Locating the anomalous scatterers within a crystal using dual-wavelength X-ray diffraction measurements

Richard L. Kingston

Institute of Molecular Biology, Howard Hughes Medical Institute, University of Oregon, Eugene, OR 97403, USA, and Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA

Correspondence e-mail:
richard@uoxray.uoregon.edu

A simple algebraic expression is used to calculate the structure-factor amplitudes of the anomalous scatterers within a protein crystal, using X-ray diffraction measurements made at two wavelengths and given a single type of anomalous scatterer within the crystal. The expression is exact and is readily derived from the standard algebraic equations for analyzing multi-wavelength anomalous diffraction data. Evaluation of the expression requires estimates for $^\lambda f'$ and $^\lambda f''$, the real and imaginary components of the anomalous scattering, respectively. When these are not known at the wavelengths in question, a statistical procedure is suggested, applicable when the crystal space group has centrosymmetric projections, which allows the expression to be used regardless. The method may provide a useful alternative to existing procedures when $^\lambda f'$ and $^\lambda f''$ are unknown or when data have been measured at a limited number of wavelengths.

1. Introduction

In crystallographic phase determination using the multi-wavelength anomalous diffraction (MAD) method, the initial problem is to locate the positions of the anomalous scatterers within the crystal. Usually this is performed by interpreting a Patterson synthesis or by using a direct-methods procedure. In either case, estimates for the structure-factor amplitudes of the anomalous scatterers are required.

The theoretical and practical aspects of MAD have been reviewed elsewhere (Hendrickson & Ogata, 1997; Ramakrishnan & Biou, 1997; Fourme et al., 1996; Hendrickson, 1991; Karle, 1989a). The practical consequences of anomalous scattering are that reflections with indices $+\mathbf{h}$ and $-\mathbf{h}$ (Friedel pairs) have differing amplitudes. Additionally, the amplitude of any given reflection ($+\mathbf{h}$ or $-\mathbf{h}$) is wavelength-dependent. Hence, in a dual-wavelength experiment, for a general reflection with index $\mathbf{h}$ there are four different experimentally measurable quantities $|^{\lambda_1}F(+)|$, $|^{\lambda_1}F(-)|$, $|^{\lambda_2}F(+)|$, $|^{\lambda_2}F(-)|$ (Fig. 1). These quantities would all be identical in the absence of anomalous scattering.

In protein crystallography, there are usually a large number of atoms within the crystal for which the anomalous scattering is negligible and a very few atoms of a single type for which the anomalous scattering is significant. In this case, it is useful to partition the total structure factor into two components, one arising from the normal scatterers (N) within the crystal and the other from the anomalous scatterers (A). Hence, a vector equation can be written,

$$^\lambda\mathbf{F} = {}^\circ\mathbf{F}_N + {}^\lambda\mathbf{F}_A. \tag{1}$$

Wavelength-dependent quantities are denoted with the superscript λ and wavelength-independent quantities with the

superscript ∘. Given that the significant anomalous scatterers within the crystal are of a single type, the wavelength-dependent structure factor can be further expanded in terms of the components of the scattering factor of the anomalously scattering group ($^\circ f$, $^\lambda f'$, $^\lambda f''$),

$$^\lambda \mathbf{F} = {}^\circ \mathbf{F}_N + {}^\circ \mathbf{F}_A \left(1 + \frac{{}^\lambda f'}{{}^\circ f} + i \frac{{}^\lambda f''}{{}^\circ f} \right). \qquad (2)$$

The situation is shown geometrically in Fig. 1. $^\circ \mathbf{F}_A$ is the contribution from the normal scattering component of the anomalously scattering atoms. It is the quantities $|^\circ F_A|$, the wavelength-independent structure-factor amplitudes of the anomalous scatterers, that we wish to estimate. If these are known, then the positions of the anomalous scatterers can generally be determined and phases calculated for the diffraction data using standard procedures.

Some early approaches to the problem were those of Okaya *et al.* (1955) and Rossmann (1961) and are applicable even when only single-wavelength measurements have been made. In particular, Rossmann (1961) in essence suggested

$$\left[\left(\frac{{}^\lambda f''}{{}^\circ f} \right) |^\circ F_A| \right]^2 \simeq \frac{1}{4} \left[|^\lambda F(+)| - |^\lambda F(-)| \right]^2 \qquad (3)$$

and used these terms as the coefficients of a Patterson synthesis. This approximation (discussed from varying perspectives by Kartha & Parthasarathy, 1965; Karle, 1980; Cascarano & Giacovazzo, 1984) is still frequently used to determine the positions of anomalous scatterers in protein crystals (see, for example, Howell *et al.*, 2000).

When diffraction measurements have been made at more than one wavelength, it is possible in principle to determine $|^\circ F_A|$ without approximation. The standard algebraic equations for analyzing MAD data were formulated by Karle and Hendrickson (see Karle, 1980, 1989*b*). A well known least-squares method which determines $|^\circ F_A|$ is based directly on this algebraic analysis and is implemented in the program *MADLSQ* (Hendrickson & Ogata, 1997). Another analytical method was described by Fan *et al.* (1993) (see also Woolfson *et al.*, 1997) and is implemented in the program *REVISE* within the *CCP4* program suite (Collaborative Computational Project, Number 4, 1994). Terwilliger (1994) describes a Bayesian statistical procedure for estimating $|^\circ F_A|$, which has been implemented in the program *SOLVE* (Terwilliger & Berendzen, 1999).

