# THE DIRECT DETERMINATION OF THE MOLECULAR STRUCTURES OF NATURAL PRODUCTS

A. McL. Mathieson
*C.S.I.R.O. Chemical Research Laboratories, Melbourne, Australia*

## INTRODUCTION

When I came to look again at the title of my talk, it appeared to me that it had not quite succeeded in defining the scope of my subject adequately. Let me say, therefore, that my purpose is to show that, given the empirical formula of a suitable derivative of a natural product in a satisfactory crystalline form, it is possible to arrive at the three-dimensional molecular structure in a direct manner and with only minimum assumptions, these being of a general type.

The task of preparing this lecture was influenced, although not made easier, by the increase in the numbers of moderately complex molecules which have been subjected to X-ray analysis in the last few years. It appears probable that this increased interest (particularly in natural products) is not a passing phase, but a consistent change in attitude which may lead, by its results, to a reconsideration of the rôle of X-ray analysis in this field. Prospects exist for the growth of specialist groups to determine structures of natural products by the application of this technique.

There is little need to dwell on the reasons for determining molecular structure—such information is basic to any understanding of how molecules can be chemically modified with intention and how they may be interrelated biogenetically. Although two-dimensional structural formulae have served well in the past, it is clear from the extensive discussion of the influence of conformation and the increased interest in absolute configuration that a fairly detailed three-dimensional picture of the molecule is now required. This being so, the determination of molecular structure in such detail is the essential first stage in the study of a new compound and one which must be solved in as direct and rapid a way as is consistent with accuracy. A solution by classical methods through chemical degradation, and then confirmation by synthesis, offers an attractive and exciting task for many organic chemists; but even for moderately complex compounds, the number of man-years involved in the first step, *i.e.* solution of the structure, is becoming burdensome, particularly for those compounds with inter-linked ring systems—*e.g.* the recent description of the chemical constitution of limonin which involved 14 authors[1]. In this field, X-ray structure analysis may occupy a unique rôle since, with certain conditions satisfied, it can be of general application. It is one in which the latent power of computers may be very effectively placed at the service of organic chemistry.

I want to illustrate, by selected examples, how X-ray analysis has been of help in this aspect of organic chemistry, with particular reference to natural products[2–5], and how the co-ordination of facilities has greatly increased the prospect of this assistance in the future. To do so, we need first to deal with certain of the features which render the X-ray method peculiarly suited to this task.

The physical methods of determining molecular structure fall into two general types, and it is useful to appreciate the clear distinction between them. The first group deals with " fragmentary " information, *i.e.* it gives information primarily about " bits " of the molecule (groups or radicals). Of such techniques, the main representative is that which measures changes in energy between the states of a molecule—ranging from ultra-violet through to microwave absorption. In many respects, these techniques resemble chemical degradative methods which also yield " fragments ", recognition of which permit these " fragments " to be fitted together to derive the molecular structure. Perhaps because of this similarity, the organic chemist has been quick to appreciate and accept these techniques and make wide application of them. The second type of physical method is the diffraction technique which, by contrast with the first group, does not offer partial information—it represents an all-or-none approach to molecular structure. Wide use has not been made of this technique in the past but, as a result mainly of the availability of large electronic computers, this attitude may be changing. For this reason now is an appropriate time to indicate the scope of the technique.

## X-RAY ANALYSIS

### Outline of the X-ray method

With regard to this aspect of the all-or-none approach, we may consider the diffraction of X-rays by a single crystal, taken over all orientations. Experimentally, we end up with an array on films of spots of varying intensity, each spot being related to a set of parallel planes in the crystal referred to by their Miller indices, *hkl*, and its intensity being related to an amplitude value, $|F|$. So we have an extensive list of measurements, $|F(hkl)|$. These arise from the periodic smooth distribution of electron density within the crystal, $\rho(xyz)$, where $x$, $y$, and $z$ are fractions of the repeat distances of the unit cell $a.bxc$. There are two relationships, (1) and (2), between these quantities.

$$\rho(XYZ) = \frac{1}{V}\sum h \sum_{-\infty}^{+\infty} k \sum l\, F(hkl) \exp\left\{-2\pi i(hX + kY + lZ)\right\} \qquad (1)$$

$$F(hkl) = V_c \iiint \rho(xyz) \exp\left\{2\pi i(hx + ky + lz)\right\} \mathrm{d}x\mathrm{d}y\mathrm{d}z \qquad (2)$$

$$\equiv \sum_r f_r(hkl) \exp\left\{2\pi i(hx_r + ky_r + lz_r)\right\} \qquad (2a)$$

It is not our intention to explain here in any detail the two distributions, $\rho(xyz)$ and $F(hkl)$, except to point out that their relationship is a mutually

dependent or reciprocal one. Thus, in a strict sense, to find out about one structure factor, $F(hkl)$, we need to know all about $\rho(xyz)$, *i.e.* the contents of the unit cell, with reasonable precision. Conversely, to determine the value of the electron density $\rho$ at any given point $x,y,z$ in the unit cell, we need all the $F(hkl)$ values. It is evident, therefore, in the approach imposed by this technique, that the available experimental information—the $|F|$ values—cannot be treated piecemeal. We must determine all atom positions to be completely satisfied with, and certain of, our end-result. Fortunately, the X-ray method does not demand a one-step deduction of the completely correct structure, but is amenable to a gradual approach to the true solution by iterative processes.

So far, then, the procedure appears straightforward. It is a matter of collecting the $|F|$ data and combining them according to equation (1). However, this conceals the fact that each $F$ can be a complex quantity, $|F|\cos\alpha + i\,|F|\sin\alpha$, and two parameters are necessary to define $F$, namely the modulus, $|F|$, and the phase angle, $\alpha$. We can measure $|F|$, but, in general, cannot determine $\alpha$ experimentally for a single crystal using one radiation. Should the crystal contain a centre of symmetry, then $\alpha$ reduces to 0 or $\pi$, *i.e.* $F$ may be $+F(\cos 0°)$ or $-F(\cos\pi)$. If, however, the structure is asymmetric, as is extremely probable for a natural product, then $\alpha$ may range from 0 to $2\pi$, and it is obvious that the situation becomes complex. The possibility of a practicable solution of the crystal structure is then closely bound up with the size of the molecule and the number of molecules constituting the asymmetric unit, *i.e.* the total number of atomic parameters that must be determined. As to the various proposals for the solution of the general phase problem, *i.e.* for any arbitrarily selected crystal, I must refer you to more extensive treatments[6], but here, if we set ourselves the more limited task of determining certain organic structures with some prospect of success, we must narrow our field to consideration of the " heavy atom " method as a means of yielding a first approximation to the phase angle. This does not mean that other methods cannot be used to solve such structures, *e.g.* hydroxydihydroeremophilone[7], but to use them must involve a certain amount of chemical information, and, further, they are not yet adequately developed to deal with complex asymmetric structures. The heavy-atom method offers the possibility of determining a structure with the minimum of assumptions and the minimum of chemical knowledge, and, because of these conditions, parts of the process can be made automatic. In principle, the restrictive conditions of the heavy-atom method mean that we cannot tackle any crystal offered, but must select a suitable derivative which contains one or more (but not too many) atoms of atomic number higher than those constituting the main bulk of the molecule, *i.e.* C, N and O. For the small sacrifice of the added complexity of the heavy atom(s), we gain a more certain starting point in analysis.

## The relative weight of the heavy atom

In considering the general strategy for the structure determination of a compound of natural origin, the question arises as to what is the most suitable weight of the heavy atom with respect to the remainder of the

molecule. Choice is frequently restricted by chemical considerations or the crystallizability of derivatives, but some numerical estimate is useful. There are several factors involved. The value of $|F|$ is formed of a vector sum of the atomic contributions (see equation (2) ) which, for convenience, we may consider in two parts—that due to the $N$ heavy atoms $NA_H$, and that due to the $n$ light atoms $nA_L$ in which we are mainly interested since they constitute the molecular skeleton(s). The use of a very heavy atom will, by itself, give a very close approximation to the " true " phase angle, but due to the error inherent in the measurement of $|F_0|$ (5–10 per cent), we will not gain very exact information on the contribution of $A_L$. On the other hand, if $A_H$ is not heavy enough, then there may arise uncertainties during the analysis which may retard progress—in the extreme, the analysis may not be able to start since it may not be possible to locate $A_H$ with certainty in the first stage. A useful measure for selection in compounds of moderate complexity is given by $\overset{N}{\Sigma} Z_H^2 / \overset{n}{\Sigma} Z_L^2$, where $Z_H$ and $Z_L$ refer to the atomic number of the " heavy " and " light " atom types respectively. Values for typical analyses[4] cover a wide range with a mean value near 1·0. Values above 2·0 have been used, but, for the vitamin $B_{12}$—SeCN derivative[8], a value of 0·46 (based on Co + Se) sufficed, and an even lower value of 0·22 (for S) in tosyl-prolyl-hydroxyproline[9] permitted a successful analysis. Given the choice, it is preferable to veer on the low side, 0·6–0·8. The heaviness index does not provide a complete answer, since the heavy-atom technique depends also on the change in distribution of the magnitudes of $|F|$ with increasing number of atoms (see *Figure 2* in a review by the author[4]). This factor becomes of greater importance as the molecular size increases, and its significance is clear in the case of haemoglobin[10] and myoglobin[11] heavy-atom derivatives for which the heaviness index is extremely small. However, for the range of molecular size being considered here, the index is a useful guide*.

## Experimental conditions

A successful analysis is not dependent only on the choice of a suitable derivative as indicated by the heaviness index, but is intimately tied up with the range and quality of the diffraction data. This aspect need be mentioned only insofar as it influences the analysis both as to speed of solution of the structure and the accuracy of the refinement. It is useful to revert to equations (1) and (2) to note a second aspect of their reciprocal relationship, *i.e.* between the shapes of atoms in real space (in terms of the electron density, $\rho$) and the distribution of the diffraction data, $F(hkl)$. For an isolated atom, the distribution in diffraction space is compared with the shape of an isolated atom in real space in *Figure 1*. The broader the peak in diffraction space which corresponds to more extensive diffraction data, the sharper is the atom peak in real space, and hence it may be more readily and accurately located. Conversely, where the range of data is limited, the peak in real space is very diffuse. For adjacent atoms, the ability to site them accurately depends on the ability to differentiate them into separate peaks (see *Figure 1* (*a*) (*iii*) and (*b*) (*iii*) ). For most organic

* More elaborate discussions of the influence of the heavy atom in determining phase (or signs) have been given by Luzzati[12], Woolfson[13] and Sim[14].

