

Lead Article

Acta Cryst. (1989). **A45**, 765-781

Direct Methods in Protein Crystallography*

BY JEROME KARLE

*Laboratory for the Structure of Matter, Naval Research Laboratory,
Washington, DC 20375-5000, USA*

(Received 23 February 1989; accepted 19 June 1989)

Abstract

It is pointed out that the 'direct methods' of phase determination for small-structure crystallography do not have immediate applicability to macromolecular structures. The term 'direct methods in macromolecular crystallography' is suggested to categorize a spectrum of approaches to macromolecular structure determination in which the analyses are characterized by the use of two-phase and higher-order-phase invariants. The evaluation of the invariants is generally obtained by the use of heavy-atom techniques. The results of a number of the more recent algebraic and probabilistic studies involving isomorphous replacement and anomalous dispersion thus become valid subjects for discussion here. These studies are described and suggestions are also presented concerning future applicability. Additional discussion concerns the special techniques of filtering, the use of non-crystallographic symmetry, some features of maximum entropy and attempts to apply phase-determining formulas to the refinement of macromolecular structure. It is noted that, in addition to the continuing remarkable progress in macromolecular crystallography based on the traditional applications of isomorphous replacement and anomalous dispersion, recent valuable advances have been made in the application of non-crystallographic symmetry, in particular, to virus structures and in applications of filtering. Good progress has also been reported in the application of exact linear algebra to multiple-wavelength anomalous-dispersion investigations of structures containing anomalous scatterers of only moderate scattering power.

Introduction

The term direct methods, when applied to small-molecule crystallography, has a fairly generally understood meaning. It developed from a methodology for the determination of the phases of scattered X-ray amplitudes. Direct methods implies

the initiation of a process for phase determination that does not involve the introduction of previously known structural information. The absence of previously known structural information distinguishes direct methods from Patterson methods since the latter are initially concerned with obtaining structural rather than phase information. Techniques that have been especially valuable in investigating the structures of macromolecules are isomorphous replacement and anomalous dispersion. They depend upon the presence of heavy atoms in the structures of interest and have been distinguished from the usual direct methods not only by their own special features, but also by the fact that in applications to date structural information concerning the locations and occupancies of the heavy atoms has been required before the determination of the phases could proceed.

In recent investigations concerning the evaluation of triplet phase invariants from isomorphous replacement data, it was found that such invariants may be evaluated without the use of any structural information including the chemical identification of the replacement atoms. The same is true for triplet phase invariants from anomalous-dispersion data, namely that atomic location is not required. In this case, however, information concerning the chemical identification and the proportion of different types of anomalous scatterers is needed. With values for triplet phase invariants, a parallel exists with direct methods for small molecules, since such analyses are also based on estimated values for triplet phase invariants. The question arises concerning whether the triplet phase invariants for macromolecules, evaluated from the special heavy-atom techniques, can lead to reliable and efficient procedures for phase determination, or whether there are more effective ways to use the same data. Some opinions on the latter question will be forthcoming.

If we attempt to apply in a fairly strict fashion the concept of direct methods to macromolecules, as it applies to small molecules, then we are left with little to discuss. It may be desirable, however, to have an umbrella term such as direct methods in macromolecular crystallography to categorize a spectrum of approaches to macromolecular structure

* *Editorial note:* This invited paper is one of a series of comprehensive Lead Articles which the Editors invite from time to time on subjects considered to be timely for such treatment.

determination in which the analyses are characterized by the use of two-phase and higher-order-phase invariants even though the analyses may be facilitated in part by other techniques that make use of the presence of heavy atoms. I suggest that we use this as our definition. It will permit us then to discuss a number of newer developments and some traditional ones under the title of this paper.

Isomorphous replacement and anomalous dispersion are complementary heavy-atom techniques that have played a most important role in macromolecular structure analysis by greatly facilitating phase determination. They will be discussed separately for the most part, although the value of both techniques is usually enhanced when they are used in combination. The traditional widely used and well documented procedures for employing isomorphous replacement and anomalous dispersion will not be treated here. Rather, special or novel features will be highlighted.

Isomorphous replacement

Crystals that have the same unit-cell geometry but differ in chemical composition are called isomorphous. Isomorphous replacement may be described by

$$F_{R+X} + F_{Y-X} = F_{R+Y} \quad (1)$$

where F_{R+X} is the structure factor for a structure consisting of invariant atoms R and replaceable atoms X , F_{R+Y} is a structure factor for a structure consisting of the same invariant atoms R and replaceable atoms Y and F_{Y-X} is the structure factor for the configuration of the difference between atoms Y and X . In macromolecules, a heavy-atom moiety, Y , is usually added to a native structure, R , consisting of invariant atoms. This reduces (1) to $F_R + F_Y = F_{R+Y}$, which is often written

$$F_P + F_H = F_{PH} \quad (2)$$

with P , H and PH replacing R , Y and $R + Y$, respectively.

Direct methods and the location of heavy atoms

Small-molecule direct methods have been used to locate heavy atoms in isomorphous macromolecular crystals since the work of Steitz (1968). By this time there have been quite a few studies and applications. It is therefore appropriate to take the opportunity not only to describe the nature of such applications but also to characterize briefly the features of small-molecule direct methods.

Traditionally, direct methods for small molecules are initiated with only a very small amount of phase information. This consists of some permitted specifications related to origin assignments and, when appropriate, the specification of enantiomorph or axis

direction or both (Karle, 1988), and a few symbolic assignments or trial numerical values. Phase assignments are associated with large magnitudes of the structure factors. At times additional initial phase information is obtained from the use of special formulas. Phase relationships and probability measures are then applied to the initial phase information in order to extend and refine it. Simple phase-determining formulas, such as the sum-of-angles formula,

$$\varphi_h \approx \varphi_k + \varphi_{h-k}, \quad (3)$$

or a simple tangent formula,

$$\tan \varphi_h \approx \sin(\varphi_k + \varphi_{h-k}) / \cos(\varphi_k + \varphi_{h-k}), \quad (4)$$

are employed in a stepwise fashion, in the initial stages of phase determination. The origin of these formulas is the determinantal inequality theory based on the non-negativity of the electron density distribution in a crystal [(Karle & Hauptman, 1950); see *e.g.* inequality (34)]. The justification for the use of (3) and (4) derives from the inherent probabilistic implications of the inequalities. Probabilistic investigations gave quantitative measure to the reliability of the phase-determining formulas (Hauptman & Karle, 1953; Woolfson, 1954; Cochran, 1955). It has been shown in later years that when the predominant atoms are not too different in atomic number, the probability formulas for phase determination can be expressed in terms of the determinants that occur in the determinantal inequalities (Tsoucaris, 1970; Karle, 1971, 1978). As a phase determination proceeds, the many indications for the value of the phase of a particular reflection are combined by use of a more general tangent formula [see (33)] (Karle & Hauptman, 1956). An alternative procedure involves the use of numerous sets of 40–50 phases that are given initial random values for further extension and refinement by use of the tangent formula (Yao Jia-xing, 1981). For structures containing up to about 50 atoms, a successful refinement often occurs within about 100 starting sets. More complex structures can require many more starting sets.

The application of direct methods to the determination of the locations and occupancies of heavy atoms in isomorphous crystals is based on the presumption that the function

$$|\Delta|F|| = ||F_{PH} - F_P|| \quad (5)$$

is a sufficiently accurate representation of the structure-factor magnitudes of the heavy atoms, where $|F_P|$ is the magnitude of the structure factor for the native protein and $|F_{PH}|$ is the magnitude of a corresponding structure factor for an isomorphous derivative. An analysis based on $F_{PH} = F_P + F_H$ shows that $|\Delta|F||$ is equal to $|F_H|$, the magnitude of the structure factor for the heavy atoms, when the F_P and F_{PH} are among the larger magnitudes for a macromolecule and represent centric reflections. For acentric reflections, (5)

is an approximation. In practice, errors in measuring diffraction intensities and deviations from isomorphism add to the inaccuracies in the use of the $|\Delta|F|$. The inaccuracies appear to have been quite tolerable in a number of successful analyses that have been achieved with the use of the $|\Delta|F|$ as a source of normalized structure-factor magnitudes for the application of direct methods.

Normalized structure-factor magnitudes, $|E_{h\Delta}|$ are formed from the $|\Delta|F_h|$, corrected for positional disorder and scale, by means of

$$|E_{h\Delta}|^2 = |\Delta|F_h|^2 / \varepsilon \sum_{j=1}^N f_{jh}^2 \quad (6)$$

where f_{jh} is the atomic scattering factor for the j th atom in a unit cell containing N heavy atoms and ε is a factor dependent upon the type of reflection and the space group (Stewart, Karle, Iwasaki & Ito, 1977). Direct methods can be expected to play an especially useful role when the occupation of numerous sites by heavy atoms would make the interpretation of Patterson maps rather difficult.

Direct methods were used to evaluate phases of reflections associated with centrosymmetric projections in the first application by Steitz (1968). The applications involved three derivatives of carboxypeptidase A, one of which contained four heavy atoms per protein molecule (8 Hg atoms/unit cell). In structure determinations of small molecules, chemical insights often play an important role in the selection of a correct structure among ambiguous alternatives. Generally, no such insights are available for choosing a correct arrangement for heavy atoms in a macromolecule. Selection of the correct solutions for carboxypeptidase A was made by comparing $E_{h\Delta}$ maps with the use of protein phases determined by multiple isomorphous replacement.

The locations of the heavy atoms in an isomorphous osmium derivative of yeast initiator transfer RNA were determined by Patterson and direct methods applied to centrosymmetric projections, followed by least-squares refinement, by Schevitz *et al.* (1972). Data at 6 Å resolution were employed. The primary site was found to have an occupancy of 2/3, but efforts to establish subsidiary sites in projection were not successful. It was suggested that the use of three-dimensional data may improve the results. Subsequent investigations confirmed this view.

Direct methods were used to evaluate the phases of difference coefficients for three-dimensional reflections at 3.5 Å resolution for three heavy-atom derivatives of the enzyme elastase by Neidle (1973). The metal atoms were located with good accuracy in two of the three derivatives and with somewhat less in the third. A study of one derivative at low resolution indicated that the methods may be applicable at 6 Å resolution and even lower. This investigation demon-

strated that three-dimensional data could play a useful role in the location of heavy atoms.

Heavy-atom positions in numerous derivatives of yeast hexokinase were located by use of direct methods by Steitz, Fletterick & Hwang (1973) and by Anderson, Fletterick & Steitz (1974). All the derivatives had multiple heavy-atom binding sites. It was not possible to find the heavy-atom sites in all the derivatives from high-resolution projection maps alone. In the investigation by Anderson *et al.* (1974), use was made of three-dimensional difference coefficients. This resulted in the location of the four heavy atoms per asymmetric unit in a mersalyl derivative with data at 7 Å resolution. As a consequence of these studies, the authors suggested that low-resolution three-dimensional difference coefficients may be more effective than higher-resolution projection difference coefficients.

