

RECOVERING PHASE INFORMATION FROM INTENSITY DATA

Nobel lecture, 9 December, 1985

by

JEROME KARLE

Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, D. C. 20375-5000, U.S.A.

The concept of a crystal is that of a solid body in which the atomic or molecular units are so arranged as to form an array having three dimensional periodicity. Because of the periodicity, it is possible to describe the arrangements of the atomic composition by means of Fourier series. The type of Fourier series that is used in crystal structure analysis represents the electron density distribution in a crystal. This is indeed equivalent to representing the structure of a crystal since the atomic locations are represented by the regions of highest electron density in the electron distributions.

The experimental technique used for examining the structure of crystals is called diffraction. In a diffraction experiment, rays are made to impinge on a crystalline substance of interest and, given the proper geometric conditions, the rays are scattered as if they were bouncing off large numbers of different planes imagined to be cutting through the crystal. The collected intensities of scattering (often 5000-10000 in number) are called a scattering pattern or diffraction pattern and comprise the experimental data from which the structure of the crystal of interest is to be elucidated. The most commonly used rays are Roentgen rays or X-rays, as they are usually called. Other rays, composed of neutrons or electrons, also have their purposes.

It is possible to obtain an insight into the character and sense of a diffraction experiment by imagining some experimental circumstances on the macroscopic scale. Let us suppose that we would like to probe the shape of some large object, hidden from view, by using balls that are hurled in a precise way at the object of interest and interact with the surface with essentially perfect restitution. Let us also suppose that it is possible to minimize and correct for the gravitational effects on the impinging balls. We assume that we can observe the results of the bouncing of the balls from the surface of the object. If a large area were scanned perpendicular to the direction in which the balls were hurled and each time the bouncing pattern were essentially parallel but in the opposite direction to that of the impinging balls, we would conclude that the object had a face with a high degree of flatness that was perpendicular to the direction of the impinging balls. Evidently, by varying the orientation of an arbitrarily

shaped object and observing the patterns of the bouncing balls or, equivalently, the scattering pattern, knowledge of the shape of the object could be developed in quite some detail. On the submicroscopic scale, the situation is somewhat different. Crystals are quite porous to X-rays and the interactions between the rays and the atomic composition of the crystal are different than that of a ball bouncing off a surface such that, as noted, certain geometric conditions need to be fulfilled before it appears that reflection of the rays has taken place. The nature of the interactions is understood, however, and the relevance of the comparison with the macroscopic thought experiment involving bouncing balls maintains. In the case of X-rays and a crystal, the X-rays replace the bouncing balls and the way the X-rays interact with electron density distributions within the crystal gives rise to a scattering pattern unique to each crystalline substance. The problem that the analyst faces is to be able to take the diffraction pattern and from it determine the atomic architecture of the crystal which cannot be observed directly.

There is a special problem in taking the intensity information from a diffraction pattern and calculating from it the electron density distribution of a crystal by use of the Fourier series. The coefficients in the Fourier series are, in general, complex numbers. Only the magnitudes of the complex numbers appear to be available from the measured intensities of scattering. The required phases of the complex numbers seem to be lost in an ordinary X-ray diffraction experiment. It was therefore generally thought that it was not possible to go directly from a diffraction pattern to a determination of a crystal structure. The impasse was overcome in a series of steps that involved recognition that the required phase information was contained in the experimental intensity information, the derivation of a foundation mathematics that displayed relationships between phases and magnitudes and even among phases alone and, finally, the development of practical procedures for structure determination, strategies that brought together in a more or less optimal fashion the mathematical relationships with suitably adjusted and refined experimental data.

The results of structure determinations have been playing a valuable role in a number of areas of scientific endeavor. Crystallization, for example, is a very common phenomenon and many types of substances form crystals ranging from metals and minerals to huge macromolecules such as viruses. Knowledge of structure allows one to relate structure to function, i.e., understand physical, chemical or biological properties and activities, provides the chemist with useful information for syntheses, modifications and reaction mechanisms and can also be used to identify very small quantities of scarce material. It often provides the theoretical chemist with a starting point for his calculations. Structural research provides a conceptual basis for many associated scientific disciplines and it is the opportunity to interact with a variety of such disciplines that has made structural research particularly appealing to me.

As this article proceeds, it will elaborate on a number of the items discussed in this introductory part, describe some interesting applications and discuss briefly some research paths and opportunities for the future.

ELECTRON DENSITY DISTRIBUTION

The electron density distribution, $\rho(\mathbf{r})$, is expressed in terms of the three-dimensional Fourier series

$$\rho(\mathbf{r}) = V^{-1} \sum_{\mathbf{h}}^{\infty} F_{\mathbf{h}} \exp(-2\pi i \mathbf{h} \cdot \mathbf{r}) \quad (1)$$

where V is the volume of the unit cell of the crystal, the basic structural unit from which, through three-dimensional periodicity, the crystal is formed. The coefficients

$$F_{\mathbf{h}} = |F_{\mathbf{h}}| \exp(i\phi_{\mathbf{h}}) \quad (2)$$

are the crystal structure factors associated with the planes labeled with the vectors \mathbf{h} . The \mathbf{h} have integer components, h , k , and l , the Miller indices, whose values are inversely proportional to the intercepts on the x , y and z axes, respectively, of planes cutting through the crystal. The angle $\phi_{\mathbf{h}}$ is the phase associated with $F_{\mathbf{h}}$ and \mathbf{r} labels the position of any point in the unit cell. $F_{\mathbf{h}}E$ is the amplitude of the scattered wave associated with the plane labeled by \mathbf{h} , where E is the electric field vector of the incident beam. The measured intensities of X-ray scattering are proportional to $|F_{\mathbf{h}}|^2$. If the values of the $\phi_{\mathbf{h}}$ were also obtained directly from experiment, structures could be immediately calculated from (1). The seeming absence of this information gave rise to the so-called "phase problem".

The Fourier inversion of (1) followed by the replacement of the integral by the sum of contributions from the N discrete atoms in the unit cell gives, for the Fourier coefficient,

$$|F_{\mathbf{h}}| \exp(i\phi_{\mathbf{h}}) = \sum_{j=1}^N f_{j\mathbf{h}} \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j) \quad (3)$$

where $f_{j\mathbf{h}}$ represents the amplitude of scattering of the j th atom in the unit cell and \mathbf{r}_j is its position vector.

OVERDETERMINACY

A system of simultaneous equations is formed by the definition of the crystal structure factors given by (3) since the values of the scattered intensities are measured for a large number of \mathbf{h} . The unknown quantities in (3) are the phases $\phi_{\mathbf{h}}$ and the atomic positions \mathbf{r}_j . The known quantities are the $|F_{\mathbf{h}}|$ obtained from the measured intensities and the $f_{j\mathbf{h}}$ which differ little from the theoretically calculated atomic scattering factors for free atoms. Since each equation in (3) involves complex quantities, there are really two equations, one for the real and one for the imaginary part. In order to determine the overdeterminacy, a comparison is made of the number of unknown quantities with the number of independent data available. With the use of $\text{CuK}\alpha$ radiation, the overdeterminacy can be as great as a factor of about 50 for crystals that have a

center of symmetry and about 25 for those that do not. In practice, somewhat fewer than the maximum available data are measured, but the overdeterminacy is still quite high.

SOME ATTEMPTS, SOME SUCCESSES

There were some early attempts to obtain structural information or phase information from the structure factor equations. Ott [1] made use of the structure factor equations (3) to derive relationships among the structure factors and atomic positions and he showed that in some simple cases atomic coordinates could be obtained directly from the relationships. Banerjee [2] devised a trial and error self-consistency routine based on Ott's results for finding the phases of structure factors that are centric and therefore with phases that have values that are limited to zero or π . The number of trials increased rapidly with complexity limiting applications to rather simple structures. Avrami [3] worked with equations that relate intensities to interatomic vectors (4). Solutions to these equations were given in terms of the roots of a polynomial equation whose degree increases rapidly as the complexity of the crystal increases. In all these approaches, the increase in computational demands with complexity, sensitivity to experimental errors, and inherent ambiguities in the results prevented their application to any but the simplest structures. Even though present day computational facility is enormously greater than when this work was done originally, the limitations cannot be suitably overcome even now.

A significant advance in the attempt to obtain structural information from the measured intensities was made in 1934 by A. L. Patterson. He developed a Fourier series which has as its coefficients the magnitude of the square of the structure factors rather than the structure factors themselves. The phases may be eliminated from (3) by multiplying by the corresponding complex conjugates to obtain

$$|F_{\mathbf{h}}|^2 = \sum_{j=1}^N \sum_{k=1}^N f_{j\mathbf{h}} f_{k\mathbf{h}} \exp [2\pi i \mathbf{h} \cdot (\mathbf{r}_j - \mathbf{r}_k)] \quad (4)$$

The Fourier transform of (4) is known as the Patterson function [4,5]

$$P(\mathbf{r}) = \sum_{\mathbf{h}}^{\infty} |F_{\mathbf{h}}|^2 \exp(-2\pi i \mathbf{h} \cdot \mathbf{r}) \quad (5)$$

The maxima of a Patterson function represent the interatomic vectors in a structure. Evidently the values of the coefficients are directly obtainable from the measured intensities of scattering. This function has been very useful in locating the heavier atoms in a structure, if they are not too numerous, since the interatomic vectors associated with them would predominate in a map computed from (5) and the atomic positions for them could then be readily deduced. The coordinates for the heavy atoms may be used with (3) to compute an initial set of approximate phases. Depending upon the scattering power of the heavy

atoms, such a computation may be suitable for structures containing up to a few hundred atoms. There are numerous procedures for developing a complete structure from the initial phase information obtained from the heavy atoms. The use of the Patterson function with structures containing heavy atoms has found widespread application and remains one of the major methods of crystal structure determination.

