Acta Cryst. (1967). **23**, 50

Phase Determination Using Non-crystallographic Symmetry

BY PETER MAIN

Department of Biological Sciences, Purdue University, Lafayette, Indiana, U.S.A.

(*Received* 9 *September* 1966)

A three-dimensional hypothetical structure containing four crystallographically independent but chemically identical molecules in space group $P1$ has been solved. Each molecule contained ten carbon atoms separated by distances greater than 1·2 Å. The solution of the phases required only a knowledge of the structure amplitudes and the relative orientations and positions of the molecules.

1. Introduction

In a paper by Main & Rossmann (1966) (hereafter MR) there was described a method of phase determination which depended upon having chemically identical molecules in different crystallographic environments. Equating the electron densities of such molecules places restriction on the phases which are expressed by equation (10) of MR.

$$|\mathbf{F}_p| \exp\{i\alpha_p\} = \frac{U}{V} \sum_{h=-\infty}^{+\infty} |\mathbf{F}_h| \exp\{i\alpha_h\} \sum_{n=1}^{N} G_{hpn} \exp\{2\pi i(\mathbf{p}\,.\,\mathbf{S}_n - \mathbf{h}\,.\,\mathbf{S})\} \quad (1)$$

where $|\mathbf{F}_p|$, α_p, $|\mathbf{F}_h|$ and α_h are the structure amplitudes and their phases at the reciprocal lattice points $\mathbf{p}$ and $\mathbf{h}$ in either the same or different crystals. Each molecule is enclosed in an envelope of volume U, the centres of the N molecular envelopes in the 'p' crystal being at $\mathbf{S}_n(n=1,2\ldots N)$, while $\mathbf{S}$ is the centre of a molecular envelope in the 'h' crystal. The function G_{hpn} is the magnitude of the Fourier transform of the molecular envelope which is given both in magnitude and phase by

$$G_{hpn}\{\exp i\Omega_{hpn}\} = \int_U \exp\{2\pi i(\mathbf{p}\,.\,[\mathbf{C}_n] - \mathbf{h}\,.\,[\mathbf{C}])\mathbf{x}\}d\mathbf{x}\,, \quad (2)$$

where $[\mathbf{C}_n]$ is the rotation matrix describing the orientation of the nth molecule in crystal 'p' and $[\mathbf{C}]$ is the rotation matrix corresponding to the molecule centred on $\mathbf{S}$ in crystal 'h'.

MR showed that, using these equations in a one-dimensional case, the phases α_p and α_h could be determined with sufficient accuracy for the structure to be recognized in the resulting Fourier synthesis. This paper describes the method of phase determination used here and by MR, as well as the application of the technique to the solution of a hypothetical three-dimensional structure in the space group $P1$.

2. Method of phase determination

Because of the nature of the function G_{hpn}, the largest terms on the right hand side of equation (1) will tend to have $[\mathbf{h}]\approx[\mathbf{p}]$ or, more specifically, $\mathbf{h}$ and $\mathbf{p}$ will be such that the vector $(\mathbf{p}\,.\,[\mathbf{C}_n]-\mathbf{h}\,.\,[\mathbf{C}])$ is small. Initially, the only known phase will be $\alpha_o(=2\pi)$, so that the first phases to be determined will be those for which $|\mathbf{p}|$ is small as these will have the largest interactions with $\mathbf{F}_o$. Next, equations with $|\mathbf{p}|$ a little larger are used since these will have large interactions with those phases previously determined. The equations are therefore arranged in increasing order of their Bragg angle and knowledge of the phases is gradually extended outwards in reciprocal space. This is similar in outline to the method of Rossmann & Blow (1963).

The actual process of phase determination is to take one new phase, α_p, on the edge of the known part of reciprocal space and find its 'best' value by a search procedure. This is done by choosing an equation with α_p on the left hand side and summing the right hand side over all known values of α_h. Arbitrary values of α_p at say 5° intervals are substituted into the equation and the discrepancy between the two sides is calculated for each value of α_p. This is repeated for each equation which contains α_p explicitly on the left hand side and the sum of all the discrepancies for each angle is computed. That value of α_p which gives the lowest total discrepancy is considered to be the best present estimate of the phase, subject to the error introduced by the lack of knowledge or inaccuracy of the phases α_h. In the event that $\mathbf{h}$ and $\mathbf{p}$ refer to the same crystal, terms involving α_p may occur also on the right hand side, that is, those terms for which $\mathbf{h}=\pm\mathbf{p}$. This procedure is similar to that described by Rossmann & Blow (1964).

