off: $2.5\sigma_I$
Number of independent parameters: n=309
Number of reflexions per parameter: m/n=9.0

6.2.2 STRUCTURE DETERMINATION AND REFINEMENT

Exhaustive trials with both MULTAN and the X-RAY system proved unsuccessful. 6513 quartet invariants were generated for the top 399 E-magnitudes having $|E| \geqslant 1.4$. After several cycles of invariant least-squares using different starting sets of reflexions an E-map was produced which showed two 5-membered rings and a 7-membered ring in one molecule, and a 5-membered and 7-membered ring in another. Sim-weighted Fourier and difference syntheses calculations failed to complete the structure.

Having successfully experimented with the temperature coefficient, B, for glaucolide-F, a similar procedure was adopted here. Table 2.1 summarises the results for the two values of B employed, 2.6 (Wilson plot) and 4.0. An E-map

calculated from 350 E-magnitudes ($|E| \geq 1.42$) using 5973 triplets gave 42 out of a possible 57 atomic positions. The structure was completed by using successive difference Fourier syntheses. However, the unusual spread of electron density, and concomitant high thermal parameters associated with the terminal atoms of the ester side-chain in both molecules of the asymmetric crystal unit, together with the difficulty in assigning some ^{13}C and ^{1}H NMR signals indicated the possible co-existence of isomers or homologues[4].

Structure refinement was completed using SHELX, where, after adjustment of positional and anisotropic thermal parameters, a final R value of 0.067 (R_w=0.067) was attained. Table 2.2 lists the final fractional coordinates and anisotropic thermal parameters. Table 2.3 summarises relevant bond lengths, angles and torsion angles for both molecules, while an ORTEP projection of one of the molecules in the asymmetric unit is shown in Figure 2.1.

6.2.3 DISCUSSION

This X-ray structure analysis has resolved an asymmetric crystal unit comprising two $C_{20}H_{26}O_8$ molecules linked by a water molecule, whereby each sesquiterpenoid moiety has identical conformation (within experimental error). The conformational analysis of the seven-membered rings indicates that they take on a conformation which is approximately midway between the twist-chair (C_2) and chair (C_s) extremes that differ in energy by *ca.* 1-2 kcal mol^{-1}, with the former at energy minima and the latter at energy maxima[5], characterised by the analyses originally outlined by Hendrickson[6] and subsequently by McPhail and Sim[1].

As a result of the publication by McPhail and Sim two parameters, Σ_2 and Σ_s, signifying an axis of symmetry (C_2) and a plane of symmetry (C_s), may be used, as in

Chapter 5, to denote the degree of strain-energy conferred on the sesquiterpene skeleton. Table 2.4 contains the values found for the structure under discussion and compares them to the values of the ideal boat and chair conformations as well as the ideal twist boat and twist chair forms. From the values given, neither of the molecules adopt the minimum energy form.

Molecular mechanics calculations for a series of chair and twist chair conformers[7], carried out by White et al., indicate that the most stable form of an isolated methylenecycloheptane is a twist chair form in which the approximate C_2 axis of the ring passes through the carbon atom adjacent to that bearing the exocyclic methylene group. Since this is not the minimum energy form observed in this particular case (cf. chapliatrins), the question arises as to whether the adoption of this ring conformation can be accounted for by the constraints imposed by the three 5-membered rings, B, C and D. The arguments for such constraints are set out by White et al.[8], and are summarised briefly here.

The torsion angle C(2)-C(1)-C(5)-C(4) in ring B must be ca. 0° (experimentally 9°) and, due to the presence of the cis 3,4-epoxide, the torsion angle C(10)-C(1)-C(5)-C(6) is restricted to ca. 60°. It has previously been established[8] that if the torsion angle C(11)-C(7)-C(6)-O(4) is -28° then the ring torsion angle C(5)-C(6)-C(7)-C(8) must be ca. 90°. The fact that this torsion angle is 90° allows only the C_s conformation to be adopted.

Inspection of the torsion angle geometry of the three 5-membered rings shows rings B and C to have a C_2 chair conformation, the former ring being fairly flat. Ring D has an envelope (C_s) conformation with C(10) being the out of plane atom.

The valency angles in the cycloheptane rings are all generally much greater than tetrahedral, the mean values 118.6 and 118.3° for each ring, being larger than those found in other similar sesquiterpenoid systems where the average values are in the range 114.5-115.9°. This increase in endocyclic valency angles is thought to be due to the increased strain imposed by the addition of the cis-2,10 fused 5-membered ring to the sesquiterpene mainframe.

The two molecules are linked by a water molecule, hydrogen bonded to atoms O(2) and O(2'). Intramolecular bond lengths show these contacts to be 2.852 and 2.764Å respectively.

Table 2.4

	Molecule I	Molecule II	Twist Chair	Twist Boat
Σ_2	75	76	0	215
Σ_s	53	54	141	0

REFERENCES

1. A. T. McPhail and G. A. Sim, Tetrahedron (1973) 29, pp. 1751-1758.
2. W. E. Thiessen and H. Hope, Acta Cryst. (1970) B26, 554.
3. J. D. Asher and G. A. Sim, J. Chem. Soc. 1584 (1965).
4. W. Hertz and P. S. Kalyanaraman, Phytochemistry (1975) 14, 1664.
5. D. N. J. White and M. J. Bovill, Acta Cryst. (1977) B33, 3029.
6. J. B. Hendrickson, Tetrahedron (1963) 19, 1387.
7. M. J. Bovill, M. H. P. Guy, G. A. Sim, D. N. J. White and W. Hertz, J. Chem. Soc. Perkin II, (1979) pp. 53-56.
8. T. A. Dullforce, G. A. Sim, and D. N. J. White, J. Chem. Soc.(B), (1971), 1399.

Table 2.1

(a) Distribution of E^2 with $\sin\theta/\lambda$ as B is increased.

	$\sin\theta/\lambda$									
B	0.0461	0.1008	0.1599	0.2199	0.2804	0.3406	0.4007	0.4624	0.5224	0.5836
2.6	1.7930	2.0660	0.9292	1.1426	0.9621	0.7688	0.9590	1.2220	0.8968	0.9092
4.0	1.0517	1.2512	0.5838	0.7661	0.7002	0.6237	0.8799	1.2919	1.1243	1.3718

(b) The change in starting set reflexions as B is increased.

B	
2.6	2 0 7, 1 2 0, 0 1 1, 1 1 4, 2 9 0
4.0	0 25 5, 5 0 14, 0 13 7, 2 0 7, 1 1 4, 6 11 7, 1 23 1

Table 2.2

Atomic coordinates and thermal parameters for $C_{20}H_{26}O_8.1/2H_2O$

(i) Atomic coordinates ($x10^4$)

ATOM	x/a	y/b	z/c
O(1)	6071(7)	3618(2)	9034(3)
O(2)	6687(8)	3462(2)	10070(3)
O(3)	7795(9)	4458(2)	10183(3)
O(4)	11364(8)	4919(2)	8712(3)
O(5)	13630(10)	5356(2)	8264(3)
O(6)	10463(10)	3830(2)	7578(2)
O(7)	9031(7)	2957(2)	9197(2)
O(8)	12058(32)	3361(8)	6935(9)
C(1)	9305(11)	3765(3)	9384(3)
C(2)	7135(11)	3764(3)	9578(3)
C(3)	6567(12)	4241(3)	9723(4)
C(4)	8094(13)	4570(3)	9534(3)
C(5)	9887(11)	4289(3)	9284(4)
C(6)	10563(11)	4435(3)	8654(3)
C(7)	12311(12)	4149(3)	8409(4)
C(8)	11826(14)	3717(3)	8067(4)
C(9)	11085(12)	3322(3)	8456(4)
C(10)	9294(10)	3395(2)	8870(3)
C(11)	13423(13)	4527(3)	8055(4)
C(12)	12904(13)	4980(3)	8318(4)
C(13)	14697(16)	4475(4)	7585(5)
C(14)	7361(2)	3520(3)	8544(4)
C(15)	7706(18)	5070(3)	9358(5)
C(16)	10829(42)	3723(12)	6973(6)
C(17)	9101(40)	3849(8)	6576(14)
C(18)	9251(49)	4305(9)	6223(14)
C(19)	8052(44)	3450(9)	6249(11)
C(20)	7735(48)	3077(9)	6762(12)

