Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Anomalous signal indicators in protein crystallography

P. H. Zwart‡

Basic Research Programme, SAIC-Frederick Inc., Argonne National Laboratories, Argonne, IL 60549, USA

‡ Current address: Lawrence Berkeley National Laboratory, One Cyclotron Road, Building 64R0121, Berkeley, California 94720, USA.

Correspondence e-mail: phzwart@lbl.gov

Received 8 April 2005
Accepted 23 July 2005

A Monte Carlo procedure is described that generates random structure factors with simulated errors corresponding to an X-ray data set of a protein of a specific size and given heavy-atom content. The simulated data set can be used to estimate Bijvoet ratios and figures of merit as obtained from SAD phasing routines and can be used to gauge the feasibility of solving a structure *via* the SAD method. In addition to being able to estimate results from phasing, the simulation allows the estimation of the correlation coefficient between $|\Delta F|$, the absolute Bijvoet amplitude difference, and F_A, the structure-factor amplitude of the heavy-atom model. As this quantity is used in various substructure-solution routines, the estimate provides a rough estimate of the ease of substructure solution. Furthermore, the Monte Carlo procedure provides an easy way of estimating the number of significant Bijvoet intensity differences, denoted as the measurability, and is proposed as an intuitive measure of the quality of anomalous data.

1. Introduction

Structure solution by the single-wavelength anomalous diffraction (SAD) method has become increasingly popular over the last couple of years (*e.g.* Dauter *et al.*, 2002; Dodson, 2003; Olczak *et al.*, 2003; Yang *et al.*, 2003). This increase in popularity can be largely ascribed to the improvement and increasing availability of diffraction data-collection facilities (*e.g.* Cassetta *et al.*, 1999; Fourme *et al.*, 1999; Helliwell, 1992; Hendrickson, 1999; Pohl *et al.*, 2001), new and improved data-collection procedures (*e.g.* Alkire *et al.*, 2004; Weiss, Sicker, Djinovic Carugo *et al.*, 2001) and novel theoretical developments implemented in user-friendly software (*e.g.* de La Fortelle & Bricogne, 1997; Pannu & Read, 2004; Schneider & Sheldrick, 2002; Terwilliger, 2003; Weeks & Miller, 1999). SAD phasing relies on the presence of 'anomalously' scattering atoms that cause the violation of Friedel's Law. The differences in Bijvoet-related intensities, the so-called anomalous differences, are used for substructure solution and subsequent phasing. As these differences are expected to be only a small fraction of the total signal for each reflection, accurate measurements and a proper statistical treatment of the errors are vital for a successful structure-solution process. If accurate data are available, Wang (1985) predicted that data sets with a Bijvoet amplitude ratio ($\langle\Delta F\rangle/\langle F\rangle$) as low as 0.6% would be solvable by the SAD method and this has recently been confirmed experimentally (Banumathi *et al.*, 2004; Ramagopal *et al.*, 2003). Although the expected Bijvoet amplitude ratio is a useful quantity indicating the expected amount of anomalous difference on the amplitudes, it does not show in a straightforward manner what the results of a successful phasing procedure would be. Furthermore, the

Table 1
Notation.

Symbol	Meaning
$\mathbb{E}[g(x)]_x$	Expectation value of $g(x)$, with x as the random variable
$\langle g(x)\rangle$	Population average of $g(x)$; $(1/M)\sum_{j=1}^{M} g(x_j)$
F^+, F^-	Friedel-related amplitudes
ΔF	Bijvoet amplitude difference; $F^+ - F^-$
F	Friedel averaged amplitude
F_A	Structure-factor amplitude of heavy-atom substructure; $(A_{heavy}^2 + B_{heavy}^2)^{1/2}$
I^+, I^-	Friedel-related intensities
I_{obs}^+, I_{obs}^-	Friedel-related intensities with an added random error
σ_I^2	Variance of additive error on intensities
ΔI	Bijvoet intensity difference; $I^+ - I^-$
N_{heavy}	No. of heavy atoms
N_{light}	No. of light atoms
f_{light}	Form factor of light atoms
f_{heavy}	Form factor of heavy atoms
f'_{heavy}	Dispersive correction on form factor of heavy atoms
f''_{heavy}	Anomalous correction on form factor of heavy atoms
d^*	The inverse resolution
κ	$f''_{heavy}/(f_{heavy} + f'_{heavy})$
$\gamma\cdot(d^*)$	Term accounting for geometric regularities in average intensities
σ_{light}^2	Expected intensity owing to light-atom part of total structure
σ_{heavy}^2	Expected intensity owing to heavy-atom part of total structure
$N(\mu, \sigma^2)$	The normal distribution with mean μ and variance σ^2
$B_{Wil,\ heavy}$, $B_{Wil,light}$	Wilson B value for light and heavy atoms
A_{light}, B_{light}	Real and imaginary part of the light-atoms (protein) structure factor
A_{heavy}, B_{heavy}	Real and imaginary part of the heavy-atom structure factor
$A_{heavy,\ error}$, $B_{heavy,\ error}$	A complex error vector modelling substructure errors
A_{heavy}^{obs}, B_{heavy}^{obs}	Real and imaginary part of the heavy-atom structure factor, including an error term
D	Luzzati D value
q	Heavy-atom substructure error-control parameter
$Q(d^*)$	A function describing the $\mathbb{E}(I/\sigma_I)$ behavior as a function of resolution
Measurability (this paper)	See equations (29) and (30)

crystallographic experiment records intensities rather than amplitudes, so the Bijvoet intensity ratio would be a quantity that is more closely related to the experiment. A major drawback of both the amplitude and intensity ratios is that they are relatively sensitive to experimental errors, diminishing the practical use of these indicators for assessing the strength of the anomalous signal in a given data set.

In the following sections of this report, a number of classical anomalous difference indicators are reviewed. A number of these indicators, given some characteristics of a data set, will be computed numerically *via* a Monte Carlo simulation and compared with the estimates obtained *via* their theoretical expressions. A major benefit of the simulation procedure over the theoretical expressions is that in addition to being able to directly investigate the effects of errors on various anomalous signal indicators, the expected cosine of the phase error after SAD phasing can be computed. The expected cosine of the phase error is related to the information content of the phase probability distributions and will thus give a much clearer indication whether or not a particular SAD phasing scenario is feasible. Another advantage is that the correlation coefficient between the normalized (Giacovazzo, 2001) values of ΔF and F_A can be obtained easily. This correlation coefficient can be used as an indicator of the ease of the substructure-solution process.

2. Terminology and notation

Some confusion in the (protein) crystallographic literature is present with regards to the term Bijvoet ratio. In papers from S. Parthasarathy and coworkers (*e.g.* Parthasarathy, 1967; Parthasarathy & Parthasarathi, 1973), the term 'Bijvoet ratio' was reserved for the Bijvoet intensity difference divided by the average of Friedel-related intensities. In the more recent literature, however, amplitudes are used. Furthermore, the term 'expected Bijvoet ratio' in early papers refers to the expected value of the ratio, whereas in the more recent literature the term 'expected Bijvoet ratio' is used for the ratio of expectation values. These different Bijvoet ratios are of course all related. To overcome these difficulties in nomenclature, the term Bijvoet intensity ratio or Bijvoet amplitude ratio will be used as well as a referral to the expressions in order to avoid confusion.

Expectation values will be denoted by $\mathbb{E}[g(x)]_x$, indicating that the expectation value of the function $g(x)$ is obtained by integrating the probability distribution of x multiplied by $g(x)$. Table 1 summarizes the symbols used in this text.

3. Anomalous signal indicators

A number of indicators have been developed to assess or predict the expected anomalous signal within a data set. They will be briefly reviewed in the following sections.