As estimates of more phases become known, phases determined earlier may now be redetermined with more accuracy. The determination of a batch of phases is therefore followed by a refinement of all known phases before further phase determination takes place. The refinement consists simply of substituting the present estimate of the phases in the right hand sides of the equations and performing the vector summation. The argument of each resultant is then taken as the new estimate of the phase angle appearing on the left hand side of the equation. Whenever the same phase appears explicitly on the left hand side of more than one equation, the argument of the vector sum of all the right hand sides is taken as the new estimate. The new set of phases is only accepted, however, if it reduces the residual of the equations as defined by MR.

That this process is one of refinement can be seen by examining its real space equivalent. By referring to the derivation of the equations by MR it can be seen that any one equation represents the structure factor $\mathbf{F}_p$ calculated from the assumption that all the molecules in crystal '*p*' are identical with that centred on S in crystal '*h*'. There will be approximately one such equation for each crystallographically independent molecule, so taking the vector average of the right hand side sums of these equations is the same as calculating structure factors from an electron density which is the average over all the independent molecules. It is this averaging process which produces the refinement, but since there is nothing in the procedure which forces the residual to decrease, it will rise again after several cycles and begin to oscillate.

3. Application to a trial structure

Because the equations (1) are non-linear, there immediately arose the problem of how to solve them and whether the solution would be unique. As a preliminary answer to these questions, the equations were applied to a variety of one-dimensional problems and were found to be successful (MR). Recently, a more satisfactory test of the method has been completed in which a hypothetical structure in the space group $P1$ was solved. The structure (Fig. 1) was made up of four identical molecules arranged in different orientations in the unit cell ($a=11{\cdot}8$, $b=11{\cdot}1$, $c=6{\cdot}8$ Å, $\alpha=\beta=\gamma=90{\cdot}0°$), each molecule consisting of ten equal carbon atoms separated by distances greater than 1·2 Å. The size of the molecule is immaterial as far as the method is concerned but, in order to attain atomic resolution with a minimum amount of work, the 10-atom molecule was chosen. The molecular envelope was chosen to be a sphere of radius 3·4 Å and, in order to avoid errors due to atoms spilling over into the wrong envelope, the molecules were given a larger than normal separation. This resulted in the calculated density of the crystal being 0·90 g.cm^{-3} and in this respect the crystal was rather ideal. After calculating structure factors to 1·0 Å resolution, the equations were set up using only the structure amplitudes and a knowledge of the relative positions and orientations of the molecules. In all, there were 1838 unique reflexions in the 1·0 Å limiting sphere but only the 1113 largest were used in the calculations. The magnitude of G_{hpn} was assumed to be negligible for arguments greater than $2\pi\times1{\cdot}6$ and only the 60 largest terms were considered in each equation. (A previous attempt at solving the structure, using $2\pi\times1{\cdot}1$ as the maximum argument of G_{hpn} and 40 terms in each equation, proved to be only partially successful.)

Putting the correct phases into the equations and calculating the residual produced the results shown in