In this paper, a simple algebraic expression is used to calculate $|^\circ F_A|$ given X-ray diffraction measurements made at two wavelengths and a single type of anomalous scatterer within the crystal. The expression is exact and is readily derived by explicit solution for $|^\circ F_A|$ of the standard algebraic equations for analyzing MAD data. The expression was first derived by Singh & Ramaseshan (1968*a*) in a paper that was primarily concerned with neutron anomalous scattering. However, their analysis is general and is equally applicable to the X-ray case. This approach to the estimation of $|^\circ F_A|$ provides a straightforward alternative to the approximation introduced by Rossmann (1961) or the elegant but more

complicated least-squares procedure implemented in *MADLSQ* (Hendrickson & Ogata, 1997). When $^\lambda f'$ and $^\lambda f''$ are not known at the wavelengths in question but the space group of the crystal has centrosymmetric projections, it is shown how the expression of Singh & Ramaseshan can still be used. The general approach taken is similar to earlier procedures for combining information from isomorphous replacement and anomalous scattering to locate anomalously scattering heavy atoms in protein derivatives (Kartha & Parthasarathy, 1965; Matthews, 1966; Singh & Ramaseshan, 1966).

It is emphasized that throughout this paper the anomalous corrections ($^\lambda f'$, $^\lambda f''$) to the normal atomic scattering factor ($^\circ f$) are treated as isotropic and independent of the scattering angle and are considered dependent on the wavelength alone.

2. Theory

2.1. An exact expression for $|^\circ F_A|$ given dual-wavelength X-ray diffraction measurements

The starting point for the method is the basic equation expressing the diffracted intensity from a crystal with one predominant type of anomalous scatterer. This can be written

$$|^\lambda F(\pm)|^2 = |^\circ F_N|^2 + \left[1 + 2\frac{{}^\lambda f'}{{}^\circ f} + \frac{({}^\lambda f'^2 + {}^\lambda f''^2)}{{}^\circ f^2} \right] |^\circ F_A|^2$$
$$+ 2\left(1 + \frac{{}^\lambda f'}{{}^\circ f} \right) |^\circ F_N||^\circ F_A| \cos({}^\circ \alpha_N - {}^\circ \alpha_A)$$
$$\pm 2\left(\frac{{}^\lambda f''}{{}^\circ f} \right) |^\circ F_N||^\circ F_A| \sin({}^\circ \alpha_N - {}^\circ \alpha_A). \qquad (4)$$

This equation can be readily obtained by multiplying the vector expression for $^\lambda \mathbf{F}$ (2) by its complex conjugate. Karle (1989*b*) summarizes various other ways of writing the exact algebraic equations for analyzing MAD data, which are all mathematically equivalent. For a reflection with index $\pm \mathbf{h}$, each wavelength provides two equations relating the unknown quantities. A two-wavelength experiment gives rise to four equations of this type, which is enough in principle to allow evaluation of all the unknowns.

With a suitable choice of unknowns [$|^\circ F_N|^2$, $|^\circ F_A|^2$, $|^\circ F_N||^\circ F_A| \cos({}^\circ \alpha_N - {}^\circ \alpha_A)$ and $|^\circ F_N||^\circ F_A| \sin({}^\circ \alpha_N - {}^\circ \alpha_A)$], the system of equations becomes linear and can be solved by least-squares methods for these quantities. This was Karle's preferred method of solution and was subsequently implemented by Hendrickson and coworkers in the program *MADLSQ* (although they use a related system of equations involving $|^\circ F_T|$ and $|^\circ F_A|$).

Karle (1980) also indicated the possibility of solving such a system of equations explicitly for $|^\circ F_A|$ by eliminating the unknown quantities $|^\circ F_N|$, $\cos({}^\circ \alpha_N - {}^\circ \alpha_A)$ and $\sin({}^\circ \alpha_N - {}^\circ \alpha_A)$, but did not develop this idea further. If this is performed, then a quadratic equation emerges, a result that was first given by Singh & Ramaseshan (1968*a*). The starting point of their derivation, allowing for the very different notation they use, is

seen to be the equations (4) written for two different wavelengths.

Explicit solution of these equations for $|{^\circ}F_A|$ gives the following result

$$A\left[\left(\frac{^{\lambda_2}f''}{^\circ f}\right)|{^\circ}F_A|\right]^4 + B\left[\left(\frac{^{\lambda_2}f''}{^\circ f}\right)|{^\circ}F_A|\right]^2 + C = 0, \quad (5)$$

where

$$A = k_2^4 + 2k_2^2(1 + k_1^2) + (k_1^2 - 1)^2,$$
$$B = -k_2^2({^{\lambda_1}}I_{\text{sum}} + {^{\lambda_2}}I_{\text{sum}}) - (k_1^2 - 1)({^{\lambda_1}}I_{\text{sum}} - {^{\lambda_2}}I_{\text{sum}}),$$
$$C = \frac{1}{4}({^{\lambda_1}}I_{\text{sum}} - {^{\lambda_2}}I_{\text{sum}})^2 + \frac{k_2^2}{8}\left[({^{\lambda_2}}I_{\text{diff}})^2 + \frac{({^{\lambda_1}}I_{\text{diff}})^2}{k_1^2}\right]$$

and

$$k_1 = \frac{^{\lambda_1}f''}{^{\lambda_2}f''},$$
$$k_2 = \frac{^{\lambda_1}f' - ^{\lambda_2}f'}{^{\lambda_2}f''},$$
$${^\lambda}I_{\text{sum}} = |{^\lambda}F(+)|^2 + |{^\lambda}F(-)|^2,$$
$${^\lambda}I_{\text{diff}} = |{^\lambda}F(+)|^2 - |{^\lambda}F(-)|^2.$$