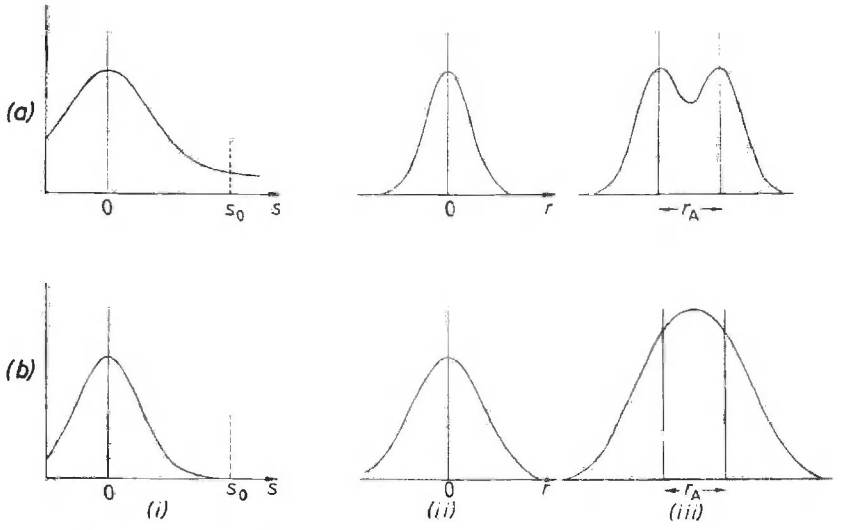


*Figure 1.* Comparison of the influence of (*a*) small and (*b*) considerable thermal motion on the distribution in (*i*) diffraction space and (*ii*) real space for an isolated atom. (*iii*) compares the corresponding distributions in real space of two atoms of the same shape as in (*ii*) separated by $r_A$. The experimental limit of recorded diffraction data is $s_0$

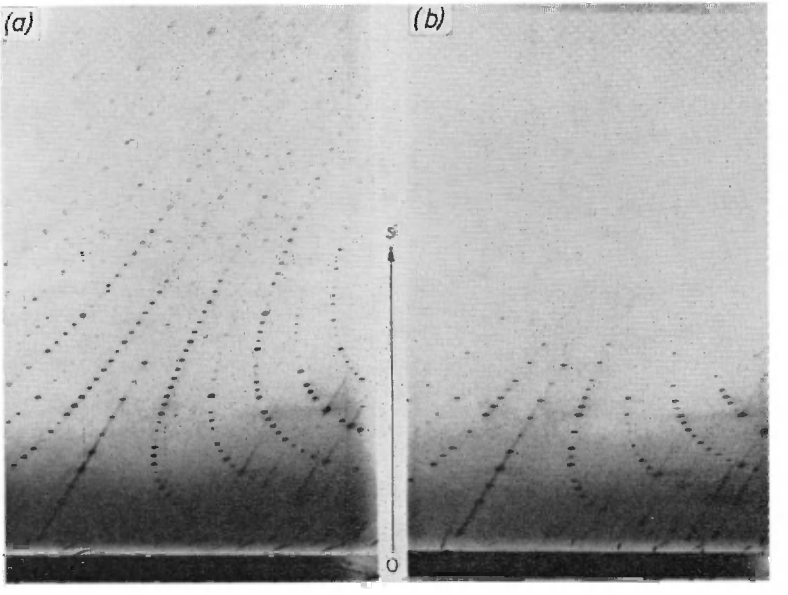


*Figure 2.* Comparison of a zero-layer Weissenberg photograph of lanostenyl iodoacetate (*a*) at − 160°C and (*b*) at 25°

compounds, the molecules and the atoms thereof are subject to considerable thermal vibration at room temperature, and the situation is more in accord with *Figure 1* (*b*), the diffraction data falling off very rapidly with increasing angle. The possibility exists of changing this situation experimentally to the more advantageous case, *Figure 1* (*a*), by lowering the temperature of the crystal, thus reducing thermal vibration of the atoms and extending the range of diffraction data. Temperatures of the order of −150 to −180°C are readily obtained by the use of liquid air[15], while lower values are feasible if liquid hydrogen or helium is available[16]. The improvement which is possible is shown dramatically in *Figure 2*.

The second aspect of the experimental data is the magnitude of the $|F_0|$ terms. For a given exposure, only a certain proportion of reflections will be recorded, all others lying below the background value or limit of visibility. To ensure that the maximum data are recorded so that amplitude termination will not influence results, the exposure must be the maximum practicable—an important factor for complex molecules if the analysis is being carried out with three-dimensional data. For a normal X-ray generator, it is not uncommon with crystals of natural product origin for each layer, of which there may be 5 to 20, to be exposed for 200–250 h, and thus the collection of the data becomes a major time-component of the analysis. To overcome this, the time factor can be reduced by the use of a high-power X-ray generator[17].

Because of the importance not only of the maximum amount of data but also of its full population over the whole observational range, a combination of a low-temperature adaptor with a high-power generator provides a considerable improvement in the range of data and the speed of collection. Thus, for the analysis of himbacine hydrobromide[18], it has been found possible to record the selected 6 layers in 5 days, each exposure being from 5 to 8 h and the percentage population recorded being 90 per cent.

So far, photographic techniques of collecting intensity data have been implied. The accuracy of measurement is of the order of 5–10 per cent and, for greater accuracy, it would be necessary to use counter techniques of measuring intensities. For complex molecules with large numbers of $|F|$ values, *e.g.* 2927 for vitamin $B_{12}$ (wet)[19], 3351 for the hexacarboxylic acid from vitamin $B_{12}$[20] and 3400 for jacobine bromohydrin[21], it would require to be an automatic or semi-automatic process. Several designs have been proposed[22], and it is very probable that with increased need there will be an increased effort to make this aspect of structure analysis automatic.

## Space groups—$P2_1$ and $P2_12_12_1$

Before considering the methods used to extract structural information from the diffraction data, it is informative for this later purpose to mention the three-dimensional spatial arrangements which molecules adopt when persuaded to crystallize. There are 230 spatial arrangements (space groups) which are theoretically possible, but natural products crystallize in only a very limited number of these. Since the molecules of natural products are generally asymmetric and, in addition, have little internal symmetry, it is found in practice that the majority crystallize in one or other of only two

space groups of the 138 theoretically available asymmetric space groups—either the monoclinic $P2_1$ or the orthorhombic $P2_12_12_1$ (see *Table 1*). The many analyses of amino-acids, peptides, sugars, purines and pyrimidines have been omitted from the list in *Table 1*. This decision is arbitrary but appears valid for the author's purpose in that, for the majority of such compounds, their molecular formulae were not in doubt, details of conformation and configuration being the principal aim of the analyses.

Since the tactical approach to the solution of the structure is influenced by the space group in which it crystallizes, and since these two space groups dominate the field, we shall draw attention to certain of their features and mention conditions imposed by their detailed symmetry. The relationships of the symmetry elements and asymmetric units are shown in *Figure 3*.

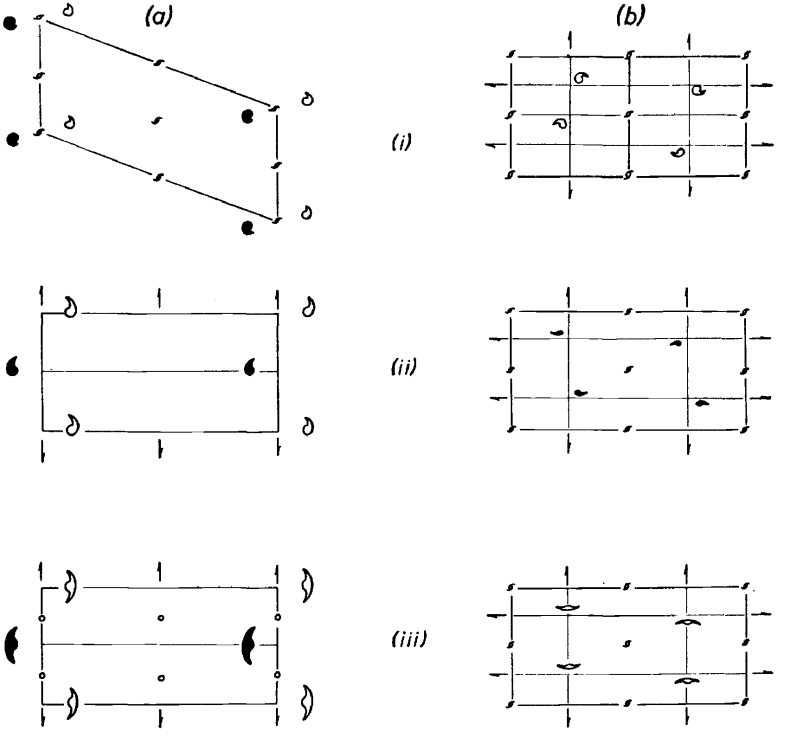


*Figure 3.* The two space groups " most popular " with crystals of natural products and their derivatives: (*a*) The monoclinic space group $P2_1$: (*i*) as viewed down the *b* symmetry axis; (*ii*) as viewed down the *a* axis; (*iii*) a symmetrized projection of (*ii*) due to phasing with a heavy atom. (*b*) The orthorhombic space group $P2_12_12_1$: (*i*) as viewed down the *a* axis; (*ii*) as viewed down the *b* axis; (*iii*) a symmetrized projection of (*ii*) due to phasing with a heavy atom in a special position

The symbols used for symmetry elements are standard (see *International Tables for X-Ray Crystallography*, Vol. 1 (1952) ). For $P2_1$, the situation is simple—deceptively so from a structural aspect. There is only one type of symmetry element and only one set of these—two-fold screw axes parallel to the *b* axis. There are only two asymmetric units in the unit cell, *i.e.* given one unit or molecule then the other unit or molecule can be described as

*Table 1.* A list of the majority of natural product derivatives for which a detailed structure analysis is available, grouped according to the space group in which they crystallize

| *Compound* | *Formula* | *No of Atoms* | *Cell dimensions* | *Space Group* |
|---|---|---|---|---|
| α-Bromocamphor[74] | $C_{10}H_{15}OBr$ | 11 | 7·38, 7·57, 9·12; $\beta=94°$ | $P2_1$ |
| Ephedrine hydrochloride[75] | $C_{10}H_{15}ON.HCl$ | 12 | 12·65, 6·09, 7·32; $\beta=102°\ 15'$ | $P2_1$ |
| Bromodilactone from jacobine[54] | $C_{10}H_{13}O_4Br$ | 14 | 9·11, 6·42, 10·84; $\beta=98·2°$ | $P2_1$ |
| Isoclovene hydrochloride[76] | $C_{15}H_{24}.HCl$ | 15 | 6·35, 13·91, 7·89; $\beta=95·3°$ | $P2_1$ |
| Ergine hydrobromide[77] | $C_{16}H_{17}ON_3.HBr$ | 20 | 9·60, 5·74, 13·79; $\beta=91·2°$ | $P2_1$ |
| α-Methodone hydrobromide[78] | $C_{21}H_{27}ON.HCl$ | 23 | 10·69, 8·74, 10·74; $\beta=94·6°$ | $P2_1$ |
| Thelepogine methiodide[30] | $C_{21}H_{34}ONI$ | 23 | 11·12, 13·04, 7·46; $\beta=109·5°$ | $P2_1$ |
| α-Methyldihydrothebaine methiodide[79] | $C_{21}H_{28}O_3NI$ | 25 | 11·06, 8·75, 10·54; $\beta=94°\ 32'$ | $P2_1$ |
| Calcium thymidylate[80] | $C_{10}H_{13}O_8N_2Ca.6H_2O$ | 26 | 14·40, 6·87, 9·81; $\beta=90·6°$ | $P2_1$ |
| 7-Bromocholesteryl bromide[81] | $C_{27}H_{44}Br_2$ | 27 | 12·05, 8·75, 12·57; $\beta=101°\ 19'$ | $P2_1$ |
| Cholesteryl iodide[25] | $C_{27}H_{45}I$ | 27 | 12·57, 9·04, 21·89; $\beta=149°$ | $P2_1$ |
| Des-(oxymethylene)-lycoctonine hydroiodide hydrate[66] | $C_{24}H_{39}O_6N.HI.H_2O$ | 34 | 12·79, 9·44, 11·00; $\beta=97·75°$ | $P2_1$ |
| *p*-Bromobenzoate diester of iresin[82] | $C_{29}H_{28}O_6Br_2$ | 35 | 28·75, 7·41, 6·40; $\beta=92°$ | $P2_1$ |
| Phyllochlorine ester[83] | $C_{33}H_{38}N_4O_2$ | 39 | 12·23, 14·84, 7·87; $\beta=94°\ 59'$ | $P2_1$ |
| Lumisterol-4-iodo-3-nitrobenzoate[84] | $C_{35}H_{46}O_4NI$ | 40 | 10·55, 7·63, 20·57; $\beta=93·5°$ | $P2_1$ |
| Jacobine bromohydrin ethanolate[21] | $(C_{18}H_{26}O_6NBr)_2.C_2H_5OH$ | 53 | 10·11, 15·53, 14·90; $\beta=113°$ | $P2_1$ |
| Epilimonol iodoacetate[26] | $(C_{26}H_{31}O_3.COCH_2I)_2$ | 74 | 15·03, 12·36, 15·93; $\beta=95·2°$ | $P2_1$ |
| Quinine sulphate dihydrate[85] | $C_{20}H_{24}O_2N_2.H_2SO_4.2H_2O$ | 30 | 15·49, 6·74, 20·46; $\beta=114°$ | C2 |
| Strychnine sulphate pentahydrate[86] | $C_{21}H_{22}O_2N_2.H_2SO_4.5H_2O$ | 34 | 35·85, 7·56, 7·87; $\beta=107°\ 21'$ | C2 |
| Kojic acid[87] | $C_5H_6O_4$ | 9 | 3·85, 18·4, 8·55; $\beta=96°$ | $P2_{1/n}$ |
| Sodium tropolonate[88] | $C_7H_5O_2Na$ | 9 | 13·91, 3·69, 11·67; $\beta=93·1°$ | $P2_{1/c}$ |
| Tropine hydrobromide[89] | $C_8H_{16}ONBr$ | 10 | 6·32, 20·2, 7·44; $\beta=95°$ | $P2_{1/n}$ |
| Nootkatin copper complex[90] | $C_{30}H_{38}O_4Cu$ | 17 | 8·40, 11·96, 15·21; $\beta=115·5°$ | $P2_{1/a}$ |
| Cryptopleurine methiodide[55] | $C_{25}H_{30}O_3NI$ | 29 | 9·95, 24·2, 9·95; $\beta=112·0°$ | $P2_{1/n}$ |

