Acta Cryst. (1984). A**40**, 32–34

The Solution of Protein Structures Containing Several Anomalous Scatterers of the Same Kind

BY M. M. WOOLFSON

Department of Physics, University of York, Heslington, York YO1 5DD, England

(*Received* 29 *April* 1983; *accepted* 5 *August* 1983)

Abstract

Many proteins contain a number of atoms of the same kind which are potential anomalous scatterers. It is shown that from two data sets, one normal and one with anomalous scattering, it is possible to extract the contribution of the anomalous scatterers and to solve for them by normal direct methods. From the phases of the structure factors of the anomalous scatterers alone the phases of the normal protein structure factors may, in principle, be found exactly.

Introduction

With the availability of synchrotron radiation much recent attention has been directed towards the use of anomalous scattering as a tool for solving crystal structures. Karle (1980) has described a technique which involves collecting data at several wavelengths and which can be made to yield a system of linear equations from which the separate contributions of each of several sets of different kinds of anomalous scatterer can be derived.

Here we shall deal with the case of protein structures containing many anomalous scatterers all of the same kind as may occur, for example, if the structure contains several disulphide bridges.

The basic idea

In Fig. 1, $F(\mathbf{h})$ and $F(\bar{\mathbf{h}})$ represent Friedel-related structure factors in the absence of anomalous scattering. We consider the case of m similar anomalous scatterers for which the *extra* component of the scattering factor is

$$f_e = f' + if''. \qquad (1)$$

The extra contribution to the structure factor of index $\mathbf{h}$ is

$$F_e(\mathbf{h}) = \sum_{j=1}^{m} f_e \exp(2\pi i \mathbf{h} . \mathbf{r}_j), \qquad (2)$$

where the summation is over all the anomalous scatterers. This may be written as

$$F_e(\mathbf{h}) = |f_e|\, e^{i\delta} \sum_{j=1}^{m} \exp(2\pi i \mathbf{h} . \mathbf{r}_j), \qquad (3)$$

where

$$\tan\delta = f''/f'$$

and this contribution makes an angle $\theta + \delta$ with $F(\mathbf{h})$ as shown in Fig. 1.

The extra contribution to the structure factor of index $-\mathbf{h}$, $F_e(\bar{\mathbf{h}})$, has the same magnitude as $F_e(\mathbf{h})$ but makes an angle $(\theta - \delta)$ with $F(\bar{\mathbf{h}})$.

Hence we may write for the structure factors of index $\mathbf{h}$ and $\bar{\mathbf{h}}$ with anomalous scatterering

$$|F_A(\mathbf{h})|^2 = |F(\mathbf{h})|^2 + |F_e(\mathbf{h})|^2 + 2|F(\mathbf{h})F_e(\mathbf{h})| \cos(\theta + \delta) \qquad (4a)$$

and

$$|F_A(\bar{\mathbf{h}})|^2 = |F(\mathbf{h})|^2 + |F_e(\mathbf{h})|^2 + 2|F(\mathbf{h})F_e(\mathbf{h})| \cos(\theta - \delta). \qquad (4b)$$

Writing $|F_e(\mathbf{h})|/|\mathrm{F}(\mathbf{h})| = g$,

$$\frac{|F_A(\mathbf{h})|^2}{|F(\mathbf{h})|^2} - 1 = \Delta(\mathbf{h})$$

and (5)

$$\frac{|F_A(\bar{\mathbf{h}})|^2}{|F(\mathbf{h})|^2} - 1 = \Delta(\bar{\mathbf{h}})$$

one obtains for equations (4)

$$g^2 + 2g\cos(\theta + \delta) - \Delta(\mathbf{h}) = 0 \qquad (6a)$$

and

$$g^2 + 2g\cos(\theta - \delta) - \Delta(\bar{\mathbf{h}}) = 0. \qquad (6b)$$

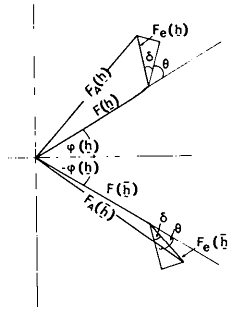

Fig. 1. The composition of the structure factors with anomalous scattering, $F_A(\mathbf{h})$ and $F_A(\bar{\mathbf{h}})$, in the terms of the normal structure factors, $F(\mathbf{h})$ and $F(\bar{\mathbf{h}})$, and the anomalous contributions, $F_e(\mathbf{h})$ and $F_e(\bar{\mathbf{h}})$.

0108-7673/84/010032-03$01.50

Since δ is fixed, and the same for all $\mathbf{h}$, then values of g and θ can be found for any pair of values $\Delta(\mathbf{h})$ and $\Delta(\bar{\mathbf{h}})$.