3.1. The Bijvoet amplitude ratio

The expectation value of the ratio of the root of the mean-square absolute Bijvoet amplitude difference and the root of the mean-square amplitude has been deduced by Hendrickson & Teeter (1981) using a zero diffraction-angle approximation (see also Hendrickson *et al.*, 1985; Hendrickson & Ogata, 1997). Dauter *et al.* (2002) introduced angular dependence into this estimate and showed a good fit to experimental data at low resolution. A difference in atomic displacement parameters between the protein atoms and the heavy-atom model as well as a modulating term describing effects arising from geometric regularities in the heavy-atom model was incorporated in the Bijvoet amplitude ratio estimate by Shen *et al.* (2003). A drawback of the expression of Shen and coworkers is the lack of the incorporation of the effects of structural regularities in the protein model and its effect on the average intensity as well as a number of issues in the published derivation, most of which have to do with the fact that the expectation value of a function of a random variable is not equal to the function of the expectation value of the random variable. A re-derivation of the expression published by Shen *et al.* (2003) results in (see Appendix *A*)

$$\mathbb{E}(|\Delta F|) = \frac{2\kappa}{\pi}[\mathbb{E}(F_{\rm A}^2)]^{1/2}. \quad (1)$$

κ is equal to the ratio of the imaginary and real parts of the form factor of the heavy atom and $F_{\rm A}$ is the structure-factor amplitude of the heavy-atom substructure without the anomalous correction (see Table 1). Taking into account the effects of geometric regularities (Zwart & Lamzin, 2004) in the heavy-atom substructure on the average intensity and assuming a Wilson distribution on $F_{\rm A}$, this results, for a substructure containing a single type of heavy atom, in (Appendix *A*)

$$\mathbb{E}(|\Delta F|) = \frac{2}{\pi^{1/2}}\{N_{\rm heavy} f''^2_{\rm heavy}[1 + \gamma_{\rm heavy}(d^*)]\}^{1/2}. \quad (2)$$

The expectation value of the average amplitude is equal to

$$\mathbb{E}(F) = \frac{\pi^{1/2}}{2}\{N_{\rm heavy} f^2_{\rm heavy}[1 + \gamma_{\rm heavy}(d^*)] + N_{\rm light} f^2_{\rm light}[1 + \gamma_{\rm light}(d^*)]\}^{1/2}, \quad (3)$$

where

$$\gamma_{\cdot} = \mathbb{E}[\sin(2\pi a d^*)/(2\pi a d^*)]_a. \quad (4)$$

a is a bond length between two atoms in the heavy-atom or protein model. The expectation value in (4) is obtained by averaging over all interatomic distances in a given model. d^* is equal to the inverse resolution.

The expected Bijvoet amplitude ratio is thus

$$\frac{\mathbb{E}(|\Delta F|)}{\mathbb{E}(F)} = \frac{4}{\pi}\left\{\frac{N_{\rm heavy} f''^2_{\rm heavy}[1 + \gamma_{\rm heavy}(d^*)]}{N_{\rm heavy} f^2_{\rm heavy}[1 + \gamma_{\rm heavy}(d^*)] + N_{\rm light} f^2_{\rm light}[1 + \gamma_{\rm light}(d^*)]}\right\}^{1/2}. \quad (5)$$

(5) differs from the expression given by Shen *et al.* (2003) by inclusion of a correlation term for both the light-atom and heavy-atom components and by the fact that *B* values of the heavy-atom substructure are assumed to be equal to those of the protein atoms. Inclusion of a term accounting for a possible *B*-value difference between protein and heavy atoms is straightforward, however, and has been omitted for clarity. In (5) equal atoms for the light atoms are assumed rather than an average form factor as used by Shen *et al.* (2003). In practice however, $N_{\rm light} f^2_{\rm light}$ is replaced by (Weeks *et al.*, 2005)

$$N_{\rm res}[5.0 f_{\rm C}^2(d^*) + 1.2 f_{\rm N}^2(d^*) + 1.5 f_{\rm O}^2(d^*) + 8.0 f_{\rm H}^2(d^*)], \quad (6)$$

where $f_{\cdot}(d^*)$ are the scattering factors for carbon, nitrogen, oxygen and hydrogen, respectively, and $N_{\rm res}$ is the number of residues.

Note that the expectation value (5) differs from the expectation value given by Hendrickson & Teeter (1981) or Hendrickson & Ogata (1997), who define the Bijvoet ratio in terms of the ratio of root mean squares rather than absolute values.

For an economical use of symbols and clarity of the resulting expressions, *B*-value corrections are omitted and equal atoms are assumed throughout the rest of this paper. However, in subsequent computations *B*-value corrections and substitution (6) are used.

3.2. The expected value of the absolute Bijvoet intensity difference

The probability distribution and expectation value of the modulus of the Bijvoet difference has been derived by Parthasarathy & Srinivasan (1964), who showed that the distribution of a normalized form of the Bijvoet intensity difference denoted by x,

$$x = \frac{|I^+ - I^-|}{4(\sum f^2_{\rm light} \sum f''^2_{\rm heavy})^{1/2}} \quad (7)$$

is distributed according to

$$p(x) = 2\exp(-2x). \quad (8)$$

The expectation value of x is then $\frac{1}{2}$ and thus (with $|\Delta I| = |I^+ - I^-|$)

$$\mathbb{E}(|\Delta I|) = 2(\sum f^2_{\rm light} \sum f''^2_{\rm heavy})^{1/2}. \quad (9)$$

Under an equal atom assumption for both the sets of light-atom and heavy-atom groups, the latter expression simplifies to

$$\mathbb{E}(|\Delta I|) = 2 f_{\rm light} f''_{\rm heavy}(N_{\rm light} N_{\rm heavy})^{1/2}. \quad (10)$$

It must be noted that the underlying assumption in this model is that the number of heavy atoms is large enough to ensure that the structure-factor amplitude of the heavy-atom substructure is distributed according to a Wilson distribution. For substructure containing a small number of heavy atoms (up to three), specific distributions are available (Parthasarathy & Srinivasan, 1964). In this work, it is assumed that the structure-factor amplitudes of the heavy-atom model follow a Wilson distribution.

3.3. Bijvoet intensity ratios

A logical use of (10) would be in the estimation of the ratio of the average absolute Bijvoet difference and the average intensity,

$$\frac{\mathbb{E}(|I^+ - I^-|)}{\mathbb{E}[\frac{1}{2}(I^+ + I^-)]} = \frac{2 f_{\rm light} f''_{\rm heavy}(N_{\rm light} N_{\rm heavy})^{1/2}}{N_{\rm light} f^2_{\rm light} + N_{\rm heavy} f^2_{\rm heavy} + N_{\rm heavy} f''^2_{\rm heavy}}. \quad (11)$$

A simplified form of (11) has been derived by Einspahr *et al.* (1985) (equation 2) on the basis of the Crick & Magdoff (1956) approximation.

The Bijvoet ratio, denoted by δ, in the definition of Zachariasen (1965) and Parthasarathy (1967), is equal to

$$\delta = \frac{|I^+ - I^-|}{\frac{1}{2}(I^+ + I^-)}. \quad (12)$$

The derivation of the expectation value of δ, the expectation value of a ratio, is less straightforward compared with the ratio of expression values (11) owing to the dependence between numerator and denominator. The derivation given by

Parthasarathy (1967) results, after an equal-atom approximation, in

$$\mathbb{E}(\delta) = \frac{4 f_{\rm light} f''_{\rm heavy} (N_{\rm light} N_{\rm heavy})^{1/2}}{N_{\rm light} f_{\rm light}^2 + N_{\rm heavy} f_{\rm heavy}^2}. \quad (13)$$

The latter expectation value has been obtained using the approximation that the contribution from $f''_{\rm heavy}$ to the mean intensity can be neglected. The cumulative distribution function of the Bijvoet ratio has been derived by Parthasarathy & Parthasarathi (1973). It is very straightforward to incorporate $\gamma(d^*)$ and B-value correction terms in (10), (11) and (13), which are omitted for brevity but are included in subsequent computations. From an experimental point of view, (13) is more useful than (5) and (11) as it indicates the expected signal in a randomly chosen reflection. Although (11) and (13) give an indication of the strength of the anomalous signal, the number of reflections with a significant Bijvoet difference might be a more useful quantity to assess the feasibility of a SAD experiment, as this is directly related to the number of reflections used in modern direct-methods programs to solve the substructure.