Table 2.2 (continued)

ATOM	x/a	y/b	z/c
O(8a)	12501(44)	3651(12)	6817(12)
C(16a)	10846(61)	3610(16)	7023(12)
C(17a)	8945(52)	3590(12)	6671(20)
C(18a)	8617(52)	4087(12)	6428(16)
C(19a)	10463(54)	4156(14)	5994(20)
C(20a)	6916(52)	4044(14)	5940(19)
O(aq)	7481(8)	2523(2)	10408(3)
O(1')	8769(7)	1846(2)	8452(3)
O(2')	8282(8)	1871(2)	9497(3)
O(3')	7281(8)	0880(2)	9449(2)
O(4')	3547(8)	0547(2)	7962(3)
O(5')	1279(10)	0136(2)	7453(3)
O(6')	4287(10)	1761(2)	7025(3)
O(7')	5724(7)	2473(2)	8752(3)
O(8')	2189(31)	2016(7)	6318(9)
C(1')	5552(10)	1635(2)	8825(3)
C(2')	7831(10)	1627(3)	8963(4)
C(3')	8400(11)	1120(3)	8992(4)
C(4')	6807(12)	0821(3)	8787(3)
C(5')	5012(11)	1112(2)	8638(3)
C(6')	4289(11)	1036(3)	7976(4)
C(7')	2497(12)	1326(3)	7786(4)
C(8')	2898(13)	1806(3)	7524(4)
C(9')	3628(13)	2158(3)	7990(4)
C(10')	5437(10)	2058(3)	8318(3)
C(11')	1434(13)	0988(3)	7378(4)
C(12')	1980(13)	0510(3)	7578(4)
C(13')	0199(18)	1068(4)	6929(5)
C(14')	7373(13)	2016(3)	8009(4)

Table 2.2 (continued)

ATOM	x/a	y/b	z/c
C(15')	7140(16)	0333(3)	8555(5)
C(16')	3794(37)	1909(11)	6444(6)
C(17')	5448(34)	1785(7)	6007(10)
C(18')	5362(44)	1263(8)	5856(12)
C(19')	5211(37)	2131(8)	5466(11)
C(20')	7240(38)	2090(9)	5254(10)
O(8b)	2431(46)	2193(10)	6411(14)
C(16b)	3785(56)	1944(16)	6445(12)
C(17b)	5758(50)	2041(11)	6114(17)
C(18b)	6423(46)	1555(11)	5921(15)
C(19b)	7898(52)	1626(13)	5377(16)
C(20b)	4677(53)	1287(13)	5602(18)

(ii) Thermal parameters ($Å^2 \times 10^3$)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
O(1)	28	67	61	04	-04	00
O(2)	35	52	69	06	13	13
O(3)	54	62	61	08	05	00
O(4)	47	40	66	-05	-03	13
O(5)	61	53	123	-13	08	32
O(6)	86	81	37	-03	07	-03
O(7)	27	34	66	00	-05	10
O(8)	190					
C(1)	27	43	38	00	-02	05
C(2)	28	49	50	01	-01	07
C(3)	28	59	57	08	02	00
C(4)	50	50	45	03	-05	-04
C(5)	31	40	54	-01	-07	02

Table 2.2 (continued)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
C(6)	32	39	48	Ø0	-Ø5	Ø7
C(7)	36	4Ø	54	-Ø2	Ø6	11
C(8)	54	56	57	11	16	12
C(9)	43	44	57	Ø4	13	1Ø
C(1Ø)	24	38	57	Ø4	Ø2	11
C(11)	4Ø	58	79	Ø1	-Ø3	23
C(12)	48	52	75	-Ø1	Ø2	21
C(13)	53	116	74	Ø2	1Ø	3Ø
C(14)	39	49	57	Ø0	-Ø7	Ø2
C(15)	76	49	81	1Ø	1Ø	Ø7
C(16)	19Ø					
C(17)	19Ø					
C(18)	19Ø					
C(19)	19Ø					
C(2Ø)	19Ø					
O(8a)	148					
C(16a)	148					
C(17a)	148					
C(18a)	148					
C(19a)	148					
C(2Øa)	148					
O(aq)	37	56	58	-Ø7	Ø7	Ø0
O(1')	29	57	74	Ø0	Ø7	Ø8
O(2')	43	62	69	12	-19	-25
O(3')	42	62	52	Ø5	-Ø2	Ø8
O(4')	5Ø	41	67	-Ø3	-Ø6	-16
O(5')	67	55	1Ø1	-13	-14	-25
O(6')	75	84	4Ø	-Ø7	-Ø2	Ø8
O(7')	31	37	74	Ø4	-Ø5	-17
O(8')	136					
C(1')	2Ø	41	45	-Ø2	Ø1	-Ø7
C(2')	19	47	56	Ø1	Ø1	-Ø7

Table 2.2 (continued)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
C(3')	30	49	50	10	-01	-03
C(4')	41	42	49	-01	-01	02
C(5')	35	34	44	-01	01	-05
C(6')	35	37	55	02	03	-10
C(7')	32	46	55	-01	-04	-09
C(8')	45	52	61	07	-14	03
C(9')	41	43	58	05	-05	-04
C(10')	20	42	49	00	00	-04
C(11')	39	62	72	-01	-13	-21
C(12')	46	62	54	01	04	-22
C(13')	88	79	94	-12	-34	-09
C(14')	43	54	54	02	07	04
C(15')	64	47	86	12	-15	-14
C(16')	136					
C(17')	136					
C(18')	136					
C(19')	136					
C(20')	136					
O(8b)	180					
C(16b)	180					
C(17b)	180					
C(18b)	180					
C(19b)	180					
C(20b)	180					
Average e.s.d.'s						
O,C	6	5	5	4	4	4

Table 2.3

Interatomic distances(Å) and angles(°) for $C_{20}H_{26}O_8.1/2H_2O$

(i) Bonded distances

		1	2
O(1)	-C(2)	1.446(9)	1.424(9)
O(1)	-C(14)	1.409(10)	1.438(10)
O(2)	-C(2)	1.403(9)	1.384(9)
O(3)	-C(3)	1.443(10)	1.426(10)
O(3)	-C(4)	1.456(9)	1.480(9)
O(4)	-C(6)	1.489(9)	1.483(9)
O(4)	-C(12)	1.368(11)	1.362(10)
O(5)	-C(12)	1.186(10)	1.199(10)
O(6)	-C(8)	1.448(11)	1.446(11)
O(6)	-C(16)	1.369(16)	1.369(17)
O(6)	-C(16a)	1.381(33)	1.403(31)
O(7)	-C(10)	1.446(8)	1.443(9)
O(8)	-C(16)	1.333(39)	1.173(32)
O(8a)	-C(16a)	1.224(50)	1.170(51)
C(1)	-C(2)	1.544(10)	1.589(10)
C(1)	-C(5)	1.560(10)	1.587(10)
C(1)	-C(10)	1.533(10)	1.543(10)
C(2)	-C(3)	1.448(11)	1.497(11)
C(3)	-C(4)	1.463(12)	1.453(11)
C(4)	-C(5)	1.563(11)	1.517(11)
C(4)	-C(15)	1.498(12)	1.495(12)
C(5)	-C(6)	1.500(11)	1.532(11)
C(6)	-C(7)	1.542(11)	1.535(11)
C(7)	-C(8)	1.474(11)	1.505(11)
C(7)	-C(11)	1.525(12)	1.496(12)
C(8)	-C(9)	1.494(12)	1.508(12)
C(9)	-C(10)	1.534(11)	1.528(11)
C(10)	-C(14)	1.542(11)	1.556(11)
C(11)	-C(12)	1.454(12)	1.477(12)
C(11)	-C(13)	1.349(14)	1.309(15)