The solutions of equations (6) are illustrated in Fig. 2 for $\delta = 30°$. This figure shows lines of constant θ and, with the exception of $\theta = 180°$, the values of g are marked at intervals of 0·1. For $\theta = 180°$ values of g first increase with distance from the origin. However, $g = 0{\cdot}9$ corresponding to $\Delta(\mathbf{h}) = \Delta(\bar{\mathbf{h}}) = -0{\cdot}75$ represents an extremum and no pair of values further from the origin in the same direction on the 180° line are possible. Subsequently, the $\theta = 180°$ line turns back along the same path, passes through the origin and then overlaps the line $\theta = 0°$. This pattern may be illustrated by the lines $\theta = 175$ and 185° which turn backwards after a tight hairpin bend; in the case of $\theta = 180°$ the turning backwards is along exactly the same path, which makes the behaviour more difficult to illustrate.

Values of Δ correspond to the fractional increases in the intensity of a reflexion when anomalous scattering occurs. Clearly $\Delta \geq -1$, which is seen in Fig. 1, but in principle Δ can take large positive values. Actually, for all but the very smallest values of $|F(\mathbf{h})|$ the values of $|\Delta|$ are unlikely to exceed 0·5 or so and will usually be much smaller. In fact the maximum value of g is

$$g_{\max} = m|f_e|/|F(\mathbf{h})| \tag{7}$$

and its r.m.s. value

$$g_{\text{r.m.s.}} = m^{1/2}|f_e|/|F(\mathbf{h})|. \tag{8}$$

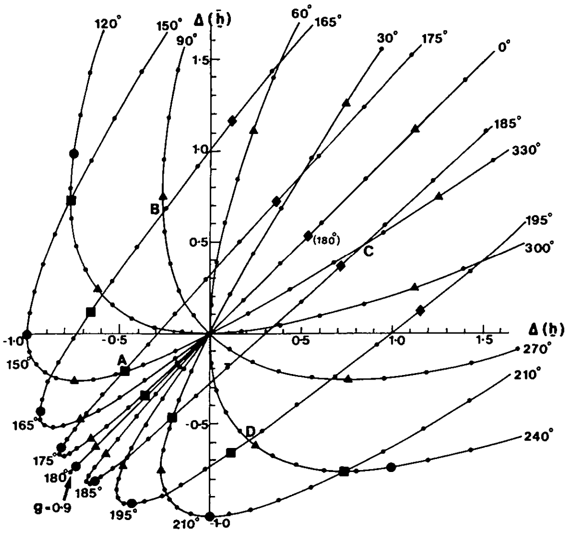

Fig. 2. Lines of constant θ for various values of $\Delta(\mathbf{h})$, $\Delta(\bar{\mathbf{h}})$ and $\delta = 30°$. Values of g are marked at intervals of 0·1 with special markings for $g = 0{\cdot}5$ ▲, $g = 1{\cdot}0$ ●, $g = 1{\cdot}5$ ■ and $g = 2{\cdot}0$ ◆. For the line $\theta = 180°$ only the special-marking points are given together with the point $g = 0{\cdot}9$ at which the line turns back on to itself.

For large m it is unlikely that a value of g much larger than $3g_{\text{r.m.s.}}$ will occur.

While Fig. 2 indicates that for each pair $\Delta(\mathbf{h})$ and $\Delta(\bar{\mathbf{h}})$ there is a pair of solutions (g, θ), in practice it will be very rare for both solutions to be possible. Even if both solutions are theoretically possible then one of them (usually the one with the smaller g) will be much more probable. This is illustrated by the pairs of solutions for the marked points in Fig. 2.

A	(0·27, 150°)	(1·50, 175°)
B	(0·45, 90°)	(1·79, 165°)
C	(0·36, 330°)	(2·05, 185°)
D	(0·47, 240°)	(1·56, 195°).

Except for very small values of $|F(\mathbf{h})|$ the first of each pair of solutions will be by far the more probable and, indeed, the second solutions may be impossible.

A procedure for structural solution

For the success of this method the number of anomalous scatterers should be sufficiently large to give measurable values of Δ but not so huge that the anomalous scatterers alone constitute an impossibly difficult structure to solve. An ideal case for a small protein might be to have forty or so anomalous scatterers in the unit cell.

Even at modest resolution the number of data from a protein will be very large and the number of values of g which can be reliably found will be much greater than is required precisely to define the positions of the anomalous scatterers. A subset of the largest $g|F(\mathbf{h})|$ values can be used as an input to a direct-methods program and an E map showing characteristic features, *e.g.* disulphide bridge atoms, should readily be recognised.