3.4. The measurability

A quantity related to the Bijvoet intensity ratio is the measurability $M(\delta_0, z_0)$ (Parthasarathy & Parthasarathi, 1974, 1976; Parthasarathy & Ponnuswamy, 1976, 1981*a*,*b*; Sekar & Parthasarathy, 1987; Velmurugan *et al.*, 1979; Velmurugan & Parthasarathy, 1984). A measurable Bijvoet difference is defined as a Bijvoet intensity ratio δ (12) larger than a given value δ_0, while the smallest intensity of the (normalized) intensity pair (I^+, I^-) is larger than a given value z_0. The percentage of reflections that fulfill both latter conditions is denoted as the measurability. A graphical interpretation of the measurability is depicted in Fig. 1. The measurability estimate differs from the Bijvoet intensity and amplitude-ratio estimates in a sense that it tries to assess more directly the feasibility of a SAD experiment in terms of estimating the number of reliably measurable Bijvoet differences. A major drawback of the published work on the measurability is the lack of incorporation of the effects of experimental errors.

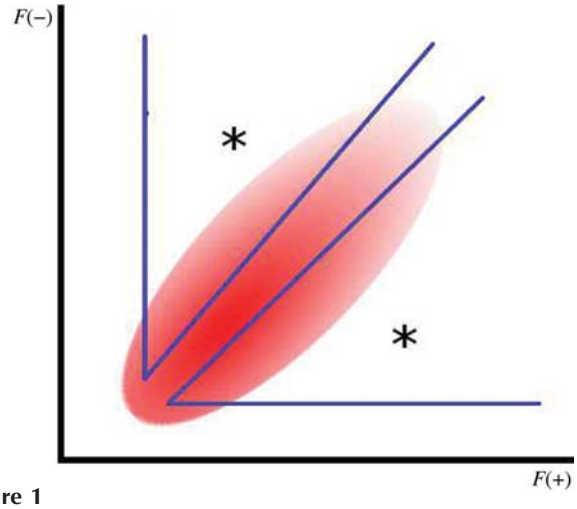

Figure 1
The measurability of Bijvoet differences is defined as the fraction of Bijvoet differences that can be measured accurately. The regions bounded by the blue lines and indicated by the asterisks have a Bijvoet intensity difference and intensities large enough to be measured accurately. Integrating the probability distribution (in red) in this area will result in an estimate of the measurability of the Bijvoet difference.

Interestingly, the measurability in the defintion by Parthasarathy & Parthasarathi (1974) was used by Rossmann in 1961 to assess the strength of the anomalous signal in early studies on locating anomalous scatterers by anomalous difference Patterson methods (Rossmann, 1961).

3.5. Post-data-processing anomalous indicators

In addition to the Bijvoet ratios and measurability estimates, other indicators are used to assess the quality and amount of anomalous signal. A prominent statistic is the observed average anomalous signal-to-noise ratio,

$$\langle |I^+ - I^-| / (\sigma_{I^+}^2 + \sigma_{I^-}^2)^{1/2} \rangle. \quad (14)$$

This statistic is often used to determine the resolution limit up to which the anomalous signal can still be considered to be significant (Mukherjee *et al.*, 1989). The anomalous signal-to-noise expression using amplitudes rather than intensities is often used as well (Usón *et al.*, 2003). The amplitude-based signal-to-noise criteria are typically a couple of percent larger than the intensity-based average anomalous signal-to-noise ratio, especially at the high-resolution limit. Both the intensity-based and amplitude-based anomalous signal-to-noise ratio can be used to judge the strength of the anomalous signal. For the amplitude-based criterion, a value of 0.8 or lower indicates no anomalous signal and amplitude-based anomalous signal-to-noise ratios larger than 1.2 indicate the presence of significant signal (Sheldrick, 2004). A quantity related to the average anomalous signal-to-noise ratio is the number or fraction of Bijvoet differences whose absolute value is larger than three times its estimated standard deviation (Hädener *et al.*, 1999). The latter criterion is very closely related to the measurability as defined by Parthasarathy & Parthasarathi (1974) as discussed in §3.4.

Another quantity frequently used to judge the strength of the anomalous signal is the correlation between the anomalous difference between multiple data sets (Buehner *et al.*, 1974; Schneider & Sheldrick, 2002). Although this measure was developed to be used for MAD data sets, the same statistic can be computed for a SAD data set by artificially splitting the collected frames into two distinct sets and computing the correlation between the intensity differences in the two half data sets (Evans, 2005). The major benefit of this criterion is that it does not depend on the estimated standard deviations of the individual intensities. The approach developed by Fu *et al.* (2004) tries to assess the strength of the anomalous signal by comparing the intensity differences between Friedel-related centric and acentric reflections. Because the intensity differences of Friedel-related centric reflections have a theoretical Bijvoet difference equal to zero, the observed

differences can be used to 'calibrate' the error model for the acentric differences. In the *HKL* suite (Otwinowski & Minor, 1997), the presence of anomalous signal in the data is judged by comparing the χ^2 values as obtained by merging the data with and without averaging Friedel pairs. If a significant anomalous signal is present, merging the Friedel-related pairs will result in χ^2 values that are significantly larger than unity. A related test to detect the presence of an anomalous signal is the normal probability plot (Howell & Smith, 1992). The normal probability plot method for the detection of anomalous signal tries to assess whether or not a set of anomalous differences normalized by their estimated standard deviation is distributed according to a Gaussian distribution with a unit variance (Evans, 2005). Significant deviation from the standard normal distribution indicates the presence of an anomalous signal.

3.6. Limitations

Although the various Bijvoet ratios predict the expected amount of anomalous signal with various degrees of accuracy, these estimates fail to quantify the success of a subsequent SAD phasing procedure. Clearly, not only the size of f'' relative to the amount of experimental error plays a role, but the normal scattering power of the substructure also influences the resulting SAD phase probability distribution. The influence of the known partial structure on the SAD phase probability distribution is clear when looking at the maximum-likelihood SAD function as given by McCoy *et al.* (2004). This function consists of two components: a term describing a symmetric bimodal phase probability arising from the (trigonometric) SAD phase ambiguity and a term describing the phase probability of the total phase given the fact that part of the structure has been located. The latter term is known as the Sim contribution (Sim, 1964) and skews the phase probability towards one of the modes of the bimodal phase distribution. An excellent graphical illustration of this principle can be found in McCoy *et al.* (2004).

An example of the importance of the Sim contribution is illustrated in phasing a protein containing S atoms at a wavelength of 1.5 Å ($f'' \simeq 0.5$) and comparing the resulting phases with the SAD phases of a protein of similar size containing the same amount of Se atoms at a wavelength of 1.0 Å ($f'' \simeq 0.5$). Because the selenium partial structure provides a larger Sim contribution than the sulfur substructure, one expects better phases for the protein containing the Se atoms.

