RESEARCH PAPERS

Acta Cryst. (1994). D**50**, 7–10

On the Application of One-Wavelength Anomalous Scattering. IV. The Absolute Configuration of the Anomalous Scatterers

BY M. M. WOOLFSON AND YAO JIA-XING

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

(*Received* 18 *June* 1993; *accepted* 2 *August* 1993)

Abstract

An essential first step in most techniques for using anomalous-scattering data for phase determination is to determine the positions of the anomalous scatterers. This is usually done by use of the anomalous differences, either as input to a direct-methods procedure or to produce a Patterson map. If the arrangement of anomalous scatterers is non-centrosymmetric then it is also necessary to find their absolute configuration and a process is described for doing this based on the properties of the P_s function [Okaya, Saito & Pepinsky (1955). *Phys. Rev.* **98**, 1857–1858]. If the arrangement of anomalous scatterers is centrosymmetric then the problem does not occur.

Introduction

Many methods have been proposed for the use of one-wavelength anomalous scattering (OAS) for solving crystal structures, *e.g.* Okaya, Saito & Pepinsky (1955), Ramachandran & Raman (1956), Fan (1965), Hendrickson & Teeter (1981), Hauptman (1982), Giacovazzo (1983), Karle (1984*b*, 1989), Fan & Gu (1985), Hao & Woolfson (1989), and Ralph & Woolfson (1991). The increasing availability of synchrotron sources with beamlines dedicated to anomalous scattering means that the collection of multi-wavelength anomalous-scattering (MAS) data is becoming more common. While scaling different data sets together is straightforward in principle (Karle, 1984*a*), in practice one often finds incompatibilities in intensities of data at different wavelengths that scaling cannot correct. Nevertheless MAS data, used with discretion, are a powerful tool for solving structures, in particular macromolecular structures.

Numbers of different methods are available for the use of MAS data, *e.g.* Karle (1980), Woolfson (1984), and Fan, Woolfson & Yao (1993). The basis of these methods can be seen in Fig. 1 which relates to a two-wavelength case where the structure contains only one kind of anomalous scatterer. The structure factor of index **h** is given by

$$F(\mathbf{h}) = F^{\circ}(\mathbf{h}) + F'(\mathbf{h}) + iF''(\mathbf{h}), \qquad (1)$$

where $F^{\circ}(\mathbf{h})$ is the normal component of the scattering, including that of the anomalous scatterers, and $F'(\mathbf{h})$ and $F''(\mathbf{h})$ are the contributions of the real and imaginary components of the anomalous scattering, respectively. Also shown in Fig. 1 is

$$F^{*}(\bar{\mathbf{h}}) = F^{\circ}(\mathbf{h}) + F'(\mathbf{h}) - iF''(\mathbf{h}). \qquad (2)$$

The subscripts 1 and 2 in Fig. 1 refer to the two wavelengths λ_1 and λ_2. The basis of the multi-