Table 2.3 (continued)

	1	2
C(16) -C(17)	1.506(38)	1.518(31)
C(17) -C(18)	1.510(36)	1.523(30)
C(17) -C(19)	1.520(37)	1.540(31)
C(19) -C(20)	1.552(36)	1.550(35)
C(16a)-C(17a)	1.510(54)	1.554(50)
C(17a)-C(18a)	1.527(49)	1.512(44)
C(18a)-C(19a)	1.588(53)	1.575(48)
C(18a)-C(20a)	1.578(52)	1.584(48)

(ii) Interbond angles

	1	2
C(14) -O(1) -C(2)	110.8(6)	111.5(6)
C(1) -C(2) -O(1)	105.4(6)	106.9(6)
C(10) -C(14) -O(1)	103.8(6)	104.3(6)
C(3) -C(2) -O(2)	110.6(6)	113.0(6)
C(2) -C(3) -O(3)	113.3(7)	110.6(6)
C(3) -C(4) -O(3)	59.3(5)	58.2(5)
C(15) -C(4) -O(3)	115.4(7)	113.4(7)
C(5) -C(6) -O(4)	107.1(6)	105.2(6)
O(5) -C(12) -O(4)	120.0(8)	121.5(8)
C(11) -C(12) -O(5)	131.3(9)	130.6(8)
C(16a)-O(6) -C(8)	114.4(18)	118.4(17)
C(9) -C(8) -O(6)	111.1(7)	110.0(7)
O(8) -C(16) -O(6)	110.3(18)	121.7(19)
O(8a) -C(16a)-O(6)	116.8(33)	118.4(28)
C(1) -C(10) -O(7)	103.8(6)	106.6(6)
C(14) -C(10) -O(7)	108.5(6)	103.6(6)
C(5) -C(1) -C(2)	106.6(6)	105.3(5)
C(3) -C(2) -C(1)	104.6(6)	106.1(6)
C(4) -C(5) -C(1)	103.9(6)	105.6(6)
C(9) -C(10) -C(1)	120.9(6)	122.2(6)

Table 2.3 (continued)

			1	2
C(4)	-C(3)	-C(2)	110.4(7)	110.9(6)
C(15)	-C(4)	-C(3)	123.6(8)	122.2(7)
C(6)	-C(5)	-C(4)	114.6(6)	112.6(6)
O(2)	-C(2)	-O(1)	109.4(6)	109.2(6)
C(3)	-C(2)	-O(1)	108.1(6)	109.7(6)
C(1)	-C(2)	-O(2)	114.7(6)	111.7(6)
C(4)	-O(3)	-C(3)	60.6(5)	60.0(5)
C(4)	-C(3)	-O(3)	60.1(5)	61.8(5)
C(5)	-C(4)	-O(3)	109.4(6)	108.8(6)
C(12)	-O(4)	-C(6)	110.4(6)	110.8(6)
C(7)	-C(6)	-O(4)	103.4(6)	103.0(6)
C(11)	-C(12)	-O(4)	108.7(7)	107.9(7)
C(16)	-O(6)	-C(8)	122.2(13)	119.8(12)
C(7)	-C(8)	-O(6)	109.1(7)	108.7(6)
C(17)	-C(16)	-O(6)	110.5(22)	108.5(19)
C(17a)	-C(16a)	-O(6)	107.0(30)	105.4(27)
C(9)	-C(10)	-O(7)	105.6(5)	105.5(6)
C(17)	-C(16)	-O(8)	130.2(22)	127.9(17)
C(10)	-C(1)	-C(2)	101.0(6)	100.3(5)
C(10)	-C(1)	-C(5)	123.9(6)	124.2(6)
C(6)	-C(5)	-C(1)	118.1(6)	116.5(6)
C(14)	-C(10)	-C(1)	100.3(6)	102.7(6)
C(5)	-C(4)	-C(3)	109.3(6)	110.7(6)
C(15)	-C(4)	-C(5)	122.6(7)	124.0(7)
C(7)	-C(6)	-C(5)	114.0(6)	115.7(6)
C(8)	-C(7)	-C(6)	116.0(7)	116.3(7)
C(11)	-C(7)	-C(8)	116.7(7)	116.8(7)
C(12)	-C(11)	-C(7)	107.9(7)	107.3(7)
C(10)	-C(9)	-C(8)	120.0(7)	121.1(7)
C(13)	-C(11)	-C(12)	123.4(9)	122.8(9)
C(18)	-C(17)	-C(16)	116.1(24)	109.4(21)
C(19)	-C(17)	-C(18)	116.4(24)	117.2(19)
C(11)	-C(7)	-C(6)	100.8(6)	101.6(6)

Table 2.3 (continued)

	1	2
C(9) -C(8) -C(7)	114.9(7)	114.3(7)
C(13) -C(11) -C(7)	128.7(9)	129.9(8)
C(14) -C(10) -C(9)	116.8(7)	114.7(6)
C(19) -C(17) -C(16)	117.4(23)	104.3(18)
C(20) -C(19) -C(17)	104.1(21)	100.9(19)
C(18a)-C(17a)-C(16a)	105.4(31)	102.9(28)
C(20a)-C(18a)-C(17a)	105.5(28)	109.4(27)
C(17a)-C(16a)-O(8a)	128.2(28)	123.5(36)
C(19a)-C(18a)-C(17a)	101.6(28)	106.2(26)
C(20a)-C(18a)-C(19a)	101.6(28)	102.1(26)

(iii) Ring torsion angles

Ring A	1	2
C(6) -C(5) -C(1) -C(10)	-2.9	-1.5
C(1) -C(5) -C(6) -C(7)	-56.0	-56.6
C(5) -C(6) -C(7) -C(8)	88.5	88.1
C(6) -C(7) -C(8) -C(9)	-72.5	-70.7
C(7) -C(8) -C(9) -C(10)	56.2	55.4
C(8) -C(9) -C(10)-C(1)	-60.5	-60.3
C(9) -C(10)-C(1) -C(5)	51.5	49.6

Ring B	1	2
C(3) -C(2) -C(1) -C(5)	-12.1	-11.9
C(2) -C(1) -C(5) -C(4)	9.1	10.1
C(1) -C(5) -C(4) -C(3)	-3.4	-4.8
C(5) -C(4) -C(3) -C(2)	-4.2	-3.0
C(4) -C(3) -C(2) -C(1)	10.2	9.5

Table 2.3 (continued)

Ring C	1	2
O(4) -C(6) -C(7) -C(11)	-28.6	-29.7
C(6) -C(7) -C(11)-C(12)	24.5	26.6
C(7) -C(11)-C(12)-O(4)	-10.4	-13.0
C(11)-C(12)-O(4) -C(6)	-9.3	-7.3
C(12)-O(4) -C(6) -C(7)	24.8	24.0

Ring D	1	2
C(1) -C(2) -O(1) -C(14)	1.6	2.7
C(2) -O(1) -C(14)-C(10)	24.8	20.7
O(1) -C(14)-C(10)-C(1)	-40.8	-35.9
C(14)-C(10)-C(1) -C(2)	40.4	35.7
C(10)-C(1) -C(2) -O(1)	-27.0	-24.7

Average e.s.d's of torsion angles is 0.8^{o}.