From the coordinates of the anomalous scatterers there may be calculated

$$Q(\mathbf{h}) = \sum_{j=1}^{m} \exp(2\pi i\mathbf{h}\,.\,\mathbf{r}_j) = |Q(\mathbf{h})| \exp\{i\psi(\mathbf{h})\} \tag{9}$$

and then the phase of $F(\mathbf{h})$ is given by

$$\varphi(\mathbf{h}) = \psi(\mathbf{h}) - \theta, \tag{10}$$

where θ is the angle associated with g in Fig. 2.

Thus, in principle, it is possible unambiguously to phase the protein reflexions.

Discussion

The extent to which this procedure will be successful will depend on the accuracy with which the data can be measured – in particular values of $\Delta(\mathbf{h})$ and $\Delta(\bar{\mathbf{h}})$. Thus an experimental procedure in which the *difference* in intensity is detected when the wavelength moves to the absorption edge would be preferred to finding differences from complete data sets with possible scaling errors.

Errors in assessing g values are probably not too critical in the sense that a sufficient number of reliably indicated values will usually be available. Any g not reliably estimated may simply be rejected. It will also be seen from Fig. 2 that even if the accuracy of $\Delta(\mathbf{h})$, $\Delta(\bar{\mathbf{h}})$ measurements is sufficient only to be in the correct quadrant of the diagram an error in θ [and hence in $\varphi(\mathbf{h})$] will result which will be quite acceptable by the standards of protein crystallography.

Although the determination of (g, θ) solutions has been explained in terms of the diagram in Fig. 2, in practice these can be deduced for any δ, $\Delta(\mathbf{h})$ and $\Delta(\bar{\mathbf{h}})$ from a simple computer program.

Tests need to be done to confirm or otherwise the effectiveness of this procedure. It is hoped to be able to report on such tests in due course.

Reference

KARLE, J. (1980). *Int. J. Quantum Chem.* 7, 357–367.

Acta Cryst. (1984). A**40**, 34–42

Dualistic Interpretation of the Symmetry of Incommensurate Structures

BY P. M. DE WOLFF

Vakgroep Vaste-stof fysica, Lab. v. Technische Natuurkunde, Lorentzweg 1, *Postbus* 5046, 2600 *GA Delft,*
The Netherlands

(*Received* 20 *June* 1983; *accepted* 3 *August* 1983)

Abstract

The symmetry of an incommensurately modulated structure may be described in terms of the basic structure (B) and the modulation pattern (M). This description contrasts with the superspace-group approach, in which the structure is defined in a space of dimension $3+d$, where d is the number of rationally independent modulation vectors. Space groups G_B and G_M are defined, consisting of symmetry operations of B and M which are simply interrelated. These groups together characterize the total symmetry; they lead to a classification which for $d = 1$ is equivalent to the superspace groups. With this dualistic approach, all symmetry operations can be based on symmetry elements in the space of the crystal, and the lattice types can be composed simply from those of G_B and G_M.

1. The dualistic approach

We shall exemplify the modulation phenomenon by a hypothetical two-dimensional compound OX, in which the X atoms are not modulated at all. The basic structure is orthogonal, plane group *pmm*, basic vectors a, b; $Z = 1$ (Fig. 1). We assume modulation of the O atoms with a wave vector $\beta\mathbf{b}^*$, where $\beta = b/\lambda$ is an irrational number, λ being the modulation wavelength. Then for a transverse displacive type of modulation, the crystal might look as in Fig. 2. The periodicity in the direction of $\mathbf{b}$ is lost – at least when 'periodicity' is taken in the usual sense. However, there is still perfect order in this direction – only there are two periods instead of one, *viz* b and λ.

The structure has, therefore, a dual character, being composed of the basic structure and a modulation pattern. The latter is indicated by the dashed wavy line in Fig. 2. Such a line becomes ambiguous as soon as a longitudinal component of the displacement is present (Fig. 3). In that case, a complete vector function is required, representing the displacement vector $\mathbf{u}$ as a periodic function of Y, the coordinate in the direction of $\mathbf{b}$. It is important to specify that $\mathbf{u}(Y_1)$ is the displacement of an atom which in the basic structure (not in the modulated one) has $Y = Y_1$.

In this way the modulated structure of Fig. 3, too, can be decomposed into two components: the basic structure and $\mathbf{u}(Y)$. Each of these is strictly periodic within the two-dimensional space considered here. Therefore the symmetry of the crystal can be

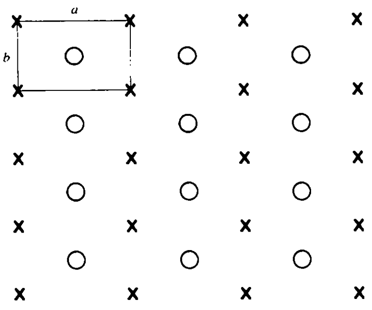

Fig. 1. Two-dimensional crystal XO, basic structure for the modulated structures in Figs. 2–5.

0108-7673/84/010034-09$01.50