The difficulty with using the Patterson function with experimental data in a general way in the absence of heavy atoms arises from the lack of resolution that occurs for the $N(N-1)$ interatomic vectors as well as inaccuracies. The Patterson function becomes somewhat accessible when it is used in combination with known atomic groupings [6-8].

The fact that the Patterson function could be used to solve simple structures made a positive contribution to the background atmosphere in which progress in phase determination was made. Once a structure was solved, it was possible to use (3) to calculate values for the phases. In effect then, phase values were determined from the measured intensities. In the case of the Patterson function, this happened through the intermediary step of first determining the structure. It was, however, conceivable that the process could be reversed, namely, to obtain phase information directly from the intensities and from that compute the structure by use of (1). This is indeed what happened. So far, except for special cases involving heavy atoms, it has turned out to be easier to obtain phase values directly from the measured intensities and then compute the structure than to obtain the structure directly from the intensities without the use of phases.

Relationships between phases and magnitudes that anticipated the later developments were the inequalities of Harker and Kasper [9]. They derived a number of inequalities by application of the Schwarz and Cauchy inequalities to the structure factor equations (3) in the presence of crystallographic symmetry. The Harker-Kasper inequalities have provided valuable insights. For example, the simple inequality formulas can provide useful phase information as shown by Kasper, Lucht and Harker [10] in their solution of the structure of decaborane. In addition, work with the inequalities indicated that they may have probabilistic characteristics. Gillis [11], for example, noted that the implication of an inequality was probably correct even when the magnitudes of the structure factors were too small to permit a definitive conclusion to be drawn. Gillis speculated that the smallness of the structure factor magnitudes may have been due to thermal effects and employed an appropriate function to increase the values of the structure factor magnitudes so that the inequalities could be applied. The probabilistic interpretation, however, remained a possible alternative, namely, that although an inequality does not quite determine the value of a phase definitively, it still does so with a high probability that the value is correct. This could be important because it would imply that the inequalities have probabilistic implications that could extend their range of usefulness.

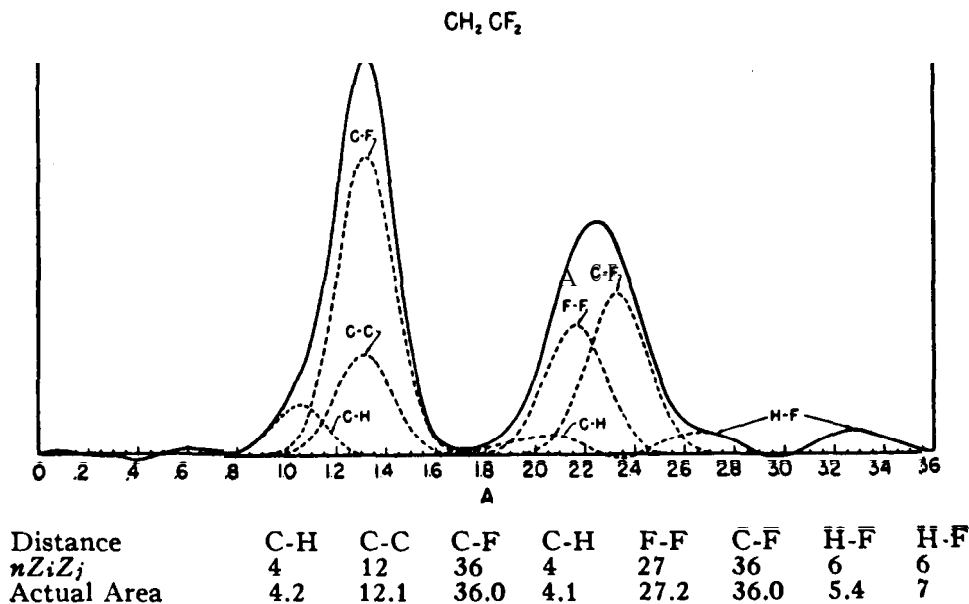


Fig. 2. The essentially non-negative radial distribution function (solid curve) for CH_2CF_2 , computed from the experimental molecular intensity data extracted by use of a properly formed background intensity curve. The solid curve is a probability density that gives the probability of finding interatomic distances in some distance interval along the horizontal axis. The dashed lines represent the decomposition of the main peaks into their component individual interatomic distances. The individual peaks have a definite width related to the internal vibrations in the molecule.

NON-NEGATIVITY AND GENERAL INEQUALITIES

The initial motivation to investigate the mathematics of crystal structure determination arose from experiences in the development of an analytical procedure for obtaining accurate radial distribution functions for determining the structures of gaseous molecules by electron diffraction. A problem arose, namely, to find an accurate background intensity so that the molecular interference intensity could be accurately extracted from the total intensity of scattering. The Fourier transform of the molecular intensity can be interpreted as representing the probability of finding interatomic distances in a molecule. Therefore, this transform must be non-negative and the non-negativity imposed a very useful constraint on the shape of the background intensity [12,13]. Figure 1 shows a radial distribution function for CH_2CF_2 [14] derived from application of the non-negativity constraint and the component distances in the molecule. The attendant accuracy of the result permitted not only equilibrium interatomic distances to be determined but also estimates of the root-mean-square amplitudes of vibration associated with the interatomic distances.

At about the time this work in gas electron diffraction was proceeding, Herbert Hauptman joined our group at the Naval Research Laboratory and, in view of the success of the non-negativity criterion, we decided to explore the

possibility that this criterion might be useful in other areas of structural research. This led us to investigate the determination of electron density distributions around free atoms [15] which found a very fine application in the determination of the electron distribution about argon by Bartell and Brockway [16].

We were also quite interested in seeing what the consequences of the application of non-negativity would be for crystal structure analysis since the electron density distribution defined in (1) is constrained not to be negative. This brought in the work of Toeplitz [17] early in this century on non-negative Fourier series and subsequent development by others. We discussed the theory in three-dimensions and wrote it in a form that would have particular relevance to crystallographic data.

The fundamental result was that the necessary and sufficient condition for the electron density distribution in a crystal to be non-negative is that an infinite system of determinants involving the crystal structure factors be non-negative. A typical determinant is [18]

$$\begin{vmatrix} F_{000} & F_{-k_1} & F_{-k_2} & \dots & F_{-h} \\ F_{k_1} & F_{000} & F_{k_1-k_2} & \dots & F_{k_1-h} \\ F_{k_2} & F_{k_2-k_1} & F_{000} & \dots & F_{k_2-h} \\ \dots & \dots & \dots & \dots & \dots \\ F_h & F_{h-k_1} & F_{h-k_2} & \dots & F_{000} \end{vmatrix} \geq 0 \quad (6)$$

The subscripts in the first column start with 0,0,0 but are otherwise arbitrary. The subscripts in the first row are the same but of opposite sign. The subscript of the element in the *i*th row and *j*th column is the sum of the subscripts of the elements of the *i*th row and first column and the first row and *j*th column. The third order inequality

$$\begin{vmatrix} F_{000} & F_{-k} & F_{-h} \\ F_k & F_{000} & F_{k-h} \\ F_h & F_{h-k} & F_{000} \end{vmatrix} \geq 0 \quad (7)$$

contains a relationship among the structure factors that has played a most important role in direct crystal structure analysis. This may be seen by rewriting (7) in the form [18],

$$\left| F_h - \frac{F_k F_{h-k}}{F_{000}} \right| \leq \frac{\begin{vmatrix} F_{000} & F_{-k} \\ F_k & F_{000} \end{vmatrix}^{1/2} \begin{vmatrix} F_{000} & F_{-h+k} \\ F_{h-k} & F_{000} \end{vmatrix}^{1/2}}{F_{000}} \quad (8)$$

For structure factors of unusually large magnitude, the right side of (8) becomes quite small and then

$$F_h \sim F_k F_{h-k} / F_{000} \quad (9)$$

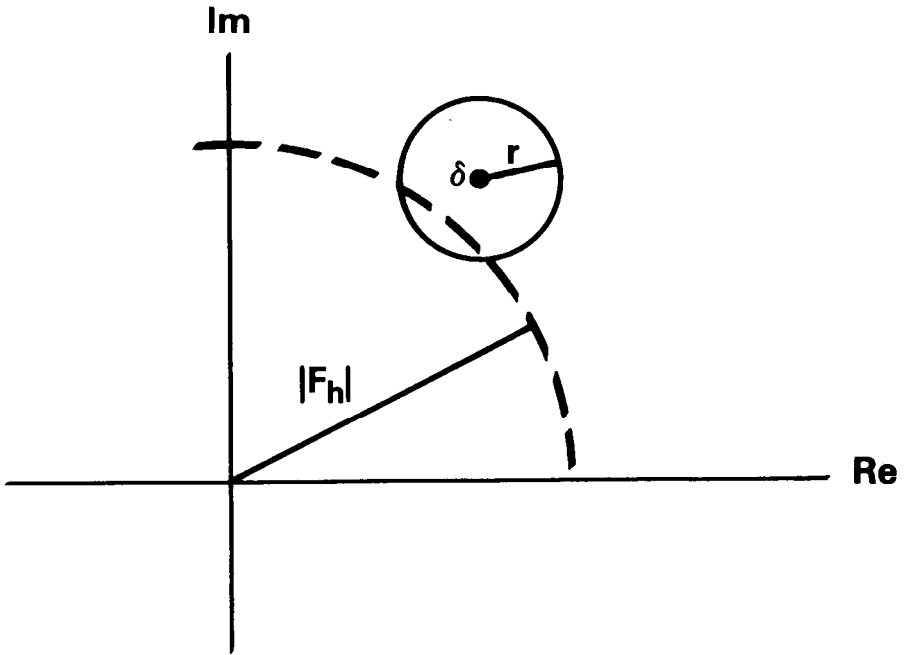


Fig. 2. The determinantal inequalities can be written in the general form $|F_h - \delta| \leq r$. This means that F_h is bounded by a circle of radius r in the complex plane centered at δ . If $|F_h|$ is known, then F_h is confined to a line within the circle.