The major contribution to the success of a SAD phasing procedure is the accuracy of the data. Clearly, the more the observed data resembles error-free data, the easier subsequent phasing procedures are (Weiss, Sicker & Hilgenfeld, 2001). The effect of errors in the data on the Bijvoet ratio as well as measurability estimates have not been investigated thoroughly, although Dauter *et al.* (2002) speculate that the effects of errors on the observed Bijvoet amplitude ratio is probably large. A similar observation was made earlier by Einspahr *et al.* (1985). Attempts to include the effects of errors on the expected Bijvoet ratios and measurability would result in relatively complicated integrations that might not be straightforward to solve *via* analytical methods. The difficulties of an analytical method can be bypassed by using a numerical approach.

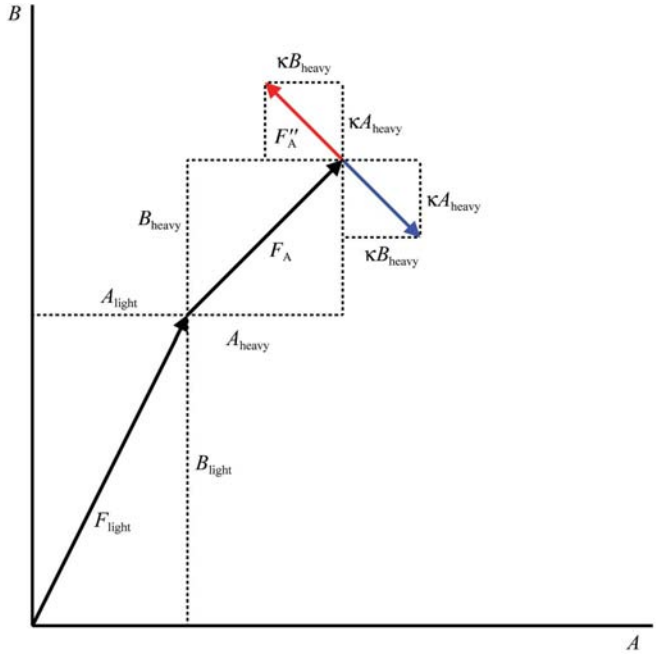

Figure 2
Real and imaginary components of the Friedel-related structure factors. Note that the f' contribution has been included in the form factor of the heavy-atom model and is not explicitly shown owing to the nature of the SAD experiment.

4. Simulated SAD data

Structure factors are usually assumed to be distributed according to a bivariate normal distribution in the complex plane (Wilson, 1942, 1949). Inclusion of anomalous scattering effects in this model is relatively straightforward and results in the joint probability distribution of Friedel-related amplitudes (Hauptman, 1982) rather than the Wilson distribution. The real and imaginary components of both Friedel mates of the total structure factor can be written as follows (see also Fig. 2),

$$A^{+}_{tot} = A_{light} + A_{heavy} - \kappa B_{heavy}, \quad (15)$$

$$B^{+}_{tot} = B_{light} + B_{heavy} + \kappa A_{heavy}, \quad (16)$$

$$A^{-*}_{tot} = A_{light} + A_{heavy} + \kappa B_{heavy}, \quad (17)$$

$$B^{-*}_{tot} = B_{light} + B_{heavy} - \kappa A_{heavy} \quad (18)$$

and

$$F^{+} = I^{+1/2} = [(A^{+}_{tot})^2 + (B^{+}_{tot})^2]^{1/2}, \quad (19)$$

$$F^{-} = I^{-1/2} = [(A^{-}_{tot})^2 + (B^{-}_{tot})^2]^{1/2}. \quad (20)$$

A_{light}, B_{light}, A_{heavy} and B_{heavy} denote the real and imaginary part of the structure factor from the protein (light-atoms) part and the heavy-atom substructure, respectively. The inequality between F^{+} and F^{-} results from the imaginary component of

the form factor of the heavy substructure, as can be seen from Fig. 2. The real and imaginary components (A, B) of the protein part and heavy-atom-substructure structure factors can be assumed to be both distributed according to a normal distribution centered on the origin, $N(0, \sigma_x^2/2)$. The variance terms for the protein part and heavy-atom substructure can be written as (Giacovazzo, 1998; Zwart & Lamzin, 2004)

$$\sigma^2_{\rm light} = N_{\rm light} f^2_{\rm light}[1 + \gamma_{\rm light}(d^*)] \exp(-B_{\rm Wil,light} d^{*2}/2), \quad (21)$$

$$\sigma^2_{\rm heavy} = N_{\rm heavy}(f_{\rm heavy} + f'_{\rm heavy})^2 \times [1 + \gamma_{\rm heavy}(d^*)] \exp(-B_{\rm Wil,heavy} d^{*2}/2). \quad (22)$$

A random structure factor can thus be generated by drawing two random numbers from $N(0, \sigma^2_{\rm light}/2)$ and two random numbers from $N(0, \sigma^2_{\rm heavy}/2)$. Using these random numbers in (15)–(18) a pair of Friedel-related structure factors is obtained from which amplitudes and intensities can be computed. A single pair of Friedel-related structure factors then behaves as if it was randomly picked from an X-ray data set from a protein structure with a heavy-atom substructure.

4.1. Measurement and substructure errors

A random structure-factor amplitude generated by the procedure outlined in the previous section can be regarded as 'error-free' with regard to the individual protein and heavy-atom-substructure components. However, an inherent limitation of experimental crystallography is the presence of measurement errors. In order to simulate the effect of experimental errors on the observable I^+, I^- pair, two random numbers from $N(0, \sigma_I)$ are drawn and added to the (I^+, I^-) pair, while making sure the resulting 'observed' intensities are positive. The latter positivity constraint is imposed to avoid the need for a Bayesian update of these intensities (French & Wilson, 1978). The variance of the simulated experimental error is chosen in such a way that $\mathbb{E}(I/\sigma_I) = Q(d^*)$, where $Q(d^*)$ is a function describing the signal-to-noise ratio as a function of the inverse resolution and, in this study, is parametrized by a exponential or polynomial function. If the variance of the experimental error is chosen to be equal to

$$\sigma_I^2 = \frac{I\pi^{1/2}}{2Q(d^*)}, \quad (23)$$

the average signal to noise is expected to be equal to the stipulated value $Q(d^*)$. Note that in this way strong reflections will have larger I/σ_I values than weak reflections, as is expected from experimental data.

The now generated pair of 'observed' intensities $I^+_{\rm obs}$, $I^-_{\rm obs}$ is used to compute $F^+_{\rm obs}$, $F^-_{\rm obs}$ by taking the square root of the intensities. Associated standard deviations are estimated *via* standard error-propagation techniques.

A fairly detailed error model has been published by Popov & Bourenkov (2003) and is in good agreement with experimental observations. Inclusion of this model will most likely provide a more realistic error model of the intensities than the rather approximate model currently adopted, but is beyond the scope of this paper.

Modelling of substructure errors is carried out by adding a complex error vector to the simulated error-free heavy-atom part of the structure factor,

$$A_{\rm heavy,error} = N[0, (1 - D^2)\sigma^2_{\rm heavy}/2], \quad (24)$$

$$B_{\rm heavy,error} = N[0, (1 - D^2)\sigma^2_{\rm heavy}/2] \quad (25)$$

and

$$A^{\rm obs}_{\rm heavy} = DA_{\rm heavy} + A_{\rm heavy,error}, \quad (26)$$

$$B^{\rm obs}_{\rm heavy} = DB_{\rm heavy} + B_{\rm heavy,error}. \quad (27)$$

For simplicity reasons, D is chosen to be equal to the classical Luzzati model (Luzzati, 1952; Read, 1986),

$$D = \exp[-2\pi^2 q^2 (d^*)^2], \quad (28)$$

where the constant q is a parameter controlling the dependence of D as a function of the inverse resolution d^*. The classic interpretation of q is related to an error in positional parameters, but should in this case rather be seen as a flexible way of modelling other errors not accounted for in the simulations but which are present in the heavy-atom refinement and subsequent phasing.