| | | | | |
|---|---|---|---|---|
| $\psi$-Conhydrine hydrobromide[91] | $C_8H_{18}ONBr$ | 10 | 15·15, 9·28, 7·72 | $P2_12_12_1$ |
| Muscarine iodide[92] | $C_9H_{20}ONI$ | 11 | 6·81, 11·75, 15·40 | $P2_12_12_1$ |
| $\beta$-Biotin (vitamin H)[93] | $C_{10}H_{16}O_3N_2S$ | 15 | 5·17, 10·33, 21·00 | $P2_12_12_1$ |
| Hydroxydihydroeremophilone[7] | $C_{15}H_{24}O_2$ | 15 | 7·5, 10·0, 19·5 | $P2_12_12_1$ |
| $\beta$-Caryophyllene bromide[94] | $C_{15}H_{25}Br$ | 15 | 8·35, 10·75, 16·41 | $P2_12_12_1$ |
| Longifolene bromide[95] | $C_{15}H_{25}Br$ | 15 | 8·51, 9·78, 16·67 | $P2_12_12_1$ |
| Allokainic acid[51] | $C_{10}H_{15}O_4N$ | 15 | 26·08, 5·83, 6·95 | $P2_12_12_1$ |
| Kainic acid (Zn salt)[51] | $C_{10}H_{13}O_4NZn.2H_2O$ | 17 | 25·50, 6·93, 7·60 | $P2_12_12_1$ |
| Glutathione[96] | $C_{10}H_{17}O_6N_3S$ | 19 | 28·06, 8·80, 5·58 | $P2_12_12_1$ |
| Annotinine bromohydrin[39] | $C_{16}H_{22}O_3NBr$ | 20 | 11·89, 13·40, 9·68 | $P2_12_12_1$ |
| Benzyl penicillin (Rb salt)[53] | $C_{14}H_{17}O_4N_2SRb$ | 21 | 9·45, 6·44, 30·2 | $P2_12_12_1$ |
| Erythraline hydrobromide[41] | $C_{18}H_{19}O_3NBr$ | 22 | 13·36, 10·46, 11·67 | $P2_12_12_1$ |
| Morphine hydroiodide dihydrate[97] | $C_{17}H_{19}O_3N.HI.2H_2O$ | 23 | 20·32, 13·02, 6·84 | $P2_12_12_1$ |
| Gelsemine hydroiodide[24] | $C_{20}H_{22}O_2N_2.HI$ | 24 | 7·80, 9·17, 26·90 | $P2_12_12_1$ |
| Codeine hydrobromide dihydrate[98] | $C_{18}H_{21}O_3N.HBr.2H_2O$ | 24 | 13·10, 20·86, 6·82 | $P2_12_12_1$ |
| Himbacine hydrobromide hydrate[18] | $C_{22}H_{36}O_2NBr.H_2O$ | 26 | 12·67, 26·32, 6·68 | $P2_12_12_1$ |
| Aspidospermine-$N$(b) methiodide[48] | $C_{23}H_{33}O_2N_2I$ | 27 | 24·3, 8·5, 11·1 | $P2_12_12_1$ |
| Strychnine hydrobromide hydrate[38] | $C_{21}H_{22}O_2N_2.HBr.2H_2O$ | 27 | 7·64, 7·70, 32·2 | $P2_12_12_1$ |
| Calycanthine hydrobromide hydrate[50] | $C_{22}H_{28}N_4Br_2.2H_2O$ | 28 | 9·61, 14·13, 16·97 | $P2_12_12_1$ |
| Colchicine methylene bromide[99] | $C_{22}H_{25}O_6N.CH_2Br_2$ | 30 | 9·1, 14·87, 18·54 | $P2_12_12_1$ |
| Aureomycin hydrochloride[49] | $C_{22}H_{23}O_8N_2Cl.HCl$ | 32 | 11·20, 12·89, 15·47 | $P2_12_12_1$ |
| Lanostenyl iodoacetate[72] | $C_{32}H_{53}O_2I$ | 34 | 7·6, 10·9, 38·6 | $P2_12_12_1$ |
| Demethanol aconinone hydroiodide hydrate[40] | $C_{24}H_{35}O_8N.HI.3H_2O$ | 36 | 14·69, 22·08, 8·33 | $P2_12_12_1$ |
| Methyl oleanolate iodoacetate[100] | $C_{33}H_{51}O_4I$ | 37 | 7·7, 16·11, 25·14 | $P2_12_12_1$ |
| Calciferol-4-iodo-3-nitrobenzoate[46] | $C_{35}H_{36}O_4NI$ | 40 | 6·98, 15·3, 31·2 | $P2_12_12_1$ |
| Vitamin $B_{12}$[19] | | | | |
| Vitamin $B_{12}$ (wet) | $C_{63}H_{84}N_{14}O_{14}PCo.nH_2O$ | 105 | 25·33, 22·32, 15·92 | $P2_12_12_1$ |
| Vitamin $B_{12}$ (dry) | $C_{63}H_{84}N_{14}O_{14}PCo.nH_2O$ | 103 | 24·35, 21·29, 16·02 | $P2_12_12_1$ |
| Vitamin $B_{12}$—SeCN | $C_{63}H_{84}N_{14}O_{14}PCoSeCN.nH_2O$ | 99 | 23·98, 21·46, 16·02 | $P2_12_12_1$ |
| Vitamin $B_{12}$ hexacarboxylic acid | $C_{49}H_{64}O_{14}N_6CoCl.2H_2O$ | 73 | 24·58, 15·52, 13·32 | $P2_12_12_1$ |
| Purpurogallin[101] | $C_{11}H_8O_5$ | 16 | 19·56, 24·6, 3·62 | Pb2a |
| Bromoamphenicol[45] | $C_{11}H_{12}O_5N_2Br_2$ | 18 | 7·4, 17·9, 22·1 | $C222_1$ |
| $\alpha$-Isosparteine hydrate[102] | $C_{15}H_{26}N_2.H_2O$ | 18 | 20·18, 10·61, 6·84 | $C222_1$ |

reproduced by the symmetry element. In $P2_12_12_1$, there is a set of two-fold screw axes parallel to each of the three axes—an asymmetric unit is reproduced four times by the operation of the symmetry elements.

For $P2_12_12_1$ with a single heavy atom in the asymmetric unit, the heavy atoms can contribute to define the phase angle of almost all reflections with, however, the presumably rare possibility that they may lie in special positions, *e.g.* at a quarter position, *Figure 3* (*b*) (*iii*), where its effectiveness as a phase-fixing agent is compromised, as in the case of gelsemine hydroiodide[24]. On the other hand, for $P2_1$, if there is only one equivalent heavy atom, then it and its symmetry-related atom can be regarded as disposed around a centre of symmetry. We have noted earlier that, where a centre of symmetry is involved, the phase angles defined by the heavy atoms degenerate to either 0° or $\pi$. Hence, in this case, so far as the heavy atoms are concerned, the crystal structure is centrosymmetric and, if the heavy atom is used to initiate the analysis, it will impose this symmetry upon the diffraction data and thus upon the apparent structure. This situation is depicted in *Figure 3* (*a*) (*iii*) in contrast with the real situation in *Figure 3* (*a*) (*ii*). The situation may be described to a first, but necessarily crude, approximation by imputing half atoms on either side of the introduced reflection plane. For this situation, we therefore face a somewhat more intractable problem—the sites of the $A_L$ atoms (if we recognize them) are now of half-weight except on the introduced planes of symmetry. Of the $2n$ sites, assuming that all are located sufficiently accurately and none are spurious, we have to distinguish which atoms are related to which to form a molecule. At this point, we must recognize that for this space group $P2_1$ with one heavy asymmetric atom we require to apply tests of chemical sensibility in terms of bond lengths and angles in the recognition of groups of stereochemical significance. The first complete analysis of a natural product—cholesteryl iodide by Carlisle and Crowfoot[25]—provides an excellent example of this dilemma. This situation does not apply to space group $P2_12_12_1$, at least in theory, although any chemical generalizations are useful guides at all stages of an analysis. However, for space group $P2_1$ they are a necessity. For such symmetrized structures, it is necessary to tilt the balance towards asymmetry in one direction or another by a conscious choice, *e.g.* in gelsemine hydroiodide[24], acceptance of the existence of the indole ring decided the balance in one direction and the remainder of the molecular skeleton was defined by gradually working outwards by iterative application of Fourier syntheses. Comparing the two space groups for an asymmetric unit of one molecule, the amount of diffraction data per molecule is the same, and hence, under comparable circumstances, it is more desirable to carry out an analysis with a structure crystallizing in $P2_12_12_1$, since ambiguities are less likely to arise. For this reason also, the comparative numbers of structures in $P2_1$ and $P2_12_12_1$ which have been analysed (see *Table 1*), may have no statistical significance in terms of their distribution in Nature, but rather are representative of the relative difficulties of analyses with the respective space groups.

For $P2_1$, the ambiguity arising from the existence of one heavy atom in the asymmetric unit may theoretically be eliminated by introducing a second heavy atom. This situation was found to exist in the case of the iodoacetate of epiliminol[26], which crystallized with two molecules in the

asymmetric unit, the two iodine atoms being at different $y$ levels. However, such a situation may well lead to further complication as was found for jacobine bromohydrin[21]. Here also there were two molecules in the asymmetric unit, but the situation was degenerate since both Br atoms were on the same $y$ level.