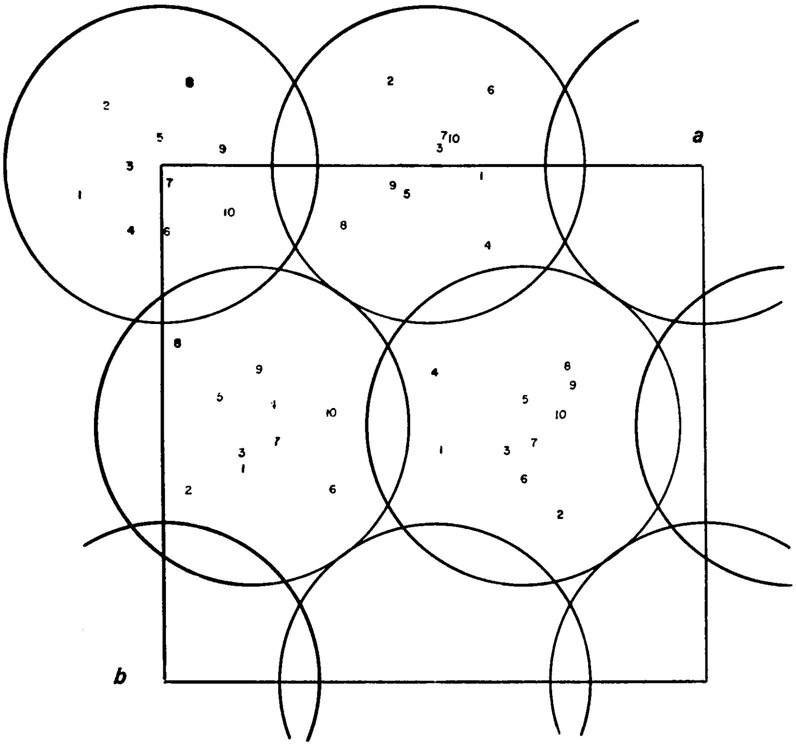

Fig. 1. Projections of the hypothetical structure onto the *ab* face of the unit cell. The unit cell contains four identical molecules in different orientations in the space group $P1$. Each molecule contains ten carbon atoms numbered 1–10.

Table 1. The overall residual was 22·7%, but doubtless this could be decreased by accepting larger arguments of G_{hpn} and including more terms in the equations. The average discrepancy in angle between the left and right hand sides of the equations was only about 10°, however.

Table 1. *Residuals obtained on substituting the correct phases into equations*

The residual R is defined as the sum of the magnitude of the lack of closure vectors of each equation divided by the sum of the magnitude of the left hand sides. The error $\Delta\alpha$ is the mean difference in angle between the two sides of the equation.

Resolution range	R	$\Delta\alpha$
∞ – 1·8 Å	18·8 %	9·4°
1·8 – 1·4	22·7	9·5
1·4 – 1·0	26·0	10·7

The method of phase determination outlined in § 2 was used, and when the calculations were terminated the average error in the 1113 phases determined was 13°. At this point the phases were still refining to more accurate values, but there seemed no need to pursue the calculations further as the structure was obviously solved and refinement was rather slow. A plot of error in phase angle against sin θ is shown in Fig. 2. The lower curve shows the errors at the termination of the calculations and the upper curve shows the errors at the stage where the phase determination had progressed to 1·35 Å resolution. Immediately after the 1·35 Å stage the accuracy of the determined phases increased dramatically. Presumably this was because a large number of atoms suddenly became resolved, changing the features of the electron density quite

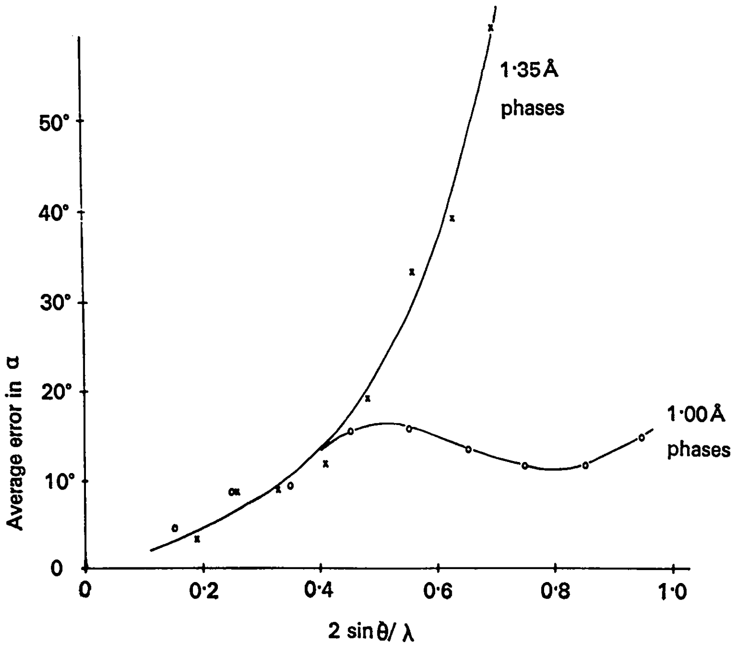

Fig. 2. Quality of phase determination as a function of resolution. The crosses show accuracy of phase determination using all equations up to a resolution 1·35 Å. An increase of resolution to 1·0 Å gave a dramatic improvement in phase determination as shown by the circles.