(5) defines $|{^\circ}F_A|$ in terms of the four experimentally measured quantities $|{^{\lambda_1}}F(+)|$, $|{^{\lambda_1}}F(-)|$, $|{^{\lambda_2}}F(+)|$ and $|{^{\lambda_2}}F(-)|$ and the components of the scattering factor of the anomalously scattering group ($^\circ f$, $^\lambda f'$, $^\lambda f''$). Singh & Ramaseshan (1968a) present this result in a slightly different form and provide a full proof, which is not repeated here.

The use of this expression in protein crystallography was suggested by Klop et al. (1989a) as part of a proposal for the estimation of triplet invariants from MAD data. An algebraically equivalent expression can be developed involving ratios of the experimentally measured quantities instead of sums and differences (Klop et al., 1989b). (5) is applicable to both the centric and acentric reflections, although in the centric case [where $|{^\lambda}F(+)| = |{^\lambda}F(-)| = |{^\lambda}F|$] some simplification is possible (Singh & Ramaseshan, 1968b).

Since (5) is quadratic in $[(^{\lambda_2}f''/^\circ f)|{^\circ}F_A|]^2$ there are two possible solutions, which are given by

$$X_\pm = \frac{-B}{2A} \pm \frac{(B^2 - 4AC)^{1/2}}{2A}. \quad (6)$$

Recalling that the values of the coefficients A, B and C are determined from the X-ray diffraction data, with unreliable data or unreasonable values for k_1 and k_2 the term $B^2 - 4AC$ may be negative and there will be no real solutions to the equation. However, in general the problem is to select the correct solution from two positive alternatives. This presents little practical difficulty. Referring to Fig. 1, for the acentric reflections the correct solution depends on the phase difference between $^\circ\mathbf{F}_N$ and $^\circ\mathbf{F}_T$. The lower solution (X_-) is relevant when $|{^\circ}\alpha_N - {^\circ}\alpha_T| < 90°$, while the upper solution (X_+) is relevant when $|{^\circ}\alpha_N - {^\circ}\alpha_T| > 90°$. Normally, the average value of $|{^\circ}F_A|$ is much smaller than that of $|{^\circ}F_N|$. Simple heuristic arguments, similar to those made by Dodson & Vijayan (1971) in the related case of isomorphous replacement, show that the

lower solution will correspond to the true value of $[(^{\lambda_2}f''/^\circ f)|{^\circ}F_A|]^2$ in almost all cases. For the centric reflections, it can be shown by a different argument (Singh & Ramaseshan, 1968b) that the lower solution will again usually be the correct choice.

It is worth considering the selection of the wavelengths of the X-ray diffraction data to be used in the calculation. Clearly, data collected at wavelengths on the low-energy side of the X-ray absorption edge should not be used, since here $^\lambda f''$, the imaginary part of the anomalous scattering factor, will be close to its minimum value. Additionally, it seems that the two wavelengths should be chosen so that the variation in $^\lambda f'$, the real part of the anomalous scattering factor, is large. The optimal choice of wavelengths for phase determination in the dual-wavelength experiment has been discussed by Gonzàlez et al. (1999), whose results are consistent with earlier predictions (see, for example, Phillips & Hodgson, 1980; Narayan & Ramaseshan, 1981; Fan et al., 1993)

A general problem with all algebraic methods for estimating $|{^\circ}F_A|$ is that measurement errors may give rise to a few
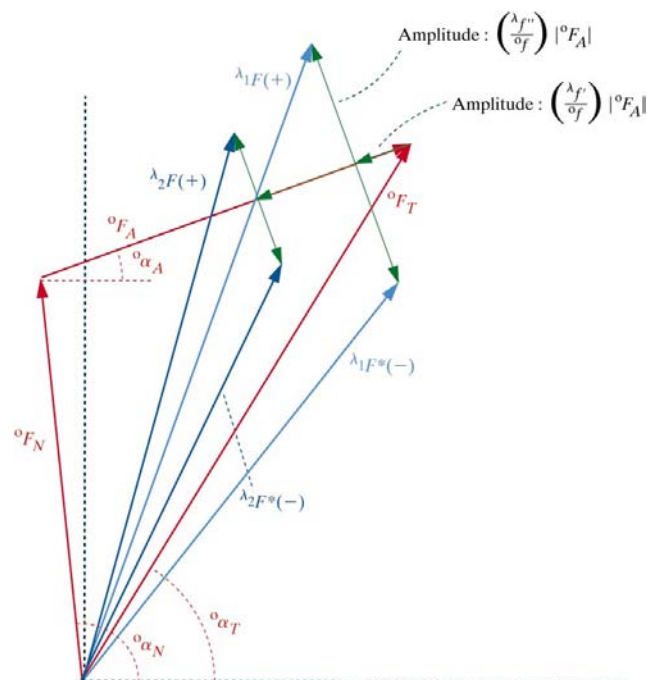