## Vector distribution

With the experimental data, one more step may be carried through without assumptions other than those involved in the well-defined operations of the space group. Because the crystal is periodic in three dimensions, we may convert our data in diffraction space into an equivalent distribution in real space by means of a Fourier synthesis whose terms are $|F|^2$. Since $|F|^2 = F.F^*$, this calculation (3) required no information regarding phase:

$$P(XYZ) = \frac{1}{V_c} \cdot \sum h \sum_{-\infty}^{+\infty} k \sum l \left|F(hkl)\right|^2 \cos 2\pi(hX + kY + lZ) \qquad (3)$$

Patterson[23], who introduced this function in 1935, showed that it gives not atomic positions but the array of vectors between every pair of atoms, all the vectors starting from the origin of the unit cell. Thus, if there are $N$ atoms in the unit cell, there are $N^2 - N$ vectors in the distribution $P$. If $N$ is 100 (say), then there are 9900 vectors. However, this complication is unavoidable in transferring to real space without information regarding phase angles. At the least, the vector distribution is equivalent to the data in diffraction space, and is more readily comprehensible in terms of bond lengths and angles and the general stereochemistry of molecules.

## Methods of analysis

We may now consider the steps by which the complete molecular structure can be extracted from the diffraction data.

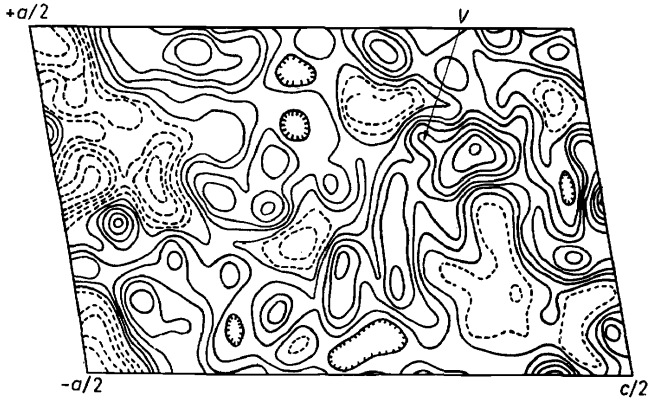


*Figure 4.* A contoured vector map showing the location of the vector peak, *V*, between two S atoms. The vector peak is not immediately obvious and internal evidence was required to reveal its significance

The first step is to determine the peaks corresponding to the $A_H$–$A_H$ vectors in the vector distribution. These may be immediately obvious; but, if the $A_H$ atoms are of inadequate weight, then tests involving internal consistency of the vector distribution may be required to distinguish the true $A_H$–$A_H$ vector peak from peaks of comparable size, due to coincidence of several minor vectors, *e.g.* in tosyl-prolyl-hydroxyproline[9] (*Figure 4*) and in cephalosporin C[27]. Confirmation of the correct selection is particularly necessary where there are several atoms of only moderate weight; *e.g.* in the carbon tetrachloride adduct of sporidesmin[28] there are seven $A_H$ atoms to locate, two S and five Cl. However, we shall assume that the $A_H$–$A_H$ vectors can be identified and that, from their location, the sites of the $A_H$ atoms in the unit cell can be deduced.

This is the basic information to be used to locate the lighter atoms, so that they may be grouped to form the molecular skeleton(s) and the atoms differentiated into C, N and O.

If the analysis is started from an empirical formula, we must have no preconceived ideas regarding the molecular shape. In the early stages, it will not be possible even to tell which regions in the unit cell enclose the molecule(s), particularly if there is more than one molecule in the asymmetric unit. It is important not to superimpose one's ideas upon the data, since it is found in practice that it is extremely easy with asymmetric structures to persuade the data in the direction we wish it to go. This arises from the flexibility of definition of the phase angles and the large numbers of parameters involved in this type of analysis. This has been especially noted in the analysis of the vitamin $B_{12}$ group of compounds by Hodgkin and her co-workers[19, 20, 29] and we have had similar experience with jacobine bromohydrin[21] and thelepogine methiodide[30].

For this reason, it is important to handle the diffraction data so that the derived information should have maximum structural significance, a minimum of spurious detail, and lead to the correct solution in a minimum number of stages. If one deviates from the right track, it will be possible to correct the error because of the inherent properties of the Fourier method, but this may require time-consuming cycles of calculation to resolve the dilemma. Since the data initially in the form of a vector distribution will be eventually unravelled, and the molecular structure presented more clearly in the form of the electron density distribution, the various approaches to the process of structure analysis by this method may be differentiated by the point at which the transfer from the vector map to the direct $\rho$ map occurs. As we have noted previously, the vector distribution contains all the structure information with the minimum of assumption, and accordingly it appears advisable to use this function to extract as much structural information as possible before transferring to the electron density representation.

Although the vector distribution was proposed by Patterson in 1935[23], a systematic method of extracting information was not developed until 1950, when Buerger[31] presented his image-seeking approach based on an earlier suggestion by Wrinch[32]. Other approaches of a similar type were proposed almost simultaneously by Beevers and Robertson[33], Clastre and Gay[34] and Garrido[35]. These processes consist of placing in the vector distribution a search-polygon which may represent a part or whole of a molecule, com-

bining the values at the vertices of the polygon (*i.e.* at atom centres) and transferring the combined values to form one value in a new derived distribution. The polygon (or molecule) is moved over all possible positions, keeping the orientation fixed, and the complete new distribution is derived. There is a variety of methods of combining the values, but, in practice, since we are dealing with a complex situation, the individual operations must be uncomplicated and capable of rapid exploitation. The only ones which have proved satisfactory in use are the simple summation and the minimum function. We can see that if the search-polygon coincides with a number of high peaks in the vector map, a high peak is reproduced in the new distribution and this result will be structurally significant. Theoretically, one can start with only two atoms as search points and, as new atom sites are revealed, these are added to the search-polygon until finally the whole molecule is involved and the derived function shows only what was put in—with no spurious features. In practice, as is usual, many problems obtrude upon this idealized scheme, some of which are dealt with in Buerger's book *Vector Space*[36].

However, all operations are carried out on the vector function, and so the image-seeking approach is in accord with our premise. The process of

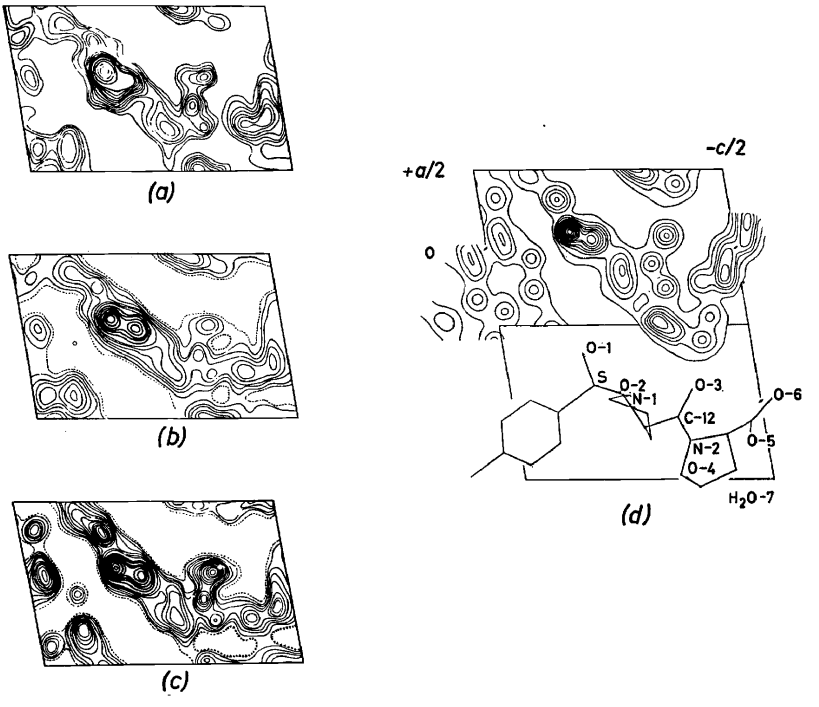


*Figure 5.* Image-seeking by calculation of Fourier syntheses using terms $|F_0|^2 \times \phi_c$. $\phi_c$ corresponds in: (*a*) to S; (*b*) to S + (N–1 + O–2); (*c*) to S + (N–1 + O–2) + the tolyl ring + O–1 + O–3 + C–12. The improvement with increasing number of atoms in the search-polygon is shown by comparison of (*a*), (*b*) and (*c*) with (*d*), the final electron-density distribution. The distribution (*c*) is calculated on the basis of knowledge of the location of approximately half the total number of atoms in the molecule

image-seeking may be carried out by inspection, somewhat more exactly by hand calculation and most expediently by an electronic computer. Alternatively, the combination of the vector distribution by summation may be replaced exactly and often more conveniently by a Fourier synthesis with terms $|F(hkl)|^2 \times \phi_c$ [37], where $\phi_c$ is the structure factor of the search-polygon. An illustration of this latter process is given in *Figure 5*, which shows the influence of increasing the number of atoms in the search-polygon. The final electron-density map is shown for comparison.

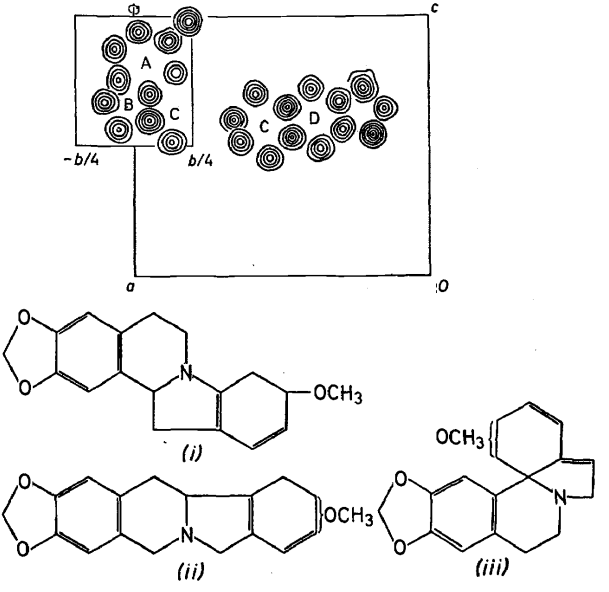


*Figure 6.* Part of the final three-dimensional electron-density distribution of erythraline hydrobromide. The group of atoms on the left forming rings A and B must be visualized as rotated 90° to join with the atomic group on the right through ring C. (*i*), (*ii*) and (*iii*) were possible structural formulae on the basis of chemical evidence

With regard to natural products, this method was first developed and applied in the analysis of strychnine hydrobromide by Beevers and Robertson[38] in 1950. Subsequently, Przybylska has made use of image-seeking by inspection in the analysis of annotinine bromohydrin[39] and demethanol aconinone hydroiodide[40]. The best example, using the $F^2 \times \phi_c$ method, has been by Nowacki and Bonsma in their analysis of erythraline hydrobromide[41]. Here the structure was derived almost completely from the first image-seeking operation. Part of the final result is shown in *Figure 6*, the three-dimensional array of atoms being in accord with structural proposals of Prelog and co-workers. The minimum image-seeking function has been less widely used in this field of natural products, a notable exception being the analysis of cellobiose[42] by Lipscomb and co-workers.