4.2. The ease of substructure solution and SAD phasing

As a result of the described simulation procedure, the structure factor of the heavy-atom substructure and a pair of 'observed' $F^+_{\rm obs}$, $F^-_{\rm obs}$ amplitudes is obtained. The structure factor of the heavy-atom model and the 'observed' amplitudes can first of all be used to compute the correlation coefficient $\rm CC_A$ between the absolute anomalous amplitude difference $|\Delta F|$ and the heavy-atom structure-factor amplitude, $F_{\rm A}$. For error-free data, this correlation coefficient is estimated to be equal to 0.692 (see Appendix *B*). Measurement errors on the data will tend to lower this correlation coefficient and possibly obscure the detection of heavy atoms if the correlation coefficient is too low. Although *SHELXD* uses a weighted correlation coefficient (Schneider & Sheldrick, 2002), weights are omitted in the calculations, corresponding to a run of *SHELXD* with the keyword CCWT 0. In practice, successful SAD phasing correlation coefficients lie roughly between 0.20 and 0.65. The successful identification of a solution produced by *SHELXD* also depends, especially for low correlation coefficients, on the contrast between a solution and clear non-solutions and possible consistency between positional parameters obtained in independent solutions (Grosse-Kunstleve & Adams, 2003). However, both these criteria of identifying a possible substructure solution cannot be obtained *via* the described simulation method. Nevertheless, a rough indication of the expected value of the correlation coefficient is useful, as it will indicate the ease of the substructure-solution process.

Predicting the results of a SAD phasing procedure is possible because the total 'observed' amplitudes and heavy-atom structure-factor components are available from the simulation. Phasing is carried out by the maximum-likelihood SAD function as outlined by McCoy *et al.* (2004). Estimates of figures of merit in resolution bins can be obtained by phasing each simulated reflection and averaging the estimated figures

of merit (Blow & Crick, 1959) over all reflections in the given resolution bin. If desired, the quality of the resulting SAD map can be quantified by the correlation coefficient, computed in reciprocal space, of the experimental map to the final map (Lunin & Woolfson, 1993).

The success of a subsequent solvent flattening to improve phases will largely depend on the solvent content and on the information content of the phase probability distribution. Another important prerequisite for the success of solvent flattening is the need for a reasonably well defined solvent mask. Again, the success of a subsequent solvent flattening is impossible to predict *via* the described simulation procedure, as density-modification techniques rely heavily on the correlation between phases. However, the figure of merit is a much clearer measure of the potential success of phasing than an estimated Bijvoet (amplitude) ratio only.

4.3. Numerical determination of Bijvoet ratios and measurability

Numerical estimates of the Bijvoet ratios are obtained by using resolution-dependent population-average equivalents of (5), (11) and (13). In order to assess the quality of a SAD data set, the definition of the measurability as described in §3.4 is modified to include the quality of the data. A Bijvoet difference is defined as 'measurable' if the following two conditions are met:

$$\frac{|I^{+}_{\rm obs} - I^{-}_{\rm obs}|}{[\sigma^{2}_{I(+)} + \sigma^{2}_{I(-)}]^{1/2}} \geq 3 \qquad (29)$$

and

$$\min[I^{+}_{\rm obs}/\sigma_{I(+)}, I^{-}_{\rm obs}/\sigma_{I(-)}] \geq 3. \qquad (30)$$

The expected measurability is equal to ratio of the number of measurable Bijvoet differences to the total number of simulated Bijvoet differences. Note that for error-free data the measurability is equal to 1. The measurability as defined above can thus be seen as a combination between the 'classical' measurability (Parthasarathy & Parthasarathi, 1974) and the anomalous signal-to-noise criterion as discussed in §3.5. Combing the condition (29) with (30) has the benefit that potential outliers are not included in the summary statistics describing the quality of the anomalous data set.

The definition of measurability that includes experimental errors is close to the criteria used in selecting Bijvoet differences in the SAD substructure-solution process (Mukherjee *et al.*, 1989; Blessing & Smith, 1999).

5. Results and discussion

The described routines have been implemented in a Python script (http://www.python.org) in such a way that the characteristics of the protein and heavy-atom model as well as parameters controlling the error model can be given upon input. In the following paragraphs, the Bijvoet ratios are determined numerically and compared with the analytical expressions and real data. Finally, results of the simulations are compared with experimental data. The $\gamma_{\rm light}$ term used in the simulations has been determined from 20 good-quality X-ray data sets, in a similar manner to that described by Zwart & Lamzin (2004).

5.1. Bijvoet intensity ratios and the effect of experimental errors

To validate (5), (11) and (13), a simulation has been carried out for a hypothetical protein containing 250 residues and six Se atoms. The B value for the protein has been set to 20 Å^2, whereas the heavy-atom substructure has a B value equal to 18 Å^2. The value of f' and f'' were set to -4 and 5.5 e, respectively. No geometric regularities in the heavy-atom substructure were assumed, resulting in $\gamma_{\rm heavy}(d^*) = 0$. The error on the intensities have been set to $\langle I/\sigma_I \rangle = 400$ throughout the resolution range. The results of the simulation are depicted in Fig. 3. Although the numerically obtained estimates of $\langle \Delta F \rangle / \langle F \rangle$ and $\langle \Delta I / I \rangle$ follow the results of (5) and (13) closely, the ratios are overestimated, especially at high resolution. The derivation of (11) involves fewer approximations than those needed to obtain (5) and (13) and matches the simulated data rather well.

In order to show the effects of errors on the expected Bijvoet ratio, four simulations have been carried out. In all four cases, the $\langle I/\sigma_I \rangle$ at 10 Å was set to 40. The $\langle I/\sigma_I \rangle$ at 2.5 Å was chosen to be 39, 10, 2 and 0.5 in the four different simulations. $\langle I/\sigma_I \rangle$ was set to decrease exponentially from the low-resolution shell to the stipulated value at 2.5 Å. The results of the simulation are shown in Fig. 4. Clearly, the effect of errors on $\langle \Delta F \rangle / \langle F \rangle$ is enormous, suggesting these types of plots are not very indicative of the amount of anomalous signal within a data set. However, a more intuitive feeling of the quality of the anomalous data is obtained from a plot of the estimated measurability (Fig. 5), correctly showing the falloff in measurable Bijvoet differences as a function of resolution.

5.2. Sulfur *versus* selenium SAD

A simulation was carried out on a hypothetical protein consisting of 250 residues and containing six methionines. The success of a possible S-SAD experiment at Cu wavelength is gauged by running the simulation assuming a nominal resolution of 2.0 Å. $\langle I/\sigma_I \rangle$ in the lowest resolution shell (at 10 Å) is assumed to be 80, whereas the $\langle I/\sigma_I \rangle$ at 2.0 Å is assumed to be equal to 4.0. The Wilson plot B value of the protein part was chosen to be equal to 20 Å^2 and the B value of the heavy-atom substructure was set to be 18 Å^2. The simulated $\langle$FOM$\rangle$ as a function of resolution is shown in Fig. 6.

Using the same set of parameters as in the previous case, but changing the heavy-atom type to selenium and assuming that data wre collected at 1.0 Å, where $f'' \simeq 0.5$, larger $\langle$FOM$\rangle$ values were found compared with the sulfur case (Fig. 6). The improvement in phasing is ascribed to the larger Sim contribution of the selenium substructure in comparison to the sulfur substructure.