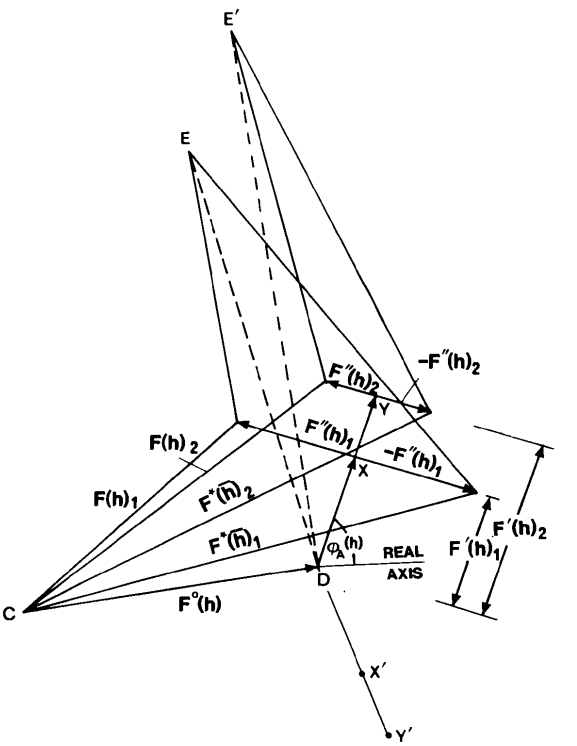

Fig. 1. The basis of the resolution of the phase ambiguity with anomalous-scattering data at two wavelengths. The common solution for the structure factor without anomalous scattering is *CD*.

ISSN 0907-4449

wavelength approach is as follows. Given the positions of the anomalous scatterers and a single wavelength, say λ_1, the $F_1'(\mathbf{h})$ and $F_1''(\mathbf{h})$ are defined in magnitude and orientation. Then, assuming that the anomalous scatterers are all of the same type,

$$F_1'(\mathbf{h}) = f_1' \sum_{j=1}^{M} \exp(2\pi i \mathbf{h}.\mathbf{R}_j)$$
$$= |F_1'(\mathbf{h})| \exp[i\varphi_A(\mathbf{h})] \quad (3)$$

and

$$F_1''(\mathbf{h}) = i(f_1''/f_1')F_1'(\mathbf{h}), \quad (4)$$

where the M anomalous scatterers have coordinates R_j, $j = 1$ to M, and f_1' and f_1'' are the real and imaginary parts of the anomalous-scattering factor, assumed independent of $\mathbf{h}$. From the measured structure amplitudes $F(\mathbf{h})|$ and $|F(\bar{\mathbf{h}})|$ triangles ABO and ABO' can be constructed, and from this information alone $F^{\circ}(\mathbf{h})$, the structure factor if anomalous scattering did not occur, can be found to be either CD or ED. If a second wavelength, λ_2, is available then $F^{\circ}(\mathbf{h})$ can be found to be either CD or $E'D$ and the ambiguity is thus resolved since there is only one common solution. With real data the position is usually less clear but by analytical means involving a best-fit, *e.g.* least-squares, procedure the ambiguity can be resolved and a best value can be found for $F^{\circ}(\mathbf{h})$ both in magnitude and phase. A Fourier synthesis calculated with the values of $F^{\circ}(\mathbf{h})$ should, in principle, show the structure but in any case usually gives a starting point either for model fitting or phase extension and refinement to provide a better map.

An important prerequisite for either an OAS or a MAS approach was mentioned above but rather glossed over, *i.e.* that the positions of the anomalous scatterers should be known. Various procedures have been suggested for doing this but commonly the values of $|\Delta F(\mathbf{h})| = ||F(\mathbf{h})| - |F(\bar{\mathbf{h}})||$ are used either to calculate a Patterson synthesis or as input into a direct-methods program (Mukherjee, Helliwell & Main, 1989). However, either of these two approaches alone is not enough if the arrangement of anomalous scatterers is non-centrosymmetric since the absolute configuration will not be defined. The effect of taking the enatiomorph of the correct configuration of anomalous scatterers can be visualized from Fig. 1. The line DXY, along $F_1'(\mathbf{h})$ and $F_2'(\mathbf{h})$, will be reflected in the real axis to $D'X'Y'$ and the whole diagram will be rotated anticlockwise by an angle $2\theta(\mathbf{h})$. Thus each $F^{\circ}(\mathbf{h})$ will be estimated with a correct magnitude but an incorrect phase and the calculated map will not show the structure. One way around this difficulty is just to calculate phases for both alternatives and then the correct one may be selected by some figure-of-merit tests or may be obvious by the appearance of the map. There is, however, an objective procedure for solving this problem, which requires only one data set and this we now describe.