Before a systematic treatment of the vector distribution had evolved, the usual method of analysis was to discard the Patterson function as soon as the heavy atoms had been located, and to use the direct representation of

the electron density. At first, the representation would only approximate crudely to the " true " distribution, since phase angles were defined by the heavy atoms alone; but as single atoms or groups of atoms were recognized, their contribution to the phase angle was incorporated, and the cycle of structure factor and electron-density calculation repeated until the conditions for a satisfactory analysis were achieved, *i.e.* all peaks structurally significant used in (1) and (2)*. For this approach, the classical work

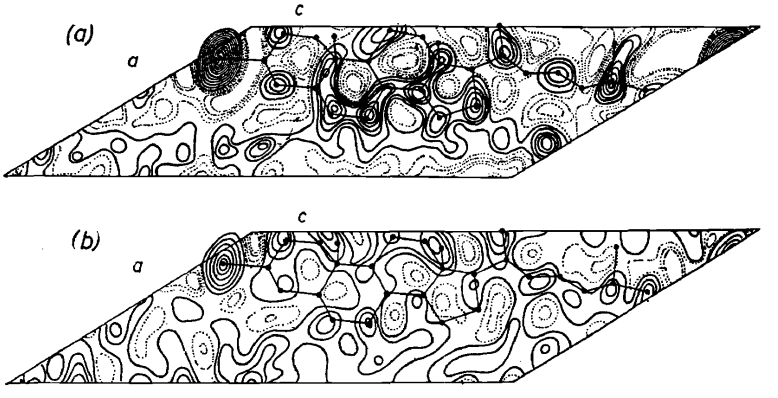


*Figure 7.* Cholesteryl iodide B. Two sections of the three-dimensional distribution of electron-density (*a*) at $y = 0{\cdot}25$, and (*b*) at $y = 0{\cdot}33$

which established the value of the heavy atom method as yielding the molecular structure without assumptions was the analysis of Pt phthalocyanine by Robertson and Woodward[44], while the first significant application to natural products was the analysis of cholesteryl iodide by Carlisle and Crowfoot[25] (see *Figure 7*). A striking example of the application of this method was given by Dunitz in 1952. Despite rather difficult experimental conditions—decomposition of the crystal to a gum under X-radiation—an analysis of bromoamphenicol[45] was completed. In this, Dunitz succeeded in showing that the first three-dimensional electron density distribution, based on the Br atoms, yielded the three-dimensional atomic array of the molecular skeleton (see *Figure 8*).

The most extensive application of the use of the direct distribution† is in the analyses of the group of vitamin $B_{12}$ compounds—wet and dry vitamin $B_{12}$, vitamin $B_{12}$—SeCN and the derived hexacarboxylic acid by Hodgkin and her co-workers[19]. Perusal of the work already reported[19, 20, 29], describing an outline of the investigation of vitamin $B_{12}$ and the crystal structure of the hexacarboxylic acid (see *Figure 9*), will give a greater appreciation of the process of approximation to the complete and correct

* To speed up the derivation of the complete structure, Sim[43] has proposed the weighting of individual $|F|$ terms by a function representing the probability of the phase angle being correct. This scheme was used in the analysis of epiliminol iodoacetate[26].

† By single-derivative single-crystal techniques, to differentiate from the multi-derivative single-crystal methods of Perutz and co-workers and Kendrew and co-workers in their analyses of haemoglobin and myoglobin respectively.

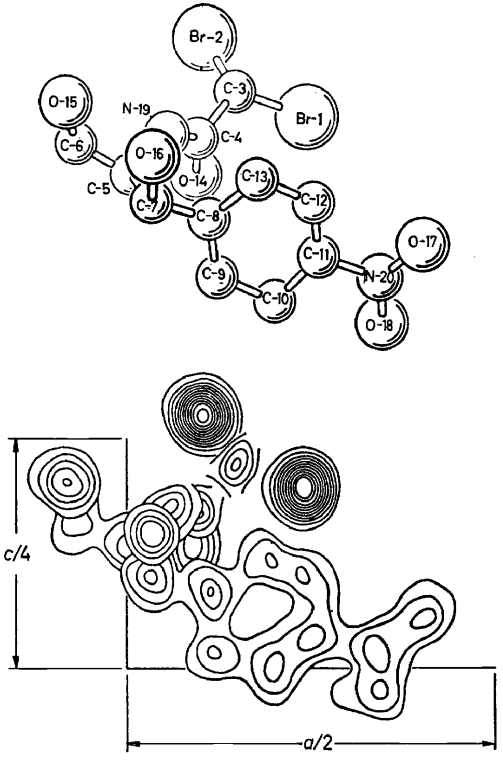


*Figure 8.* Sections of the three-dimensional electron-density distribution of bromoamphenicol from which the atomic array, shown above, was deduced

structure. For the hexacarboxylic acid, an idea of the magnitude of the calculations involved may be suggested by the observation that there are 73 atoms in the molecule and ten three-dimensional electron density distributions were computed[20]. The story of this group of compounds is not yet complete, but it is already so extensive that it would be invidious to deal with it here. There would be little time left for the smaller molecules on which I have preferred to focus attention.

At this point we may note that, where a considerable part of the molecule is known, the process of analysis is much more rapid in achieving a final answer, *e.g.* in the case of calciferol[46], which required only two cycles of three-dimensional calculation.

In the more recent analyses of natural products, *e.g.* those of iresin[47], aspidospermine[48], gelsemine[24], aureomycin[49], calycanthine[50], and kainic and allokainic acids[51], the use of the electron-density distribution as soon as the heavy atoms have been located has proved more popular than image-seeking methods. These analyses have usually involved quite a number of cycles of calculation to arrive at the final answer. It would, however, appear to be profitable to make a careful comparison of this method with that of the image-seeking methods, in particular to consider whether a combination

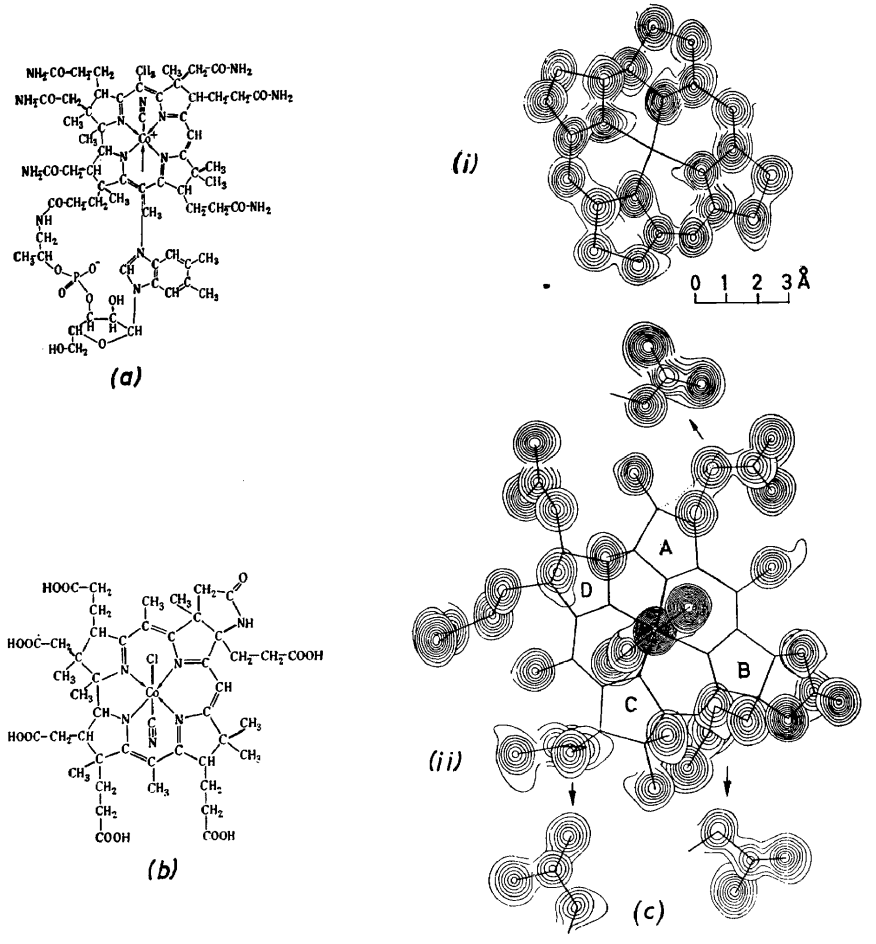


*Figure 9.* The molecular structures of (*a*) vitamin $B_{12}$ and (*b*) the hexacarboxylic acid derived from vitamin $B_{12}$, as deduced from X-ray data; and the three-dimensional structure (*c*) of the hexacarboxylic acid, the atoms being indicated by contour sections near the atom sites. For clarity, the contoured distribution is shown in two parts: (*i*) the corrin nucleus; and (*ii*) the remainder of the molecule. Where atoms overlap in this projected view, groupings of sections at right angles are displayed adjacently

of such methods would approach the correct result more rapidly. We may note that, in the analysis of strychnine hydrobromide[38], an electron density map was used as a check on the results of the image-seeking procedure, while for epilimonol iodoacetate[26], an image-seeking function by direct superposition was used as a check on the first electron-density distribution. In the analysis of jacobine bromohydrin[21] (a rather severe test-case due to the presence of two molecules in the asymmetric unit and the coincidence in $y$ levels of the Br atoms), we first used three-dimensional image-seeking by calculation[37]; but we were dissatisfied with the apparent large amount of spurious detail, so an electron-density distribution was calculated with a

high cut-off, *i.e.* only terms for which the Br atoms contributed 45 per cent or more of their maximum value were used. It was found that minor details were not coincident in the two distributions. Hence coincidences of positive peaks in both could be taken as structurally significant, and on this basis it was possible to extract 52 atoms which grouped into two chemically sensible molecules (identical to a first order)*. The resultant grouping of the molecules of jacobine bromohydrin is shown in *Figure 10*. The structure amplitude comparison for the 55 atoms in the asymmetric unit are grouped in *Table 2*. The data are grouped for constant $k$. The good agreement at

*Table 2.* The measurement of agreement, $R$, between the experimental values of $|F|$, including " unobserved " terms, and those calculated for the first set of sites for the 55 atoms derived from the combined use of image-seeking and an electron-density distribution with high cut-off

| $k$ | $\Sigma\|F_o\|$ | $\Sigma\|F_c\|$ | $\Sigma\|\Delta F\|$ | $R$* |
|---|---|---|---|---|
| 0 | 10,588·7 | 10,241·0 | 2,177·5 | $0{\cdot}20_6$ |
| 1 | 11,375·8 | 10,588·0 | 2,089·1 | $0{\cdot}18_4$ |
| 2 | 10,661·6 | 10,159·8 | 2,066·2 | $0{\cdot}19_4$ |
| 3 | 9,394·0 | 9,925·0 | 2,065·6 | $0{\cdot}22_0$ |
| 4 | 8,793·4 | 9,278·3 | 2,337·2 | $0{\cdot}26_6$ |
| 5 | 8,615·8 | 8,874·4 | 2,405·7 | $0{\cdot}27_9$ |
| 6 | 7,815·5 | 8,274·4 | 2,387·1 | $0{\cdot}30_5$ |
| 7 | 7,756·2 | 7,660·8 | 2,721·1 | $0{\cdot}35_1$ |
| 8 | 7,235·1 | 7,220·6 | 2,662·5 | $0{\cdot}36_8$ |
| 9 | 6,124·6 | 6,435·7 | 2,437·2 | $0{\cdot}39_8$ |
| 10 | 5,662·6 | 5,723·7 | 2,456·4 | $0{\cdot}43_4$ |
| $\Sigma k$ | 94,024·3 | 94,381·7 | 25,805·6 | $0{\cdot}27_4$ |

* $R = \Sigma|\Delta F| \,/\, \Sigma|F_o|$, where $|\Delta F| = |F_o| - |F_c|$

low $k$ is indicative of the general correctness of the molecular structure, while the steady trend of $R$ with increasing $k$ suggests that displacement of atoms is required in the $y$ direction, individually or a molecule as a whole. The structure arrived at is in agreement with the formulation proposed by Geissman[52], but gives in addition full details regarding conformation and configuration.

We have dealt mainly with three-dimensional methods of analysis, since these represent full use of the data, lead to a less equivocal result and are perfectly possible if one has easy access to a large-scale computer. There are, however, occasions when, for reasons of computer distance or economy, it may be advantageous to use two-dimensional or partial three-dimensional data. The diffraction data for each layer may be dealt with separately in the form of normal or generalized projections†. The normal projection has considerable restrictions for work on natural products (*e.g.* the analysis

* Although two molecules per asymmetric unit are a disadvantage to an analysis at an early stage, they provide useful supporting evidence of internal consistency during the later stages (see also epiliminol[26]).