Figure 1
Geometric relationships in the complex plane between structure factors with indices $+\mathbf{h}$ and $-\mathbf{h}$ measured at two wavelengths. The figure is only representative when the crystal contains a single type of anomalous scatterer. Vectors in red derive from normal diffraction alone and are independent of the wavelength. Vectors in green show the effect of anomalous scattering. Vectors in blue are the resultant (experimentally measured) quantities. To emphasize the geometric relationship between reflections and make comparison of magnitudes easier, the complex conjugates of $^{\lambda_1}\mathbf{F}(-)$ and $^{\lambda_2}\mathbf{F}(-)$ are shown [$^{\lambda_1}\mathbf{F}^*(-)$ and $^{\lambda_2}\mathbf{F}^*(-)$, respectively]. This corresponds to reflection of these vectors across the real (horizontal) axis. As represented in the figure, $^\lambda f'$ is negative, which is the situation for most elements at wavelengths useful for protein crystallography. Here and throughout this paper, N denotes contributions from the normally scattering atoms, A denotes contributions from the anomalously scattering atoms and T denotes the total diffraction.

large and completely discordant estimates (upper outliers). Although such outliers are usually readily identified, failure to detect and remove them from a data set can have catastrophic consequences for direct-methods or Patterson-based procedures for locating the anomalous scatterers. A completely robust method for detecting outliers has not yet been devised.

2.2. Estimating k_1^2 and k_2^2 from the X-ray diffraction data

At wavelengths close to an X-ray absorption edge, where MAD data are usually collected, the anomalous scattering factors are poorly described by theory. $^\lambda f'$ and $^\lambda f''$ can in principle be obtained from analysis of the X-ray absorption spectra of the crystal (see, for example, Hendrickson et al., 1988; Evans & Wilson, 1999). However, the rapid change in the anomalous signal close to an X-ray absorption edge requires that the wavelength can be precisely and reproducibly selected. This may not be realised in practice. Therefore, we consider how to proceed when $^\lambda f'$ and $^\lambda f''$ are not known from experiment or have been estimated only crudely.

In this case (5) cannot be used to directly determine $|^\circ F_A|$. However, it suffices to estimate $k_1^2 = (^{\lambda_1}f''/^{\lambda_2}f'')^2$ and $k_2^2 = [(^{\lambda_1}f' - {}^{\lambda_2}f')/^{\lambda_2}f'']^2$, since these ratios contain all the required unknown terms. Since $^\lambda f'$ and $^\lambda f''$ are assumed to be isotropic and resolution-independent, k_1^2 and k_2^2 are simple scalar quantities. In the following arguments, we assume that data at the two wavelengths have been put on a common scale. It is noted, however, that the determination of the relative scale factor between data collected at different wavelengths and the determination of k_1^2 and k_2^2 are not strictly separable problems and that a completely rigorous solution to the scaling problem is not currently available (see Karle, 1984 for discussion).

It is readily shown (see e.g. Singh & Ramaseshan, 1968a) that for all acentric reflections

$$k_1 = \frac{^{\lambda_1}f''}{^{\lambda_2}f''} = \frac{|^{\lambda_1}F(+)|^2 - |^{\lambda_1}F(-)|^2}{|^{\lambda_2}F(+)|^2 - |^{\lambda_2}F(-)|^2} = \frac{^{\lambda_1}I_{\mathrm{diff}}}{^{\lambda_2}I_{\mathrm{diff}}}. \qquad (7)$$

Hence, k_1 is expressible as a ratio of experimentally measured quantities. A least-squares-based method can be used to estimate k_1, noting that there are generally significant uncertainties in both the numerator and denominator of (7). Under a Gaussian error model, the problem reduces (Moreno, 1996) to finding the value of k_1 that minimizes

$$S(k_1) = \sum_{i=1}^{N} \frac{[(^{\lambda_1}I_{\mathrm{diff}})_i - k_1(^{\lambda_2}I_{\mathrm{diff}})_i]^2}{(^{\lambda_1}\sigma_i)^2 + k_1^2(^{\lambda_2}\sigma_i)^2}, \qquad (8)$$

where $^{\lambda_1}\sigma_i$ and $^{\lambda_2}\sigma_i$ are the estimated standard deviations of $(^{\lambda_1}I_{\mathrm{diff}})_i$ and $(^{\lambda_2}I_{\mathrm{diff}})_i$, respectively, and the summation is over all N available measurements. The minimization can be performed by an iterative numerical procedure.

The estimation of

$$k_2^2 = \left(\frac{^{\lambda_1}f' - {}^{\lambda_2}f'}{^{\lambda_2}f''}\right)^2$$

is not as straightforward. The approach suggested here is based on the comparison of statistics calculated from centric

and acentric reflections. Hence, the procedure is not applicable in the triclinic space group $P1$ nor in the trigonal space groups $P3$, $P3_1$, $P3_2$ and $R3$. These space groups have no centrosymmetric projections and hence contain no centric reflections.

First, the probability distributions for intensities first derived by Wilson (1949, 1950) are recalled. These provide an approximate statistical description of the normal diffraction from the anomalous scatterers (that is, for the quantities $|^\circ F_A|^2$).