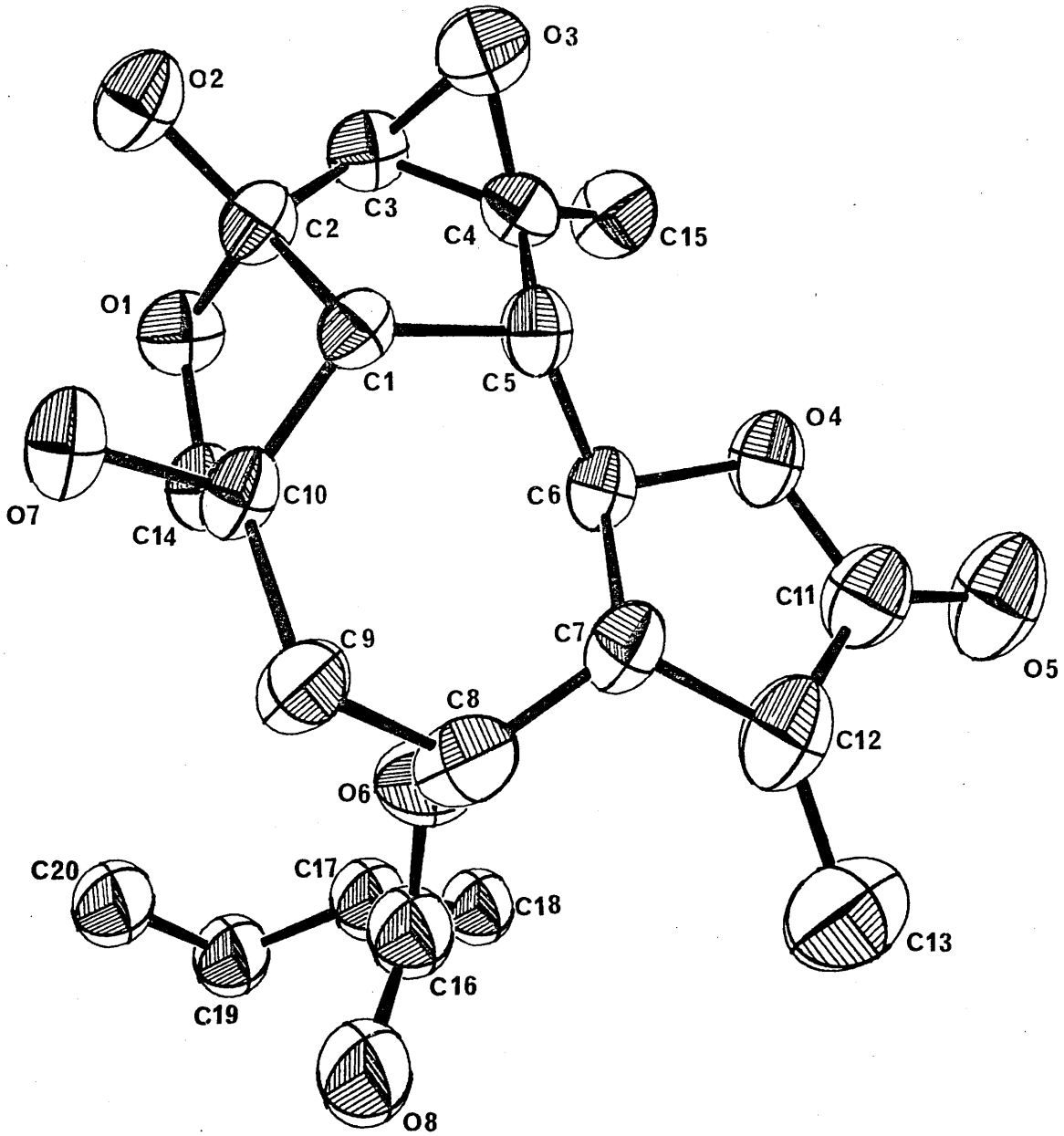

O3
O2
C3
C2
C4
C15
O1
C1
C5
O4
O7
C10
C6
C14
C11
O5
C9
C7
C8
C12
O6
C20
C17
C18
C13
C19
C16
O8

Figure 2.1

III

THE CRYSTAL STRUCTURE DETERMINATION OF OXIDOPANAMENS DIACETATE

6.3.0 INTRODUCTION

In a continuing search for novel plant anticancer agents it was found that ethanol extracts of the Panamanian plant Rondeletia panamensis DC. (Rubiaceae) exhibited cytotoxic activity[1].

Partition of a methanol extract of the wood stem and stem bark of R. panamensis indicated that cytotoxity remained in the chloroform phase. Through extensive chromatography with concomitant bioassay two cytotoxic compounds, oxidopanamensin (II) and panamensin(III), were obtained together with a third, closely related but inactive compound, rondeletin(V).

The similarity of the main features of the proton NMR and UV spectral data of the three isolates suggested that a determination of one of the structures might permit deduction of the remaining structures. Interpretation of the spectral data did not prove fruitful in terms of a unique skeleton and consequently X-ray analysis was carried out. The parent compounds were unsuitable for analysis, but crystals of the diacetate of oxidopanamensin were found to be appropriate, and it was this compound (I) which was therefore analysed.

6.3.1 EXPERIMENTAL

Crystal data

Oxidopanamens diacetate, $C_{24}H_{32}O_6$, M_r=416.5, orthorhombic, a=12.430(2), b=10.969(2), c=16.320(2) Å, U=2225.2 Å^3, D_m=1.25, D_c=1.24 Mgm^{-3}, Z=4, F(000)=896, space group $P2_12_12_1$, μ(Mo-K_α)=0.95 cm^{-1}.

Data collection

Instrument used: Enraf Nonius CAD-4
Radiation used: Mo-K_α , λ=0.7169 Å
Upper limit of data collection: $2\theta_{max}=60^o$
Number of independent reflexions: m=1910
Unobserved cut-off: $3\sigma_I$
Number of independent parameters: n=271
Number of reflexions per parameter: m/n=7.0

6.3.3 STRUCTURE DETERMINATION AND REFINEMENT

The structure was solved by direct methods using MULTAN 76. Several runs of MULTAN using default options and K-curve normalising proved unsuccessful. Closer examination of the statistics from the Wilson plot revealed an uncharacteristic deviation from the calculated least squares line. This deviation only occurred at high $\sin^2\theta$ values (Figure 3.1). NORMAL was subsequently altered to include the user option for limiting the value of $\sin^2\theta$ when accepting F's into the normalisation calculation. A further run of MULTAN with $\sin^2\theta_{max}$=0.39, but otherwise employing default options revealed the complete structure using 300 E-magnitudes having $|E| \geqslant 1.30$.

The approximate atomic coordinates were adjusted by the least-squares program CRYLSQ from the X-RAY 72 suite of

programs; with isotropic thermal parameters converging at R=Ø.131. Subsequently, a difference Fourier synthesis revealed 27 out of the 32 hydrogen atoms in the molecule. These atoms were then incorporated in the refinement procedure with isotropic thermal parameters while the remaining atoms were assigned anisotropic parameters. Refinement converged at R=Ø.Ø58 (R_w=Ø.Ø6Ø).

Atomic coordinates and thermal parameters are listed in Table 3.1 and bond lengths, bond angles, and torsion angles are in Table 3.2. The molecular structure is shown in Figure 3.2.