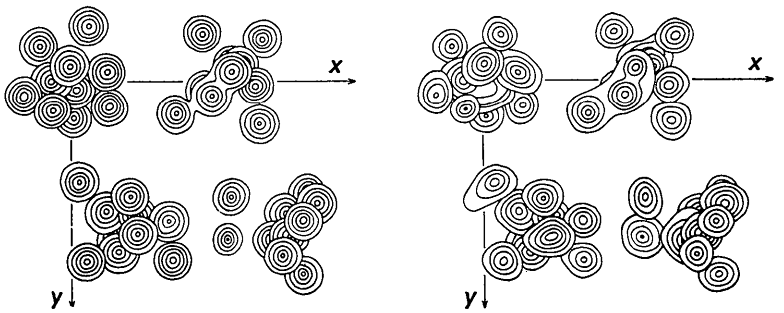

Fig. 3. Comparison of the three-dimensional electron density maps at 1·0 Å resolution viewed down the c axis. The map on the left has been computed with the correct phases, whereas the map on the right is based on the phases determined by the solution of the molecular replacement equations.

radically and allowing the residual of the equations to fall into a deep minimum.

The electron density calculated from the 1113 determined phases and the corresponding structure amplitudes, projected onto the *ab* face, is shown in Fig. 3. The true electron density calculated from all 1838 reflexions in the 1·0 Å sphere and the correct phases is also shown. All the features of the correct structure are clearly discernible in the determined structure – all the atoms are at least partially resolved, the vast majority being fully resolved, and no atom lies more than 0·15 Å away from its correct position, most of them being within 0·10 Å. The largest undesirable features in the determined structure are regions of negative electron density (some as low as $-1{\cdot}2$ e.Å^{-3}) between the molecules and outside of the molecular envelopes.

4. Discussion

An average error of only 13° in the phases of over 1000 reflexions may be attributed to the fact that the relative orientations and positions of all the molecules were known exactly, the molecular envelopes could be placed accurately and there was no experimental error in the structure amplitudes. In addition, there were four independent molecules, which is probably more than sufficient. The one-dimensional examples quoted by MR suggest that with only two independent molecules, the phase determination is likely to be less reliable and may even be quite random at high resolution.*

A number of questions remain unanswered by this trial of the molecular replacement method. It is not yet known what effect experimental error will have on the ability of the equations to determine accurate phases, although experiments with one-dimensional structures show that a small amount of error will have very little effect (MR). Errors in the placement† or shape‡ of the molecular envelopes will have an effect on the phase determination though, again, the extent of this effect has not been investigated. In the one-dimensional structures reported by MR the positions of the molecular envelopes were successfully refined by least-squares methods during phase determination, but in three dimensions this becomes an unwieldy computing problem and has not yet been programmed.

The application of the method described in this paper has been limited to the case of chemically identical but crystallographically independent molecules within the same crystal. Additional problems may exist, however, when the molecules are in different crystals. In the latter case, the structure factors $\mathbf{F}_p$ and $\mathbf{F}_h$ in equations (1) belong to different crystals, whereas in the problem treated in this paper they correspond to the same crystal.

The example given in this paper also assumes a knowledge of the absolute scale of the structure amplitudes. This is implied in the use of the $\mathbf{F}(000)$ term to determine some of the early phases. Furthermore, if this method is to be applied to a protein, we should consider the space outside the molecular envelopes to be filled uniformly by electron density representing the liquid of crystallization. As will be shown, these two factors are closely related.