One evident conclusion from (9) is

$$\phi_h \sim \phi_k + \phi_{h-k} \tag{10}$$

This states that for large structure factor magnitudes, the value of ϕ_h may be defined in terms of the values of two other phases. This may also be seen from a construction based on (8), Fig. 2, in which $\delta = F_k F_{h-k} / F_{000}$ and r is equal to the right side of (8). The form for (8) is then

$$|F_h - \delta| \leq r \tag{11}$$

It can be readily shown [18] that all determinants (6) can be written in the form (11). As the order of the determinants increases, there is a tendency for r to decrease in size, making the determination of ϕ_h rather definitive.

Formula (10) has found wide application beyond the range of usefulness of (8). This is because of the probabilistic characteristics of the inequalities [19] which imply that the most likely value of ϕ_h is that of $\phi_k + \phi_{h-k}$ and the probability decreases the farther the value of ϕ_h deviates from that of $\phi_k + \phi_{h-k}$. Therefore, even when the radius, r , of the bounding circle is large, the most likely value of ϕ_h is $\phi_k + \phi_{h-k}$.

The structure factors in (6) can be replaced by quasi-normalized structure

factors, \mathbf{E} , that represent point atoms (to an approximation when atoms of unequal atomic number are present) rather than atoms with electron distributions,

$$\mathbf{E}_h = F_h / \left(\sum_{j=1}^N f_{jh}^2 \right)^{1/2} \quad (12)$$

Structure factors representing point atoms are the type of quantity normally used in phase-determining procedures. Instead of (9), we have

$$\mathbf{E}_h \sim \mathbf{E}_k \mathbf{E}_{h-k} / E_{000} \quad (13)$$

For centrosymmetric crystals, we have

$$s\mathbf{E}_h \sim s\mathbf{E}_k \mathbf{E}_{h-k} \quad (14)$$

where s means "sign of" a plus sign implying that the phase is equal to zero and a minus sign that it is equal to π . A one-term tangent formula also follows from (13),

$$\tan\phi_h \simeq \frac{|\mathbf{E}_k \mathbf{E}_{h-k}| \sin(\phi_k + \phi_{h-k})}{|\mathbf{E}_k \mathbf{E}_{h-k}| \cos(\phi_k + \phi_{h-k})} \quad (15)$$

The tangent formula composed of the sum of terms over k both in the numerator and the denominator is another formula that has played a major role in the practical applications of the theory for structure determination.

After the set of determinantal inequalities (6) were obtained on the basis of the non-negativity of the electron density distribution in a crystal, it was of interest to investigate their relationship to the inequalities derived by Harker and Kasper [10] from use of the Schwarz and Cauchy inequalities. It was shown [18] that, when the appropriate symmetry was introduced into the third order determinantal inequality by means of certain relationships among the structure factors, the Harker-Kasper inequalities could be derived. Examination of the derivation of the Harker-Kasper inequalities shows, as would be expected, that the non-negativity of the electron density distribution is a requirement for their validity.

The variety of phase determining formulas contained within the determinantal inequalities (6) have their counterpart in probability theory, i.e., similar formulas can be derived with the use of probability theory. Their virtue is that measures of reliability can be attached to them and the judicious use of such measures was an important feature in bridging the gap between mathematical theory and practical application.

It had been pointed out that the determinants have inherent probabilistic characteristics [19] that can, in fact, be directly read out from the form (11). This was not, however, how the first probability formulas were derived. The pursuit of such formulas was motivated by the expectation that the usefulness

of the formulas from the inequalities could be extended because of the great overdeterminancy of the structure problem, the expectation that the points within the bounding circle of radius r and center δ , Fig. 2, would not be uniformly probable and the experience of Gillis [11] with structure factor magnitudes that were too small to elicit definitive conclusions from the inequalities but which, when made larger artificially, led to correct conclusions.

In order to characterize the probabilistic aspects of this subject, we initially decided to develop a facility in the use of the random walk [20], but subsequently changed to the joint probability distribution [21,22] which culminated in a monograph [23] that contained, for the first time, a set of probabilistic formulas and measures for attacking the phase problem, in this case limited to crystals that have a center of symmetry. It was in the monograph [23] also that the theory of invariants and semi-invariants was introduced for the purpose of solving the problem concerning how many and what types of phases to specify to fix the origin in a crystal and what values are permitted. The practical aspects of solving crystals that lack a center of symmetry were developed later on and it was not until 1964 that the structure of the first crystal lacking a center of symmetry was solved [24] by the "direct method" for obtaining phase information by direct use of the measured intensities of scattering. It is interesting that fairly recent developments in the mathematics of the random walk have made this technique quite accurate, stimulating revived interest in its application to the probabilistic aspects of phase determining formulas [25].

FORMULAS FOR PHASE DETERMINATION

The main formulas for phase determination are now listed. They will suffice to characterize the nature of the phase determining procedures. There are additional formulas that play a variety of helpful roles and may be found in the referenced literature of this article.

Centrosymmetric crystals

The Σ_2 formula is [23],

$$sE_{\mathbf{h}} \approx s \sum_{\mathbf{k}_r} E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}} \quad (16)$$

where s means the "sign of" and \mathbf{k}_r , represents restricted values of \mathbf{k} for which the corresponding $|E_{\mathbf{k}}|$ and $|E_{\mathbf{h}-\mathbf{k}}|$ values are large. A plus sign refers to a phase of zero and a minus sign to a phase of π , the only two values possible for a centrosymmetric crystal when an origin in the crystal is properly chosen. The quantities, E , are normalized structure factors which arise as appropriate quantities to use with probability theory and are the same as the quasi-normalized structure factors, ϵ , except for a reweighting [26] of certain subsets of the E . The treatment of the intensity data to obtain normalized structure factors [27] arises from the work of Wilson' [28,29], the earliest application of probability methods to crystal structure analysis. Formula (16) is the probability equivalent of the set of inequalities (8) as \mathbf{k} varies over the set \mathbf{k}_r . The appropriate probability function, $P+(\mathbf{h})$, which represents the probability that

the sign of $E_{\mathbf{h}}$ be positive, was given in the monograph [23]. It is conveniently applied in the form derived by use of the central limit theorem by Woolfson [30] and Cochran and Woolfson [31]

$$P_+(\mathbf{h}) \approx 1/2 + 1/2 \tanh \sigma_3 \sigma_2^{-3/2} |E_{\mathbf{h}}| \sum_{\mathbf{k}} E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}} \quad (17)$$

where

$$\sigma_n = \sum_{j=1}^N Z_j^n \quad (18)$$

and Z_j is the atomic number of the j th atom in the unit cell containing N atoms.

Noncentrosymmetric crystals

The sum of angles and tangent formulas are, respectively,

$$\phi_{\mathbf{h}} \approx \langle \phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}} \rangle_{\mathbf{k}_r} \quad (19)$$

$$\tan \phi_{\mathbf{h}} \approx \frac{\sum_{\mathbf{k}} |E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}| \sin(\phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}})}{\sum_{\mathbf{k}} |E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}| \cos(\phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}})} \quad (20)$$

Formulas (19) and (20) are comparable to (10) and (15), respectively, and result from combining a number of individual terms as \mathbf{k} varies over some chosen set. An appropriate measure of the reliability of (19) and (20) is a variance, V , [32] given by

$$V = \frac{\pi^2}{3} + \left[I_0(\alpha) \right]^{-1} \sum_{n=1}^{\infty} (I_{2n}(\alpha)/n^2) - 4 \left[I_0(\alpha) \right]^{-1} \sum_{n=0}^{\infty} \left[I_{2n+1}(\alpha)/(2n+1)^2 \right] \quad (21)$$

where I_n is a Bessel function of imaginary argument [33]

$$\alpha = \left\{ \left[\sum_{\mathbf{k}_r} \kappa(\mathbf{h}, \mathbf{k}) \cos(\phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}}) \right]^2 + \left[\sum_{\mathbf{k}_r} \kappa(\mathbf{h}, \mathbf{k}) \sin(\phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}}) \right]^2 \right\}^{1/2} \quad (22)$$

and

$$\kappa(\mathbf{h}, \mathbf{k}) = 2\sigma_3 \sigma_2^{-3/2} |E_{\mathbf{h}} E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}| \quad (23)$$

Expression (21) gives the variance of $\phi_{\mathbf{h}}$ as determined from a given set of $\phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}}$. This variance formula has its origin in a probability distribution (in somewhat different notation) of Cochran [34] for $\phi_{\mathbf{h}}$, given a particular $\phi_{\mathbf{k}} + \phi_{\mathbf{h}-\mathbf{k}}$ and the accompanying $|E|$ values. The tangent formula (20) can be derived

in many ways. It has arisen, for example, in theoretical investigations of noncentrosymmetric space groups by use of the joint probability distribution [35] and can be shown to occur [32] in a generalization of the Cochran formula [34] for a particular $\Phi_{\mathbf{h}}$ to take into consideration a set consisting of several or more $\Phi_{\mathbf{k}} + \Phi_{\mathbf{h}-\mathbf{k}}$ rather than just one [32,34]. The average in (19) is to be taken in the context of maximum clustering i.e. a minimum deviation of the contributions of individual addition pairs, $\Phi_{\mathbf{k}} + \Phi_{\mathbf{h}-\mathbf{k}}$ from the average value. All ϕ are kept in a range $-\pi < \phi \leq \pi$ and maximum clustering requires the addition of 0, 2π , or -2π to each addition pair. A practical method for effecting appropriate clustering has been described [32].