In additional studies, Navia & Sigler (1974) obtained trial heavy-atom positions for isomorphous derivatives of concanavalin A and formylmethionine transfer RNA by direct methods applied to projection data. Three-dimensional sets of starting phases for use with the tangent formula were computed from the trial heavy-atom positions. The correct positions were determined from this application of the tangent formula. Wilson (1978) investigated the applicability of direct methods to the location of heavy atoms in isomorphous derivatives of four proteins and discussed the relative values of various functions for selecting the best phase set from alternative sets obtained from use of the tangent formula.

Westbrook, Piro & Sigler (1984) illustrated the usefulness of direct methods when a large number of sites are occupied by heavy atoms in isomorphous derivatives. Two heavy-atom derivatives of Δ^5 -3-ketosteroid isomerase were investigated, a K_2PtCl_4 derivative having 14 sites and a uranyl acetate derivative having 11 sites. Not all sites were found from use of direct methods. The remaining ones were found by analysis of difference maps. This was a particularly complex problem because the unit cell is hexagonal (space group $P6_122$) and contains four identical protomers in each of its 12 asymmetric units.

Two-phase invariants in isomorphous replacement

It readily follows from (2) that

$$\begin{aligned} \cos(\varphi_{hP} - \varphi_{hH}) \\ = (|F_{hPH}|^2 - |F_{hP}|^2 - |F_{hH}|^2) / 2|F_{hP}||F_{hH}| \end{aligned} \quad (7)$$

and

$$\begin{aligned} \cos(\varphi_{hPH} - \varphi_{hH}) \\ = (|F_{hPH}|^2 - |F_{hP}|^2 + |F_{hH}|^2) / 2|F_{hPH}||F_{hH}|. \end{aligned} \quad (8)$$

The symbols P , H and PH have been defined in connection with (2). The added subscript h defines a

specific plane. Information concerning the positions of the heavy atoms permits the evaluation of the φ_{hH} and $|F_{hH}|$. The $|F_{hPH}|$ and $|F_{hP}|$ are obtained from experimental measurements on the isomorphous crystals. When the right-hand sides of (7) and (8) have values in the vicinity of ± 1 , φ_{hP} or φ_{hPH} are approximately equal to or π away from φ_{hH} . In these circumstances and with the known values for φ_{hH} , values for φ_{hP} or φ_{hPH} can then be obtained. This has the potential for affording initial values for numerous φ_{hP} and φ_{hPH} . Test calculations (Karle, 1986) were performed with data for cytochrome c550.PtCl₄²⁻ (Timkovich & Dickerson, 1976) for the purpose of gaining some insight into the usefulness of (7) and (8). The calculations, summarized in Table 4 of the cited reference, have shown that for average errors of about 10% in the structure-factor magnitudes, at least half of the available data could be phased with an average error of less than 0.6 rad. This suggests that in this way the two-phase invariants could be a valuable source of initial phase values.

Probability theory and triplet phase invariants for isomorphous replacement

Probability distributions that can lead to evaluations of three-phase or triplet phase invariants when two crystals are related by an isomorphous replacement have been derived by Hauptman (1982a). The probabilistic theory expresses the joint probability distribution of six structure factors, three related to a phase triplet for one of the crystals and three additional ones related to the corresponding triplet for the isomorphous mate. The joint distribution of the six structure factors gives rise to conditional distributions for triplet phase invariants Ω of the form $P(\Omega) = (1/K) \exp(A \cos \Omega)$, on the assumption that the magnitudes of the six structure factors are known. The results are expressed in terms of A_n ($n = 0, 1, 2, 3$), with the implication that when the A_n are large and positive certain triplet phase invariants have values close to zero and when the A_n are large and negative the appropriate triplet phase invariants have values close to π . The values of A_n are expressed in terms of β_n ($n = 0, 1, 2, 3$) among other quantities. In order to calculate values for the β_n , information concerning the number and occupancy factors of the heavy atoms is required. In the case that the latter information is not available, it may be possible to make an estimate of the β_n from the measured intensities. The latter possibility has not been explored.

If, as is likely, it is not generally possible to determine the number and occupancy of the heavy atoms without also locating their positions, two alternative circumstances prevail. If the required information is not available, the probability distribution cannot be applied as defined. On the other hand, if the heavy-

atom structure is known, then, as will be described below, more accurate probability distributions become accessible. It is also then possible to obtain much initial phase information for the macromolecular structure from (7) and (8).

Test calculations of the probability theory were performed by Hauptman, Potter & Weeks (1982) on error-free diffraction data for cytochrome c550 from *Paracoccus denitrificans* and a single isomorphous PtCl₄²⁻ derivative which crystallized in space group $P2_12_12_1$ (Timkovich & Dickerson, 1976). Estimates (0 or π) were made of triplet phase invariants and compared with known values. It was found that thousands of triplet phase invariants having values close to 0 or π could be evaluated with high reliability by use of the probability theory.

Conditional probability distributions for triplet phase invariants have been obtained by Fortier, Weeks & Hauptman (1984) for the case of a native structure and two isomorphous derivatives in which the heavy replacement atoms for each derivative are located in different positions in the unit cell. The distributions for the triplet phase invariants, represented by Ω , again have the form $P(\Omega) = (1/K) \exp(A \cos \Omega)$, where the parameters K and A are functions of the nine structure-factor magnitudes that enter the analysis and the occupancy and chemical identification of the heavy-atom sites. As may be anticipated, a test calculation with cytochrome c550 shows that the triple of isomorphous structures affords more accurate estimates of the triplet phase invariants than a pair of the same isomorphous structures.

Algebraic analysis and triplet phase invariants

An algebraic analysis of triplet phase invariants for isomorphous crystals affords a number of insights that lead to a simple rule (Karle, 1983) for estimating the values (0 or π) of appropriate triplet phase invariants. In order to apply this rule, it is not necessary to know the chemical identity, nor the number or occupancy of the heavy atoms. In other words, no knowledge concerning the heavy atoms is required and, as will be seen, the evaluation can be made by inspection of the signs of the differences between appropriate structure-factor magnitudes.

For two reflections labeled with \mathbf{h} and \mathbf{k} , we are concerned with the three phases forming a triplet invariant in the crystal of the native macromolecule, φ_{hP} , φ_{kP} and $\varphi_{(\bar{h}+\bar{k})P}$, and the three phases forming a corresponding triplet invariant in the isomorphous substituted crystal, φ_{hPH} , φ_{kPH} and $\varphi_{(\bar{h}+\bar{k})PH}$. We therefore have eight different triplet phase invariants arising in the mathematics as sums of three phases composed of φ_{hP} or φ_{hPH} plus φ_{kP} or φ_{kPH} plus $\varphi_{(\bar{h}+\bar{k})P}$ or $\varphi_{(\bar{h}+\bar{k})PH}$. This is also true for the probabilistic approach.

By use of (2), it follows immediately that

$$F_{hH}F_{kH}F_{(\bar{h}+\bar{k})H} \\ = (F_{hPH} - F_{hP})(F_{kPH} - F_{kP})(F_{(\bar{h}+\bar{k})PH} - F_{(\bar{h}+\bar{k})P}). \quad (9)$$

For the larger values of $|F_{PH}|$ and $|F_P|$ and the largest values of $||F_{PH}| - |F_P||$, the phase of F_P , φ_P , will differ little in value from the phase of F_{PH} , φ_{PH} . It is then possible with good approximation to replace the eight triplet phase invariants associated with (9) with an average expression, $\langle \varphi_h + \varphi_k + \varphi_{\bar{h}+\bar{k}} \rangle$, and (9) becomes

$$|F_{hH}F_{kH}F_{(\bar{h}+\bar{k})H}| \exp [i(\varphi_{hH} + \varphi_{kH} + \varphi_{(\bar{h}+\bar{k})H})] \\ \simeq (|F_{hPH}| - |F_{hP}|)(|F_{kPH}| - |F_{kP}|) \\ \times (|F_{(\bar{h}+\bar{k})PH}| - |F_{(\bar{h}+\bar{k})P}|) \\ \times \exp [i\langle \varphi_h + \varphi_k + \varphi_{\bar{h}+\bar{k}} \rangle]. \quad (10)$$

When the product of normalized structure factors associated with the product on the left of (9), for the rather simple heavy-atom structures occurring in macromolecules, is among the larger ones in magnitude, the corresponding triplet phase invariant is close to zero and the left-hand side of (9) is essentially a real positive number. We can thus make the following interpretation when the triple product on the right-hand side of (10) has large differences:

Rule, R_{iso} : *If the sign of the product of the magnitude differences on the right-hand side of (10) is plus, the value of the average invariant, $\langle \varphi_h + \varphi_k + \varphi_{\bar{h}+\bar{k}} \rangle$, is close to zero, and if the sign of the product is minus, the value of the average invariant is close to π .*

This, in effect, assigns the estimate to all eight triplet phase invariants contained in (9). As an alternative to the rule, the estimates of zero or π may be restricted to only the larger products of structure-factor magnitudes among the eight triplet phase invariants. Calculations indicate that this restriction would enhance the accuracy of the estimates to some small extent but may be relatively meaningless when errors in the data are taken into account. The restricted application of the estimates would also appear to make the use of the triplet invariants in phase evaluation, extension and refinement more difficult to apply. A calculation has been made (Karle, 1986, Table 3) of the average of the magnitudes of the discrepancies of sets of eight triplet phase invariants from their average values for 2.5 Å data from cytochrome c550 and a single isomorphous PtCl_4^{2-} derivative (Timkovich & Dickerson, 1976). The average of the magnitudes of the discrepancies for about 100 000 average triplet phase invariants was 21°. Tests of the simple rule (Karle, 1983), based on the magnitude differences on the right-hand side of (10), have given results of high

reliability for both small structures and the protein cytochrome.

Known heavy-atom structures, isomorphous replacement and triplet phase invariants

The accuracy and applicability of the theory for triplet phase invariants are enhanced by a knowledge of the heavy-atom structure. This applies for both the probabilistic and algebraic approaches. The enhancement of applicability is manifested by the fact that knowledge of the heavy-atom structure permits the evaluation of the cosines of triplet phase invariants at any value in the interval -1 to $+1$ instead of merely in the vicinity of ± 1 . Fortier, Moore & Fraser (1985) developed the conditional probability distributions for triplet phase invariants when the heavy-atom structure is known. An algebraic theory has also been described (Karle, 1983) and further investigated (Karle, 1986). Test calculations of the probability and algebraic formulas in the cited references illustrate the increase in accuracy that derives from having structural information for the heavy atoms.

Probabilistic resolution of ambiguities in two- and three-phase invariants

Probabilistic formulas have been developed by Fan Hai-fu, Han Fu-son, Qian Jin-zi & Yao Jia-xing (1984) with the objective of resolving the ambiguities in the phase differences (two-phase invariants) that arise both in single-isomorphous-replacement experiments and one-wavelength-anomalous-dispersion experiments. In the analyses, it is assumed that the heavy-atom structure is known. The probability distributions are based on the Cochran (1955) distribution for a triplet phase invariant in which, for single isomorphous replacement, the three phases and magnitudes refer either to the native or substituted macromolecule. This distribution is combined with information concerning the phase differences. Formulas are obtained for the probability of the sign of the phase differences. In a later article, Fan Hai-fu & Gu Yuan-xin (1985) have extended the earlier work by introducing the product of the Cochran (1955) and Sim (1959) distributions in order to incorporate partial structural information.