Let us assume that all $\mathbf{F}_h$'s have been placed on a roughly absolute scale. Although it is possible to calculate $\mathbf{F}_h(000)$ from the total number of electrons in the cell, nevertheless, this value should be adjusted to be $k \,.\, \mathbf{F}_h(000)$ in order to bring it more accurately onto the same scale as the other structure amplitudes. Let ϱ_s be the average electron density of liquid between the molecular envelopes. We may now see that

$$\mathbf{F}_p = \sum_{n=1}^{N} \int_U \varrho(\mathbf{x}_n) \exp\{2\pi i \mathbf{p} \,.\, \mathbf{x}_n\} d\mathbf{x}_n + \int_{(V-NU)} \varrho_s \exp\{2\pi i \mathbf{p} \,.\, \mathbf{x}\} d\mathbf{x}$$

i.e.

$$\mathbf{F}_p = \sum_{n=1}^{N} \int_U \varrho(\mathbf{x}_n) \exp\{2\pi i \mathbf{p} \,.\, \mathbf{x}_n\} d\mathbf{x}_n + \int_V \varrho_s \exp\{2\pi i \mathbf{p} \,.\, \mathbf{x}\} d\mathbf{x} - \sum_{n=1}^{N} \int_U \varrho_s \exp\{2\pi i \mathbf{p} \,.\, \mathbf{x}_n\} d\mathbf{x}_n$$

$$\mathbf{F}_p = \sum_{n=1}^{N} \int_U [\varrho(\mathbf{x}_n) - \varrho_s] \exp\{2\pi i \mathbf{p} \,.\, \mathbf{x}_n\} d\mathbf{x}_n + \int_V \varrho_s \exp\{2\pi i \mathbf{p} \,.\, \mathbf{x}\} d\mathbf{x} \,.$$

The last term is zero unless $\mathbf{p}$ is zero, when this term is equal to $V\varrho_s$. Hence in all equations, apart from the $\mathbf{F}_p(000)$ equation, we simply replace $\varrho(\mathbf{x})$ by $[\varrho(\mathbf{x}) - \varrho_s]$, which is equivalent to subtracting $V\varrho_s$ from $k \,.\, \mathbf{F}_h(000)$. The equations therefore remain in exactly the same form apart from the interactions with $\mathbf{F}_h(000)$.

We may rewrite the interactions with the $\mathbf{F}_h(000)$ term as $[k(U/V)\mathbf{F}_h(000) - U\varrho_s]re^{i\varphi}$ when $|\mathbf{p}| \neq 0$, or by

---

* Since this paper was submitted, the author has shown that, using the techniques described here, phase determination with only two independent molecules does become random at high resolution, though not before useful structural information has been obtained.

† The order of magnitude of the error involved may be obtained as follows. Let us assume that the orientation error may be corrected by a rotation of $\delta\theta$ about an axis a distance r from the reference point S_n within the nth molecule. This corrects $\mathbf{S}_n$ to $\mathbf{S}_n + \delta\mathbf{S}_n$, where $\delta S_n = r \,.\, \delta\theta$ and the argument of the corresponding coefficient in the equation will be changed by $2\pi\mathbf{p} \,.\, \delta S_n$ radians. If we assume the magnitudes of all the coefficients with the same h are equal then the total error involved in the phase determined by that equation (i.e. α_p) is $2\pi\mathbf{p} \,.\, \sum_n \delta S_n$ radians. The value of $\delta\theta$ may be estimated from the rotation function and r will be approximately the distance between the non-crystallographic rotation axis and the centre of the molecule.

‡ A satisfactory envelope can be chosen for the monomer in R3 insulin from packing and other considerations. This envelope can be further refined from initial poor electron density maps. The degree of accuracy required in the first guess is, however, unknown and will undoubtedly vary from structure to structure.