PRACTICAL PHASE DETERMINATION

In this part, the various aspects of practical phase determination will be outlined in terms of the first procedure that had broad practical applications to both centrosymmetric and noncentrosymmetric crystals, the symbolic addition procedure [24,32,36,37]. It arose mainly from the efforts of my wife, Dr. Isabella Karle, to bridge the gap between the mathematics of phase determination and the world of experimental data and practical application. At about 1956, we acquired apparatus for carrying out X-ray diffraction experiments with crystals and Isabella Karle taught herself with the aid of a book written by Martin Buerger how to collect and interpret diffraction photographs. Working with crystal diffraction data was quite different then than now; computer-aided collection of diffraction data was essentially nonexistent and, by present standards, computers were very primitive. The thousands of diffraction data collected in the the experiments were measured by eye with the use of a calibrated comparison strip. At the present time, data are collected by automatic diffractometers that measure and record the data after some modest, preliminary human intervention and the largest computers are many orders of magnitude superior in speed and capacity than thirty years ago. Technical advances in the recent past, as exemplified by synchrotron sources of X-radiation and area detectors, promise even greater advances in the near future in terms of the speed and facility with which data can be collected. They provide experimental opportunities that otherwise could not be contemplated.

Once the intensity data are collected, they are transformed into normalized structure factor magnitudes defined by,

$$|E_{\mathbf{h}}| = |F_{\mathbf{h}}| / (\epsilon \sum_{j=1}^N f_{j\mathbf{h}}^2)^{1/2} \quad (24)$$

where ϵ reweights certain subsets of the data [26]. A procedure for doing this is described in International Tables for X-ray Crystallography [38].

It is apparent on examining (16) (19) and (20) that it is necessary to know the values of some phases before additional ones can be evaluated. There are several sources of such information, from certain phase specifications associated with establishing an origin in the crystal [38], the assignment of symbols to some phases for later evaluation, and the use on occasion of auxiliary formulas, such as Σ_1 and Σ_3 , that define individual phases in terms of structure factor

magnitudes alone [38]. The number and types of phases to be used for specifying the origin in a crystal has been determined by use of the theory of invariants and semi-invariants that was developed for this purpose. Depending upon the type of space group involved, the number can vary from none at all to as many as three. Suitable tables [38] are available for carrying out this task.

The phase determining procedure is a stepwise one with few contributors to (16) or (19) at the start. Use at the start of phases associated with the largest possible values of the normalized structure factor magnitudes, $|E|$, will assure that the probability measures, (17) and (21), will be as large as possible. The large overdeterminacy of the problem helps to ensure that initial probabilities will be large enough to proceed in a stepwise fashion to build up a sufficiently reliable set of phase values to effect a solution to the structure problem. Because the nature of phase determination is inherently probabilistic and contingent in a stepwise and interdependent fashion, the problem of establishing optimal procedures based on experimental data was not at all straightforward. There are a very large number of paths through a phase determination. Among many of them are pitfalls in which there arise, for example, temptations to take a path in which the interconnections between phase evaluations flow easily at the expense somewhat of the probability measures. Such paths are more likely to lead to missteps and cumulative errors that could damage or defeat a phase determination than ones that are based only on the highest values of the probability measures.

There is also an ambiguousness inherent in procedures for phase determination which is controlled by the use of symbols. The symbols can assume more than one value. For centrosymmetric crystals, they can have only two phase values, zero or π . For noncentrosymmetric crystals, experience has shown that whereas phase values for the general reflections can have any value in the range from $-\pi$ to π , it is usually sufficient to use only four possible values for the symbols spaced $\pi/2$ apart. One of the virtues of using symbols is that, as the phase determination develops, relationships develop among the symbols reducing the number to be assigned values. Here again, one must proceed with caution so that reliable relationships are distinguished from those that are not. With centrosymmetric crystals, the entire phase determination is carried through before the remaining symbols are given alternative numerical values and tested to see which set yields a Fourier series that makes good chemical sense and reproduces the measured intensities well. For noncentrosymmetric crystals the symbolic addition procedure initially makes use only of (19). After about 100 phases have been evaluated, the remaining symbols are given alternative numerical values after which (20) is applied to further extend the phase set. Again, the correct phase set is the one whose Fourier series makes good chemical sense and yields a structure that is in line agreement with the measured intensities. When a satisfactory result is not obtained, it is appropriate to try an alternative path through the phase determination.

An additional specification, whose character also derives from the theory of invariants and semi-invariants, is required for most noncentrosymmetric space groups. In making the specification, a choice is made of enantiomorph or axis

direction or both. A good way in which this specification is achieved in the symbolic addition procedure is to find that a symbolic representation of a phase value most likely has a magnitude that differs significantly from 0 and π . The specification is accomplished by assigning a plus or minus sign arbitrarily to the magnitude of the phase. The enantiomorph ambiguity for noncentrosymmetric crystals arises from the fact that such a crystal gives the same diffraction pattern as its mirror image in the absence of detectable anomalous dispersion effects.

In the course of phase determinations, particularly in the case of noncentrosymmetric crystals, it may turn out that only a partial structure will appear in a Fourier map based on the determined phase values. This occurs because the phase determination went somewhat awry but not altogether so. The partial structure would be recognized as possibly being a correct fragment because it would make chemical sense, i.e., the fragment would have connectedness and acceptable interatomic distances and angles. A method has been developed for deriving a complete structure from a partial structure [39] that is based on the use of the tangent formula (20) and has been used in many applications. It may happen that a partial structure is not placed correctly with respect to a proper origin in the unit cell of a crystal. Under such circumstances, the use of a translation function [40-43] to place the fragment properly may be helpful.

The symbolic addition procedure [24,32,36,37] was developed as an outgrowth of the experiences in applying the procedure for phase determination for centrosymmetric crystals described in the 1953 monograph [23]. In the monograph, the procedures proposed for sign determination involved the initial use of auxiliary phase formulas such as Σ_1 and Σ_3 [23]. The application of preliminary phase information obtained from these formulas along with probability measures associated with individual indications from the phase-determining formulas facilitated the use of the Σ_2 formula (16). The combination of auxiliary formulas and probability theory greatly reduced to a practicable level the ambiguousness that would be obtained in the employment of the Σ_2 relation by itself along with some rather insensitive and often misleading criterion such as internal consistency.

In the course of applying the procedures of the monograph [23], two important features were found that ultimately played an important role in the symbolic addition procedure. The first feature was that if probability measures were carefully employed at each step of a phase determination, it was possible to carry out the procedure with a small set of starting phases. It was also apparent that the use of symbols could greatly increase the efficiency of the procedure by carrying along in their alternative values a residual ambiguity that could not be easily overcome. A sufficient number of reliable relationships among the symbols usually developed in the course of a phase determination to reduce the alternative possible sets of phases to consider to just a few. A further reduction could be obtained, if desired, by applying auxiliary phase determining formulas at the end of a phase determination to help evaluate the remaining symbols. If an evaluation were incorrect, as occasionally happened with the auxiliary formulas, no great harm would be done. The alternative value could

be tried without the necessity of repeating the phase determination. The symbolic addition procedure for centrosymmetric crystals is thus the procedure of the monograph [23] facilitated by the use of symbols and the application of auxiliary formulas at the end, if needed, instead of at the beginning of a phase determination.

The symbolic addition procedure for centrosymmetric crystals has several features in common with the procedure of Zachariasen [44]. The main distinction from the latter procedure was the application of probability measures to guide each step of the phase determination, especially in the beginning, the consequent use of a minimal number of symbols and the resulting minimization of the ambiguousness of the determination and the optimization of the reliability of the result. There were efforts by others in the 1950's, e.g. Rumanova [45] and Cochran and Douglas [46] that met with some success. Rumanova developed a systematic method for using symmetry relations in centrosymmetric space groups with (16). She used it in connection with Zachariasen's procedure [44]. Cochran and Douglas used a variant of (16) based on a formula derived by Sayre [47] in a procedure that generated a very large number of sets of signs from which the correct one had to be selected under highly ambiguous circumstances.

The procedural features of the symbolic addition procedure for centrosymmetric crystals were extended, in the main by the efforts of Isabella Karle, to noncentrosymmetric crystals [24] with the use of (19), (20), and (21). Several problems arose in developing the technique for phase determination for noncentrosymmetric crystals concerning, for example, the assignment and handling of symbols, the use of the probability measure (21), the number of possible values to assign to the symbols that represent phase values that range continuously from $-\pi$ to π , the combined use of (19) and (20) for phase determination, the proper use of the tangent formula for the processes of phase refinement and phase extension, the development of techniques for specifying an enantiomorph or axis directions or both, and special considerations such as the avoidance of certain troublesome triplet phase invariants involving one and two-dimensional centric reflections. These various aspects of the symbolic addition procedure are to be found in references 32 and 38 and in the papers concerning the applications described further on.

A considerable virtue of the symbolic addition procedure is that, because of its efficiency, a main part of the procedure for phase determination can be carried through by hand. For many years, the procedure for centrosymmetric crystals was carried out in our laboratory completely by hand. In the case of noncentrosymmetric crystals, the first stage, which involved the use of formula (19), was performed by hand until about 100 phases were evaluated and useful relationships developed among the symbols. Only after selected numerical values were assigned to the few remaining symbols was the tangent formula (20) applied with the aid of a computer. The benefits of this aspect of the efficiency of symbolic addition have been the opportunity for those with modest computing facilities to carry out structure determinations, the possibility of close interaction with the phase determination as it progresses, and the educa-

tional value for those newly learning about phase determination to be able to witness and carry through the procedure by hand.