For the acentric reflections

$$P(I) = (1/S)\exp(-I/S), \qquad (9)$$

while for the centric reflections

$$P(I) = [1/(2\pi I S)^{1/2}]\exp(-I/2S). \qquad (10)$$

Here, S is a distribution parameter that can be written as $S = \varepsilon\Sigma$. The quantity Σ is the sum of the squared normal scattering factors of the atoms in the structure. The factor ε corrects for differences in expected intensity owing to crystallographic symmetry in certain zones and rows of the reciprocal lattice and also for the effect of lattice centering (see Wilson, 1950; Rogers, 1965; Iwasaki & Ito, 1977). Full details are given in Srinivasan & Parthasarathy (1976). For our purposes it is convenient to rewrite these probability density functions in terms of the transformed variable $\hat{I} = I/\varepsilon$, since this puts all reflections on a common statistical footing. Thus, for the acentric reflections

$$P(\hat{I}) = (1/\Sigma)\exp(-\hat{I}/\Sigma), \qquad (11)$$

while for the centric reflections

$$P(\hat{I}) = [1/(2\pi\hat{I}\Sigma)^{1/2}]\exp(-\hat{I}/2\Sigma). \qquad (12)$$

Note that both the centric and acentric probability distributions are gamma distributions, having the characteristic form

$$P(x) = \frac{x^{\alpha-1}\exp(-x/\theta)}{\theta^\alpha \Gamma(\alpha)}, \qquad (13)$$

where $\Gamma(\alpha)$ is the gamma function, α is the shape parameter of the distribution and θ is the scale parameter. Considering transformed intensities $\hat{I}$ (11 and 12), for the centric distribution $\alpha = \frac{1}{2}$ and $\theta = 2\Sigma$, while for the acentric distribution $\alpha = 1$ and $\theta = \Sigma$. In both cases, the theoretical mean of the distribution $\mu = \alpha\theta = \Sigma$.

Applying (11) and (12) to describe the distribution of the quantities $(1/\varepsilon)|^\circ F_A|^2$, two expressions for the corresponding quantity $^\circ\Sigma_A$ will be developed involving the centric and acentric reflections, respectively. These will be used to estimate the ratio k_2^2.

For the centric reflections, the geometric situation shown in Fig. 1 is greatly simplified. $^\circ\mathbf{F}_N$ and $^\circ\mathbf{F}_A$ are collinear and Friedel's law is obeyed, so that $|^\lambda F(+)| = |^\lambda F(-)| (= |^\lambda F|)$. Hence, $\sin(^\circ\alpha_N - {}^\circ\alpha_A) = 0$ and $|\cos(^\circ\alpha_N - {}^\circ\alpha_A)| = 1$ and the general expression for the diffracted intensity (4) simplifies to

$$\begin{cases} |^{\lambda}F|^2 = \left[|^{\circ}F_N| + \left(1 + \frac{^{\lambda}f'}{^{\circ}f}\right)|^{\circ}F_A|\cos(^{\circ}\alpha_N - {^{\circ}}\alpha_A) \right]^2 \\ \qquad + \left(\frac{^{\lambda}f''}{^{\circ}f}\right)^2 |^{\circ}F_A|^2 \\ |\cos(^{\circ}\alpha_N - {^{\circ}}\alpha_A)| = 1 \end{cases} \quad (14)$$

Since $|^{\circ}F_N|$ is in general much larger than $|^{\circ}F_A|$, the contribution from the second term $(^{\lambda}f''/^{\circ}f)^2|^{\circ}F_A|^2$ to the diffracted intensity is usually negligible compared with the contribution $\{|^{\lambda}F_N| + [1 + (^{\lambda}f'/^{\circ}f)]|^{\circ}F_A|\cos(^{\circ}\alpha_N - {^{\circ}}\alpha_A)\}^2$ from the first term.

Now consider the values taken by $(|^{\lambda_1}F| - |^{\lambda_2}F|)^2$. In general, the resultant quantity depends on the magnitude and phase of both $^{\circ}\mathbf{F}_N$ and $^{\circ}\mathbf{F}_A$. It can be assumed that these quantities are uncorrelated, since $^{\circ}\mathbf{F}_N$ and $^{\circ}\mathbf{F}_A$ are the sum of numerous atomic contributions from two mutually exclusive sets of atoms. It is then easy to show that for almost all centric reflections

$$(|^{\lambda_1}F| - |^{\lambda_2}F|)^2 \simeq \left(\frac{^{\lambda_1}f' - {^{\lambda_2}}f'}{^{\circ}f}\right)^2 |^{\circ}F_A|^2.$$

The relation is approximate because the term $(^{\lambda}f''/^{\circ}f)^2|^{\circ}F_A|^2$ in (14) has been ignored. For the remainder of the centric reflections

$$(|^{\lambda_1}F| - |^{\lambda_2}F|)^2 \simeq \left[2|^{\circ}F_A| + \left(\frac{^{\lambda_1}f' + {^{\lambda_2}}f'}{^{\circ}f}\right)|^{\circ}F_A| - 2|^{\circ}F_N| \right]^2.$$

This latter case only occurs when $^{\circ}\mathbf{F}_N$ and $^{\circ}\mathbf{F}_A$ are of opposite phase [hence $\cos(^{\circ}\alpha_N - {^{\circ}}\alpha_A) = -1$] and

$$\left(\frac{^{\circ}f}{^{\circ}f + {^{\lambda_1}}f'}\right)|^{\circ}F_N| < |^{\circ}F_A| < \left(\frac{^{\circ}f}{^{\circ}f + {^{\lambda_2}}f'}\right)|^{\circ}F_N|$$

(where it is assumed, without loss of generality, that $^{\lambda_2}f' < {^{\lambda_1}}f'$).