† This does not contradict the statement made earlier in the text, as the data are not used piecemeal but in specific related groups (layers).

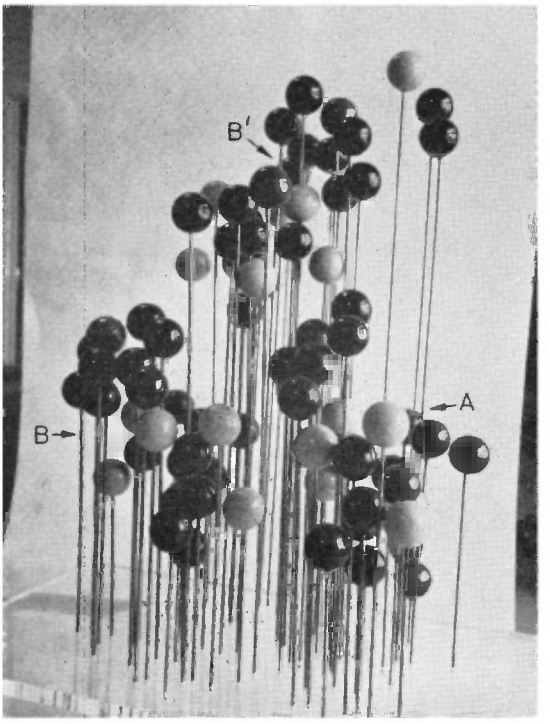


*Figure 10.* Grouping of molecules in the crystal structure of jacobine bromohydrin ethanol adduct. Molecules A and B are not related by symmetry; B and B′ are crystallographically related by a diad screw axis

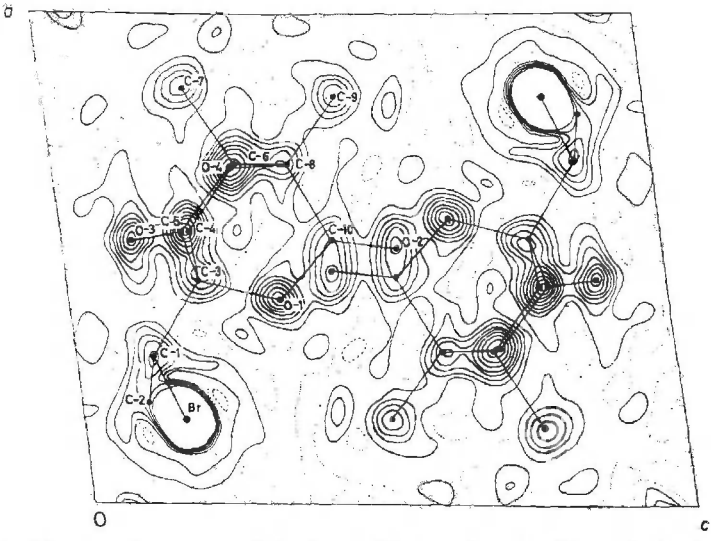


*Figure 11.* The crystal structure of the bromodilactone from jacobine. Projection of the electron-density distribution down the *b* axis with line diagrams of the two symmetry-related molecules superimposed

of benzyl penicillin[53], which was mainly two-dimensional in the early stages), but it may still be of assistance for smaller molecules. A case in point is that of the bromodilactone from jacobine[54] (see *Figure 11*), the structure analysis of which provides further corroborative evidence of the correctness of Geissman's deductions regarding this group of alkaloids from *Senecio jacobaea* L.[52]. For normal projections, the $x,y$ parameters are determined by the position of the peak while, for the corresponding generalized

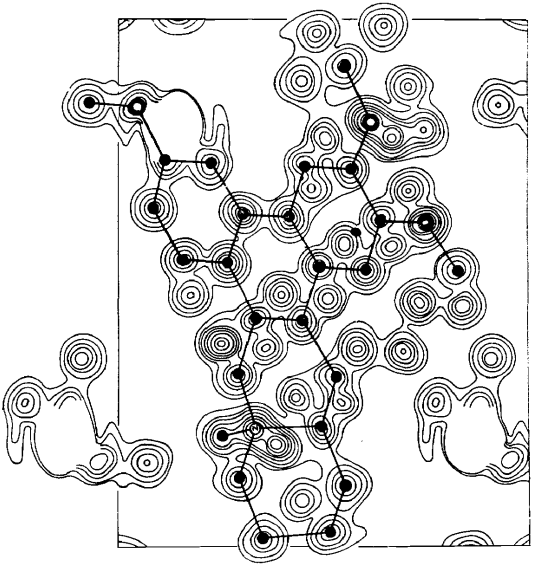

*Figure 12.* A combination of the normal projection and modulus projections for first- and fifth-layer data to produce clear resolution of the atoms. A line diagram of one molecule ( (±)-cryptopleurine methiodide) is superimposed

projections, the third parameter $z$ is derived from the magnitude and sign of the peak value in the cosine and sine components. Because these latter distributions are two-dimensional in essentials, they can be handled more rapidly than three-dimensional distributions while it is possible to combine the two components, $C_l$ and $S_l$, of any layer to yield $|\rho_l| = (C^2_l + S^2_l)^{\frac{1}{2}}$, a modulus projection which, with certain reservations, should be the same as the normal $\rho_0$ projection. Use was made of this property of modulus projections* in the analysis of (±)-cryptopleurine methiodide[55], an interesting example since the structure first determined by X-rays (see *Figure 12*) was later confirmed by chemical synthesis by two independent groups[56, 59]. Use has been made of generalized projections in the determination of the configuration of the α-prodine molecule by Ahmed, Barnes and Kartha[58],

* The distribution of spurious detail in $\rho_0$ and any $|\rho_l|$ does not coincide, since both are derived from different sets of data. As a result, it is possible to extract the coincident peaks as significant structurally.

while we have utilized both generalized and modulus projections in the analysis of himbacine hydrobromide[18] (see *Figure 13*) and thelepogine methiodide[73].

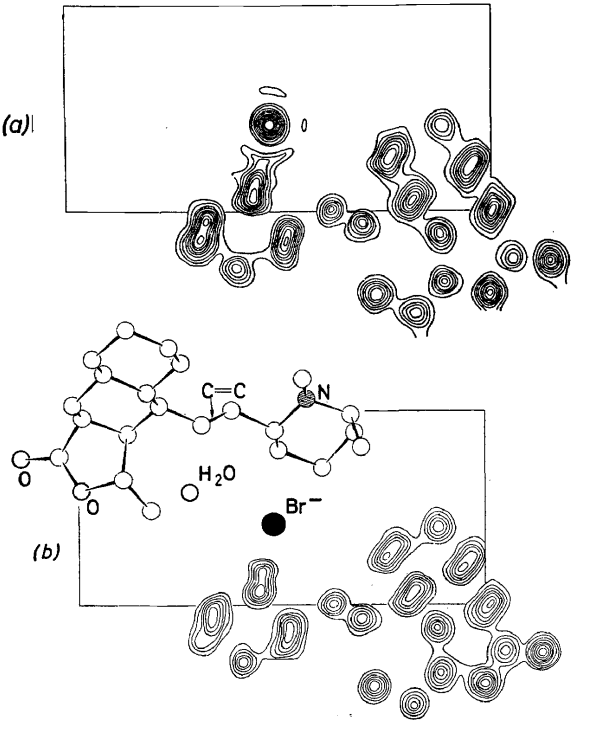


*Figure 13.* A comparison of (*a*) the normal projection and (*b*) the modulus projection from first-layer data for himbacine hydrobromide. The absolute configuration of the molecular skeleton is shown in (*b*)

One invaluable technique which is auxiliary to all the methods discussed above is the difference method, developed and systematized by Cochran[59]. The difference between the distributions based on $F_0$ terms and on $F_c$ terms is calculated by equation (5):

$$\Delta\rho = \rho_0 - \rho_c = \frac{1}{V_c} \sum h \sum_{-\infty}^{+\infty} k \sum l \, (F_0(hkl) - F_c(hkl)) \exp\left\{-2\pi i(hX + kY + lZ)\right\} \quad (5)$$

In this new function, errors and discrepancies in the calculation of structure amplitudes are accentuated, and can be identified and corrected. It has been used to determine the asymmetric vibrations of atoms and molecules and to reveal errors in scattering curves. For our purposes, it can be used: (1) to differentiate light atoms, *i.e.* C, N and O; (2) to refine atom positions; and (3), most important, to assist in the location of lost or wrongly-placed

atoms. The method, under the name " error synthesis ", was used by Bunn and Turner-Jones during the analysis of benzyl penicillin[53]. Two examples of the application of difference methods in a complex structure are shown in *Figure 14*.

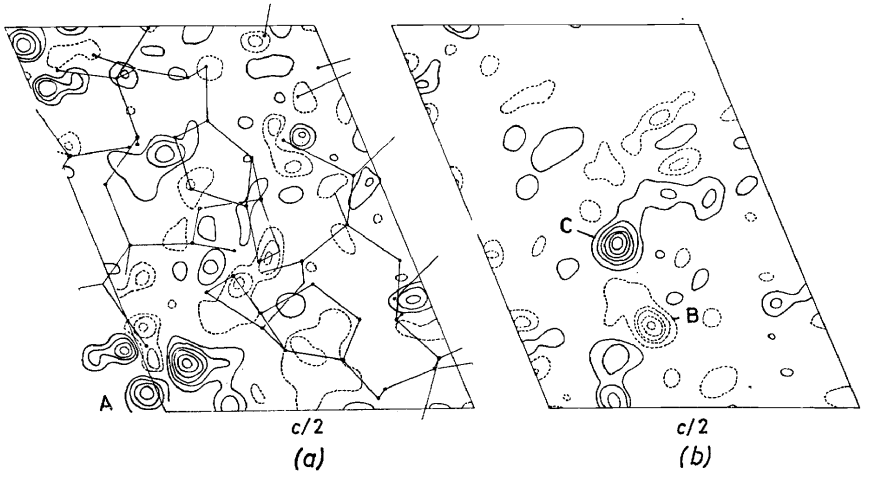


*Figure 14.* Applications of " error " or difference methods in locating atoms and correcting wrongly-placed atoms. (*a*) Contoured difference map of a zero-layer projection after the 52 atoms constituting the two molecules of jacobine have been subtracted. The ethanol molecule is revealed by the contoured positive region, A. The projection is down 15·5 Å. (*b*) Contoured difference map after the 55 atoms of the asymmetric unit were subtracted. In calculating the structure factors, the $z$ parameter of one atom was misread as 0·219 instead of $0{\cdot}421_0$. The difference map reveals this slip by the negative (dashed) region at 0·219, B, and a positive region at 0·422, C, to which the atom site had to be shifted. Despite the fact that 55 atoms are involved, these two complementary features dominate the distribution

The most striking use of difference methods has, however, been to show the location of hydrogen atoms, *e.g.* as shown in *Figure 15*[60]. This requires fairly accurate intensity data, and is not indicative of the normal analysis of a natural product. For definite location of hydrogen, other techniques, *e.g.* N.M.R., are more rapid once the skeleton has been defined. However, we must note that, if the X-ray analysis has been carried out with data of adequate accuracy and range and has been refined sufficiently, so that atom positions are accurately fixed, then it will be possible to place the hydrogen atoms on the basis of the stereochemistry of the molecule.