In an application with exact data to avian pancreatic polypeptide and its Hg derivative by Yao Jia-xing & Fan Hai-fu (1985) to single isomorphous replacement in which the replacement atoms have a centrosymmetric arrangement, it was shown that the correct sign for the difference between the phase for the structure and that for the heavy-atom partial structure could be found in most instances. The method involved sign determination among several random starting sets combined with a measure of the figure of merit. Depending upon the number of reflections included (largest 200–800 differences), sets with

the highest figures of merit had correct signs that ranged from about 95 to 70% correct. Test calculations are also presented for the case that the heavy atoms are non-centrosymmetrically placed with use of exact isomorphous replacement data for insulin and its Pb derivative (Fan Hai-fu, Han Fu-son, Qian Jin-zi & Yao Jia-xing, 1984). A test on experimental one-wavelength anomalous-dispersion data was made on the Hg derivative of avian pancreatic polypeptide (Fan Hai-fu & Gu Yuan-xin, 1985). In the calculations performed the signs of 86% of the phase differences were correctly determined.

An investigation of a conditional probability distribution for triplet phase invariants in single isomorphous replacement that also involves knowledge of the structure of the replacement atoms has been carried out by Klop, Krabbendam & Kroon (1987). The conditional probability distribution is formed from the joint probability distribution of Hauptman (1982*a*) for triplet phase invariants in single isomorphous replacement, intensity information, and phase difference information obtained from the structure-factor magnitudes associated with the structure formed by the replacement atoms and the structures of the isomorphous pair. The phase-difference information concerns the differences of phases from the isomorphous pairs as well as differences that include those for a member of the isomorphous pair and those for the structure of the replacement atoms. The use of enantiomorph-sensitive distributions and a statistical analysis has led to formulas for resolving the ambiguity in triplet phase invariants from single isomorphous replacement. Comparisons are made with the features of the distributions of Fan Hai-fu, Han Fu-son, Qian Jin-zi & Yao Jia-xing (1984) and Fortier, Moore & Fraser (1985).

A filtering method for resolving the phase ambiguity in single isomorphous replacement and single-wavelength anomalous dispersion

A filtering method for resolving the phase ambiguities in single isomorphous replacement (SIR) or single-wavelength anomalous dispersion (SAS) has been developed by Wang (1985). Several successful applications of the methodology are described in the latter reference. In the procedure for SIR, for example, a map is made with the use of the sum of the contributions from two structure factors for each reflection, one that contains the correct phase and the other an ambiguous alternative that is incorrect. Wang (1985) points out that the result can be considered as the superposition of two Fourier maps, one the correct map and the other a map composed from incorrect phases. When the heavy atoms in the isomorphous derivative are not in special positions or related by a center of symmetry, the Fourier map produced by the incorrect phases contains no struc-

tural information and can be expected to produce features of generally lower magnitude than those of the correct electron density. These characteristics form the basis for the filtering process for extracting the correct answer.

The filtering procedure involves the definition of the molecular boundary, raising of the densities within the protein region by a constant value along with the removal of very weak and negative densities and, outside the boundary, the smoothing of the density to a constant level. This enhances the electron density within the boundary relative to the disordered solvent region. The modified region within the boundary is used as a partial structure for obtaining new values for the phases for the calculation of a new Fourier map. The process is repeated until convergence is obtained. Thus, the method is called the iterative single isomorphous replacement (ISIR) method. A similar analysis is also applicable to single-wavelength anomalous-dispersion data and is correspondingly labeled ISAS.

An application of maximum entropy

With a test example based on a protein fragment, Bryan & Banner (1987) showed that by use of a maximum-entropy calculation with single-isomorphous-replacement data, a significant improvement was obtained over a conventional 'best' electron density map. The best results were obtained by use of both the intensity data sets from the pair of isomorphous structures as constraints.

Anomalous dispersion

Anomalous dispersion is associated with absorption processes in atoms. The effect is strongest when the incident X-ray wavelength occurs in the vicinity of an absorption edge of the constituent anomalous scatterers of a substance of interest. In general, the experimental objective is to take advantage of the much stronger anomalous scattering of heavier atoms. The anomalous scattering of the light atoms is often negligible.

The effect of anomalous dispersion is expressible in terms of the atomic scattering factor, f , which may be defined as

$$f = f^n + f' + if'' \quad (11)$$

where f^n is the normal or non-anomalous scattering factor that is obtained from computations in which it is assumed that the frequency of the radiation is much larger than the absorption frequencies of the subject atoms. The quantities f' and f'' are the real and imaginary parts of the correction to f^n , representing the effects of the absorption processes that lead to anomalous dispersion. When f'' is not negligible, an important consequence is that the intensity

measured for an acentric reflection \mathbf{h} is, in general, different from that for $\bar{\mathbf{h}}$.

Bijvoet (1954) recognized that anomalous dispersion could be used to determine the absolute configuration of a molecule and with colleagues also showed that it could be used for phase determination (Peerde-man & Bijvoet, 1956). Much progress has been made in developing the technique of anomalous dispersion for use in phase determination. Its applications have usually been made in combination with isomorphous replacement but it is now developing into a technique that is, at times, applied alone. A book of reviews on anomalous dispersion concerning work performed during the early and middle 1970's has been published (Ramaseshan & Abrahams, 1975) in which, among other topics, there are discussions of multiple-wavelength experiments and application to macromolecular structure determination.

A number of developments since the middle 70's account for the increasing interest in the anomalous-dispersion technique. They concern advances in theory, instrumentation and the character of the experiments. A system of exact linear simultaneous equations facilitates the evaluation of two-phase and also three-phase invariants. There has also been good progress in evaluating triplet phase invariants by probabilistic and additional algebraic means. The high intensity and tunability of synchrotron radiation sources and the development of area detectors have greatly facilitated the collection of data.

Exact algebraic analysis of multiple-wavelength-anomalous-dispersion data

The unknown magnitudes and phase differences that arise in multiple-wavelength-anomalous-dispersion experiments may be expressed, without approximation, in terms of linear variables in a system of simultaneous equations (Karle, 1980). One kind of unknown quantity occurs in the form of intensities of scattering for the individual types of atoms present as if each type were present in isolation. The quantities that vary with wavelength, the corrections to the atomic scattering factors, occur in the simultaneous equations as coefficients of the unknown quantities. Their values are available from tables or experimental measurement.

The existence of the intensities of scattering for individual types of anomalous scatterers as unknown quantities to be evaluated by use of the simultaneous equations provides a useful means for solving for the structures of the individual types of anomalous scatterers. The evaluated intensities could be used to compute a Patterson function. If the structure is complicated and not readily amenable to analysis by means of a Patterson function, it is still possible to consider structure determination by direct methods. Once the structure of an anomalous scatterer is

known, the information provided by the simultaneous equations leads directly to the solution of the entire structure by providing phase values associated with the structure of the native protein.

A simple result that illustrates the characteristics described above concerns the case of a structure composed of atoms that scatter normally and one type of atoms that scatter anomalously. The appropriate equation is

$$\begin{aligned} |F_{\lambda\mathbf{h}}|^2 = & |F_{1,\mathbf{h}}^n|^2 + \alpha_{\mathbf{h}} |F_{2,\mathbf{h}}^n|^2 \\ & + \beta_{\mathbf{h}} |F_{1,\mathbf{h}}^n| |F_{2,\mathbf{h}}^n| \cos(\varphi_{1,\mathbf{h}}^n - \varphi_{2,\mathbf{h}}^n) \\ & + \gamma_{\mathbf{h}} |F_{1,\mathbf{h}}^n| |F_{2,\mathbf{h}}^n| \sin(\varphi_{1,\mathbf{h}}^n - \varphi_{2,\mathbf{h}}^n) \end{aligned} \quad (12)$$

where

$$\alpha_{\mathbf{h}} = 1 + (f_{\lambda 2}^a / f_{2,\mathbf{h}}^n) [(f_{\lambda 2}^a / f_{2,\mathbf{h}}^n) + 2 \cos \delta_{\lambda 2}] \quad (13)$$

$$\beta_{\mathbf{h}} = 2[1 + (f_{\lambda 2}^a / f_{2,\mathbf{h}}^n) \cos \delta_{\lambda 2}] \quad (14)$$

$$\gamma_{\mathbf{h}} = 2(f_{\lambda 2}^a / f_{2,\mathbf{h}}^n) \sin \delta_{\lambda 2} \quad (15)$$

$$f_{\lambda 2}^a = (f_{\lambda 2}'^2 + f_{\lambda 2}''^2)^{1/2} \quad (16)$$

$$\delta_{\lambda 2} = \tan^{-1}(f_{\lambda 2}'' / f_{\lambda 2}'). \quad (17)$$

$|F_{\lambda\mathbf{h}}|$ is a known structure-factor magnitude whose value is obtained from a measurement of the intensity at a particular wavelength, λ , for a given reciprocal vector, \mathbf{h} , $|F_{1,\mathbf{h}}^n|$ is the magnitude of the corresponding structure factor for the non-anomalous scattering atoms, $|F_{2,\mathbf{h}}^n|$ is the magnitude of the corresponding structure factor for the anomalously scattering atoms as if there were no anomalous scattering and $\varphi_{1,\mathbf{h}}^n - \varphi_{2,\mathbf{h}}^n$ is the difference between the phases associated with $|F_{1,\mathbf{h}}^n|$ and $|F_{2,\mathbf{h}}^n|$, respectively. Evidently, the subscript 1 refers to the non-anomalous scattering atoms and the subscript 2 refers to the anomalously scattering ones.

A system of simultaneous equations can be formed from (12) by performing anomalous-dispersion experiments at various wavelengths and adding an equation for $\bar{\mathbf{h}}$ which differs from the one for \mathbf{h} . The sign of the last term on the right-hand side of (12) becomes minus when $\bar{\mathbf{h}}$ replaces \mathbf{h} . The equations are linear if the unknown quantities are chosen to be $|F_{1,\mathbf{h}}^n|^2$, $|F_{2,\mathbf{h}}^n|^2$, $|F_{1,\mathbf{h}}^n| |F_{2,\mathbf{h}}^n| \cos(\varphi_{1,\mathbf{h}}^n - \varphi_{2,\mathbf{h}}^n)$ and $|F_{1,\mathbf{h}}^n| |F_{2,\mathbf{h}}^n| \sin(\varphi_{1,\mathbf{h}}^n - \varphi_{2,\mathbf{h}}^n)$. Values for the phase differences can be obtained by taking the arctangent of the ratio of the fourth and third unknown quantities listed above. The system of simultaneous equations can be augmented by adding a quadratic equation based on the fact that $\sin^2 \chi + \cos^2 \chi = 1$.

A completely general system of equations has been derived for the case of any number and type of anomalous scatterers (Karle, 1980). Precisely the same types of unknown quantities occur in the general system of equations as occur in (12), individual normal intensities of scattering and phase differences.