As the application of direct phase determination began to increase during the 1960's and structure determination became more and more a part of research programs, there began to be developed at the end of the 1960's "program packages", software for determining structures from X-ray diffraction data. Among the ones that are widely used, alternative numerical values have been used instead of symbolic phases in the case of noncentrosymmetric crystals, although there are some programs that retain the use of symbols for such types of crystals. For those programs that use alternative numerical values for phases, large numbers of alternative phase sets are generated by use of the tangent formula (20) and the selection of the most likely solutions is dependent upon the use of an elaborate set of probability formulas, auxiliary formulas and acceptance criteria. Other computational techniques have also evolved. For example, random sets of phases have been used as starting sets to be refined by application of the tangent formula (20). By considering large numbers of alternative starting sets it is often practicable to obtain a correct answer, although a large amount of computing is involved. There are also special programs for specific purposes such as the development of a structural fragment into a complete structure.

The computer programs are quite successful with centrosymmetric crystals and also do fairly well with noncentrosymmetric crystals having up to 100 independent (nonhydrogen) atoms to be placed in the unit cell. On occasion an answer will not be forthcoming from the use of a program package. In that case, crystallographers may pursue the problem with special techniques and the application of insights and acumen that have been too special to be found in current programs.

Some names that have been associated with the preparation and dissemination of computer programs for various aspects of automated, direct structure determination are Ahmed, Andrianov, Beurskens, Germain, Gilmore, Hall, Main, Schenk, Sheldrick, Stewart, Viterbo and Woolfson. Among them, some of their programs have enjoyed a broader range of popularity than others. Insight into the contents and philosophy of the programs can be obtained from several publications of the Commission on Crystallographic Computing of the International Union of Crystallography [48,49,50,51,52].

APPLICATIONS

This section will be devoted to illustrating the broad range of applications that have been made accessible by the development of direct structure determination. The examples will be mainly taken from my laboratory but will be seen to be representative of activities that now produce thousands of structural investigations each year.

The earliest applications after the publication of the monograph (23) were collaborations with colleagues at the U.S. Geological Survey on colemanite [53] and meyerhofferite [54]. This was followed by the initial investigations based on the experimental work of Isabella Karle, for example, on p,p'-

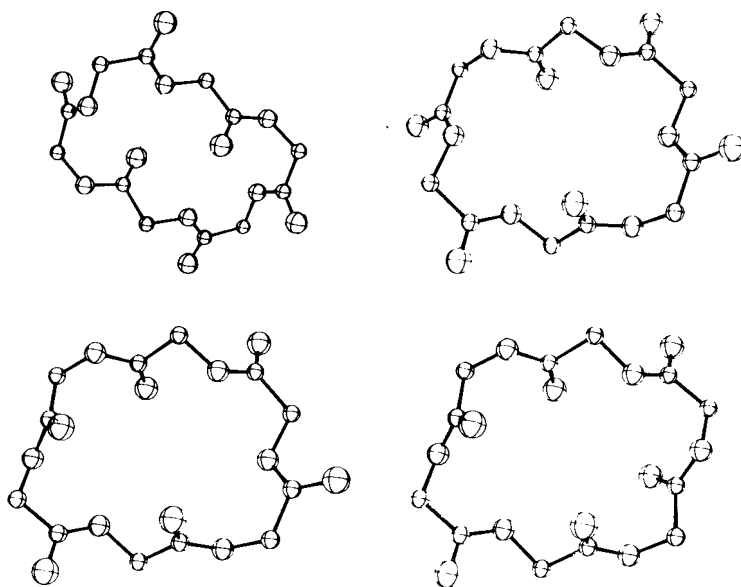


Fig. 3. The four cocrystallizing conformers of cyclohexaglycyl. The conformer at the upper left occurs four times in the unit cell, the one at the lower left occurs twice and the other two occur once.

dimethoxybenzophenone [55] and *N*-benzyl-1,4-dihydronicotinamide [56] that, in time, led to the symbolic addition procedure.

Early applications of the symbolic addition procedure

The first application of the symbolic addition procedure concerned cyclohexaglycyl, a synthetic polypeptide in which six glycine residues form an eighteen-membered ring. The polypeptide crystallizes as a hemihydrate in the triclinic space group $\overline{P}1$ with eight molecules in the unit cell [36], Fig. 3. An interesting characteristic of this structure is that the eighteen-membered ring occurs in the unit cell in four different conformations. Another feature is that all hydrogen atoms capable of forming hydrogen bonds are so involved.

Noncentrosymmetric crystals are quite common among substances of biochemical interest. The determination of the structure of the amino acid *L*-arginine dihydrate [24], which crystallizes in space group $P2_12_12_1$, is the first noncentrosymmetric structure determined by the direct method. As seen in the diagram of the contents of the unit cell, Fig. 4, the hydrogen-bonding indicated by the dotted lines is extensive. As a consequence of this hydrogen-bonding the molecules of arginine form an infinite chain. In addition, water molecules also form hydrogen bonded infinite chains which are perpendicular to the plane of the figure.

The stereochemistry of reserpine, a drug that has been used to control hypertension and nervous disorders, had been determined by chemical means. It was of interest, however, to establish the spatial arrangement of the atoms in the molecule. Reserpine was found to crystallize in the noncentrosymmetric

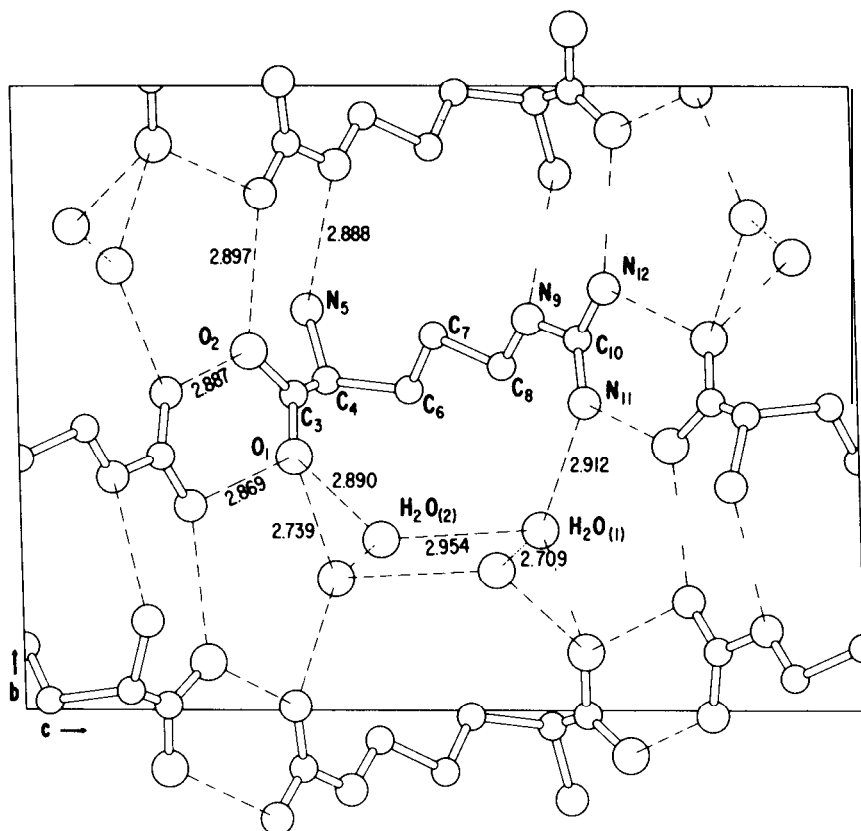


Fig. 4. The structure of L-arginine dihydrate viewed along the a axis. Hydrogen bonding is shown by the dashed lines.

space group, $P2_1$ [57]. In this investigation, the symbolic addition procedure yielded a partial structure which was developed into a complete structure, Fig. 5, by use of the recycling procedure [39] involving the tangent formula. Fig. 5 shows sections from an electron density map projected down the b -axis. The trimethoxybenzoxy group at the bottom is nearly perpendicular to the rest of the molecule. The indole group at the top of the diagram is planar and the dihedral angle between least-squares planes for the indole group and the benzoxy group is 82° .

Identification and stereoconfiguration

The problems concerning the determination of molecular formula and stereoconfiguration can become especially acute when the amount of sample available is very small, when the chemical linkages are new and unusual or when the number of asymmetric centers is large. Under such circumstances, the use of crystal structure analysis can be not only quite helpful but also essential. An example of this is provided by batrachotoxin, a powerful neurotoxin that can be