One possible explanation as to the difficulty encountered by MULTAN when using the complete data set may be that the Phillips high intensity X-ray tube was used during data collection. In general, this tube makes available high ordered reflexions which are not normally available with standard X-ray tubes. The usefulness of these extra reflexions is possibly subject to scrutiny when their structure factors are normalised. Since they occur at very high sinθ values they are subject to small Lorentz polarisation corrections ($1/L_p$). From Section 2.1, Chapter 1, during the calculation of E-magnitudes reflexions having high sinθ values employ the largest temperature corrections, B. The net effect of these two corrections is that their E-magnitudes are very much larger than their real values and any small variation in B can induce a considerable change in the E-magnitudes.

6.3.4 DISCUSSION

The analysis establishes that the compound has composition and relative stereochemistry I.

Ring A has an envelope-like conformation with atom C(5)

at the flap. Ring B has a distorted chair conformation with torsion angles ranging from 41 to 61°; the maximum puckering of the ring is at C(8) and the minimum at C(5). In ring C the departure from ideal chair geometry is much less pronounced, the range of torsion angles being 51-58°, and here the smallest pucker is at C(13).

In the epoxide group, the C-C bond (1.480Å) is slightly longer than the C-O bonds (1.440, 1.446Å) and the C-O-C angle (61.7°) is slightly larger than the O-C-C angles (59.0, 59.4°). Moreover, the C-C(epoxide)-C(epoxide) angles (114.4 and 126.6°) are larger than the corresponding C-C(epoxide)-O(epoxide) angles (111.1 and 116.3°), and the C-C(epoxide)-O(epoxide)-C(epoxide) torsion angles (106.6 and 118.8°) are larger than the C-C(epoxide)-C(epoxide)-O(epoxide) torsion angles (100.9 and 101.6°). This pattern is a feature of other terpenoid epoxides[2,3].

REFERENCES

1. R. I. Geran, N. H. Greenberg, M. M. MacDonald, A.A. Schumacher and B. J. Abbott, Cancer Chemother. Rep. (1972) 3(1), 2.
2. M. E. Cradwick, P. D. Cradwick and G. A. Sim, J. Chem. Soc. Perkin Trans II (1972) 404.
3. M. B. Hossian, D. van der Helm, J. A. Matson and A. J. Weinheimer, Acta Cryst (1979) B35, 660.

Table 3.1

Atomic coordinates and thermal parameters for oxidopanamens

(a) Fractional coordinates ($x10^4$)

ATOM	x/a	y/b	z/c
O(1)	3418(3)	2397(4)	6745(2)
O(2)	1686(3)	1911(5)	6919(2)
O(3)	3119(4)	0859(3)	5436(3)
O(4)	2204(3)	1576(3)	3983(2)
O(5)	1570(3)	5234(3)	3515(2)
O(6)	0710(3)	4345(4)	2466(2)
C(1)	3185(4)	4052(5)	5837(3)
C(2)	3188(4)	2864(5)	5961(3)
C(3)	3029(4)	1955(5)	5320(3)
C(4)	2792(4)	2452(4)	4465(3)
C(5)	2383(4)	3766(4)	4437(3)
C(6)	2360(4)	4254(4)	3540(3)
C(7)	3430(4)	4773(4)	3220(2)
C(8)	3916(3)	5704(4)	3828(2)
C(9)	4153(3)	5098(4)	4660(2)
C(10)	3046(3)	4620(4)	5006(2)
C(11)	4588(4)	6097(5)	5233(3)
C(12)	5600(4)	6730(5)	4893(3)
C(13)	5422(4)	7285(6)	4030(3)
C(14)	6495(6)	7704(9)	3675(4)
C(15)	7294(6)	7343(15)	3545(6)
C(16)	4706(6)	8406(6)	4057(5)
C(17)	3327(5)	1779(5)	3793(3)
C(18)	1209(4)	3701(5)	4750(3)
C(19)	4988(4)	4057(5)	4602(3)
C(20)	4920(4)	6306(5)	3459(3)
C(21)	2589(5)	1860(5)	7165(3)
C(22)	2989(5)	1283(7)	7938(7)
C(23)	0754(4)	5126(5)	2970(3)
C(24)	-0076(5)	6090(6)	3097(4)

Table 3.1 (continued)

(b) Thermal parameters ($Å^2 x 10^4$)

	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
O(1)	630	972	543	-146	-109	322
O(2)	660	1407	661	-146	-038	159
O(3)	1222	567	800	203	121	181
O(4)	882	483	830	-087	-080	-140
O(5)	449	485	561	-004	-176	-054
O(6)	619	1024	535	027	-173	-142
C(1)	579	616	395	-103	037	003
C(2)	552	742	515	-071	-025	130
C(3)	621	502	624	-011	047	106
C(4)	586	440	534	-003	-058	-060
C(5)	427	446	437	-030	-006	-014
C(6)	515	420	436	-027	-108	-055
C(7)	524	542	321	-110	-067	-011
C(8)	410	462	340	-038	-052	-034
C(9)	400	472	320	-024	003	-026
C(10)	411	452	393	-020	012	-022
C(11)	481	722	347	-077	-012	-064
C(12)	491	880	419	-264	-050	-012
C(13)	526	915	429	-291	-053	-050
C(14)	782	2077	569	-693	067	-132
C(15)	606	3174	1128	-009	084	-293
C(16)	1081	545	1026	-226	-232	-019
C(17)	947	518	604	089	030	-084
C(18)	543	523	645	-072	-006	-030
C(19)	478	653	584	164	-009	088
C(20)	537	720	346	-105	-029	011
C(21)	683	643	500	-088	047	055
C(22)	786	926	610	018	085	307
C(23)	456	609	457	-055	-026	083
C(24)	551	775	857	053	-072	072
Average e.s.d.'s						
O,C	28	25	20	26	28	21

Table 3.1 (continued)

(c) Hydrogen atom coordinates ($x10^3$) and isotropic thermal parameters

ATOM	x/a	y/b	z/c	U_{iso}
H(1)	323(4)	457(5)	621(3)	0.045
H(6)	204(3)	372(4)	321(2)	0.031
H(71)	338(4)	523(5)	268(3)	0.068
H(72)	409(5)	405(5)	314(3)	0.098
H(8)	340(4)	632(4)	390(3)	0.052
H(10)	264(4)	535(4)	501(3)	0.034
H(111)	470(5)	581(6)	577(4)	0.133
H(112)	404(3)	665(4)	533(3)	0.061
H(121)	623(5)	599(5)	486(3)	0.082
H(122)	572(6)	750(6)	514(4)	0.107
H(14)	665(7)	844(9)	360(5)	0.164
H(161)	397(5)	817(6)	421(4)	0.119
H(162)	454(5)	860(5)	343(4)	0.094
H(163)	475(6)	937(7)	417(5)	0.218
H(171)	384(6)	079(7)	381(4)	0.064
H(172)	354(4)	217(5)	338(3)	0.039
H(181)	103(4)	456(4)	484(3)	0.102
H(182)	122(4)	321(4)	536(3)	0.069
H(183)	080(5)	326(5)	441(3)	0.088
H(201)	548(4)	561(4)	334(3)	0.066
H(202)	477(3)	679(4)	293(3)	0.051
H(221)	286(9)	163(10)	842(6)	0.130
H(222)	248(7)	077(7)	811(5)	0.209
H(241)	-022(7)	637(8)	248(6)	0.068
H(242)	-057(5)	571(5)	350(3)	0.109
H(243)	013(5)	663(7)	330(4)	0.112