Increasing f'' to 5.5 e for selenium and to 1.4 e for sulfur gives substantially better phases (Fig. 6). Note that the

expected figure of merit is increased at the resolution where the protein Wilson plot is at a minimum ($d^* \simeq 1/6$ Å^{-1}). This increase can be explained by the notion that at that particular resolution range the Sim contribution of the partial structure is increased because $\gamma_{\rm light}(d^*)$ is at a minimum. The sharp drop in $\langle$FOM$\rangle$ at $d^* \simeq 1/4.5$ Å^{-1} is explained in a similar fashion: $\gamma_{\rm light}(d^*)$ is large, which effectively decreases the fractional contribution of the substructure to the total expected intensity, resulting in a decrease in the Sim contribution.

5.3. Test data set Jia-peak

An X-ray data set, denoted by Jia-peak, of thioesterase II crystals from *Escherichia coli* (Li *et al.*, 2000) collected at the Se-peak wavelength was used in a comparison with the results of a simulation. The protein consists of 572 residues and has a total of eight Se atoms. The values of f' and f'' were estimated to be −3.6 and 5.4 e, respectively. The data extend to 2.5 Å and have a Wilson B value equal to 24 Å^2 (Morris *et al.*, 2004). The average B value of the Se atoms was assumed to be equal to 20 Å^2. The $\langle I/\sigma_I\rangle$ was 40 at 10 Å and 10 at 2.5 Å. A polynomial curve was fitted to describe the behavior of $\langle I/\sigma_I\rangle$ as a function of resolution. A simulation with parameters specifying the global statistics of the data set and the protein and substructure content was carried out. Resulting measurability and $\langle|\Delta I|\rangle/\langle I\rangle$ plots are shown in Figs. 7 and 8. Although the measurability estimates are reasonable, a significant disagreement between the observed and predicted values is present at low resolution. This is most likely to be a consequence of the simplistic nature of the assumed error model on the intensities in the simulation. Note that the 'dip' in $\langle|\Delta I|\rangle/\langle I\rangle$ is located at the resolution where the Wilson plot has a local maximum, as discussed in the previous section.

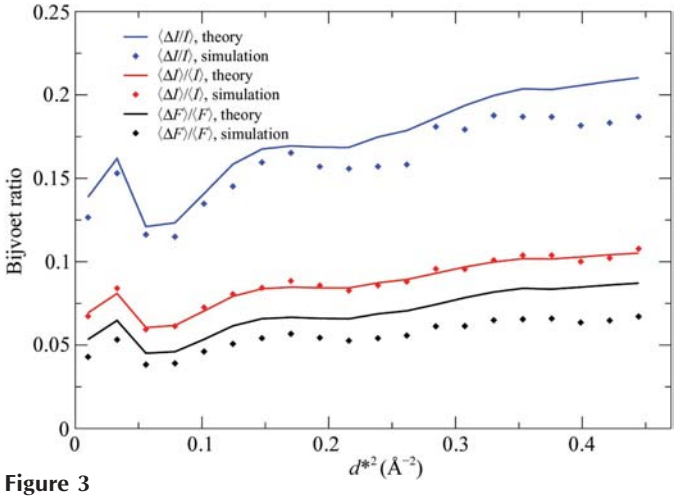

Figure 3
Bijvoet amplitudes and intensity ratios for error-free data as obtained numerically and through (5), (11) and (13). See text for details.

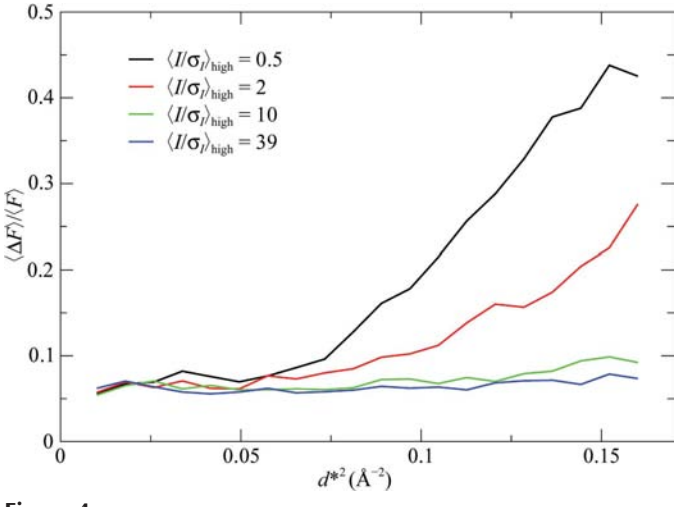

Figure 4
$\langle|\Delta F|\rangle/\langle F\rangle$ for four data sets with increasing error. The $\langle I/\sigma_I\rangle$ at 10 Å was set to 40 for all data sets and diminishes exponentially to 39 (in blue), 10 (in green), 2 (in red) and 0.5 (in black) at 2.5 Å.

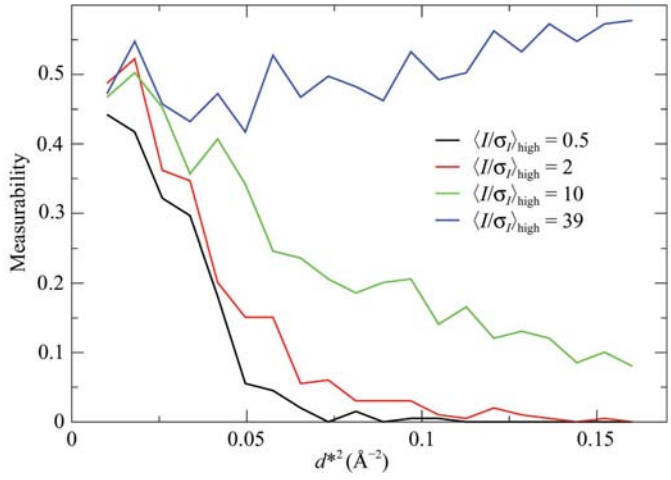

Figure 5
Measurability for four data sets with increasing error. The $\langle I/\sigma_I\rangle$ at 10 Å was set to 40 for all data sets and diminishes exponentially to 39 (in blue), 10 (in green), 2 (in red) and 0.5 (in black) at 2.5 Å.

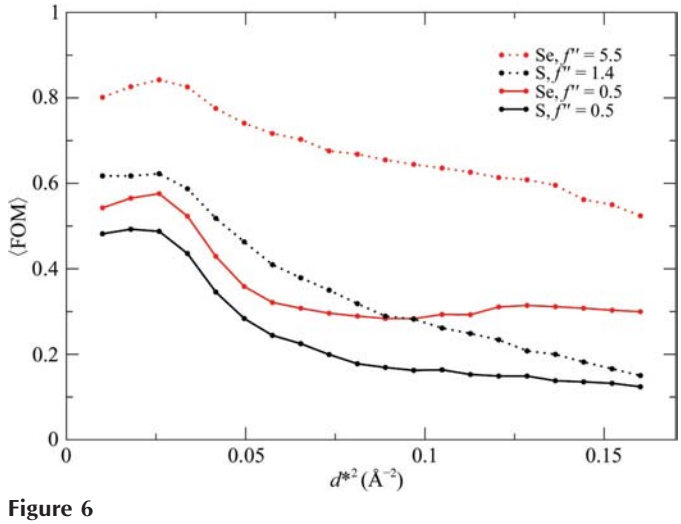

Figure 6
$\langle$FOM$\rangle$ for a 250-residue protein containing six heavy atoms, either S or Se, with different values of f''. For $f'' = 0.5$, the Se-SAD phasing is more successful than the S-SAD phasing owing to the larger contribution of the substructure to the total scattering power. See text for details.