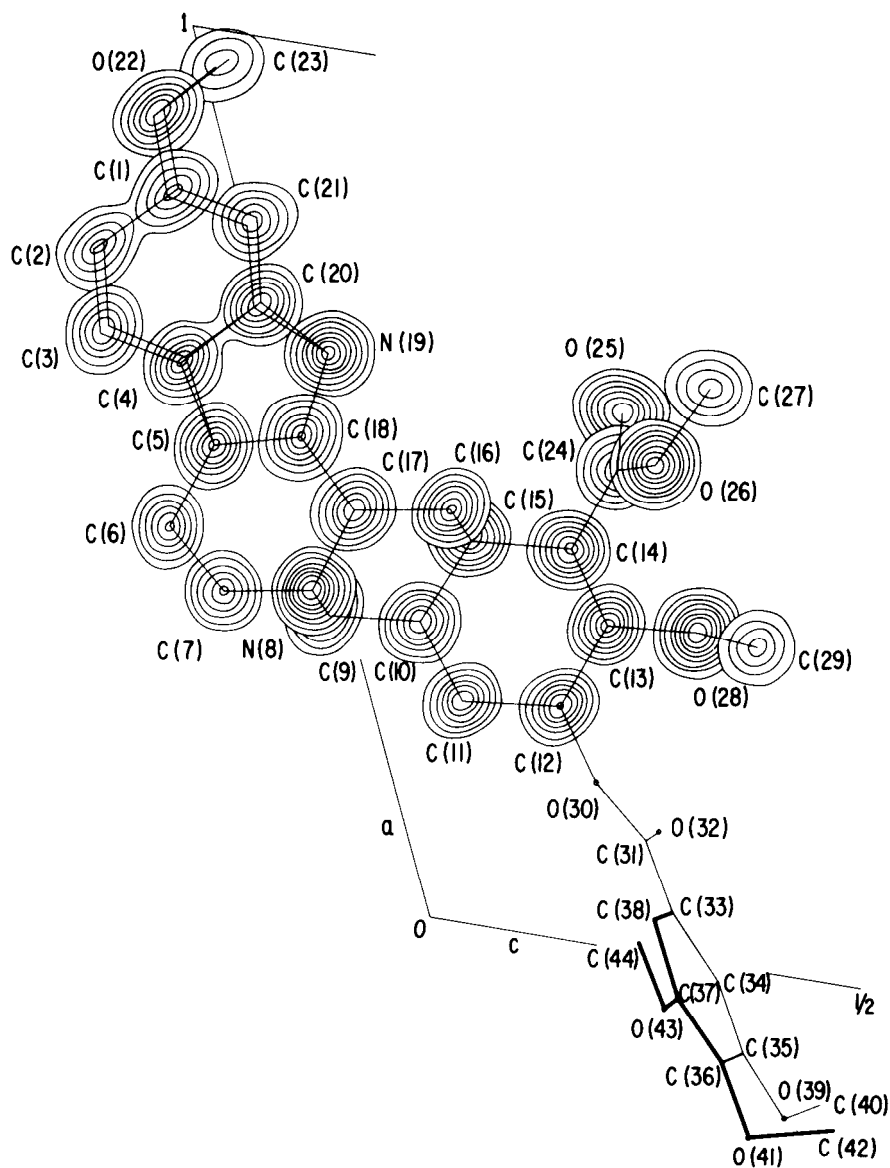


Fig. 5. Sections from a three-dimensional electron density map for the alkaloid reserpine projected down the b axis. The electron density contours are equally spaced at $1e.\text{\AA}^{-3}$ starting with the $1e.\text{\AA}^{-3}$ contour

extracted from the skin of frogs, *Phyllobates aurotaenia*, from tropical America. It is used by native Indians to tip blow darts for hunting. Purified congeners were obtained from ethanol extracts of the skins of about 8000 frogs. The amounts were insufficient to permit the determination of the structural formulae by standard methods. A few, very small crystals of the *o*-*p*-bromobenzoate derivative of batrachotoxinin A, one of the congeners, were grown and one crystal was

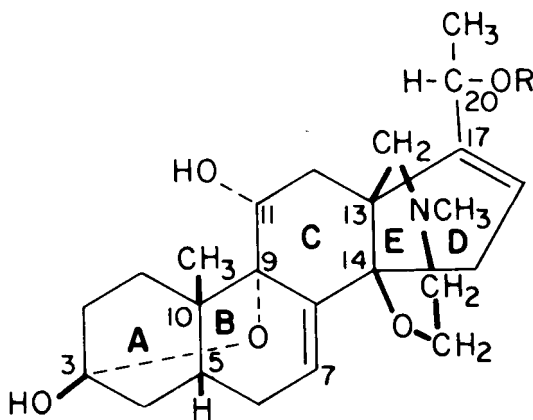


Fig. 6. The molecular formula and absolute stereoconfiguration for batrachotoxinin A (R=H) (heavy solid lines are above plane of paper and dashed ones behind).

selected for structure analysis. It was found to crystallize in space group $P2_12_12_1$. The specialized position of the Br atom (at $1/5, 0, 0$) and the restricted amount of data available limited the amount of information derivable from knowledge of the location of the Br atom. The structure was determined by use of the tangent formula to develop partial structural information into a complete structure. It showed that batrachotoxinin A was a steroidal alkaloid with several novel features [58], Fig. 6. Other information obtained was the stereoconfiguration at the nine asymmetric centers, the conformations of the six rings including the seven-membered ring containing the alkaloid function, values of bond lengths, bond angles and torsional angles, the location of intermolecular hydrogen bonds, and the packing of the models in the unit cell. The absolute configuration was established by measurement of the anomalous scattering of the Br atom [59]. From the information obtained from the structural analysis, the structure and absolute configuration of batrachotoxin and many other congeners were readily deduced [60].

Another small frog, *Dendrobates aurotaenia*, occurring in Columbia and Ecuador, secretes from its skin defensive substances among which are two major toxic alkaloids, histrionicotoxin and dihydroisohistrionicotoxin. These alkaloids are quite unique having a spiropiperidine system with acetylenic and allenic moieties, Fig. 7. The molecular structures, stereoconfigurations and absolute configurations were established by crystal structure analyses of a hydrochloride and hydrobromide of the histrionicotoxin and a hydrochloride of the dihydroiso compound [61,62]. Space groups $C2$ and $P2_12_12_1$ were involved in the analyses. The histrionicotoxins appear to offer the first examples of the occurrence of acetylene and allene moieties in the animal kingdom. Other congeners that occur in smaller quantities were shown, subsequently, by means of mass and NMR spectra to differ only in the saturation of the two side chains [63 and references therein]. The Spiro ring system, with the internal $NH...O$ hydrogen bond, remained unchanged.

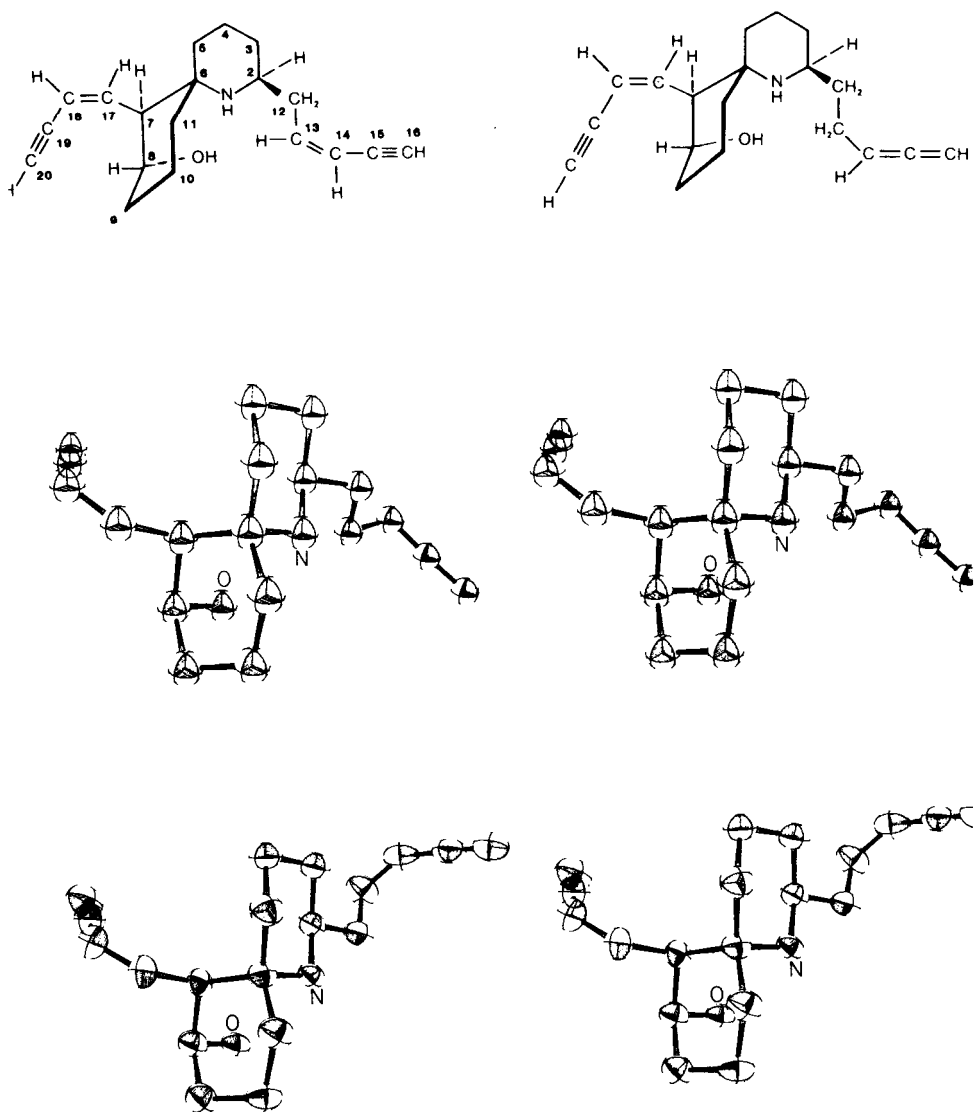


Fig. 7. Histrionicotoxin is shown on the left top and in the corresponding stereodiamgram on the bottom. Dihydroisohistrionicotoxin is shown on the right top and in the stereodiamgram in the middle.