Applications of the linear algebraic equations are beginning to be made and will be noted further on.

Cascarano, Giacovazzo, Peerdeman & Kroon (1982) have carried out an analysis to obtain expressions for the structure-factor magnitudes for identical anomalous scatterers among a majority of normal scatterers. Evaluations of such quantities also follow from the application of (12). They suggest the use of a two-wavelength technique in the vicinity of an absorption edge. With the structure-factor magnitudes at hand in a test application with ferredoxin, the authors used direct methods to find the iron atoms. Errors were introduced to simulate experimental conditions.

Woolfson (1984) has presented a further discussion of the case of a single type of anomalous scatterer in terms of a two-wavelength experiment in which significant anomalous scattering occurs at one wavelength but not at the second one. Presumably the second wavelength is far from an absorption edge. The latter would have the effect of reducing the observed pair of quantities in (12) for \mathbf{h} and $\bar{\mathbf{h}}$ to $|F_{\lambda\mathbf{h}}|^2 = |F_{\lambda\bar{\mathbf{h}}}|^2 = |F_{\mathbf{t},\mathbf{h}}|^2$ where $|F_{\mathbf{t},\mathbf{h}}|^2$ is the magnitude squared of the structure factor representing the total normal scattering for the structure of interest and is equal to $|F_{1,\mathbf{h}}|^2 + |F_{2,\mathbf{h}}|^2 + 2|F_{1,\mathbf{h}}||F_{2,\mathbf{h}}| \cos(\varphi_{1,\mathbf{h}} - \varphi_{2,\mathbf{h}})$.

One-wavelength data analysis

The possibility of using the system of equations represented by (12) in a one-wavelength experiment has been considered (Karle, 1985*b*, 1989). The data from a one-wavelength experiment generate two equations in (12), one for \mathbf{h} and one for $\bar{\mathbf{h}}$. A third equation is based on $\sin^2\chi + \cos^2\chi = 1$. As pointed out, for the case of a structure having non-anomalous scattering atoms and one type of anomalously scattering atoms, there are four unknown quantities in (12) when they are defined in a fashion to produce linearity.

In order to proceed, it is necessary to reduce by at least one the number of unknown quantities. This may be done by determining the value of $|F_{2,\mathbf{h}}|^2$ from statistical considerations and holding it fixed or determining the structure of the anomalous scatterers or measuring the data for the unsubstituted structure or a combination of the latter two possibilities. Knowledge of the structure of the anomalous scatterers gives values for the $|F_{2,\mathbf{h}}|^2$ and $\varphi_{2,\mathbf{h}}^n$ and measurement of the data for the unsubstituted structure gives values for the $|F_{1,\mathbf{h}}|^2$. With knowledge of the structure of the anomalous scatterers, potentially accurate information is obtainable from a one-wavelength experiment concerning the values of the desired $\varphi_{1,\mathbf{h}}^n$, the phases of the unsubstituted structure (Karle, 1989). If only one of $|F_{1,\mathbf{h}}|^2$ and $|F_{2,\mathbf{h}}|^2$ is a known quantity, the system of equations contains a twofold ambiguity. The selection of the appropriate solution has been found to

be generally achievable by solving the equations with the use of a least-squares technique in which the process is initiated with statistically reasonable starting values for the unknown quantities.

Relative scaling in multiple-wavelength experiments

In the application of the systems of simultaneous equations to anomalous-dispersion data, it is important that the simultaneous equations be properly relatively scaled. The containment of macromolecular crystals in capillaries may add to the difficulties in making appropriate absorption corrections. To effect appropriate scaling, use of the following relationship was suggested (Karle, 1984*a*):

$$|F_{\mathbf{h}}^n| \approx 0.5 W (|F_{\lambda\mathbf{h}}| + |F_{\lambda\bar{\mathbf{h}}}|), \quad (18)$$

where $|F_{\mathbf{h}}^n|$ is the structure-factor magnitude for the crystal of interest in the absence of anomalous scattering and

$$W = \left(\sum_{j=1}^{N_{\text{non}}} f_{j\mathbf{h}}^2 + \sum_{j=1}^{N_{\text{ano}}} f_{j\mathbf{h}}^{n^2} \right)^{1/2} \times \left\{ \sum_{j=1}^{N_{\text{non}}} f_{j\mathbf{h}}^2 + \sum_{j=1}^{N_{\text{ano}}} [(f_{j\mathbf{h}}^n + f_j')^2 + f_j'^2] \right\}^{-1/2}. \quad (19)$$

The quantities N_{non} and N_{ano} are the number of normally scattering atoms and anomalously scattering atoms in the unit cell, respectively. Tests have shown (Karle, 1984*a*) that the right-hand side of (18) is rather independent of wavelength, even when the relative amount of anomalous scattering exceeds that normally encountered in macromolecular crystallography. Relative scaling among various wavelengths can be achieved by rescaling the sums of the measured quantities on the right of (18) so that they would all agree, for example, with that at the shortest wavelength used in the measurements. If the latter is short enough to give a negligible correction for absorption, the scaling corrects for absorption at all the other wavelengths.

Applications of linear multiple-wavelength theory

Several multiple-wavelength investigations have made use of the exact linear equations (12) or an equivalent form. The applications to data have been in the nature of experimental tests on known structures and investigations of unknown structures. The focus of the investigations on macromolecules has been mostly on structures that have relatively weak anomalous scatterers as inherent parts of the structure. In one case a relatively weak anomalous scatterer was added. None of the stronger anomalous scatterers was present in the studies. In the circumstances, careful experimentation was employed involving such equipment as tunable high-intensity synchrotron sources and area detectors.

A test experiment has been performed on lamprey hemoglobin (Hendrickson, Smith, Phizackerley & Merritt, 1988). Investigations on new structures by Hendrickson and colleagues have concerned ferredoxin from *Clostridium acidi-urici* (Murthy, Hendrickson, Orme-Johnson, Merritt & Phizackerley, 1988) and streptavidin from *Streptomyces avidinii* (Hendrickson, Pähler, Smith, Satow, Merritt & Phizackerley, 1989). For the lamprey hemoglobin data set that was 87% complete out to 3 Å spacings, the average of the magnitudes of the discrepancies between the phases from multiple-wavelength anomalous dispersion and those from the least-squares-refined atomic model was 50.5°. The analysis of ferredoxin was limited because of weak data to 5 Å resolution, but the phase values obtained from the data were comparable to those of a previously determined related ferredoxin. For streptavidin, three-wavelength data measured at the Photon Factory from the selenobiotinyl complex produced a fine map at 3.3 Å resolution. An atomic model fitted to this map has been refined to $R = 0.17$ with 2 Å data collected with Cu $K\alpha$ radiation (Pähler, Hendrickson & Satow, 1989). Hendrickson anticipates the possibility of developing general techniques for the application of multiple-wavelength anomalous dispersion in which methionine is replaced by selenomethionine in proteins and bromination is applied to nucleic acids.

An application of the multiple-wavelength-anomalous-dispersion technique, with the use of (12), to the structure determination of cucumber basic protein has been reported by several collaborative groups (Guss, Merritt, Phizackerley, Hedman, Murata, Hodgson & Freeman, 1988). This is another example of a study in which the anomalous effects from the anomalous scatterer, an inherent Cu atom, were relatively weak. Full advantage was taken of current experimental techniques that involve high-intensity tunable X-ray sources and area detectors for data collection.

The use of L edges in anomalous-dispersion experiments

Templeton, Templeton, Phillips & Hodgson (1980) presented a study in which they emphasized the value of making use of L absorption edges in anomalous-dispersion experiments. They found, for example, that near the L_3 edge of caesium f' varied between -26.7 and -13.9 and f'' between 4.0 and 16.1 electrons in a wavelength interval of 0.008 Å. They also found that the experimental measurements for f' and f'' contain details that do not appear in the smoother theoretical curves, owing to certain approximations in the calculations. This suggests the virtue of making experimental determinations of f' and f'' in the vicinity of an absorption edge when it is considered important to take advantage of their detailed variations. Theoretical values should be sufficiently accurate for

applications as the wavelength departs from the absorption edge. Since values for f' and f'' are tabulated only for $K\alpha$ radiation from various specific sources, the use of most edges, L or otherwise, would require their determination by experimental means.

In an investigation concerned with the application of L edges in multiple-wavelength-anomalous-dispersion experiments, Chapuis, Templeton & Templeton (1985) applied the technique to crystals of $\text{NaHo}(\text{edta}) \cdot 8\text{H}_2\text{O}$ and $\text{NaSm}(\text{edta}) \cdot 8\text{H}_2\text{O}$. With the use of the exact linear algebra in a form equivalent to (12), both structure determinations were successful. Especially high accuracy for the phases (average error of 5°) was obtained from synchrotron radiation.

In order to facilitate multiwavelength experiments, an apparatus for making simultaneous measurements of anomalous-dispersion data over a range of wavelengths has been described by Arndt, Greenough, Helliwell, Howard, Rule & Thompson (1982).

Role of isomorphous replacement in the algebraic equations

It may occur that, in isomorphous replacement, the native substance scatters essentially normally and anomalous dispersion arises from a single type of anomalous scatterer in the heavy-atom derivative. In that case the system of equations arising from (12) would apply with the added advantage that intensity data measured for the native substance would provide values for the $|F_{1,h}^n|^2$ that occur in (12), thereby decreasing the number of unknown quantities to be evaluated. If the positions of the heavy atoms have been determined, the only unknown quantity remaining in the equations for $|F_{\lambda,h}|^2$ and $|F_{\lambda,\bar{h}}|^2$ is $\varphi_{1,h}^n$, which may then be readily evaluated.

In case the isomorphous structures each contain significant anomalous scatterers, it would be necessary to make more detailed use of (12) by appropriately including the effects of the various anomalous scatterers. If there is more than one type of anomalous scatterer in any structure, the general form of the theory mentioned earlier (Karle, 1980) would come into consideration.

Triplet phase invariants from anomalous-dispersion data: probabilistic results

One objective of formulas for triplet phase invariants based on anomalous-dispersion data is to provide suitable mathematics for proceeding with phase determination when the structure of the anomalous scatterers is not known and, presumably, does not yield to Patterson or direct-methods analysis.