Recalling that the theoretical mean value of the quantities

$$\left(\frac{^{\lambda_1}f' - {^{\lambda_2}}f'}{^{\circ}f}\right)^2 \frac{1}{\varepsilon}|^{\circ}F_A|^2$$

is

$$\left(\frac{^{\lambda_1}f' - {^{\lambda_2}}f'}{^{\circ}f}\right)^2 {^{\circ}}\Sigma_A,$$

it will be a good approximation that

$$\left(\frac{^{\lambda_1}f' - {^{\lambda_2}}f'}{^{\circ}f}\right)^2 {^{\circ}}\Sigma_{A[\text{centric}]} \simeq \left\langle (1/\varepsilon)(|^{\lambda_1}F| - |^{\lambda_2}F|)^2 \right\rangle. \quad (15)$$

The Wilson probability distributions (11 and 12) are derived under the assumption that the atomic scattering factors are constant. As the effective atomic scattering factors (incorporating the effects of atomic displacements) are strongly resolution-dependent, so too is the quantity $^{\circ}\Sigma_A$. Hence, $^{\circ}\Sigma_A$ must be determined in a number of thin concentric shells in reciprocal space, over each of which it can be assumed approximately constant.

A practical difficulty with using (15) directly is that the mean quantity $\langle (1/\varepsilon)(|^{\lambda_1}F| - |^{\lambda_2}F|)^2 \rangle$ is not robust to the

presence of outliers. A single large and aberrant measurement can severely bias the estimate of $^{\circ}\Sigma_A$ in a given resolution shell. Hence, we replace the simple arithmetic mean in (15) with a trimmed mean multiplied by an unbiasing factor, a robust statistic for estimating the mean of a gamma-distributed random variable suggested by Kimber (1983).

Given k_1^2 and k_2^2, we can estimate the same quantity $^{\circ}\Sigma_A$ from the acentric reflections as follows

$$\left(\frac{^{\lambda_1}f' - {^{\lambda_2}}f'}{^{\circ}f}\right)^2 {^{\circ}}\Sigma_{A[\text{acentric}]} = k_2^2 \left\langle \frac{1}{\varepsilon}\left(\frac{^{\lambda_2}f''}{^{\circ}f}\right)^2 |^{\circ}F_A|^2 \right\rangle, \quad (16)$$

where $(^{\lambda_2}f''/^{\circ}f)^2|^{\circ}F_A|^2$ on the right-hand side of this expression is evaluated through (6). As before, we replace the simple arithmetic mean on the right-hand side of the expression with a robust estimator based on the trimmed mean.

Now we have two estimates for the same quantity $[(^{\lambda_1}f' - {^{\lambda_2}}f')/^{\circ}f]^2{^{\circ}}\Sigma_A$. One is calculated directly from the centric reflections *via* (15). The other is calculated from the acentric reflections *via* (6) and (16) and assuming values for k_1^2 and k_2^2. A good estimate for k_1^2 is already available through use of (8). We can find the 'best' value of k_2^2, in a least-squares sense, by minimizing the squared difference between the two estimates for $[(^{\lambda_1}f' - {^{\lambda_2}}f')/^{\circ}f]^2{^{\circ}}\Sigma_A$.

The validity of the Wilson probability distributions depends on the random distribution of a sufficiently *large number* of scatterers in *general positions* in the unit cell. If either of these conditions is strongly violated (*e.g.* if there are a very small number of anomalous scatterers in the unit cell or the anomalous scatterers occupy special positions of the space group), the method described for determining k_2^2 may perform unpredictably.

3. Application

We describe the application of this procedure to a real problem, the structure determination of the N-terminal domain of the Rous sarcoma virus (RSV) capsid protein (154 amino acids). Full details of the protein expression, purification and crystallization and the collection of X-ray diffraction data have been given elsewhere (Kingston *et al.*, 2000)

Briefly, using a crystal of a selenomethionine (SeMet) substituted protein maintained at 113 K, MAD data were collected at wavelengths near the selenium K edge. The crystal space group was $P2_12_12_1$, with unit-cell parameters $a = 40.5$, $b = 64.5$, $c = 108.9$ Å and two molecules within the asymmetric unit. Each molecule contained five selenomethionine residues. Data were collected at five wavelengths, four of which were very close to the X-ray absorption edge (Table 1). The inverse-beam method was used to accumulate the Friedel pairs at each wavelength. Data collected at each wavelength were put on the same relative scale using the program *FHSCAL* (Tickle, 1991). The X-ray absorption spectra of the frozen crystal was aberrant and did not show the pronounced peak expected at energies above the absorption edge. It is possible this effect arose from photo-oxidation of the Se atoms (Smith & Thompson, 1998; Sharff *et al.*, 2000). The crystal used for

Table 1
Multiple-wavelength anomalous diffraction data for the N-terminal domain of the RSV capsid protein.

Data were collected from a single frozen crystal at the Advanced Photon Source experimental station 14-BM-d.

	λ_1	λ_2	λ_3	λ_4	λ_5
Wavelength† (Å)	0.907, 'remote'	0.978, 'peak'	0.978, 'inflexion'	0.978, 'peak'	0.978, 'inflexion'
Outer resolution limit for phasing (Å)	2.7	2.7	2.7	2.7	2.7
Total observations	34201	35789	36048	35721	35725
Unique observations	8822	8883	8855	8887	8889
Completeness (%)	97	98	98	98	98
$R_{measure}$‡: centric observations	0.034	0.037	0.036	0.034	0.035
$R_{measure}$‡: acentric observations	0.048	0.056	0.038	0.064	0.046

† Difficulties in calibrating the monochromator prevented accurate measurement of the wavelength. ‡ $|^\lambda F(+)|$ and $|^\lambda F(-)|$ were treated as equivalent when calculating these statistics. Hence, differences between the centric observations [where $|^\lambda F(+)| = |^\lambda F(-)|$] and the acentric observations [where $|^\lambda F(+)| \neq |^\lambda F(-)|$] reflect the size of the anomalous signal. $R_{measure}$ is the robust indicator of data consistency introduced by Diederichs & Karplus (1997). It is similar to the more commonly reported R_{merge}, but is not dependent on the redundancy of the data.