Table 3.2

Interatomic distances(Å) and angles(°) for oxidopanamens

(a) Bonded distances

O(1) -C(2)	1.409(6)	C(5) -C(18)	1.548(7)
O(1) -C(21)	1.370(7)	C(6) -C(7)	1.534(7)
O(2) -C(21)	1.193(7)	C(7) -C(8)	1.537(6)
O(3) -C(3)	1.222(6)	C(8) -C(9)	1.541(6)
O(4) -C(4)	1.440(6)	C(8) -C(20)	1.536(6)
O(4) -C(17)	1.446(8)	C(9) -C(10)	1.576(6)
O(5) -C(6)	1.457(6)	C(9) -C(11)	1.538(6)
O(5) -C(23)	1.354(5)	C(9) -C(19)	1.546(7)
O(6) -C(23)	1.188(6)	C(11)-C(12)	1.541(7)
C(1) -C(2)	1.319(8)	C(12)-C(13)	1.541(6)
C(1) -C(10)	1.502(6)	C(13)-C(14)	1.524(9)
C(2) -C(3)	1.458(7)	C(13)-C(16)	1.518(9)
C(3) -C(4)	1.526(7)	C(13 -C(20)	1.553(7)
C(4) -C(5)	1.546(7)	C(14)-C(15)	1.090(12)
C(4) -C(17)	1.480(7)	C(21)-C(22)	1.497(8)
C(5) -C(6)	1.558(6)	C(23)-C(24)	1.494(8)
C(5) -C(10)	1.556(6)		

Average C-H bond distance: 0.980Å

Table 3.2 (continued)

(b) Interbond angles

C(2) -O(1) -C(21)	117.2(4)	C(9) -C(8) -C(20)	112.1(3)
C(4) -O(4) -C(17)	61.7(3)	C(8) -C(9) -C(10)	107.0(3)
C(6) -O(5) -C(23)	117.3(4)	C(8) -C(9) -C(11)	107.2(4)
C(2) -C(1) -C(10)	123.2(4)	C(8) -C(9) -C(19)	113.1(3)
C(1) -C(2) -O(1)	119.9(5)	C(10)-C(9) -C(11)	109.0(3)
C(3) -C(2) -O(1)	115.5(5)	C(10)-C(9) -C(19)	111.3(4)
C(1) -C(2) -C(3)	124.5(5)	C(11)-C(9) -C(19)	109.1(4)
C(2) -C(3) -O(3)	123.3(5)	C(1) -C(10)-C(5)	110.5(4)
C(4) -C(3) -O(3)	120.7(5)	C(1) -C(10)-C(9)	111.2(3)
C(4) -C(3) -C(2)	115.9(4)	C(5) -C(10)-C(9)	116.7(3)
C(3) -C(4) -C(5)	115.4(4)	C(9) -C(11)-C(12)	112.9(4)
C(3) -C(4) -O(4)	111.1(4)	C(11)-C(12)-C(13)	112.8(4)
C(5) -C(4) -O(4)	116.3(4)	C(12)-C(13)-C(14)	109.8(4)
O(4) -C(4) -C(17)	59.4(4)	C(12)-C(13)-C(16)	112.1(5)
C(3) -C(4) -C(17)	114.4(4)	C(12)-C(13)-C(20)	109.3(5)
C(5) -C(4) -C(17)	126.6(4)	C(14)-C(13)-C(16)	106.3(6)
C(4) -C(5) -C(6)	110.0(4)	C(14)-C(13)-C(20)	109.4(5)
C(4) -C(5) -C(10)	111.9(4)	C(16)-C(13)-C(20)	109.9(5)
C(4) -C(5) -C(18)	105.1(4)	C(13)-C(14)-C(15)	139.6(12)
C(6) -C(5) -C(10)	111.3(4)	O(4) -C(17)-C(4)	59.0(3)
C(6) -C(5) -C(18)	108.0(4)	C(8) -C(20)-C(13)	112.8(4)
C(18)-C(5) -C(10)	109.2(4)	O(1) -C(21)-O(2)	121.3(5)
C(5) -C(6) -O(5)	107.1(3)	O(1) -C(21)-C(22)	115.5(5)
C(7) -C(6) -O(5)	107.5(4)	O(2) -C(21)-C(22)	128.0(5)
C(5) -C(6) -C(7)	114.9(4)	O(5) -C(23)-O(6)	123.5(5)
C(6) -C(7) -C(8)	112.2(3)	O(5) -C(23)-C(24)	111.4(4)
C(7) -C(8) -C(9)	110.4(4)	O(6) -C(23)-C(24)	125.1(5)
C(7) -C(8) -C(20)	110.8(3)		

Table 3.2 (continued)

(c) Torsion angles

Ring A

C(1Ø)-C(1) -C(2) -C(3)	-1.8
C(1) -C(2) -C(3) -C(4)	-3.Ø
C(2) -C(3) -C(4) -C(5)	-19.7
C(3) -C(4) -C(5) -C(1Ø)	45.Ø
C(4) -C(5) -C(1Ø)-C(1)	-47.5

Ring B

C(1Ø)-C(5) -C(6) -C(7)	4Ø.8
C(5) -C(6) -C(7) -C(8)	-5Ø.4
C(6) -C(7) -C(8) -C(9)	61.2
C(7) -C(8) -C(9) -C(1Ø)	-6Ø.7
C(8) -C(9) -C(1Ø)-C(5)	54.3
C(6) -C(5) -C(1Ø)-C(9)	-44.2

Ring C

C(2Ø)-C(8) -C(9) -C(11)	58.3
C(8) -C(9) -C(11)-C(12)	-57.7
C(9) -C(11)-C(12)-C(13)	56.2
C(11)-C(12)-C(13)-C(2Ø)	-5Ø.6
C(12)-C(13)-C(2Ø)-C(8)	51.7
C(9) -C(8) -C(2Ø)-C(8)	-57.9

Epoxide

C(3) -C(4) -O(4) -C(17)	-1Ø6.6
C(3) -C(4) -C(17)-O(4)	1Ø0.9
C(5) -C(4) -O(4) -C(17)	118.8
C(5) -C(4) -C(17)-O(4)	-1Ø1.6

The standard deviations of the torsion angles are ca. Ø.5°.