$k(U/V)\mathbf{F}_h(000)$ when $|\mathbf{p}|=0$. Here $re^{i\varphi}$ represents the sum over n in equation (1). The constant quantity $[k(U/V)\mathbf{F}_h(000)-U\varrho_s]$ will appear in every equation except when $|\mathbf{p}|=0$, and may therefore be estimated by averaging the lack of closure of each equation. Also $k(1-U/V)\mathbf{F}_h(000)$ may be found from the $\mathbf{F}_p(000)$ equation. Hence k and ϱ_s can be determined, given reasonable initial estimates of these quantities. Since the size of the interaction $re^{i\varphi}$ decreases as $|\mathbf{p}|$ increases, the effect of salt concentration or an inaccurate knowledge of the absolute scale becomes rapidly less significant as we go out in reciprocal space.

The author is indebted to Dr M. G. Rossmann for his help and inspiration in this work, to Mr John Steele at Purdue University's Computer Sciences Center for giving valuable advice, and to Mrs Margaret Wonacott for aid in the preparation of the manuscript. This work was supported by N.I.H. grant GM 10704-03 and by N.S.F. grant GB 02905.

References

Main, P. & Rossmann, M. G. (1966). *Acta Cryst.* **21**, 67.
Rossmann, M. G. & Blow, D. M. (1963). *Acta Cryst.* **16**, 39.
Rossmann, M. G. & Blow, D. M. (1964). *Acta Cryst.* **17**, 1474.

Acta Cryst. (1967). **23**, 54

Least-Squares Weighting Schemes for Diffractometer – Collected Data

By R. C. G. Killean

Department of Physics, The University, St. Andrews, Scotland

(*Received* 7 *October* 1966 *and in revised form* 2 *March* 1967)

An analysis is made of various diffractometer techniques for obtaining data, and expressions are obtained for the weighting functions that may be used in subsequent least-squares refinement. It is shown that for a constant-time diffractometer experiment the weighting function is independent of the magnitude of the counts obtained and dependent on the diffractometer geometry. The results for the constant-count diffractometer experiment are compared.

Introduction

The most accurate set of parameters that can be obtained by the least-squares technique of minimizing the function

$$R=\sum_{\mathbf{h}} w(\mathbf{h})|\Delta(\mathbf{h})|^2$$

is when $w(\mathbf{h})$ is the reciprocal of the variance of $|\Delta(\mathbf{h})|$. The two forms of $|\Delta(\mathbf{h})|$ which have been most used in refining crystal structures are

$$|\Delta_1(\mathbf{h})|=\big||F_o(\mathbf{h})|-|F_c(\mathbf{h})|\big|$$

and

$$|\Delta_2(\mathbf{h})|=\big||F_o(\mathbf{h})|^2-|F_c(\mathbf{h})|^2\big|\,.$$

If the standard deviation of each $\big||F_o(\mathbf{h})|-|F_c(\mathbf{h})|\big|$ is small in comparison with $|F_o(\mathbf{h})|$ then

$$4|F_o(\mathbf{h})|^2w_2(\mathbf{h})=w_1(\mathbf{h})$$

where

$$R_1=\sum_{\mathbf{h}} w_1(\mathbf{h})|\Delta_1(\mathbf{h})|^2$$

and

$$R_2=\sum_{\mathbf{h}} w_2(\mathbf{h})|\Delta_2(\mathbf{h})|^2\,.$$

It is often assumed that the value of $w(\mathbf{h})$ is dependent only on the statistical errors in the experimentally observed structure factor, the non-random errors such as those associated with inaccurate instrument setting or the scattering factors making an unknown, and assumed negligible contribution to the standard deviation of $|\Delta(\mathbf{h})|$. This assumption can be shown to be valid provided

$$\sum w(\mathbf{h})|\Delta(\mathbf{h})|^2=m-n\,,$$

where m is the number of structure factors included in the summation, n is the number of variables and $w(\mathbf{h})$ are the weights on the absolute scale for the equation being minimized (*Computing Methods in Crystallography*, 1965).

It is possible to perform at least two different types of diffractometer experiment. An experiment may be performed in which an integrated count is obtained for each reflexion over a constant time and in some instruments such as the linear diffractometer of Arndt and Philips there is, owing to the present construction of the instrument, no option but to use this technique. Alternatively, with more versatile instruments, either this technique or a constant count technique in which the intensity is proportional to the inverse of the time taken to achieve this constant count may be used (Lonsdale, 1948).