MAD data collection had previously been exposed to high-intensity X-rays and subsequently stored at liquid-nitrogen temperatures. At the time of initial exposure, the X-ray absorption spectra of the crystal appeared normal.

To calculate the structure-factor amplitudes of the anomalously scattering Se atoms, measurements made at λ_1 and λ_2 were used (Table 1), for which there is a large contrast in the real part ($^\lambda f'$) of the anomalous scattering factor. Because of difficulties calibrating the monochromator, $^\lambda f'$ and $^\lambda f''$ were
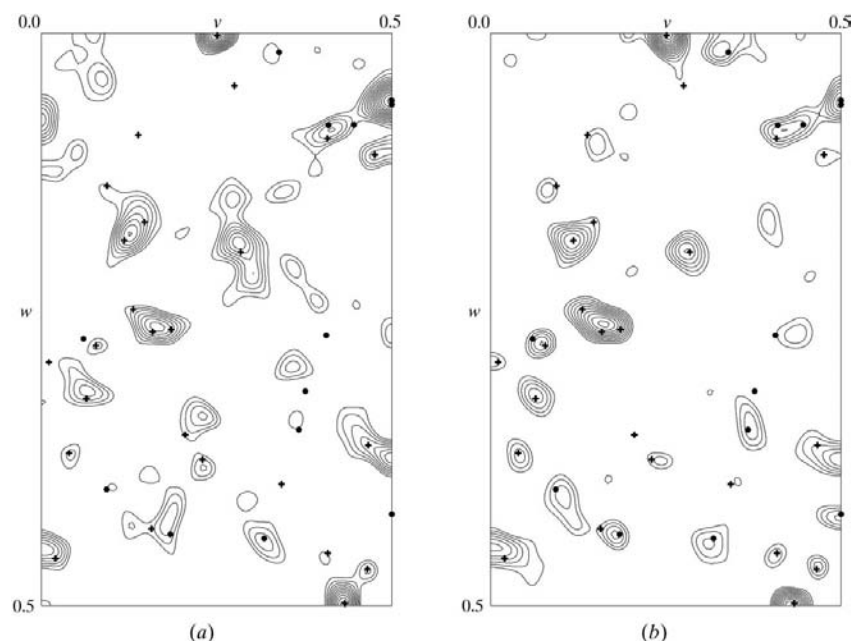

Figure 2
A comparison of Patterson syntheses. A Harker section ($u = 0.5$) of each synthesis is displayed. (a) Calculated with coefficients $[|^{\lambda_2}F(+)| - |^{\lambda_2}F(-)|]^2$ (Rossmann, 1961). (b) Calculated with coefficients $(^{\lambda_2}f''/^\circ f)^2|^\circ F_A|^2$, estimated as described in this paper. The positions of the predicted vector peaks resulting from the Se atoms, lying within 2.0 Å of the displayed section, are also indicated. Self-vectors are marked with solid circles and cross-vectors with solid crosses. Data to 3.2 Å resolution were used to compute the syntheses, which were contoured in intervals of 0.5σ, beginning at 1σ (where σ is the standard deviation of the synthesis).

not accurately known at both of these wavelengths. Using the procedures described in §2.2 above, we estimated $k_1^2 = 0.57$ and $k_2^2 = 2.63$ directly from the X-ray diffraction data and used (6) to calculate $(^{\lambda_2}f''/^\circ f)^2|^\circ F_A|^2$ for each reflection. This resulted in estimates for 94% of the possible reflections out to 3.2 Å resolution.

For λ_1 (= 0.907 Å), which is remote from the absorption edge, the values of the anomalous scattering factor will be adequately described by theory (Cromer, 1983). Taking $^{\lambda_1}f'' = 3.3$ e and with $k_1 = {}^{\lambda_1}f''/^{\lambda_2}f'' = 0.75$, the estimated value for $^{\lambda_2}f''$ at the second wavelength is $^{\lambda_2}f'' = 4.4$ e. This value, which is only slightly larger than the theoretical value at the selenium K edge (3.8 e), is consistent with the absence of pronounced white-line enhancement seen in the X-ray absorption spectra of the crystal.

Using the estimates for $(^{\lambda_2}f''/^\circ f)^2|^\circ F_A|^2$, the positions of nine of the ten Se atoms were readily located using a direct-methods procedure coupled with phase annealing (Sheldrick, 1990), as implemented in the program *SHELXL*97. In an alternative approach, the same sites were located sequentially using a Patterson-based translation function within the program *AMoRe* (Navaza & Saludjian, 1997) and a single atom as a search model (see Vagin & Teplyakov, 1998).