A formula for the sine of an average triplet phase invariant has been derived by Heinerman, Krabben-dam, Kroon & Spek (1978) from probabilistic

considerations. The result is

$$\sin \bar{\varphi}_{\mathbf{hk}} = (|\tau_{\mathbf{hk}}|^2 - |\tau_{\bar{\mathbf{h}\bar{\mathbf{k}}}}|^2) \times \{4\tau_{\mathbf{hk}}''[\frac{1}{2}(|\tau_{\mathbf{hk}}|^2 + |\tau_{\bar{\mathbf{h}\bar{\mathbf{k}}}}|^2) - |\tau_{\mathbf{hk}}''|^2]\}^{-1/2} \quad (20)$$

where $\bar{\varphi}_{\mathbf{hk}} = 0.5(\varphi_{\mathbf{hk}} - \varphi_{\bar{\mathbf{h}\bar{\mathbf{k}}}}$, $\varphi_{\mathbf{hk}}$ is the phase of $\tau_{\mathbf{hk}} = F_{\mathbf{h}}F_{\mathbf{k}}F_{\bar{\mathbf{h}+\bar{\mathbf{k}}}}$ and $\tau_{\mathbf{hk}}''$ is the contribution of the imaginary part of $\tau_{\mathbf{hk}}$. It is defined approximately for the case of two identical anomalous scatterers in the structure by

$$\tau_{\mathbf{hk}}'' \approx 2f''[f(\mathbf{h})f(\mathbf{k}) + f(\mathbf{h})f(\bar{\mathbf{h}} + \bar{\mathbf{k}}) + f(\mathbf{k})f(\bar{\mathbf{h}} + \bar{\mathbf{k}})] \times [1 + K(|E_{\mathbf{h}}|^2 + |E_{\mathbf{k}}|^2 + |E_{\bar{\mathbf{h}+\bar{\mathbf{k}}}}|^2 - 3)], \quad (21)$$

where $f(\mathbf{h})$ is the total real part of the atomic scattering factor, *i.e.* the sum of the non-anomalous contribution and the real part of the anomalous contribution, f'' is the imaginary part of the anomalous contribution,

$$K = Z^2\sigma_2/(\sigma_2^2 - \sigma_4). \quad (22)$$

Z is the atomic number of the anomalously scattering atoms,

$$\sigma_n = \sum_{j=1}^n Z_j^n \quad (23)$$

and Z_j is the atomic number of the j th atom in a unit cell containing N atoms. The E are normalized structure factors.

Test calculations were performed with (20) and gave encouraging results for the accuracy of the invariants. The problem of choosing between φ and $\pi - \varphi$ was handled by selecting the alternative that was closer to zero.

Hauptman (1982*b*) derived the conditional probability distribution for triplet phase invariants, given the values of $|E_H|$, $|E_K|$, $|E_L|$, $|E_{\bar{H}}|$, $|E_{\bar{K}}|$, $|E_{\bar{L}}|$, with the use of the joint probability distribution of the corresponding six normalized structure factors, where $H + K + L = 0$. Giacovazzo (1983) published a similar theory. The result obtained by Hauptman is

$$P_j(\Omega_j) = (1/K_j) \exp [A_j \cos (\Omega_j - \omega_j)] \quad (24)$$

where Ω_j represents the value of a triplet phase invariant, and j ranges over eight values, each representing one of the eight triplet phase invariants that can be composed from the six E 's listed above. The definitions of A_j , K_j and ω_j are fairly complex and so the reader is referred to the publication for details. Since the K_j 's and A_j 's are positive, the maximum of the right-hand side of (24) occurs when $\Omega_j = \omega_j$. Thus, ω_j is an estimate of Ω_j and, the larger the value of A_j , the more reliable it is. The estimate is unique in the entire interval $(-\pi, +\pi)$ and is explicitly expressed in terms of the complex scattering factors, f_{jH} , f_{jK} , f_{jL} , presumed to be known, and the observed magnitudes $|E_H|$, $|E_K|$, $|E_L|$, $|E_{\bar{H}}|$, $|E_{\bar{K}}|$, $|E_{\bar{L}}|$. No knowledge of the positions of the anomalous scatterers is required, nor do the anomalous scatterers need to be identical. It is necessary, however, to know

the types of anomalous scatterers, the number of atoms in each type and the occupancy of their sites.

A test calculation was performed on a PtCl_4^{2-} derivative of cytochrome c550. The atoms Pt, Cl, Fe and S were treated as anomalous scatterers at the wavelength of Cu $K\alpha$ radiation. Test groups of triplet invariants ranged from 100 to 60 000 with corresponding average values for A ranging from 6.01 to 2.15. The average magnitude of the error for calculations based on exact data ranged from 28 to 40°.

Fortier, Fraser & Moore (1986) have analyzed sources of error in probabilistic formulas for the evaluation of triplet phase invariants from anomalous-dispersion data. They have also noted the enhanced accuracy that accrues from making use of information concerning the structure of the anomalous scatterers and indicate how this may be effected in the use of the conditional joint probability distributions.

A probabilistic approach to the development of formulas for evaluating triplet phase invariants composed of a mixture of phases defined for anomalous-dispersion data at two different wavelengths or for anomalous-dispersion and isomorphous-replacement data has been presented by Pontenagel, Krabbendam, Peerdeman & Kroon (1983).

Bryan (1988) has discussed the derivation of probability distributions for isomorphous replacement and anomalous dispersion by use of a maximum-entropy formalism. Use was made of two density maps, one representing a native structure and the other representing heavy atoms or anomalous scatterers. A series expansion of the total Shannon-Jaynes entropy of both maps gave either the single-isomorphous-replacement or anomalous-dispersion probability distributions. For isomorphous replacement, for example, the assumption of known heavy-atom positions was not made. The total entropy of both maps was maximized subject to the constraints imposed by the intensities of both the native and derivative structures. The single-anomalous-scattering result was obtained similarly.

Triplet phase invariants from anomalous-dispersion data: algebraic results

In a manner similar to that used for isomorphous replacement, algebraic analyses were performed to evaluate triplet phase invariants for anomalous-dispersion data. Evaluations were based on physical and mathematical criteria and some probabilistic considerations that led to rules for evaluating triplet phase invariants having particular values (Karle, 1984*b, c*). A general rule was developed that included the results from both isomorphous replacement and anomalous dispersion and a variety of combinations that could lead to evaluations of triplet phase invariants at many points between $-\pi$ and π (Karle,

1984*d*). A more detailed combination of isomorphous replacement and anomalous dispersion led to results of potentially high accuracy (Karle, 1985*a*). An algebraic analysis that depends upon knowledge of the structure of the anomalous scatterers, or heavy atoms in isomorphous replacement, led to the possibility of determining values for triplet phase invariants anywhere between $-\pi$ and π with high accuracy (Karle, 1986). It has also been shown how triplet phase invariants for anomalous-dispersion data can be readily formed and evaluated from the two-phase invariants that arise in the exact linear algebraic analysis (Karle, 1984*e*). If measurements of intensity data are made at several wavelengths, the additional information has the potential for increasing the reliability of the estimates of the triplet phase invariants.

By starting with the structure-factor equation for anomalous dispersion

$$F_{\lambda h} = F_h^n + F_{\lambda h}^a \quad (25)$$

we can write

$$F_{\lambda h} - F_{\lambda \bar{h}}^* = F_h^n + F_{\lambda h}^a - F_{\bar{h}}^{n*} - F_{\lambda \bar{h}}^{a*} = F_{\lambda h}^a - F_{\lambda \bar{h}}^{a*} \quad (26)$$

where $F_{\lambda h}$ is the structure factor for the total scattering from all the atoms including the anomalously scattering ones, F_h^n is the structure factor that would obtain if all atoms scattered normally and $F_{\lambda h}^a$ is the structure factor that represents the scattering from only the anomalous corrections to the atomic scattering factors.

It has been shown (Karle, 1984*b*) that

$$F_{\lambda h}^a - F_{\lambda \bar{h}}^{a*} = 2i \sum_{m=1}^p (f_{\lambda m}^a / f_{m,h}^n) \sin \delta_{\lambda m} F_{m,h}^n \quad (27)$$

where

$$f_{\lambda m}^a = (f_{\lambda m}^{\prime 2} + f_{\lambda m}^{\prime \prime 2})^{1/2} \quad (28)$$

$$\delta_{\lambda m} = \tan^{-1} (f_{\lambda m}^{\prime \prime} / f_{\lambda m}^{\prime}). \quad (29)$$

The scattering factor $f_{m,h}^n$ is the normal part of the scattering factor for the m th type of atom and $F_{m,h}^n$ is the normal structure factor for the m th type of atom in a structure containing p types of atoms. For those types that make only an insignificant contribution to the anomalous scattering, the corresponding $f_{\lambda m}^a$ could be set equal to zero.

To illustrate the use of (26) and (27), it is now assumed that only one type of atom scatters anomalously and that the remainder do not do so to a significant extent. By combining (26) and (27), one may write

$$\begin{aligned} & -8i (f_{\lambda q}^a / f_{q,h}^n f_{q,k}^n f_{q,\bar{h}+\bar{k}}^n) \sin^3 \delta_{\lambda q} |F_{q,h}^n F_{q,k}^n F_{q,\bar{h}+\bar{k}}^n| \\ & \simeq (|F_{\lambda h}| - |F_{\lambda \bar{h}}|) (|F_{\lambda k}| - |F_{\lambda \bar{k}}|) \\ & \quad \times (|F_{\lambda(\bar{h}+\bar{k})}| - |F_{\lambda(h+k)}|) \exp(i\Phi_{hk}), \end{aligned} \quad (30)$$

where Φ_{hk} represents the average of eight triplet phase invariants formed from $\varphi_{\lambda h}$ or $-\varphi_{\lambda \bar{h}}$, $\varphi_{\lambda k}$ or $-\varphi_{\lambda \bar{k}}$,

and $\varphi_{\lambda(h+k)}$ or $-\varphi_{\lambda(\bar{h}+\bar{k})}$. The basis for taking the average is that for the largest values of $||F_{\lambda h}| - |F_{\lambda \bar{h}}||$, the phase of $F_{\lambda h}$, $\varphi_{\lambda h}$, will differ little in value from the phase of $F_{\lambda \bar{h}}^*$, $-\varphi_{\lambda \bar{h}}$. The structure associated with only those atoms that scatter anomalously to a significant extent is usually quite simple so that if the $|F_{q,h}^n F_{q,k}^n F_{q,\bar{h}+\bar{k}}^n|$, where q labels the type of atom that scatters anomalously, are within their larger magnitude range, the associated triplet phase invariants are equal to zero, to good approximation. This accounts for the absence of an expression for a triplet phase invariant from the left-hand side of (30). Evidently (30) will be most accurate when the magnitude of the triple product of F on the left-hand side of (30) and the differences on the right-hand side are among the largest. Relation (30) may be compared with (10) for isomorphous replacement.

The quantity on the left-hand side of (30) is a pure imaginary. This implies that, to good approximation, Φ_{hk} has a value of either $\pi/2$ or $-\pi/2$ and leads to the following interpretation when the triple product on the right-hand side of (30) has large differences (Karle, 1984*b*):

Rule, $R_{\text{ano},1}$: *If the sign of the product of the magnitude differences, $(|F_{\lambda h}| - |F_{\lambda \bar{h}}|)(|F_{\lambda k}| - |F_{\lambda \bar{k}}|) \times (|F_{\lambda(\bar{h}+\bar{k})}| - |F_{\lambda(h+k)}|)$, is the same as the sign of $\sin \delta_{\lambda q}$ (or, equivalently, the sign of $f_{\lambda q}^{\prime \prime}$), the value of the average invariant, Φ_{hk} , is close to $-\pi/2$ and when the signs are opposite, the value is close to $\pi/2$.*

This estimate is, in effect, assigned to all eight triplet phase invariants from which the average is formed. Alternatively, the estimate may be assigned only to those triplet phase invariants associated with the larger products of structure-factor magnitudes occurring among the eight possibilities. This would be expected to increase the accuracy somewhat, but would appear to complicate seriously the practical utility of the result. It would seem that the use of average triplet invariants with their redundant interrelationships in phase determination would offer a great simplification over the use of the larger number of individual invariants. Data accuracy may also not justify the use of any but average triplet phase invariants.