## The determination of absolute configuration

So far we have been concerned with the derivation of the structure of the molecule from which details of conformation and the relative configuration can be defined. The final step, the determination of the absolute configuration, is a relatively minor step experimentally, but it has taken a rather long time to arrive at what is now obvious—a fairly persistent pattern in research. In 1930, Koster, Knol and Prins[61] showed experimentally that, using X-radiation whose wavelength was just short of the absorption edge for Zn, it was possible to distinguish the Zn side from the S side of the ZnS layer structure by the breakdown in Friedel's law, $F^2(hkl) = F^2(\overline{hkl})$.

It was, therefore, possible to establish the absolute orientation of any ZnS crystal or any other crystal of similar type. Twenty-one years passed before the significance of this observation for organic chemistry was recognized by Bijvoet, who put the idea to the test in a classical experiment with NaRb tartrate and ZrK $\alpha$-radiation[62]. It was later shown by Peterson[63] (in Bijvoet's group) that, although careful choice of radiation with respect to the anomalous scatterer accentuated the effect, it is possible to measure

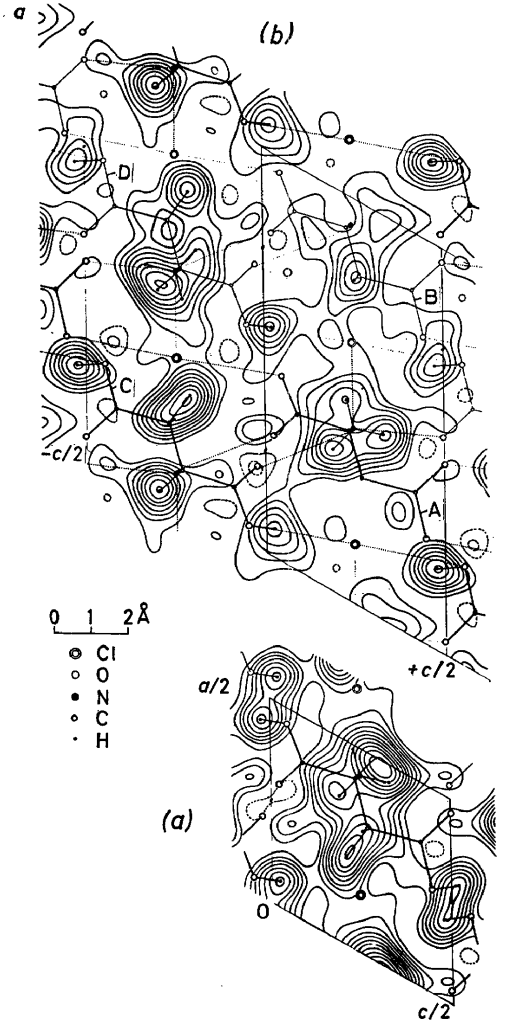


*Figure 15.* The application of difference methods in the location of hydrogen atoms. The crystal structure is (±)-aspartic acid hydrochloride. (*a*) The zero-layer difference map from which the aspartic acid skeleton and the $Cl^-$ ion have been subtracted, leaving the distribution corresponding to the hydrogen atoms. (*b*) A combination of the zero-layer and first-layer cosine and sine generalized difference maps to pick out the hydrogen atoms at different levels of the third (*y*) dimension. Line diagrams of the four molecular skeletons in the unit cell, which have been subtracted, are superimposed

the differences in intensity even when the heavy atom absorption edge lies a considerable way from the wavelength of the radiation used. This observation enabled the technique to be applied without special experimental conditions. Thus, it was shown that it was possible to define the absolute configuration of strychnine hydrobromide[64] using Cu radiation. We have used this technique to define the absolute configuration of himbacine hydrobromide[18] and jacobine bromohydrin[21].

For a heavy atom such as I with Cu radiation, the effect of anomalous dispersion is more marked than for Br, since $f'' = +7{\cdot}2$ for CuK radiation ($f = f_0 + f' + if''$, where $f_0 = 53{\cdot}0$). This is reflected in the larger differences in intensity, often readily visible to the eye; and, under such conditions, Przybylska and Marion have defined the absolute configuration[65] of (+)-des-(oxymethylene)-lycoctonine hydroiodide[66] and (+)-demethanol-aconionone hydrioiodide[40] (see *Figure 16*).

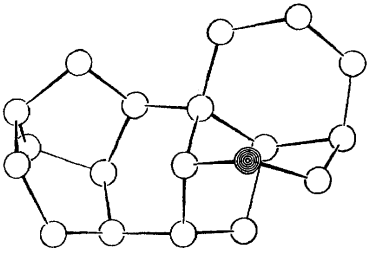

*Figure 16.* The absolute configuration of the molecular skeleton common to des-(oxymethylene)-lycoctonine and demethanolaconinone

Although of lesser interest, an alternative method more akin to the chemical approach is the introduction of a known absolute centre with the heavy atom[67]. With this technique, the analysis proceeds normally and the final structure contains its own reference standard. An example is the analysis of (+)-*S*-methyl-L-cysteine-*S*-oxide[68]. The analysis of vitamin $B_{12}$ illustrates both methods, since the absolute configuration of the molecule in the earlier stages of the analysis was defined by reference to the D-ribose component[19], while later this decision was confirmed by reference to the anomalous dispersion of the Co atom in the hexacarboxylic acid[20].

The definitive experiment of Bijvoet and his co-workers and its implications have already had considerable influence on the definition of configuration in organic chemistry, and have evoked a new set of rules by Cahn, Ingold and Prelog[69] to overcome inconsistencies in nomenclature.

That the anomalous dispersion effect may be of greater significance than the definition of configuration, that it may be of direct use in structure analysis, was realized early. However, since the intensity differences are small, the necessity for very accurate intensity measurement is one which has not yet been made completely practicable for complex molecules. Peterson[63] illustrated its feasibility and Srinivasan[70] has applied it to an analysis of L-tyrosine. Okaya, Saito and Pepinsky[71] have suggested a neat method of unravelling the vector distribution by this technique (see also Pepinsky[5]).

It is clearly a technique with great possibilities, but the experimental problems of very accurate intensity measurements for a large number of reflections remain, a point which was made earlier.

## CONCLUSION

I have attempted to indicate the peculiar advantages of this technique in the natural product field, in that it is not dependent on a mass of associated chemical evidence but is based on a few well-founded principles. Thus it would appear that the rôle of the X-ray diffraction method may lie, not in confirming fairly well-defined or worked-over structures, but rather in entering the field at an early stage—before the bone has been well chewed—and deriving the structure rapidly. To have on hand even partial chemical

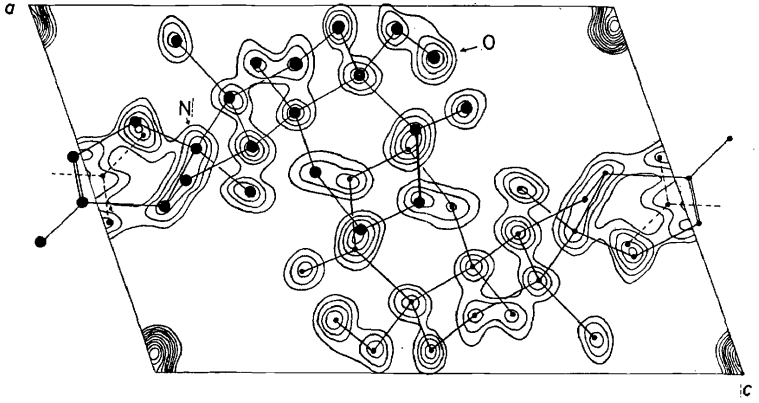


*Figure 17.* The molecular structure of thelepogine methiodide determined by X-rays. The electron-density distribution projected down the *b* axis (13·0 Å) is shown with line diagrams of the two symmetry-related molecular skeletons superimposed. The N and O atoms are indicated and the double bonds deduced from the detailed stereochemistry marked in the line diagram

information may give false confidence in steps and decisions which may later be found invalid—they will be corrected, but with expenditure of time and effort. It has appeared to us important to show that it is possible to work without supporting information. This was the main theme in our earlier work on lanostenol[72] and cryptopleurine[55], but in the analysis which we have just completed we have come nearest to the ideal conditions which we postulated. I will not give much detail, as these are to be given later, but indicate only the starting point and the end result.

The alkaloid thelepogine[73], from a grass, was available only in very small amount. A microanalysis of the alkaloid and of its methiodide (now regarded as $C_{21}H_{34}ONI$) were available, together with an estimation indicating only one C—$CH_3$ group. The X-ray analysis of the methiodide was begun towards the end of last year, and the result so far is shown in *Figure 17.* The molecular skeleton has been located, the N and O differentiated and the bond lengths and angles sufficiently well-defined to determine the location of double-bonds and the most probable disposition of the hydrogen atoms. Not only are the configurations of the six asymmetric

centres in the molecule placed on a relative basis, but the absolute configuration shown in *Figure 17* has been defined by the use of the anomalous dispersion of the I atom to the $CuK_\alpha$ radiation used in the analysis. One point which was unexpected and may be of chemical interest is the existence within the molecular skeleton of a pyrrolizidine ring system which has the same absolute configuration as found in jacobine bromohydrin.

The deduction from the evidence which we have presented is that, in the range of molecules up to 100 atoms (excluding H), and probably beyond, the structure can be solved in a time which compares favourably with other physical methods, provided that certain preliminary conditions are satisfied and a computer of reasonable storage and speed is available. Nor need this work necessarily involve a large team—the three analyses, of himbacine hydrobromide[18], jacobine bromohydrin[21] and thelepogine methiodide[30], some details of which I have used to illustrate various points, were initiated and solved in a period of less than two years. Because of the mass of detail required for an analysis and the extensive data derived from the work, however, it is important to use this technique for compounds either of great significance in themselves or those which provide a structural key to a group of compounds.

*I wish to express grateful thanks to my colleague, Mr J. Fridrichsons, for his collaboration and unflagging support in the work which has formed the basis of opinions expressed here; to Drs J. R. Price, W. D. Crow and C. C. J. Culvenor of Melbourne, and Drs E. Ritchie and W. C. Taylor of Sydney, for, without their willing and generous co-operation (a) in allowing us to tackle problems which they might well have desire to retain for their personal study, and (b) in providing excellent crystals of the derivatives, we would not have had such interesting compounds to work on; and finally, to the invaluable co-operation of the respective staffs of the electronic computers SILLIAC (University of Sydney) and UTECOM (University of New South Wales).*

## References

[1] D. Arigoni, D. H. R. Barton, E. J. Corey, O. Jeger, L. Caglioti, S. Dev, P. G. Ferrini, E. R. Glazier, A. Melera, S. K. Pradhan, K. Schaffner, S. Sternhell, J. F. Templeton and S. Tobinaga. *Experientia*, **16**, 41 (1960)
[2] D. Crowfoot. *Ann. Rev. Biochem.*, **1948**, 115
[3] J. M. Robertson. Third Henderson Memorial Lecture, *Royal Institute of Chemistry Lectures, Monographs and Reports*, **No. 6** (1954)
[4] A. McL. Mathieson. *Revs. Pure and Appl. Chem.* (*Australia*), **5**, 113 (1955)
[5] R. Pepinsky. *Record Chem. Progr.* (*Kresge-Hooker Sci. Lib.*), **17**, 145 (1956)
[6] H. Lipson and W. Cochran. *The Crystalline State*, Vol. 3, Bell & Sons, London (1953)
[7] D. F. Grant and D. Rogers. *Chem. & Ind.* (*London*), **1956**, 275
D. F. Grant. *Acta Cryst.*, **10**, 498 (1957)
[8] C. Brink, D. C. Hodgkin, J. Lindsey, J. Pickworth, J. H. Robertson and J. G. White. *Nature*, **174**, 4 (1954)
[9] A. F. Beecham, J. Fridrichsons and A. McL. Mathieson. *J. Am. Chem. Soc.*, **80**, 4739 (1958)
[10] M. F. Perutz, M. G. Rossman, A. F. Cullis, H. Muirhead, G. Will and A. C. T. North. *Nature*, **185**, 416 (1960)
[11] J. C. Kendrew, R. E. Dickerson, B. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips and V. C. Shore. *Nature*, **185**, 422 (1960)
[12] V. Luzzati. *Acta Cryst.*, **6**, 142 (1953)
[13] M. M. Woolfson. *Acta Cryst.*, **9**, 804 (1956)