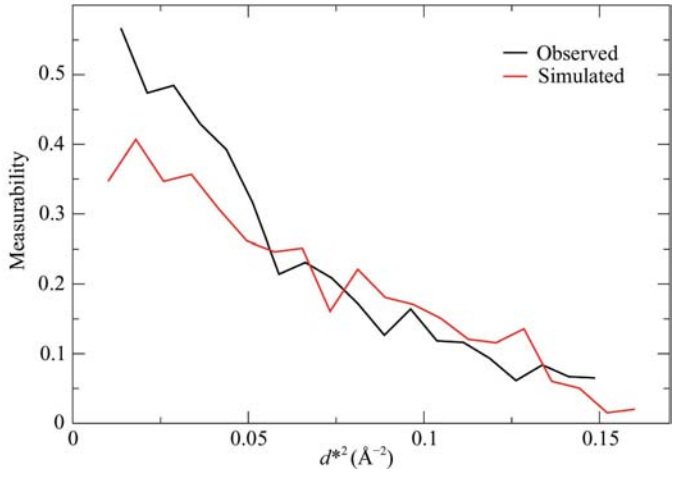

Figure 7
Measurability as a function of the squared inverse resolution for the Jia data as observed (in black) and sd estimated by the described simulation method (in red).

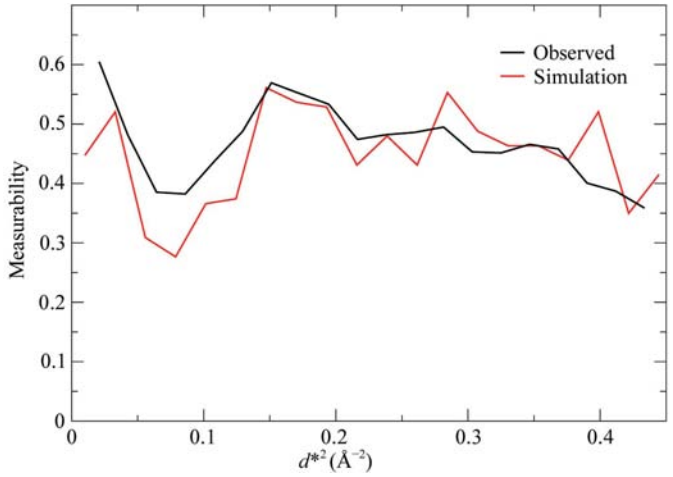

Figure 10
Measurability for the elastase data as observed in (in black) and as estimated *via* the described simulation procedure (in red). See text for details.

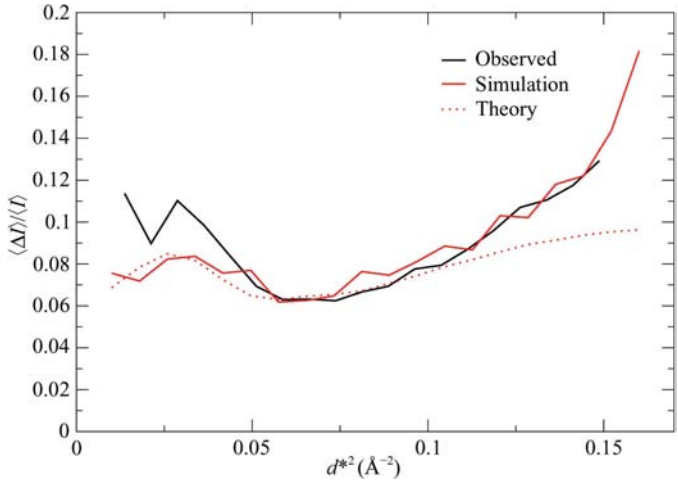

Figure 8
Estimated (red solid line) and observed (black solid line) $\langle|\Delta I\rangle/\langle I\rangle$ values for the Jia-peak data set. The red dotted line indicates the expected value of $\langle|\Delta I\rangle/\langle I\rangle$ for error-free data.

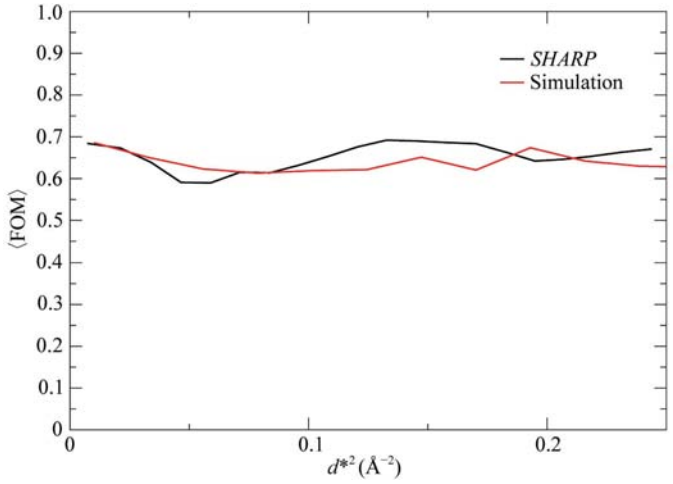

Figure 11
⟨FOM⟩ for the elastase data as obtained from *SHARP* (in black) and *via* the described simulation procedure (in red). See text for details.

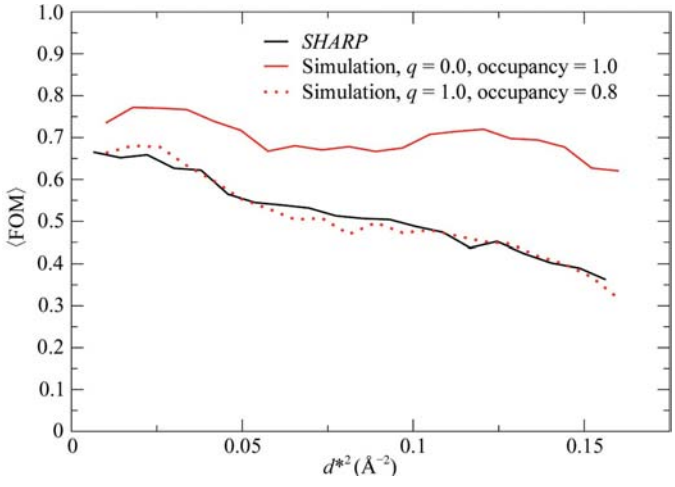

Figure 9
⟨FOM⟩ for the Jia-peak data as obtained from *SHARP* (in black) and *via* the described simulation procedure (in red). See text for details.

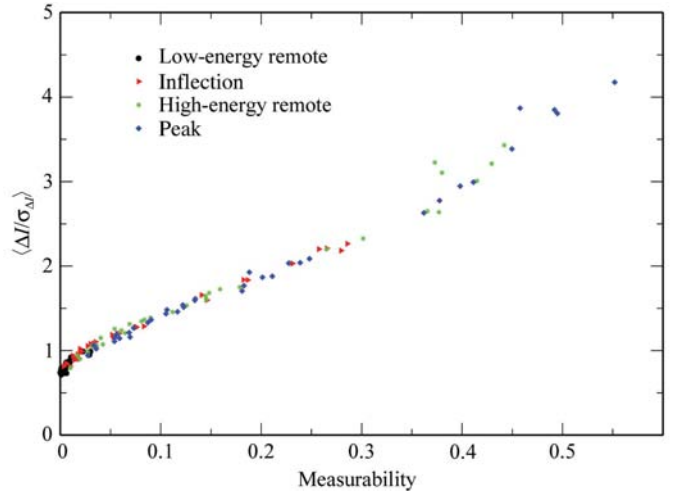

Figure 12
Observed measurabilities *versus* the observed average intensity-based anomalous signal-to-noise ratio for four data sets. At $\langle|\Delta I|/\sigma_{\Delta I}\rangle$ = 1.2 the measurability is approximately 0.06. See text for details.