The P_s function

The properties of the P_s map, defined as

$$P_s(\mathbf{u}) = (1/V)\sum_{\mathbf{h}}[|F(\mathbf{h})|^2 - |F(\bar{\mathbf{h}})|^2] \sin(2\pi \mathbf{h}.\mathbf{u}), \quad (5)$$

were first described by Okaya, Saito & Pepinsky (1955). This antisymmetric map shows positive peaks corresponding to vectors between anomalous scatterers and non-anomalous scatterers and negative peaks in the opposite directions. For other than trivial structures there is some loss of information as a result of the cancellation of positive and negative peaks but, nevertheless, Hao & Woolfson (1989) showed that the P_s function could be used to solve a protein structure if OAS data were available. Ignoring the loss of information by cancellation it is clear that the positive regions of the P_s map will show the sum of M images of the complete structure, less the anomalous scatterers, each image having a different anomalous scatterer at the origin. The Hao & Woolfson (1989) procedure involved the calculation of a sum map, based on the assumed known values of the anomalous scatterers, which was of the form

$$Q(\mathbf{r}) = (1/V)\sum_{\mathbf{h}} S(\mathbf{h})[|F(\mathbf{h})|^2 - |F(\bar{\mathbf{h}})|^2] \sin(2\pi \mathbf{h}.\mathbf{u}), \quad (6)$$

where

$$S(\mathbf{h}) = \sum_{j=1}^{M} \exp(2\pi i \mathbf{h}.\mathbf{R}_j). \quad (7)$$

Negative regions of the map were then changed to zero and the Fourier transform of the resultant map gave phase estimates from which the structure could subsequently be derived by standard methods. Hao & Woolfson tried various other procedures, for example, removing negativity before the sum-function calculations and also using a minimum function rather than a sum function but these added to the total effort and gave inferior results. Because of the loss of information by peak cancellation the sum function is found to be the best process.

Our method of finding the correct configuration of the anomalous scatterers is based on this work.

Finding the correct configuration

The situation we consider is that a set of coordinates for the anomalous scatterers $\{\mathbf{R}\}$ is available but that it is possible that the correct coordinates are $\{-\mathbf{R}\}$. If the sum function $Q(\mathbf{r})$ has been calculated for the correct configuration then it will contain a positive image of the structure with weight M plus sundry

other positive and negative images of lesser weight. Let us now consider the value of

$$C_n = \int Q(\mathbf{r})^n \mathrm{d}V \tag{8}$$

taken over the whole unit cell for various values of n. Since $P_s(\mathbf{u})$ is an antisymmetric function, and has an average value zero over the whole cell, then clearly any sum of such functions, with whatever displacements, will also equal zero. Hence $C_1 = 0$. With $n = 2$ the integral cannot be zero but the value of C_2 will not be very informative. If the wrong configuration is chosen then a build up of M negative images is formed and the function obtained, $Q'(\mathbf{r})$ is related to the correct function by

$$Q'(\mathbf{r}) = -Q(-\mathbf{r}) \tag{9}$$

and the value of C_2 will not depend on which configuration, correct or incorrect, is used for calculating the sum function. It is the value of C_3 which enables the correct configuration to be found. If it is positive then the sum function was performed with the correct configuration, otherwise the other configuration is correct. This can be visualized by reference to Fig. 2. In Fig. 2(*a*) there is shown a simple structure with three anomalous scatterers (the minimum to have a non-centrosymmetric arrangement) and three non-anomalous scatterers. Fig. 2(*b*) shows the point P_s function and Fig. 2(*c*) a threefold sum function taking the anomalous scatterers with the correct configuration. The model has been taken so that there is no cancellation of peaks. The result in Fig. 2(*c*) is as follows: 18 single-weight positive peaks, three triple-weight positive peaks, 17 single-weight negative peaks; five double-weight negative peaks. We can now calculate that

$$C_1 = (18 \times 1) + (3 \times 3) - (17 \times 1) - (5 \times 2) = 0$$

$$C_2 = (18 \times 1) + (3 \times 9) + (17 \times 1) + (5 \times 4) = 82$$

$$C_3 = (18 \times 1) + (3 \times 27) - (17 \times 1) - (5 \times 8) = 52.$$

If the incorrect configuration had been used the values would have been 0, 82 and -52, respectively.