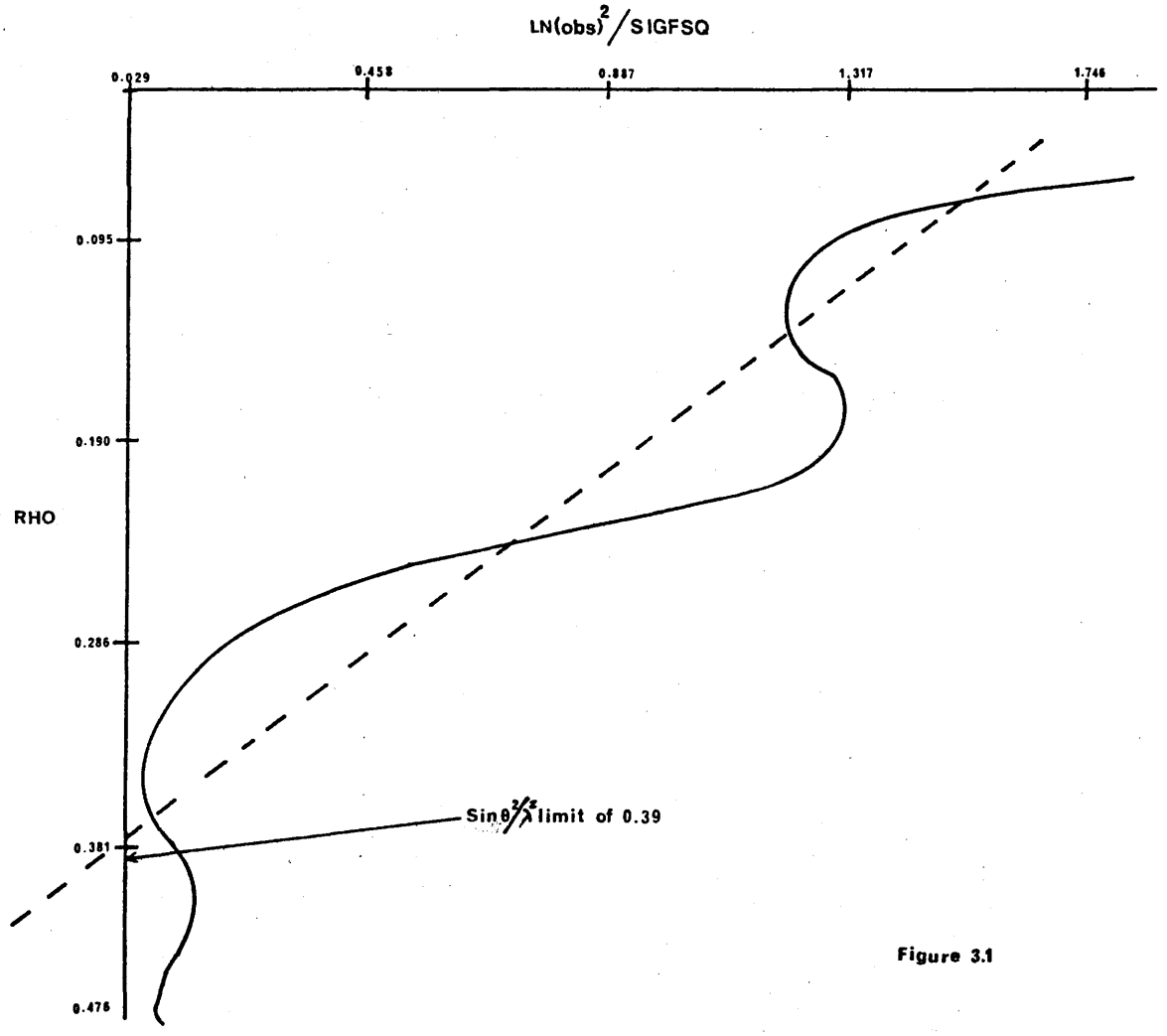

LN(obs)^2/SIGFSQ
0.029
0.458
0.887
1.317
1.746
0.095
0.190
0.286
0.381
0.476
RHO
Sinθ/λ limit of 0.39

Figure 3.1

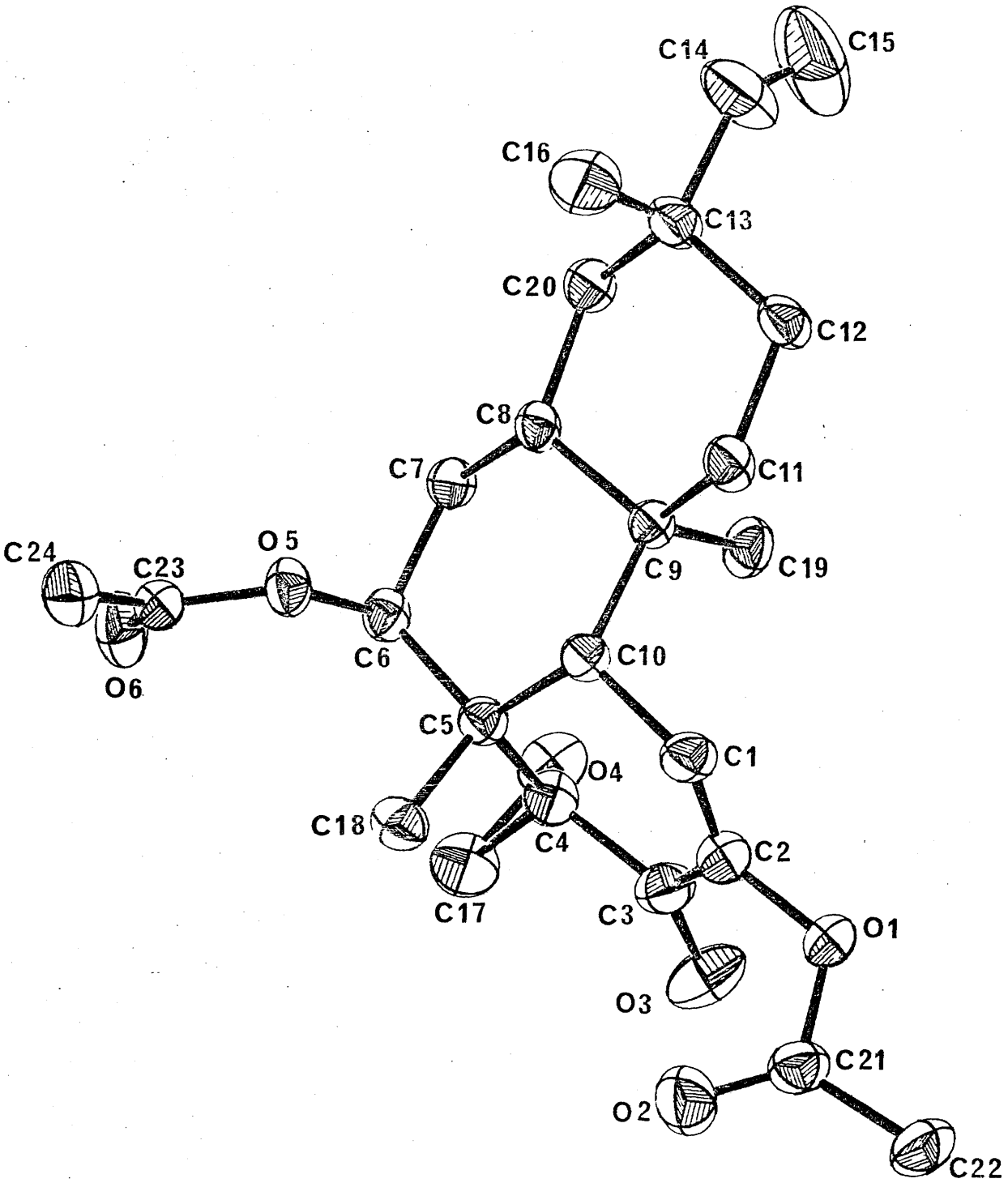

C14
C15
C16
C13
C20
C12
C8
C7
C11
C9
C19
C24
C23
O5
C6
C10
O6
C5
C1
O4
C18
C4
C2
C17
C3
O1
O3
C21
O2
C22

Figure 3.2

CONCLUSIONS

This thesis has shown that the use of higher invariants in both active and passive roles in the MULTAN program provides a useful alternative when standard runs of the package using triplets alone are unsuccessful. In this respect the technique can be compared with the alternatives MAGIC and YZARC suggested by Declercq, Germain and Woolfson[1]. Perhaps the most unexpected result was the effect of varying the temperature coefficient, B, and subsequent success in the structure determination of the two natural products in Chapter 6.

In themselves none of these developments is a panacea when difficulties are encountered in direct methods, but they offer different methods of attacking otherwise intractable problems with good prospects of success. The use of negative quartets in particular never seems to degrade the performance of MULTAN and often enhances it considerably, especially in symmorphic space groups. Positive quartets pose more difficult problems because of correlation with triplets, and whereas the weighting scheme proposed here (Chapter 4) is successful it probably does not fully exploit the phase information contained in these invariants, nor does it solve the difficulties inherent in the correlations between the quartets themselves. These are problems of a theoretical nature that need to be resolved.

A logical extension to these applications of four- and five-phase invariants is to include them in both an active and passive way in the magic integer- Ψ-map program MAGIC and the random phase set/linear equation system RANDOM where they offer the potential of further enhancing these techniques.