Another example of the application of X-ray crystal structure analysis to the determination of structural formula and stereoconfiguration of an unknown substance is illustrated by the investigation of an alkaloid derived from an *Ormosia* plant, jamine, which crystallizes in the triclinic space group, $P\bar{1}$ [64]. The configuration determined for the molecule is shown in Fig. 8. The connectivity was found to comprise six six-membered rings, five of which were in the chair configuration and one in the boat configuration. There were six asymmetric carbon atoms and the stereoconfiguration about each was, evidently, readily

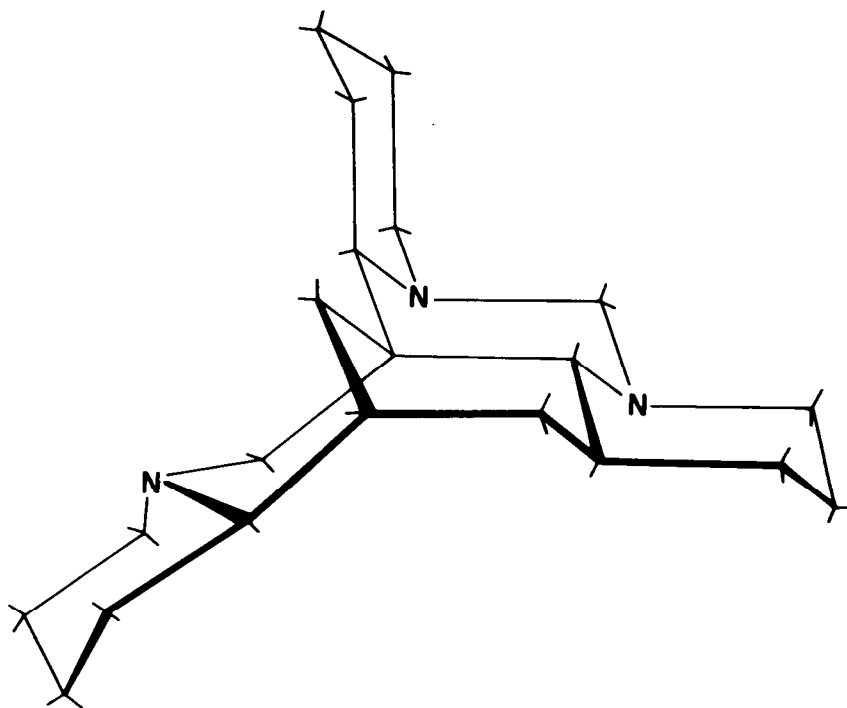


Fig. 8. The stereoconfiguration of jamine.

apparent. The identification of the nitrogen atoms was also readily made in the course of the structure determination.

Brassinolide [65-67] is a very potent plant growth promoter. It is active in very small amounts, 1-10 ng/plant. Brassinolide generates cell enlargement and stimulates cell division in many food plants and plants grown in arid regions that yield oil and other energy related materials. The yield from extraction from plant pollen is very low and therefore large-scale testing awaited chemical synthesis. Evidently, chemical synthesis would be facilitated by structural information. A few crystals of brassinolide became available for crystal structure determination. The crystallization took place in space group $P2_1$, and the analysis [66] established that the steroid nucleus of the molecule had a seven atom B-ring lactone, an unprecedented feature for a natural steroid, Fig. 9. The presence of the lactone appears to be responsible for promoting the plant growth. Subsequent synthesis has shown that compounds analogous to brassinolide are easier to prepare in large quantities and have adequate plant-growth regulating properties. Field evaluations of these regulators are currently underway.

There are very large numbers of examples that could be discussed under the heading of this section. It is noteworthy, however, that the implications of only a small number of definitive structure determinations can be extended far beyond the immediate results of the particular determinations. An outstanding

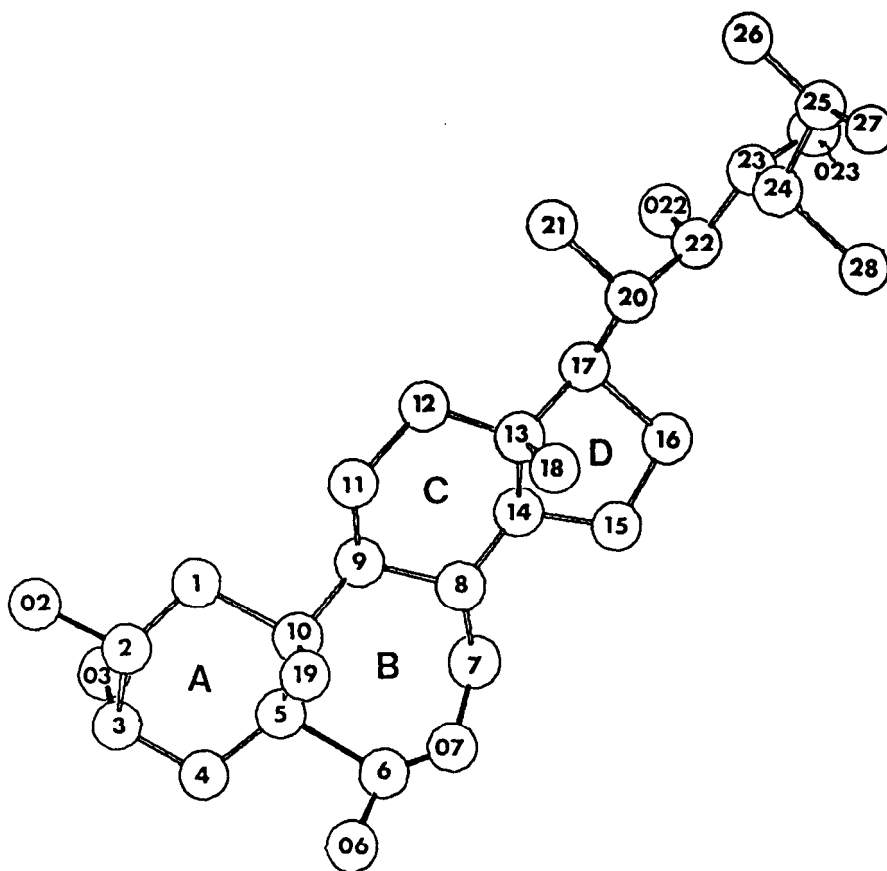


Fig. 9. The molecular formula and stereoconfiguration of brassinolide. Note that the B ring is a seven-membered lactone ring

example is given by some fundamental structural investigations on frog toxins which have provided the information on which was based the subsequent establishment of molecular formulas and stereoconfigurations of over 200 related frog toxins [68,69].

REARRANGEMENTS

In the case of rearrangement reactions, crystal structure determination can again play a particularly useful role because many rearrangement reactions give products that are the result of vast and unanticipated changes in the starting materials. The following examples provide illustrations of such changes.

A photorearrangement reaction in which a major rearrangement takes place is illustrated by the reaction shown in Fig. 10. A crystal structure investigation of a single optically active crystal, selected from a racemic conglomerate established the structural formula and configuration of the photoproduct [70,71]. The substance crystallized in space group $P2_12_12_1$. The structure

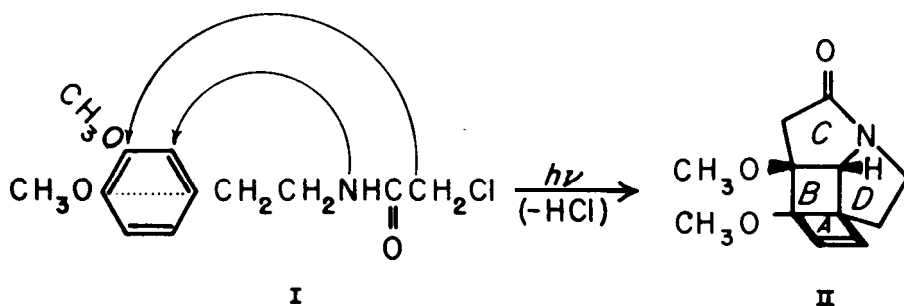


Fig. 10. The photorearrangement of *N*-chloroacetyldimethoxyphenethylamine to a fused ring system consisting of two four- and two five-membered rings.

analysis showed that the photoproduct consisted of four ring systems, two five-membered and two four-membered rings.

Ultraviolet irradiation of *N*-chloroacetyltyramine, where there is a hydroxyl group on the phenyl ring in contrast to the two methoxy groups in the previous example, yields entirely different photorearrangement products. HCl was eliminated and two unusual photodimers, shown in Fig. 11, were produced. Their molecular formulas and stereoconfigurations have been identified by use of crystal structure analysis [72]. It is interesting to note that dimer (II) is the more stable since it is produced from dimer (I) by use of additional ultraviolet radiation. Dimer (I) crystallizes in space group $P2_1/c$ and dimer (II) crystallizes in space group $Pbca$. Dimer (I) is seen to have a central cage bounded by four six-membered rings and two four-membered rings. Each four-membered ring is puckered, with the torsion angles around the ring bonds having values of about 20° . The six-membered rings assume distorted boat conformations. Dimer (II) is seen to have a more complex, partially open, cage bounded by one three-, two five-, two six-, and one seven-membered ring. The six-membered rings are again in a distorted boat conformation. Once the structural characteristics of the photoproducts are known, it is possible to consider possible reaction mechanisms that describe the intermediate changes that occur in the rearrangements of the initial materials resulting in the final products. Postulated mechanisms for the formation of dimers (I) and (II) have been presented [73].

Irradiation with ultraviolet light of a solution containing 3-methyl-5,6-diaza-2,4-cyclohexadien-1-one and 2-propenol (I) produces a cyclobutane addition compound (II) which, upon further irradiation, opens and switches bonds to form a tricyclic molecule (III), Fig. 12. Crystal structure analysis [74] of the material which crystallizes in space group PI established that the final photoproduct was composed of a fused ring system, one three-membered ring and two five-membered rings.