Similar to the circumstances with R_{iso} , it is possible, simply by visual inspection of the magnitude differences in $R_{\text{ano},1}$, to evaluate triplet phase invariants when there is one predominant type of anomalous scatterer. It is not necessary to know the positions or occupancies of the anomalous scatterers. The presence of more than one type of anomalous scatterer brings in a more complex form for (27) whose calculation requires knowledge of the types, positions and occupancies of the anomalous scatterers.

Additional algebraic formulas that include anomalous dispersion and isomorphous replacement were also derived (Karle, 1985a). In the test calculations made with these formulas on cytochrome c550.PtCl₄²⁻, the average errors were quite small, ranging from 6 to about 30°. Heavy-atom information also enhances the accuracy of the triplet phase invariants (Fortier, Fraser & Moore, 1986; Karle, 1986).

Triplet phase invariants from exact linear equations for anomalous-dispersion data

Triplet phase invariants can also be evaluated from use of the exact linear algebraic equations (Karle, 1984a). As described in connection with (12), solution of these equations gives values for numerous $\varphi_{1,h}^n - \varphi_{2,h}^n$. It is possible to form sums of suitable phase differences, such as

$$\varphi_{1,h}^n + \varphi_{1,k}^n + \varphi_{1,(\bar{h}+\bar{k})}^n - \varphi_{2,h}^n - \varphi_{2,k}^n - \varphi_{2,(\bar{h}+\bar{k})}^n = A_{hk}, \quad (31)$$

where A_{hk} would be known from the evaluation of the three individual phase differences in (31). The first set of three phases comprises a triplet phase invariant for the structure consisting of non-anomalous scattering atoms and the second for the anomalously scattering atoms. Because the structures of the predominant anomalously scattering atoms in macromolecules are generally simple, the triplet phase invariants associated with the larger products $|F_{2,h}^n F_{2,k}^n F_{2,(\bar{h}+\bar{k})}^n|$ for such structures may be quite reliably set equal to zero. Under these circumstances, we get

$$\varphi_{1,h}^n + \varphi_{1,k}^n + \varphi_{1,(\bar{h}+\bar{k})}^n \approx A_{hk}, \quad (32)$$

a result which illustrates the evaluation of triplet phase invariants for the structure of the normally scattering atoms, e.g. an unsubstituted native protein. A further discussion of this formula and some others that have been presented in this paper may be found in a review (Karle, 1987).

It may be noted that probability distributions may be readily associated with the various algebraic formulas for evaluating triplet phase invariants. If the values for the triplet phase invariants are considered to be expected values, we may readily employ the central limit theorem. This also entails an evaluation of the variance, presenting an opportunity to introduce a measure of the effect of errors in the data as well as other sources of uncertainty.

Refinement

Some of the techniques for the refinement of macromolecular structures are based on formulas that had their origin in the development of direct methods for small structures. The techniques concern use of the

tangent formula (Karle & Hauptman, 1956; Karle, 1968), the equation of Sayre (1952, 1972, 1974) and the higher-order determinants (Tsoucaris, 1970, 1980; Karle, 1971) associated with the determinantal inequalities that arise from the non-negativity of the electron density distributions in crystals (Karle & Hauptman, 1950). These techniques have not achieved widespread use up to this time, presumably because there are alternative techniques that investigators have found to be preferable.

Tangent formula

The tangent formula is given in terms of the normalized structure factors, $|E|$, and their associated phases, φ , by

$$\tan \varphi_h = \left(\sum_k |E_k E_{h-k}| \sin(\varphi_k + \varphi_{h-k}) \right) \times \left[\sum_k |E_k E_{h-k}| \cos(\varphi_k + \varphi_{h-k}) \right]^{-1}. \quad (33)$$

With small molecules, the tangent formula has been used for direct phase determination, phase extension and phase refinement. Phase extension and refinement has also been carried out with macromolecules with the use of (33). Applications have been made to cytochrome c by Weinzierl, Eisenberg & Dickerson (1969), to carboxypeptidase A by Reeke & Lipscomb (1969) and to sperm whale myoglobin by Coulter (1971) with rather limited success. Some improvement in the results is obtained from use of the tangent formula when information from isomorphous-replacement or anomalous-dispersion techniques is closely involved in the analysis. Investigations including the latter have been performed, for example, on carp muscle calcium-binding protein (Hendrickson & Karle, 1973) and are further described in a review article by Hendrickson (1973).

An investigation of the structure of avian pancreatic polypeptide, a small (36-residue) globular protein hormone, was carried out by Blundell, Pitts, Tickle, Wood & Wu (1981). In this study, phases for data at 2.1 Å resolution were calculated from a combination of single isomorphous replacement with anomalous dispersion. These phase values were used with (33), modified by Hull & Irwin (1978) to give a reweighting of individual terms to extend phase values to 1.4 Å resolution. The resulting electron density map was of high quality and led readily to a solution of the structure.

A modification of the tangent formula described by Olthof, Sint & Schenk (1979) is based on an estimate of the expected values of the magnitudes of the triplet phase invariants occurring in the tangent formula as a function of their associated structure-factor magnitudes. Thus, the triplet phase invariants are biased toward their expected values. The signs of the expected values may be obtained from knowledge

of the approximate values of the phases. Some calculations with this modification of the tangent formula have been applied by Olthof & Schenk (1982) to phase extension and refinement of metmyoglobin.

Sayre formula

A least-squares phase refinement has been presented by Sayre (1972) which is based on a formula published earlier (Sayre, 1952). The function $S(\Phi)$ is minimized with respect to the phases represented collectively by Φ , where

$$S(\Phi) = \frac{1}{2} \sum_{\mathbf{h}} \left| \left[\left(\sum_{\mathbf{k}} f_{\mathbf{k}} f_{\mathbf{h}-\mathbf{k}} \right) / f_{\mathbf{h}} \right] F_{\mathbf{h}} - \sum_{\mathbf{k}} F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}} \right|^2. \quad (34)$$

Tests have been made on the protein rubredoxin by Sayre (1974) and the method has been applied to insulin by Cutfield *et al.* (1975). Previous to the phase refinement, heavy-atom phases were extended from 2.5 and 1.9 Å resolution, respectively, to 1.5 Å resolution in both applications. Maps computed from the refined phases at 1.5 Å were readily interpretable in most regions.

Determinants

The non-negativity of the electron density in a crystal was expressed in terms of non-negativity of an infinite set of determinants of increasing order whose elements are the structure factors (Karle & Hauptman, 1950). This result was a development of the early work of Toeplitz (1911) on non-negative Fourier series. It was shown by Tsoucaris (1970) for these determinants that the values of the phases of the elements in the last row and column are most probably those that maximize the value of a determinant when the values of the elements in the remainder of the determinant are known. It has been thought that the maximum-determinant rule is more general and would apply to the complete set of phase values for all the elements in a determinant, although this has not been proved. Some supporting theoretical evidence has come from the work of Heinerman, Krabbendam & Kroon (1979), Heinerman, Kroon & Krabbendam (1979) and analyses by Karle (1978) and by Tsoucaris (1980). Additional indications have come from computations of Taylor, Woolfson & Main (1978).

Another approach to the use of the determinants of various orders is in the context of the refinement of phases. This could be done by making use of the definition of the value of a single structure factor in terms of the values of the remaining structure factors in a determinant. The method for doing this with respect to determinants of any order has been illus-

trated (Karle & Hauptman, 1950). It was suggested on this basis (Karle, 1971) that an iterative phase refinement could be carried out with the use of determinants of various orders. The phases were explicitly defined in terms of a generalized tangent formula based on determinants of various orders including the high-order ones. It was shown that the tangent formula (33) derives from the third-order determinant. Efforts have been made to make use of the higher-order determinants for phase extension and refinement of macromolecules. A test application has been made by de Rango, Mauguen, Tsoucaris, Dodson, Dodson & Taylor (1985) on experimental data from insulin in which phase extension was applied to data at 1.9 Å resolution from isomorphous replacement to obtain phase values at 1.5 Å resolution. In five cycles of refinement on data that were extended from 2390 reference data having $|E| > 1.0$ at 1.9 Å resolution to 5050 data having $|E| > 1.0$ at 1.5 Å resolution, the average of the magnitudes of the errors in the phases decreased from 48.9 to 38.3°. Comparable improvements were obtained in a variety of samplings.

Another application of the determinants arising from the non-negativity of the electron density distribution in a crystal is the phasing of low-resolution data from yeast initiator tRNA by Podjarny, Schevitz & Sigler (1981). The starting phase set consisted of 107 reflections in the resolution range 32–14 Å which had been phased by multiple isomorphous replacement. This starting set was used with the determinants to phase 28 strong low-resolution (100–19 Å) structure factors that had not been phased with multiple isomorphous replacement. The authors report that the extension improved the electron density map, permitting the establishment of a well defined molecular boundary and providing the basis for a successful structure determination to 4.0 Å resolution. The accuracy of the phases determined by use of the determinants compared favorably with that of the starting set.

It is possible that lower-order determinants may be useable for phase refinement in the form of the tangent formulas that are based on the determinants (Karle, 1971). Such tangent formulas are generally to be used in their determinantal form, but the tangent formula based on the fourth-order determinant contains few enough terms to permit them to be readily written explicitly. This is done here. If, in obtaining the result from the fourth-order determinant, quasi-normalized structure factors, \mathcal{E} , are replaced by normalized ones, E , and a weighting factor for each term which emphasizes the importance of large $|E|$ values is omitted, we obtain

$$\begin{aligned} \tan \varphi_l = & [\text{Im}(t_l) - (n_{\mathbf{h}\mathbf{k}} E_{000})^{-1} \text{Im}(q_l)] \\ & \times [\text{Re}(t_l) - (n_{\mathbf{h}\mathbf{k}} E_{000})^{-1} \text{Re}(q_l)]^{-1} \end{aligned} \quad (35)$$

where

$$t_l = \sum_h E_h E_{l-h} \quad (36)$$

$$q_l = \sum_h E_{l-h} \sum_k E_k E_{h-k} \quad (37)$$

and n_{hk} is the number of terms for the sum over k for a given h . Equations (35), (36) and (37) may be compared with equations (3), (4), (5) and (6) of Debaerdemaeker, Tate & Woolfson (1988) (DTW). The sum terms are the same, but the coefficients are differently defined and somewhat difficult to compare. The tangent formula of DTW was derived from the introduction of the condition that the resulting phases tend towards satisfying the Sayre equation. Woolfson & Yao Jia-xing (1988) have used the tangent formula of DTW to extend 129 reflections for a small protein, avian pancreatic polypeptide, to 1371 additional ones. Random values were given to the 1371 new phases followed by phase refinement. In 20 trials, 11 gave mean phase errors less than 34° for all 1500 reflections.