1. J. P. Declercq, G. Germain and M. M. Woolfson, Acta Cryst. (1979) A35, pp. 622-626.

$$P_{13}^{\pm} = \frac{1}{M} Z_{13}^{\pm}$$

$$Z_{13}^{\pm} = e^{\mp B} . \cosh R_{23}(R_2R_3 \pm R_1R_4).\cosh R_{31}(R_3R_1 \pm R_2R_4) \times$$

$$\left[\exp\left(\frac{-2}{N}(R_1R_2 \pm R_3R_4)R_5R_6\right).\cosh R_{12}(R_1R_2 \pm R_3R_4 + R_5R_6)\times\right.$$

$$\cosh R_{15}(R_1R_5 + R_2R_6).\cosh R_{25}(R_2R_5 + R_1R_6) \times$$
$$\cosh R_{35}(R_3R_5 \pm R_4R_6).\cosh R_{45}(R_4R_5 \pm R_3R_6) +$$
$$\exp\left(\frac{2}{N}(R_1R_2 \pm R_3R_4)R_5R_6\right).\cosh R_{12}(R_1R_2 \pm R_3R_4 - R_5R_6)\times$$
$$\cosh R_{15}(R_1R_5 - R_2R_6).\cosh R_{25}(R_2R_5 - R_1R_6) \times$$
$$\left.\cosh R_{35}(R_3R_5 \mp R_4R_6).\cosh R_{45}(R_4R_5 \mp R_3R_6)\right]$$

$$M = Z_{13}^{+} + Z_{13}^{-}$$

APPENDIX II

$$P_{2/13} \simeq \frac{1}{K}\exp\{-2B_{1234}\cos\phi_{34} - 2B_{1256}\cos\phi_{56} - 2B_{3456}\cos(\phi_{34} - \phi_{56})\}$$

$$\times I_o\left(\frac{2\sigma_3 R_{12}X_{12}}{\sigma_2^{3/2}}\right) I_o\left(\frac{2\sigma_3 R_{23}X_{23}}{\sigma_2^{3/2}}\right)$$

$$\times I_o\left(\frac{2\sigma_3 R_{31}X_{31}}{\sigma_2^{3/2}}\right) I_o\left(\frac{2\sigma_3 R_{15}X_{15}}{\sigma_2^{3/2}}\right)$$

$$\times I_o\left(\frac{2\sigma_3 R_{25}X_{25}}{\sigma_2^{3/2}}\right) I_o\left(\frac{2\sigma_3 R_{3\bar{5}}X_{3\bar{5}}}{\sigma_2^{3/2}}\right)$$

$$\times I_o\left(\frac{2\sigma_3 R_{4\bar{5}}X_{4\bar{5}}}{\sigma_2^{3/2}}\right),$$

where

$$B_{1234} = \frac{3\sigma_3^2 - \sigma_2\sigma_4}{\sigma_2^3} R_1R_2R_3R_4$$

$$B_{1256} = \frac{3\sigma_3^2 - \sigma_2\sigma_4}{\sigma_2^3} R_1R_2R_5R_6$$

$$B_{3456} = \frac{3\sigma_3^2 - \sigma_2\sigma_4}{\sigma_2^3} R_3R_4R_5R_6$$

and

$$X_{12} = [R_1^2R_2^2 + R_3^2R_4^2 + R_5^2R_6^2 + 2R_1R_2R_3R_4\cos\phi_{34}$$

$$+ 2R_1R_2R_5R_6 + 2R_3R_4R_5R_6\cos(\phi_{34} - \phi_{56}]^{1/2}$$

$$X_{23} = [R_2^2R_3^2 + R_1^2R_4^2 + 2R_1R_2R_3R_4\cos\phi_{34})]^{1/2}$$

$$X_{31} = [R_3^2R_1^2 + R_2^2R_4^2 + 2R_1R_2R_3R_4\cos\phi_{56}]^{1/2}$$

$$X_{15} = [R_1^2R_5^2 + R_2^2R_6^2 + 2R_1R_2R_5R_6\cos\phi_{56}]^{1/2}$$

$$X_{25} = [R_2^2R_5^2 + R_1^2R_6^2 + 2R_1R_2R_5R_6\cos\phi_{56}]^{1/2}$$

$$X_{3\bar{5}} = [R_3^2R_5^2 + R_4^2R_6^2 + 2R_3R_4R_5R_6\cos(\phi_{34} - \phi_{56})]^{1/2}$$

$$X_{4\bar{5}} = [R_4^2R_5^2 + R_3^2R_6^2 + 2R_3R_4R_5R_6\cos(\phi_{34} - \phi_{56}]^{1/2}$$

and K is a suitable normalising parameter independent of ϕ_{34} and ϕ_{56} and not relevant for the present purpose.

The $P_{1/13}$ formula is obtained from $P_{2/13}$ by integrating with respect to ϕ_{56} from $\emptyset$ to 2π :

$$P_{1/13} = \int_0^{2\pi} P_{2/13}\, d\phi_{56}$$

Appendix III

$$P^{\pm}_{1|15}=\frac{1}{K}Z^{\pm}$$

where

$$K=Z^{+}+Z^{-}$$

and

$$Z^{\pm}=\exp(\pm T)\sum_{\eta_{12},\ldots,\eta_{45}=\pm 1}^{1024}\exp(U\pm V),$$

$$T=\frac{1}{\sigma_2^{9/2}}(15\sigma_3^3-10\sigma_2\sigma_3\sigma_4+\sigma_2^2\sigma_5)R_1R_2R_3R_4R_5,$$

$$\begin{aligned}U=\frac{\sigma_3}{\sigma_2^{3/2}}&(\eta_{12}R_1R_2R_{12}+\eta_{13}R_1R_3R_{13}+\eta_{14}R_1R_4R_{14}\\&+\eta_{15}R_1R_5R_{15}+\eta_{23}R_2R_3R_{23}+\eta_{24}R_2R_4R_{24}\\&+\eta_{25}R_2R_5R_{25}+\eta_{34}R_3R_4R_{34}+\eta_{35}R_3R_5R_{35}\\&+\eta_{45}R_4R_5R_{45}),\end{aligned}$$

$$\begin{aligned}V=\frac{\sigma_3}{\sigma_2^{3/2}}&[(\eta_{23}\eta_{45}R_{23}R_{45}+\eta_{24}\eta_{35}R_{24}R_{35}\\&+\eta_{25}\eta_{34}R_{25}R_{34})R_1+\eta_{13}\eta_{45}R_{13}R_{45}\\&+\eta_{14}\eta_{35}R_{14}R_{35}+\eta_{15}\eta_{34}R_{15}R_{34})R_2\\&+(\eta_{12}\eta_{45}R_{12}R_{45}+\eta_{14}\eta_{25}R_{14}R_{25}\\&+\eta_{15}\eta_{24}R_{15}R_{24})R_3+(\eta_{12}\eta_{35}R_{12}R_{35}\\&+\eta_{13}\eta_{25}R_{13}R_{25}+\eta_{15}\eta_{23}R_{15}R_{23})R_4\\&+(\eta_{12}\eta_{34}R_{12}R_{34}+\eta_{13}\eta_{24}R_{13}R_{24}\\&+\eta_{14}\eta_{23}R_{14}R_{23})R_5]\\&-\left(\frac{3\sigma_3^2-\sigma_2\sigma_4}{\sigma_2^3}\right)(\eta_{45}R_{45}R_1R_2R_3+\eta_{35}R_{35}R_1R_2R_4\\&+\eta_{34}R_{34}R_1R_2R_5+\eta_{25}R_{25}R_1R_3R_4\\&+\eta_{24}R_{24}R_1R_3R_5+\eta_{23}R_{23}R_1R_4R_5\\&+\eta_{15}R_{15}R_2R_3R_4+\eta_{14}R_{14}R_2R_3R_5\\&+\eta_{13}R_{13}R_2R_4R_5+\eta_{12}R_{12}R_3R_4R_5),\end{aligned}$$