Tests so far, on real data for two structures, have shown that this is a reliable indicator of the correct configuration. The first of these structures core streptavidin (Hendrickson, Pähler, Smith, Satow, Merritt & Phizackerley, 1989) is a 125–127 residue protein, space group $I222$ with $a = 95.2$, $b = 105.6$, $c = 47.4$ Å. The anomalous scatterers are two Se atoms in each asymmetric unit. The wavelength used for the data was 0.9795 Å for which $f' = -6.203$ and $f'' = 3.663$. The value of C_3 was 1.156×10^6 and the mean phase error deduced from the minimum-function map produced by the correct configuration was 73.2°. Using the incorrect configuration gave $C_3 = -1.156 \times 10^6$ and the minimum-function map gave a mean phase error of 90.1°.

The second structure was that of RNA (Dodson, Sevcik, Dodson & Zelinka, 1987). This contains two molecules, each with 96 amino-acid residues per asymmetric unit, in a cell with space group $P2_12_12_1$ with $a = 64.9$, $b = 78.3$ and $c = 38.8$ Å. The anomalous scatterers are Pt atoms of which there are six in each asymmetric unit although they are distributed on 20 sites with partial occupancy. The value of C_3,

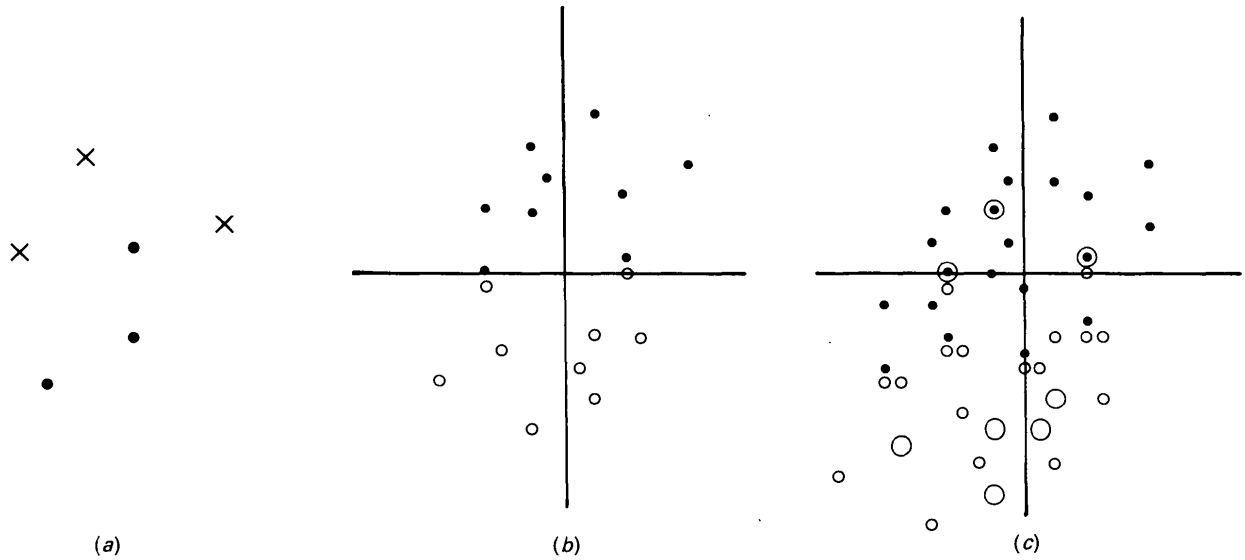

Fig. 2. (*a*) A simple model structure with three anomalous scatterers ● and three non-anomalous scatterers ×. (*b*) The P_s function for the model structure. Positive peaks are denoted by ● and negative peaks by ○. (*c*) A sum function based on the correct configuration of the three anomalous scatterers. The symbols indicate: ● single-weight positive peak; ◉ triple-weight positive peak; ○ single-weight negative peak; ◯ double-weight negative peak.