[14] G. A. Sim. *Acta Cryst.*, **10**, 177, 536 (1957)
[15] J. Fridrichsons and A. McL. Mathieson. *Rev. Sci. Instr.*, **29**, 784 (1958)
[16] J. H. Robertson. *J. Sci. Instr.*, **37**, 41 (1960)
[17] A. Taylor. *J. Sci. Instr. and Phys. in Ind.*, **26**, 225 (1949)
D. A. Davies, A. McL. Mathieson and G. M. Stiff. *Rev. Sci. Instr.*, **30**, 488 (1959)
[18] J. Fridrichsons and A. McL. Mathieson. *Acta Cryst.*, in the press
[19] D. C. Hodgkin, J. Kamper, J. Lindsey, M. MacKay, J. Pickworth, J. H. Robertson, C. B. Shoemaker, J. G. White, R. J. Prosen and K. N. Trueblood. *Proc. Roy. Soc. (London)*, **242A**, 228 (1957)
[20] D. C. Hodgkin, J. Pickworth, J. H. Robertson, R. J. Prosen, R. A. Sparks and K. N. Trueblood. *Proc. Roy. Soc. (London)*, **251A**, 306 (1959)
[21] J. Fridrichsons, A. McL. Mathieson and D. J. Sutor. *Tetrahedron Letters*, **No. 23**, 35 (1960)
[22] W. L. Bond. *Acta Cryst.*, **8**, 741 (1955)
U. W. Arndt and D. C. Phillips. *Brit. Pat. Application* No. 31197/57 (October)
A. McL. Mathieson. *Acta Cryst.*, **11**, 433 (1958)
E. Prince and S. C. Abrahams. *Rev. Sci. Instr.*, **30**, 581 (1959)
F. Langdon and B. D. Frazer. *Rev. Sci. Instr.*, **30**, 997 (1959)
J. Ladell and K. Lowitzsch. *Acta Cryst.*, **13**, 205 (1960)
[23] A. L. Patterson. *Z. Krist.*, **90**, 517 (1935)
[24] F. M. Lovell, R. Pepinsky and A. J. C. Wilson. *Tetrahedron Letters*, **No. 4**, 1 (1959)
[25] C. H. Carlisle and D. Crowfoot. *Proc. Roy. Soc. (London)*, **184A**, 64 (1945)
[26] S. Arnott, A. W. Davie, J. M. Robertson, G. A. Sim and D. G. Watson. *Experientia*, **16**, 49 (1960)
[27] S. Abrahamsson, D. C. Hodgkin and E. N. Maslen. To be published
[28] R. L. M. Synge and E. P. White. *Chem. & Ind. (London)*, **1959**, 1546
[29] D. C. Hodgkin, M. J. Kamper, K. N. Trueblood and J. G. White. *Z. Krist.*, **113**, 30 (1960)
[30] J. Fridrichsons and A. McL. Mathieson. *Tetrahedron Letters*, **No. 26**, 18 (1960)
[31] M. J. Buerger. *Acta Cryst.*, **3**, 87 (1950)
M. J. Buerger. *Acta Cryst.*, **4**, 531 (1951)
[32] D. M. Wrinch. *Phil. Mag.*, **27**, 98 (1939)
[33] C. A. Beevers and J. H. Robertson. *Acta Cryst.*, **3**, 164 (1950)
[34] J. Clastre and R. Gay. *Compt. rend.*, **230**, 1976 (1950)
J. Clastre and R. Gay. *J. phys. radium*, **11**, 75 (1950)
[35] J. Garrido. *Compt. rend.*, **231**, 297 (1950)
[36] M. J. Buerger. *Vector Space*, Wiley, New York (1959)
[37] D. McLachlan and I. D. Thomas. *Acta Cryst.*, **5**, 301 (1952)
[38] C. A. Beevers and J. H. Robertson. *Acta Cryst.*, **4**, 270 (1951)
[39] M. Przybylska and L. Marion. *Can. J. Chem.*, **35**, 1075 (1957)
M. Przybylska and F. R. Ahmed. *Acta Cryst.*, **11**, 718 (1958)
[40] M. Przybylska and L. Marion. *Can. J. Chem.*, **37**, 1116 (1959)
[41] W. Nowacki and G. F. Bonsma. *Z. Krist.*, **110**, 89 (1958)
[42] R. A. Jacobson, J. A. Wunderlich and W. N. Lipscomb. *Nature*, **184**, 1719 (1959)
[43] G. A. Sim. *Acta Cryst.*, **12**, 813 (1959)
[44] J. M. Robertson and I. Woodward. *J. Chem. Soc.*, **1940**, 36
[45] J. D. Dunitz. *J. Am. Chem. Soc.*, **74**, 995 (1952)
[46] D. C. Hodgkin, M. S. Webster and J. D. Dunitz. *Chem. & Ind. (London)*, **1957**, 1148
[47] M. G. Rossman and W. N. Lipscomb. *Tetrahedron*, **4**, 275 (1958)
[48] J. F. D. Mills and S. C. Nyburg. *J. Chem. Soc.*, **1960**, 1458
J. F. D. Mills and S. C. Nyburg. *Tetrahedron Letters*, **No. 11**, 1 (1959)
[49] S. Hirokawa, Y. Okaya, F. M. Lovell and R. Pepinsky. *Z. Krist.*, **112**, 439 (1959)
[50] T. A. Hamor, J. M. Robertson, H. N. Shrivastava and T. V. Silverton. *Proc. Chem. Soc.*, **1960**, 78
[51] I. Nitta, H. Watase and Y. Tomiie. *Nature*, **181**, 761 (1958)
H. Watase and I. Nitta. *Bull. Chem. Soc. Japan*, **30**, 889 (1957)
H. Watase. *Bull. Chem. Soc. Japan*, **31**, 932 (1958)
[52] T. A. Geissman. *Australian J. Chem.*, **12**, 247 (1959)
[53] D. Crowfoot, C. W. Bunn, B. W. Rogers-Low and A. Turner-Jones. *A Chemistry of Penicillin*, p. 310, Princeton University Press (1949)

[54] J. C. Taylor. To be published
[55] J. Fridrichsons and A. McL. Mathieson. *Nature*, **173**, 732 (1954)
J. Fridrichsons and A. McL. Mathieson. *Acta Cryst.*, **8**, 761 (1955)
[56] C. K. Bradsher and H. Berger. *J. Am. Chem. Soc.*, **80**, 930 (1958)
[57] P. Marchini and B. Belleau. *Can. J. Chem.*, **36**, 581 (1958)
[58] F. R. Ahmed, W. H. Barnes and G. Kartha. *Chem. & Ind. (London)*, **1959**, 485
[59] W. Cochran. *Acta Cryst.*, **4**, 408 (1951)
[60] B. Dawson. To be published
[61] D. Koster, K. S. Knol and J. A. Prins. *Z. Physik*, **63**, 345 (1930)
[62] J. M. Bijvoet, A. F. Peerdeman and A. J. van Bommel. *Nature*, **168**, 271 (1951)
[63] S. W. Peterson. *Nature*, **176**, 395 (1955)
[64] A. F. Peerdeman. *Acta Cryst.*, **9**, 824 (1956)
[65] M. Przybylska and L. Marion. *Can. J. Chem.*, **37**, 1843 (1959)
[66] M. Przybylska and L. Marion. *Can. J. Chem.*, **34**, 185 (1956)
[67] A. McL. Mathieson. *Acta Cryst.*, **9**, 317 (1956)
[68] R. Hine and D. Rogers. *Chem. & Ind. (London)*, **1956**, 1428
[69] R. S. Cahn, C. K. Ingold and V. Prelog. *Experientia*, **12**, 81 (1956)
[70] R. Srinivasan. *Proc. Indian Acad. Sci.*, **50**, 19 (1959)
[71] Y. Okaya, Y. Saito and R. Pepinsky. *Phys. Rev.*, **98**, 1857 (1955)
[72] R. G. Curtis, J. Fridrichsons and A. McL. Mathieson. *Nature*, **170**, 321 (1952)
J. Fridrichsons and A. McL. Mathieson. *J. Chem. Soc.*, **1959**, 2159
[73] W. D. Crow. To be published
[74] E. W. Wiebanga and C. J. Krom. *Rec. trav. chim.*, **65**, 663 (1946)
[75] D. C. Phillips. *Acta Cryst.*, **7**, 159 (1954)
[76] J. S. Clunie and J. M. Robertson. *Proc. Chem. Soc.*, **1960**, 82
[77] J. L. de Vries and R. Pepinsky. *Nature*, **168**, 431 (1951)
[78] A. W. Hanson and F. R. Ahmed. *Acta Cryst.*, **11**, 669 (1958)
[79] D. W. Smits and R. Pepinsky. *Acta Cryst.*, **7**, 653 (1954)
[80] P. Horn and V. Luzzati. *Nature*, **183**, 880 (1959)
[81] H. Bürki and W. Nowacki. *Z. Krist.*, **108**, 206 (1956)
[82] M. G. Rossman and W. N. Lipscomb. *Tetrahedron*, **4**, 275 (1958)
[83] W. Hoppe and G. Will. *Z. Krist.*, **113**, 104 (1960)
[84] D. C. Hodgkin and D. Sayre. *J. Chem. Soc.*, **1952**, 4561
[85] H. Mendel. *Proc. Koninkl. Ned. Akad. Wetenschap.*, **58B**, 132 (1955)
[86] C. Bokhoven, J. C. Schoone and J. M. Bijvoet. *Acta Cryst.*, **4**, 275 (1951)
[87] H. A. McKinstry, P. E. Eiland and R. Pepinsky. *Acta Cryst.*, **5**, 285 (1952)
[88] Y. Sasada and I. Nitta. *Acta Cryst.*, **9**, 205 (1956)
[89] J. W. Wisser, J. Manassen and J. L. de Vries. *Acta Cryst.*, **7**, 288 (1954)
[90] R. B. Campbell and J. M. Robertson. *Chem. & Ind. (London)*, **1952**, 1266
[91] H. S. Yanai and W. N. Lipscomb. *Tetrahedron*, **6**, 103 (1959)
[92] F. Jellinek. *Acta Cryst.*, **10**, 277 (1957)
[93] W. Traub. *Nature*, **178**, 649 (1956)
[94] J. M. Robertson and G. Todd. *Chem. & Ind. (London)*, **1953**, 437
[95] R. H. Moffett and D. Rogers. *Chem. & Ind. (London)*, **1953**, 916
[96] W. B. Wright. *Acta Cryst.*, **11**, 632 (1958)
[97] M. MacKay and D. C. Hodgkin. *J. Chem. Soc.*, **1955**, 3261
[98] J. M. Lindsey and W. H. Barnes. *Acta Cryst.*, **8**, 227 (1955)
[99] J. L. de Vries, M. V. King and R. Pepinsky. *Acta Cryst.*, **5**, 437 (1952)
[100] A. M. Abd El Rahim and C. H. Carlisle. *Chem. & Ind. (London)*, **1954**, 279
[101] J. D. Dunitz. *Nature*, **169**, 1087 (1952)
[102] M. Przybylska and W. H. Barnes. *Acta Cryst.*, **6**, 377 (1953)