Because of the increasing restrictiveness of the determinants as their order increases, it may be worthwhile to consider moderately higher-order determinantal tangent formulas (Karle, 1971) as well as very high-order ones as possible sources for phase refinement and extension. As noted, the determinantal tangent formulas are readily computed in that form.

Molecular replacement

Molecular-replacement techniques have served as a means of refinement for a number of macromolecular structures. The techniques make use of the occurrence of a structural unit in more than one crystallographic environment. An important example is the use of non-crystallographic symmetry. Another is the use of a molecular structure that occurs in more than one crystal form. Techniques have been developed for applications in both direct and reciprocal space.

It is apparent that the placement in space of similar structural units needs to be properly described. For this purpose, appropriate rotation matrices and translation vectors must be determined. In this, the Patterson function plays a central role. Rotational parameters are usually readily found. Translational parameters present greater difficulties.

The techniques have been generally used with, at least, single isomorphous replacement, although occasionally homologous structures have been used. The heavy-atom positions obtained from the heavy-atom isomorphous derivative facilitate the location of the molecular symmetry axes and the isomorphy helps in defining the molecular boundaries. The heavy-atom derivative also facilitates the establishment of an enantiomorph.

The earliest work on non-crystallographic symmetry was carried out by Rossmann & Blow (1962) in which they investigated the rotational relationship between molecules. The determination of translational vectors was investigated by Rossmann, Blow, Harding & Collier (1964). Theoretical and practical aspects of the molecular replacement technique as well as a description of many applications have been presented in an extensive review article by Argos & Rossmann (1980) and a further discussion of this subject has been given by Rossmann (1982).

Investigations of viral structure have benefited particularly by the use of non-crystallographic symmetry because of its extensive occurrence in virus structures.

The potential to solve a virus structure by sole use of non-crystallographic symmetry may possibly be realized at some future time. Rossmann (1982, p. 171) suggests the 'possibility for experimentation, such as perhaps *ab initio* determination of phases using a noncentric distribution of molecular envelopes filled with uniform density as a starting structure. Isomorphous replacement would then become superfluous in the structure determination of viruses where there are numerous noncrystallographic asymmetric units per crystallographic asymmetric unit.'

Restrained refinement

Techniques for phase refinement based on formulas arising from direct methods for small structures have had little impact on the practice of macromolecular crystallography. In contrast, the structure-refinement method known as restrained least-squares refinement, developed for application to macromolecules, has provided accurate coordinate refinements that approach those for small molecules in several instances. It was introduced by Konnert, Hendrickson & Karle (1975) and developed further by Sussman, Holbrook, Church and Kim (1977) and Hendrickson & Konnert (1979). An additional development in restrained refinement has involved the inclusion of molecular mechanics and dynamics to facilitate convergence (Brünger, Kuriyan & Karplus, 1987). A computer program, *X-PLOR*, includes this development.

Some observations

For the most part, this paper has been concerned with some of the more recent theoretical and practical developments in the techniques of isomorphous replacement and anomalous dispersion. The latter heavy-atom techniques have served well since the 1950's in the remarkable accomplishments of macromolecular crystallography. A variety of developments have been discussed in this paper and several have already proven their value. Optimal use of the

individual techniques will generally occur in combination with others. We consider here some of the ways that the newer developments may play a role in future analyses.

Most of the mathematical studies have concerned two-phase and three-phase (triplet-phase) invariants. Evidently, in evaluating their applicability, their characteristics are a prime consideration. An inspection of (10) and (30) for isomorphous replacement and anomalous dispersion, respectively, shows that the exponential function of the triplet phase invariants for the heavy-atom structure (or anomalous scatterers) on the left-hand side of the equations (when not set equal to zero) may be moved to the right-hand side of the argument of the exponential. This shows that the same values as are assigned to the triplet phase invariants by Rule R_{iso} and Rule $R_{ano,1}$ can also be assigned to the sum of three-phase differences. The latter point has also been noted by Yang, Xu, Furey & Wang (1984) and Xu, Yang, Furey, Sax, Rose & Wang (1984). This is a clear example of the ambiguousness associated with triplet phase invariants and the fact that the most internally consistent set of phases may not be the correct one. These are properties that are well known to workers in the crystallography of small structures where triplet phase invariants play a key role in the application of the direct-methods procedures (e.g. Karle, 1961; Karle & Karle, 1964). In fact, a key problem in the application of direct methods to small structures is the overcoming of the ambiguities. Another type of ambiguity that may occur for small structures leads to a Fourier map with a single large peak unrelated to the position of any atom in the structure. This can arise when only essentially equal non-hydrogen atoms are present and is clearly unrelated to the presence of a heavy atom. Such a calculation could be quite misleading in applications to isomorphous replacement by implying a false location for a heavy atom. In small-structure applications, chemical knowledge plays a significant role in the recognition of the correct structure among a number of ambiguous alternatives. The generally low resolution of diffraction data for macromolecular structure limits the usefulness of chemical knowledge in resolving the ambiguities of triplet phase invariants. These various problems which would be further exacerbated by errors in the data suggest that the use of triplet phase invariants for *ab initio* phase determination, whether their evaluation comes from probabilistic or algebraic analyses, may not be the method of choice.

There are alternative approaches based on the two-phase invariants that obviate the ambiguity problem. In this connection, it is worthwhile to determine initially the heavy-atom positions. This may be achieved in a number of ways from use of isomorphous-replacement or anomalous-dispersion data, e.g. by use of difference Patterson maps, the determi-

nation of the intensity of scattering appropriate to an isolated structure of anomalous scatterers by use of anomalous dispersion followed by the calculation of a Patterson map or the application of direct methods, or the use of non-crystallographic symmetry when available. With information concerning heavy-atom positions and occupancy, a number of algebraic analyses become available for phase evaluation from two-phase invariants without ambiguity.

In extreme cases in which only single-isomorphous-replacement or one-wavelength-anomalous-dispersion data are available, it is still possible to proceed since it is possible to obtain a number of phase values from simple algebraic formulas, e.g. (7) and (8), which can then be refined and extended by filtering, perhaps by use of probabilistic formulas for resolving ambiguities, by use of non-crystallographic symmetry when suitable circumstances prevail or, perhaps, by use of higher-order determinantal tangent formulas, by use of maximum entropy or by use of triplet phase invariants. Phase extension and refinement may be a useful application of triplet phase invariants.

There is one circumstance in which the use of triplet phase invariants may be the method of choice. This would occur in single isomorphous replacement when it may not be possible to locate the heavy atoms and anomalous-dispersion experiments are not accessible. Such a circumstance does not appear to be a frequently occurring one.

When it is difficult to make isomorphous heavy-atom derivatives but there may be inherent anomalous scatterers in a macromolecule, the techniques of multiple-wavelength anomalous dispersion or single-wavelength anomalous dispersion combined with filtering can play a valuable role in structure determination. It is also worthwhile to consider the use of heavy-atom derivatives even though they are not isomorphous since anomalous-dispersion techniques are available. Recombinant DNA procedures may provide a method for obtaining appropriate derivatives in addition to the more common ways of producing them. We note that the new applications of the linear algebraic theory for multiple-wavelength-anomalous-dispersion data have been confined to structures with anomalous scatterers of modest power. This bodes well for future applications with stronger anomalous scatterers.

The great potential complexity in the analysis of virus structures has been overcome in several instances by the occurrence and use of non-crystallographic symmetry. The latter technique has greatly facilitated the determination of such structures.

Analytical techniques continue to be developed and the extent of their usage will depend upon a number of the usual considerations such as applicability, capability, accuracy and convenience. Whatever the ultimate fate of the current mathematical and procedural techniques, there is no doubt that

macromolecular crystallography will continue to make extensive fundamental and practical contributions to molecular biology and medicine. Some of its broad manifestations for the public good will be recognizable from medical advances, the facilitation of the design of new and valuable pharmaceuticals, and the production of modified macromolecules that are destined to play an important role in a variety of industrial applications.