4.213×10^4, again indicated the correct enantiomorph; the minimum-function overlap gave a mean phase error of 77.3°, which became 90.5° if the wrong enantiomorph was taken to produce the overlap function. It should be stated that RNA was not a favourable case, especially for the use of the P_s function method and a much lower mean phase error could be obtained by an alternative procedure (Ralph & Woolfson, 1991).

The procedure as explained above is quite simple; the function $Q(\mathbf{r})$ is calculated and then $Q(\mathbf{r})^3$ is summed over all the grid points at which the function has been calculated. The sign of the summation then indicates which is the correct configuration. The sensitivity of the test depends on the proportion of positive to negative contribution to C_3, which is affected by peak cancellation. It will be seen in the simple numerical example we used that the ratio of positive to negative contribution was 99:47 or about 2:1 and this should not depend too sensitively on the complexity of the structure. We might also notice that peak cancellation removes both negative and positive peaks equally. It seems likely that the test is a robust one.

If the arrangement of peaks is centrosymmetric then the problem of finding the absolute configuration of the anomalous scatterers does not arise. In this case the sum-function map will, like the original P_s function, be antisymmetric and the value of C_3 will be zero. However, the positive regions of the sum-function map will correspond to the correct absolute configuration of the total structure and can be used as a starting point for the structure determination.

We wish to express our appreciation for the support of the Wellcome foundation which enabled this work to be carried out as part of a larger study of anomalous scattering and its applications.

References

DODSON, C. G., SEVCIK, J., DODSON, E. & ZELINKA, J. (1987). *Metabolism of Nucleic Acids Including Gene Manipulation*, pp. 33–36. Prague: Slovak Academy of Sciences.

FAN, H. F. (1965). *Acta Phys. Sin.* **21**, 1105–1113.

FAN, H. F. & GU, Y. X. (1985). *Acta Cryst.* A**41**, 280–294.

FAN, H. F., WOOLFSON, M. M. & YAO, J. (1993). *Proc. R. Soc. London Ser. A*, **442**, 12–32.

GIACOVAZZO, C. (1983). *Acta Cryst.* A**39**, 585–592.

HAO, Q. & WOOLFSON, M. M. (1989). *Acta Cryst.* A**45**, 794–797.

HAUPTMAN, H. (1982). *Acta Cryst.* A**38**, 632–741.

HENDRICKSON, W. A., PÄHLER, A., SMITH, J. L., SATOW, Y., MERRIT, E. A. & PHIZACKERLEY, R. P. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 2190–2194.

HENDRICKSON, W. A. & TEETER, M. M. (1981). *Nature (London)*, **290**, 107–113.

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

KARLE, J. (1984*a*). *Acta Cryst.* A**40**, 1–4.

KARLE, J. (1984*b*). *Acta Cryst.* A**40**, 4–11.

KARLE, J. (1989). *Acta Cryst.* A**45**, 303–307.

MUKHERJEE, A. K., HELLIWELL, J. R. & MAIN, P. (1989). *Acta Cryst.* A**45**, 715–718.

OKAYA, Y., SAITO, Y. & PEPINSKY, R. (1955). *Phys. Rev.* **98**, 1857–1858.

RALPH, A. C. & WOOLFSON, M. M. (1991). *Acta Cryst.* A**47**, 533–537.

RAMACHANDRAN, J. N. & RAMAN, S. (1956). *Curr. Sci.* **25**, 348–351.

WOOLFSON, M. M. (1984). *Acta Cryst.* A**40**, 32–34.