CONFORMATION

The conformations of molecules can be importantly related to their chemical and physiological behavior. Crystal structure investigations can be helpful in

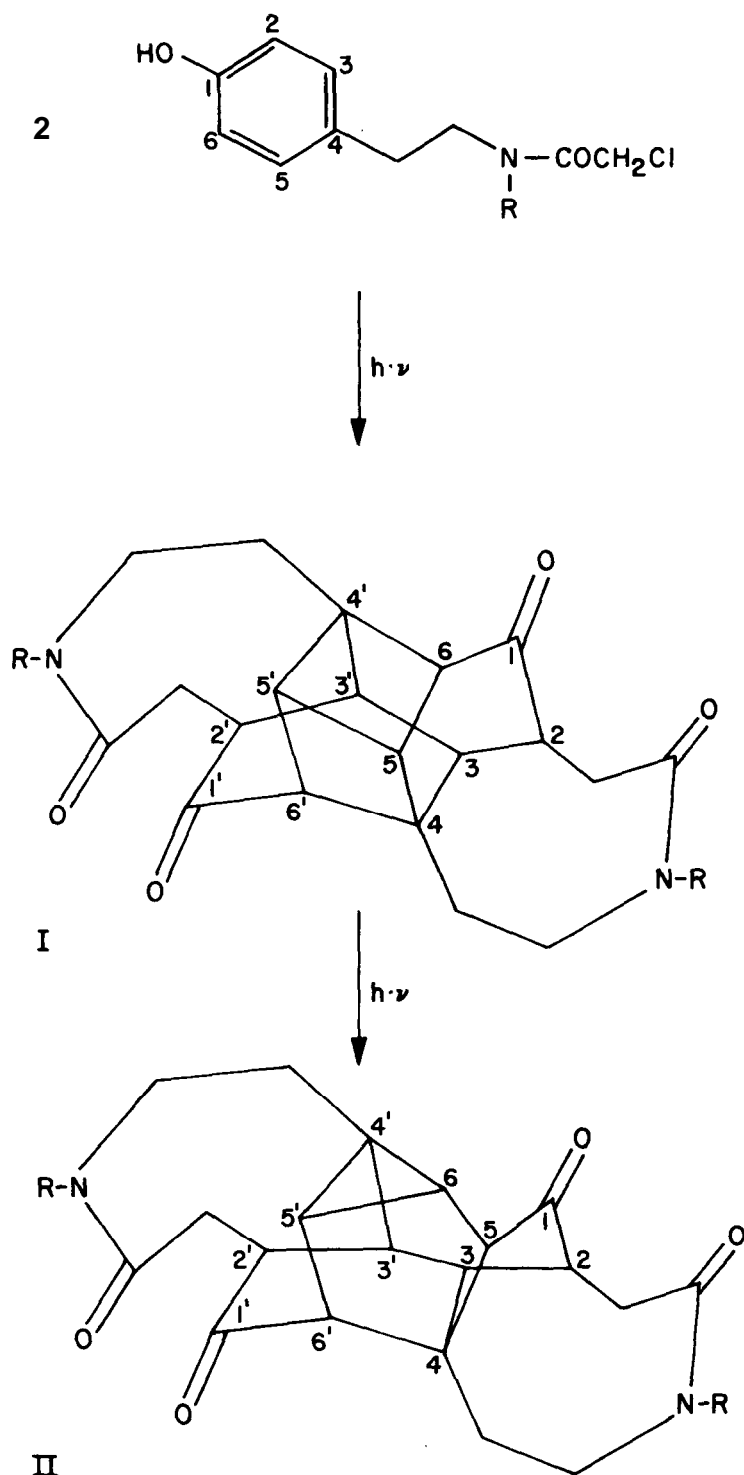


Fig. 11. The configurations of two dimers formed by photorearrangement from N-chloroacetyltyramine. Dimer (II) is seen to be formed from dimer (I) by continued irradiation with ultraviolet light.

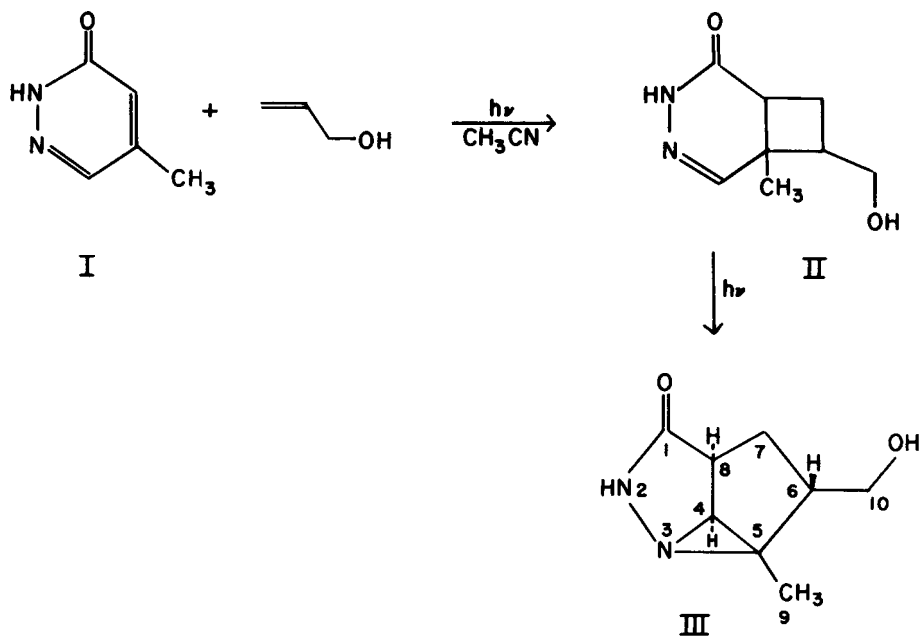


Fig. 12. Irradiation with ultraviolet light of a solution of 3-methyl-5,6-diaza-2,4-cyclohexadien-1-one (I) and 2-propenol leads to a cyclobutane addition product (II). On further irradiation, (II) rearranges to a tricyclic product (III).

providing conformational information. It may be argued, and rightly so, that biologically active materials may assume conformations in the crystalline state that they would not assume in solution. There are, however, numerous instances of conformational studies in which the results of crystal structure analyses are either highly suggestive or rather definitive. The following examples are such instances.

One way in which the crystalline state can imitate the circumstances found in solution is to include in the crystallization relatively large amounts of solvent. Such a crystal is formed by [Leu⁵]enkephalin (Tyr-Gly-Gly-Phe-Leu) grown from *N,N*-dimethylformamide (DMFA)/water. Endogenous enkephalin is a linear pentapeptide that occurs in the brain as [Leu⁵]enkephalin and [Met⁵]enkephalin in varying proportions depending upon the species [75]. Both are quite flexible and the structural study of [leu⁵]enkephalin was undertaken in the hope that insight into probable conformations would be so derived. The crystallization took place in space group *P2*₁ and it turned out that there were four molecules of the peptide in the asymmetric unit of the unit cell, each having a different conformation [76], Fig. 13. The large amount of solvent surrounding the molecules suggests that the conformations may be relevant to the circumstances in solution. The four conformers with extended backbones form an infinite antiparallel β -sheet. β -sheets related by the twofold screw axis are separated by a 12 Å spacing. Side groups protrude above and below the β -sheets and are entirely immersed in a thick layer of solvent that fills the volume

replacement method (SIR) and if several isomorphous structures are available it is called the multiple isomorphous replacement method (MIR). If two isomorphous structures are centrosymmetric, the signs of the structure factors for both crystals can be determined. In practice, some signs would remain indeterminate because of experimental errors and the presence of structure factors of small magnitude. When two noncentrosymmetric crystals form an isomorphous pair, analysis shows that a twofold ambiguity occurs in the evaluation of the phases. It can be readily shown, however, that in the case of noncentrosymmetric crystals many individual phases can be unambiguously evaluated when the structure of the substituted atoms is known [84]. Such phases are either close to the value of the corresponding phases for the substituted atoms or π away from those values. There are a number of ways to resolve the ambiguity from isomorphous replacement. They involve, for example, the use of a second isomorphous substitution or, instead, the additional use of anomalous dispersion data.

Multiple isomorphous substitution with the occasional additional use of anomalous dispersion data has been the technique that has led to the enormous progress in macromolecular structure research of biological importance for the past 25 years. The early theoretical work that supported the advances in this field is that of Patterson [4,5] and Bijvoet [85]. With the types of heavy atoms that are usually substituted into macromolecular structures, there can be significant effects from anomalous dispersion, so that data collection for both anomalous dispersion and isomorphous replacement can be, in effect, combined.

My interest in the techniques of anomalous dispersion and isomorphous replacement was rekindled by a development in my laboratory that demonstrated the considerable potential of the anomalous dispersion technique in structure determination. In an application of the anomalous dispersion technique to the structure determination of the protein, crambin, Hendrickson and Teeter [86] solved the structure by use of the anomalous scattering from six sulfur atoms occurring in three disulfide bridges. The experiment was performed as a single wavelength experiment with $\text{CuK}\alpha$ radiation which is far from the absorption edge for S at 5.02 Å. The success with the use of the relatively weak anomalous scattering from the S atoms coupled with the great span in wavelength between that of the absorption edge and the wavelength of $\text{CuK}\alpha$ radiation stimulated me to initiate a program to investigate various theoretical aspects of the single and multiple wavelength anomalous dispersion techniques and also, in time, isomorphous replacement.

One investigation concerned an exact algebraic analysis of single and multiple wavelength anomalous dispersion [87] by use of the structure factor equations (3) in which the atomic scattering factors are complex according to (25). The appropriate structure factor equations are treated as simultaneous equations and the key to obtaining the final form in which the equations were expressed was the separation of the contributions from the normal atomic scattering factors f'' from those of the real and imaginary corrections to f'' , f' and f'' , respectively.

approximate but simple algebraic manipulations that were based on certain mathematical and physical characteristics of the two techniques [95-96]. This was followed by algebraic analyses [84,97,98] that indicated a potential for good accuracy for many of a large number of formulas proposed.

The development of optimal strategies for the use of triplet phase invariants awaits practical experience. If the heavy-atom (anomalously scattering) substructure is known, there are a number of techniques for obtaining values for individual phases [84,87,89]. In that case, it is quite possible that the triplet phase invariants would be used mainly for phase extension and refinement. In those circumstances when the heavy-atom (anomalously scattering) substructure is not known and resists analysis, the triplet phase invariants may be able to play a primary role in reaching a solution to the structure since the triplet phase invariants can be evaluated without knowledge of the substructure.

The enhanced opportunities to carry out multiple wavelength anomalous dispersion experiments with high intensity and tunable synchrotron X-ray sources should facilitate the development of applications of the new theoretical formulations. Coupled with the major advances in computing, the new theories have the potential to improve the facility with which macromolecular structure research is performed.