References

- ANDERSON, W. F., FLETTERICK, R. J. & STEITZ, T. A. (1974). *J. Mol. Biol.* **86**, 261-269.
- ARGOS, P. & ROSSMAN, M. G. (1980). In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. LADD & R. A. PALMER, pp. 361-417. New York: Plenum.
- ARNDT, U. W., GREENHOUGH, T. J., HELLIWELL, J. R., HOWARD, J. A. K., RULE, S. A. & THOMPSON, A. W. (1982). *Nature (London)*, **298**, 835-838.
- BIJVOET, J. M. (1954). *Nature (London)*, **173**, 888-891.
- BLUNDELL, T. L., PITTS, J. E., TICKLE, I. J., WOOD, S. P. & WU, C.-W. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 4175-4179.
- BRÜNGER, A. T., KURIYAN, K. & KARPLUS, M. (1987). *Science*, **235**, 458-460.
- BRYAN, R. K. (1988). *Acta Cryst.* **A44**, 672-677.
- BRYAN, R. K. & BANNER, D. W. (1987). *Acta Cryst.* **A43**, 556-564.
- CASCARANO, G., GIACOVAZZO, C., PEERDEMAN, A. F. & KROON, J. (1982). *Acta Cryst.* **A38**, 710-717.
- CHAPUIS, G., TEMPLETON, D. H. & TEMPLETON, L. K. (1985). *Acta Cryst.* **A41**, 274-278.
- COCHRAN, W. (1955). *Acta Cryst.* **8**, 473-478.
- COULTER, C. L. (1971). *Acta Cryst.* **B27**, 1730-1740.
- CUTFIELD, J. F., DODSON, E. J., DODSON, G. G., HODGKIN, D. C., ISAACS, N. W., SAKABE, K. & SAKABE, N. (1975). *Acta Cryst.* **A31**, S21.
- DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1988). *Acta Cryst.* **A44**, 353-357.
- FAN HAI-FU & GU YUAN-XIN (1985). *Acta Cryst.* **A41**, 280-284.
- FAN HAI-FU, HAN FU-SON, QIAN JIN-ZI & YAO JIA-XING (1984). *Acta Cryst.* **A40**, 489-495.
- FORTIER, S., FRASER, M. E. & MOORE, N. J. (1986). *Acta Cryst.* **A42**, 149-156.
- FORTIER, S., MOORE, N. J. & FRASER, M. E. (1985). *Acta Cryst.* **A41**, 571-577.
- FORTIER, S., WEEKS, C. M. & HAUPTMAN, H. (1984). *Acta Cryst.* **A40**, 646-651.
- GIACOVAZZO, C. (1983). *Acta Cryst.* **A39**, 585-592.
- GUSS, J. M., MERRITT, E. A., PHIZACKERLEY, R. P., HEDMAN, B., MURATA, M., HODGSON, K. O. & FREEMAN, H. C. (1988). *Science*, **241**, 806-811.
- HAUPTMAN, H. (1982a). *Acta Cryst.* **A38**, 289-294.
- HAUPTMAN, H. (1982b). *Acta Cryst.* **A38**, 632-641.
- HAUPTMAN, H. & KARLE, J. (1953). *Am. Crystallogr. Assoc. Monogr.* No. 3. Western Springs: Polycrystal Book Service.
- HAUPTMAN, H., POTTER, S. & WEEKS, C. M. (1982). *Acta Cryst.* **A38**, 294-300.
- HEINERMAN, J. J. L., KRABBENDAM, H. & KROON, J. (1979). *Acta Cryst.* **A35**, 101-105.
- HEINERMAN, J. J. L., KRABBENDAM, H., KROON, J. & SPEK, A. L. (1978). *Acta Cryst.* **A34**, 447-450.
- HEINERMAN, J. J. L., KROON, J. & KRABBENDAM, H. (1979). *Acta Cryst.* **A35**, 105-107.
- HENDRICKSON, W. A. (1973). *Trans. Am. Crystallogr. Assoc.* **9**, 61-63.
- HENDRICKSON, W. A. & KARLE, J. (1973). *J. Biol. Chem.* **248**, 3327-3334.
- HENDRICKSON, W. A. & KONNERT, J. H. (1979). In *Biomolecular Structure, Conformation, Function and Evolution*, edited by R. SRINIVASAN, pp. 43-57. New York: Pergamon Press.
- HENDRICKSON, W. A., PÄHLER, A., SMITH, J. L., SATOW, Y., MERRITT, E. A. & PHIZACKERLEY, R. P. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 2190-2194.
- HENDRICKSON, W. A., SMITH, J. L., PHIZACKERLEY, R. P. & MERRITT, E. A. (1988). *Proteins*, **4**, 77-88.
- HULL, E. & IRWIN, M. J. (1978). *Acta Cryst.* **A34**, 863-868.
- KARLE, I. L. (1961). *Acta Cryst.* **14**, 497-502.
- KARLE, I. L. & KARLE, J. (1964). *Acta Cryst.* **17**, 1356-1360.
- KARLE, J. (1968). *Acta Cryst.* **B24**, 182-186.
- KARLE, J. (1971). *Acta Cryst.* **B27**, 2063-2065.
- KARLE, J. (1978). *Proc. Natl Acad. Sci. USA*, **75**, 2545-2548.
- KARLE, J. (1980). *Int. J. Quantum Chem. Symp.* **7**, 357-367.
- KARLE, J. (1983). *Acta Cryst.* **A39**, 800-805.
- KARLE, J. (1984a). *Acta Cryst.* **A40**, 1-4.
- KARLE, J. (1984b). *Acta Cryst.* **A40**, 4-11.
- KARLE, J. (1984c). *Acta Cryst.* **A40**, 366-373.
- KARLE, J. (1984d). *Acta Cryst.* **A40**, 374-379.
- KARLE, J. (1984e). *Acta Cryst.* **A40**, 526-531.
- KARLE, J. (1985a). *Acta Cryst.* **A41**, 182-189.
- KARLE, J. (1985b). *Acta Cryst.* **A41**, 387-394.
- KARLE, J. (1986). *Acta Cryst.* **A42**, 246-253.
- KARLE, J. (1987). In *Chemical Crystallography with Pulsed Neutrons and Synchrotron X-rays*, edited by M. A. CARRONDO & G. A. JEFFREY, pp. 370-418. Dordrecht: Reidel.
- KARLE, J. (1988). *Comput. Math. Appl.* **16**, 397-406. Reprinted in *Crystal Symmetries*, edited by I. HARGITTAI & B. K. VAINSHTEIN. Oxford: Pergamon Press.
- KARLE, J. (1989). *Acta Cryst.* **A45**, 303-307.
- KARLE, J. & HAUPTMAN, H. (1950). *Acta Cryst.* **3**, 181-187.
- KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635-651.
- KLOP, E. A., KRABBENDAM, H. & KROON, J. (1987). *Acta Cryst.* **A43**, 810-820.
- KONNERT, J. H., HENDRICKSON, W. A. & KARLE, J. (1975). *Proc. Third East Coast Protein Crystallography Workshop*, Eastover, Massachusetts, USA.
- MURTHY, H. M. K., HENDRICKSON, W. A., ORME-JOHNSON, W. H., MERRITT, E. A. & PHIZACKERLEY, R. P. (1988). *J. Biol. Chem.* **263**, 18430-18436.
- NAVIA, M. A. & SIGLER, P. B. (1974). *Acta Cryst.* **A30**, 706-712.
- NEIDLE, S. (1973). *Acta Cryst.* **B29**, 2645-2647.
- OLTHOF, G. J. & SCHENK, H. (1982). *Acta Cryst.* **A38**, 117-122.
- OLTHOF, G. J., SINT, L. & SCHENK, H. (1979). *Acta Cryst.* **A35**, 941-946.
- PÄHLER, A., HENDRICKSON, W. A. & SATOW, Y. (1989). Personal communication.
- PEERDEMAN, A. F. & BIJVOET, J. M. (1956). *Acta Cryst.* **9**, 1012-1015.
- PODIARNY, A. D., SCHEVITZ, R. W. & SIGLER, P. B. (1981). *Acta Cryst.* **A37**, 662-668.
- PONTENAGEL, W. M. G. F., KRABBENDAM, H., PEERDEMAN, A. F. & KROON, J. (1983). Eighth Eur. Crystallogr. Meet., 8-12 August, Abstr. 4.01-P, p. 257.
- RAMASESHAN, S. & ABRAHAMS, S. C. (1975). Editors. *Anomalous Scattering*. Copenhagen: Munksgaard.
- RANGO, C. DE, MAUGUEN, Y., TSOUCARIS, G., DODSON, E. J., DODSON, G. G. & TAYLOR, D. J. (1985). *Acta Cryst.* **A41**, 3-17.
- REEKE, G. N. & LIPSCOMB, W. N. (1969). *Acta Cryst.* **B25**, 2614-2623.
- ROSSMANN, M. G. (1982). In *Computational Crystallography*, edited by D. SAYRE, pp. 159-173. Oxford: Clarendon Press.
- ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24-31.
- ROSSMANN, M. G., BLOW, D. M., HARDING, M. M. & COLLIER, E. (1964). *Acta Cryst.* **17**, 338-342.

- SAYRE, D. (1952). *Acta Cryst.* **5**, 60–65.
 SAYRE, D. (1972). *Acta Cryst.* **A28**, 210–212.
 SAYRE, D. (1974). *Acta Cryst.* **A30**, 180–184.
 SCHEVITZ, R. W., NAVIA, M. A., BANTZ, D. A., CORNICK, G., ROSA, J. J., ROSA, M. D. H. & SIGLER, P. B. (1972). *Science*, **177**, 429–431.
 SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.
 STEITZ, T. A. (1968). *Acta Cryst.* **B24**, 504–507.
 STEITZ, T. A., FLETTERICK, R. J. & HWANG, K. J. (1973). *J. Mol. Biol.* **78**, 551–561.
 STEWART, J. M., KARLE, J., IWASAKI, H. & ITO, T. (1977). *Acta Cryst.* **A33**, 519.
 SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S.-H. (1977). *Acta Cryst.* **A33**, 800–804.
 TAYLOR, D. J., WOOLFSON, M. M. & MAIN, P. (1978). *Acta Cryst.* **A34**, 870–883.
 TEMPLETON, D. H., TEMPLETON, L. K., PHILLIPS, J. C. & HODGSON, K. O. (1980). *Acta Cryst.* **A36**, 436–442.
 TIMKOVICH, R. & DICKERSON, R. E. (1976). *J. Biol. Chem.* **251**, 4033–4046.
 TOEPLITZ, O. (1911). *Rend. Circ. Mat. Palermo*, pp. 191–192.
 TSOUCARIS, G. (1970). *Acta Cryst.* **A26**, 492–499.
 TSOUCARIS, G. (1980). In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. LADD & R. A. PALMER, pp. 287–360. New York: Plenum Press.
 WANG, B. C. (1985). In *Diffraction Methods in Biological Macromolecules*, edited by H. WYCKOFF, pp. 90–112. New York: Academic Press.
 WEINZIERNER, J. E., EISENBERG, D. & DICKERSON, R. E. (1969). *Acta Cryst.* **B25**, 380–387.
 WESTBROOK, E., PIRO, O. E. & SIGLER, P. B. (1984). *J. Biol. Chem.* **259**, 9096–9103.
 WILSON, K. S. (1978). *Acta Cryst.* **B34**, 1599–1608.
 WOOLFSON, M. M. (1954). *Acta Cryst.* **7**, 61–64.
 WOOLFSON, M. M. (1984). *Acta Cryst.* **A40**, 32–34.
 WOOLFSON, M. M. & YAO JIA-XING (1988). *Acta Cryst.* **A44**, 410–413.
 XU, Z. B., YANG, D. S. C., FUREY, W. JR, SAX, M., ROSE, J. & WANG, B. C. (1984). *Am. Crystallogr. Assoc. Meet.*, 20–25 May, Abstr. PC2, p. 50.
 YANG, D. S. C., XU, Z. B., FUREY, W. JR & WANG, B. C. (1984). *Am. Crystallogr. Assoc. Meet.*, 20–25 May, Abstr. PC1, p. 50.
 YAO JIA-XING (1981). *Acta Cryst.* **A37**, 642–644.
 YAO JIA-XING & FAN HAI-FU (1985). *Acta Cryst.* **A41**, 284–285.

Acta Cryst. (1989). **A45**, 781–786

From a Partial to the Complete Crystal Structure. III. A New Procedure and its Applications

BY M. C. BURLA

Dipartimento di Scienze della Terra, Università, 06100 Perugia, Italy

G. CASCARANO

Dipartimento Geomineralogico, Università, Campus Universitario, 70124 Bari, Italy

V. FARES

ITSE, CNR, CP 10, 00016, Monterotondo Stazione, Roma, Italy

C. GIACOVAZZO

Dipartimento Geomineralogico, Università, Campus Universitario, 70124 Bari, Italy

G. POLIDORI

Dipartimento di Scienze della Terra, Università, 06100 Perugia, Italy

AND R. SPAGNA

Istituto di Strutturistica Chimica 'G. Giacomello', CNR, CP 10, 00016 Monterotondo Stazione, Roma, Italy

(Received 23 February 1989; accepted 1 June 1989)

Abstract

The procedure described in the first two papers of this series [Giacovazzo (1983). *Acta Cryst.* **A39**, 685–692; Camalli, Giacovazzo & Spagna (1985). *Acta Cryst.* **A41**, 605–613] for recovering from a partial crystal structure the complete one has been reconsidered. Several modifications have been introduced which make the procedure more efficient. Furthermore, a new method is described which is able to combine prior information on a located fragment with the second representation formula *P10*. Experimental

tests show that very small fragments are sufficient to recover the complete crystal structure.

Symbols and abbreviations

Throughout the paper a number of symbols will find frequent application. For the sake of simplicity they are here listed together.

f(**h**): atomic scattering factor. The thermal factor is included; anomalous dispersion is not considered.

N: number of atoms in the unit cell.