In Fig. 2, a Patterson synthesis with coefficients $(^{\lambda_2}f''/^\circ f)^2|^\circ F_A|^2$, estimated as described above, is compared with a synthesis with coefficients $[|^{\lambda_2}F(+)| - |^{\lambda_2}F(-)|]^2$, the approximation suggested by Rossmann (1961). Also shown are the anticipated positions of the interatomic vectors resulting from the Se atoms. As expected, the effect of using the exact expression introduced by Singh & Ramaseshan (1968a) (equation 6) in place of the approximation suggested earlier by Rossmann (1961) (equation 3) is to suppress artifacts in the synthesis.

4. Summary

A method to recover the structure-factor amplitudes of the anomalous scatterers from MAD data collected at two wavelengths is presented. The expression used is that of Singh & Ramaseshan (1968a) and it is indicated how it is derived from the standard algebraic equations for analyzing MAD data. It is shown how to proceed when the anomalous contributions $^\lambda f'$ and $^\lambda f''$ to the atomic scattering factor are unknown. The calcula-

tions described have been implemented in a short computer program which will be made available through the *CCP*4 program suite.

References

Cascarano, G. & Giacovazzo, C. (1984). *Acta Cryst.* A**40**, 305–306.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cromer, D. T. (1983). *J. Appl. Cryst.* **16**, 437.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
Dodson, E. & Vijayan, M. (1971). *Acta Cryst.* B**27**, 2402–2411.
Evans, G. & Wilson, K. S. (1999). *Acta Cryst.* D**55**, 67–76.
Fan, H.-F., Woolfson, M. M. & Yao, J. (1993). *Proc. R. Soc. London. Ser. A*, **442**, 12–32.
Fourme, R., Shepard, W. & Kahn, R. (1996). *Prog. Biophys. Mol. Biol.* **64**, 167–199.
Gonzàlez, A., Pédelacq, J.-D., Solà, M., Gomis-Rüth, F. X., Coll, M., Samama, J.-P. & Benini, S. (1999). *Acta Cryst.* D**55**, 1449–1458.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
Hendrickson, W. A., Smith, J. L., Phizackerly, R. P. & Merrit, E. A. (1988). *Proteins*, **4**, 77–88.
Howell, P. L., Blessing, R. H., Smith, G. D. & Weeks, C. M. (2000). *Acta Cryst.* D**56**, 604–617.
Iwasaki, H. & Ito, T. (1977). *Acta Cryst.* A**33**, 227–229.
Karle, J. (1980). *Int. J. Quantum Chem. Quantum Biol. Symp.* **7**, 357–367.
Karle, J. (1984). *Acta Cryst.* A**40**, 1–4.
Karle, J. (1989a). *Phys. Today*, **42**(6), 22–29.
Karle, J. (1989b). *Acta Cryst.* A**45**, 303–307.
Kartha, G. & Parthasarathy, R. (1965). *Acta Cryst.* **18**, 745–749.
Kimber, A. C. (1983). *Appl. Statist.* **32**, 7–14.
Kingston, R. L., Fitzon-Ostendorp, T., Eisenmesser, E. Z., Schatz, G. W., Vogt, V. M., Post, C. B. & Rossmann, M. G. (2000). *Structure*, **8**, 617–628.
Klop, E. A., Krabbendam, H. & Kroon, J. (1989a). *Acta Cryst.* A**45**, 203–208.
Klop, E. A., Krabbendam, H. & Kroon, J. (1989b). *Acta Cryst.* A**45**, 609–613.
Matthews, B. W. (1966). *Acta Cryst.* **20**, 230–239.
Moreno, C. (1996). *Meas. Sci. Technol.* **7**, 137–141.
Narayan, R. & Ramaseshan, S. (1981). *Acta Cryst* A**37**, 636–641.
Navaza, J. & Saludjian, P. (1997). *Methods Enzymol.* **276**, 581–594.
Okaya, Y., Saito, Y. & Pepinsky, R. (1955). *Phys. Rev.* **98**, 1857–1858.
Phillips, J. C. & Hodgson, K. O. (1980). *Acta Cryst.* A**36**, 856–864.
Ramakrishnan, V. & Biou, V. (1997). *Methods Enzymol.* **276**, 538–557.
Rogers, D. (1965). *Computing Methods in Crystallography*, edited by J. S. Rollet, pp. 117–148. Oxford: Pergamon Press.
Rossmann, M. G. (1961). *Acta Cryst.* **14**, 383–388.
Sharff, A. J., Koronakis, E., Luisi, B. & Koronakis, V. (2000). *Acta Cryst.* D**56**, 785–788.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Singh, A. K. & Ramaseshan, S. (1966). *Acta Cryst.* **21**, 279–280.
Singh, A. K. & Ramaseshan, S. (1968a). *Acta Cryst.* B**24**, 35–39.
Singh, A. K. & Ramaseshan, S. (1968b). *Acta Cryst.* B**24**, 1701–1702.
Smith, J. L. & Thompson, A. (1998). *Structure*, **6**, 815–819.
Srinivasan, R. & Parthasarathy, S. (1976). *Some Statistical Applications in X-ray Crystallography*. Oxford: Pergamon Press.
Terwilliger, T. C. (1994). *Acta Cryst.* D**50**, 11–16.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Tickle, I. J. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 87–95. Warrington: Daresbury Laboratory.
Vagin, A. & Teplyakov, A. (1998). *Acta Cryst.* D**54**, 400–402.
Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
Wilson, A. J. C. (1950). *Acta Cryst.* **3**, 258–261.
Woolfson, M. M., Yao, J.-X. & Fan, H.-F. (1997). *Acta Cryst.* D**53**, 673–681.