To compare the simulated $\langle \mathrm{FOM} \rangle$ values and the correlation coefficient between the normalized $|\Delta F|$ and F_{A} values, the substructure was solved with *SHELXD* (Schneider & Sheldrick, 2002) and refined with *SHARP* (de La Fortelle & Bricogne, 1997). Substructure solution was straightforward using a resolution truncation at 3.0 Å and resulted in a correlation coefficient of 0.53. The estimated correlation coefficient *via* the described simulation was equal to 0.50. Phasing with *SHARP* resulted in an average figure of merit equal to 0.48. The occupancies refined to around 0.8 and B values around 28 Å^2, slightly larger than the estimated Wilson B value.

A simulation run with parameters specified for the determination of the Bijvoet ratios and measurability estimates and $q = 0$ resulted in an overestimation of the figure of merit (Fig. 9). Increasing the B value and decreasing the occupancy to 0.8 as well as increasing q to 1.0 gave a better correspondence to the *SHARP* results (Fig. 9).

5.4. Elastase + 2 Au

An atomic resolution X-ray data set of an elastase crystal soaked in a solution containing $KAu(CN)_2$ has an $\langle I/\sigma_I \rangle$ equal to 32 at 10 Å. The data set was truncated at 2.0 Å, where $\langle I/\sigma_I \rangle$ was equal to 23. Elastase contains 245 residues and this particular derivative contains two Au atoms with an average occupancy equal to 0.8. The B value for protein was equal to 10 Å^2, whereas the heavy-atom model refined to 8.0 Å^2 on average. A comparison of the estimated and observed measurability is shown in Fig. 10. The predictions of the Bijvoet ratios are better than for the Jia-peak test case (results not shown). As in the Jia-peak case, at the resolution where the Wilson plot has a local maximum the measurability is decreased significantly, indicating the importance of the γ_{light} term in the analyses.

5.5. The relation between the measurability and the average anomalous signal-to-noise ratio

As can be expected from (29), (30) and (14), the measurability and the average anomalous signal-to-noise ratio are closely related. Although the exact relation between these two quantities can probably be determined analytically given a particular (approximate) error model on the intensities, it is more straightforward to visualize the dependence by plotting observed measurabilities estimated in resolution bins against the intensity-based average anomalous signal-to-noise ratio. This has been performed for four data sets of an Se-MAD experiment (Li *et al.*, 2000) (Fig. 12). It is clear that a resolution-cutoff criterion based on the average anomalous signal-to-noise ratio is equivalent to that of the measurability. For instance, cutting data at the resolution where $\langle |\Delta I|/\sigma_{\Delta I} \rangle$ is equal to 1.2 would be equivalent to cutting the data at the resolution limit where the measurability is 6%.

5.6. What is the limit?

It has been stated that the limit of SAD phasing is 0.6% or lower for $\langle |\Delta F| \rangle / \langle F \rangle$. This limit, known as Wang's limit, originates from an example from error-free data (Wang, 1985). The key parameter to a successful SAD phasing lies in the accuracy of the ΔI or ΔF values rather than on their expected absolute values. In the unrealistic limit of perfect data, a large protein and very small substructure, the average cosine of the phase difference only depends on the phase difference between the heavy-atom structure-factor component and the total structure factor. Assuming that both these phases are distributed uniformly on $(0, 2\pi)$ and that they are independent, the expected cosine of the absolute phase difference (ignoring the Sim contribution) is approximately 0.6. However, the value of f'' and the size of the substructure does determine the expected value of $|\Delta I|$. If any errors are present, the resulting phase distribution will be less informative compared with the error-free case.

Although the presented analyses give a better clue of whether or not a particular SAD phasing scenario is feasible, a number of major factors determining the possible success or failure have not been addressed. Factors such as the number of heavy-atom sites and the solvent content cannot be taken into account owing to the assumed independence of structure factors. Another drawback is the use of the FOM to quantify the success of a heavy-atom refinement. Although the FOM does play an important part in characterizing the quality of the resulting SAD map, it is less informative than the information content of the phase probability distribution. The latter statistic, expressed for instance in terms of the average entropy (Shannon, 1948*a*,*b*) of the phase probability distributions as a function of resolution, in combination with an estimate of the solvent content, might be a better criterion to predict the behavior of a subsequent run of density modification.

6. Conclusions

The classic expressions for the expected values of the Bijvoet difference as well as various Bijvoet ratios agree reasonably well with the results from the simulation in the case of error-free data, although the expressions for $\mathbb{E}(|\Delta F|)/\mathbb{E}(F)$ and $\mathbb{E}(|\Delta I|/I)$ tend to slightly overestimate the results obtained numerically. Out of the three Bijvoet ratios investigated, the expression for $\mathbb{E}(|\Delta I|)/\mathbb{E}(I)$ agrees best with the simulation. However, if errors are present in the data then none of the expressions for the Bijvoet ratio agree with the results of the simulation. As an increase of errors in the amplitudes results in an increase in the Bijvoet ratios, a plot of the Bijvoet ratio as a function of resolution is virtually useless to identify the strength of the anomalous signal unless it is accompanied by an estimate of the Bijvoet ratio in the case of error-free data.

The early work on the measurability by S. Parthasarathy and coworkers, although thorough, lacked the incorporation of measurement errors in the analyses. Inclusion of measurement errors, as carried out in the present work, resulted in a modification of the definition of measurability. Using measurability plots based on (29) and (30) to judge the quality and amount of anomalous signal in a SAD data set is a straightforward alternative to a Bijvoet ratio plot. A plot of

the measurability *versus* the resolution can be directly linked to the number of absolute Bijvoet differences significantly larger than zero and thus gives a further indicator of the quality of the data compared with the anomalous signal-to-noise approach (14).

The simulation method presented here is a useful tool in the investigation of the contribution of individual factors governing the success of a structure solution *via* the SAD technique. Although the method is only able to predict statistics up to the stage of SAD phasing, the resulting correlation coefficient CC_A between normalized $|\Delta F|$ and F_A as well as the expected $\langle FOM \rangle$ are more indicative of the possible success than simplistic Bijvoet ratio estimates.

In principle, the method can be extended to incorporate other phasing methods such as MAD, similar to the work by Phillips & Hodgson (1980). If a suitable model is available that describes intensity changes originating from absorption or radiation damage, their effects on the success of phasing can also be investigated.

APPENDIX *A*

A1. The expectation values of $|\Delta F|$

Under the usual approximations (a small anomalous substructure in a large protein) the Bijvoet amplitude difference is approximately equal to (Kartha & Parthasarathy, 1965; Parthasarathy, 1967)

$$\Delta F \simeq 2\kappa F_A \sin(\theta), \tag{31}$$

where θ is the angle between the heavy-atom model and the total structure factor. If the substructure is small with respect to the total scattering mass, one can assume that θ is uniformly distributed on (0, 2π). However, if the substructure is a significant part of the total scattering mass, the assumption of a uniform distribution for θ is incorrect [see, for instance, Fig. 2 of Dauter *et al.* (2002) and equation (28) and Fig. 2 of Parthasarathy (1965)].

When assuming that θ is uniformly distributed, the expected value of the absolute Bijvoet amplitude difference is equal to

$$\mathbb{E}(|\Delta F|)_{F_A,\theta} \simeq 2\kappa \mathbb{E}(F_A)_{F_A} \mathbb{E}[|\sin(\theta)|]_\theta, \tag{32}$$

with

$$\mathbb{E}[|\sin(\theta)|]_\theta = \frac{1}{\pi}\int_0^\pi \sin(\theta)\, \mathrm{d}\theta \tag{33}$$

$$= \frac{2}{\pi}. \tag{34}$$

Noting that

$$\mathbb{E}(F_A) = (\pi^2/2)\sigma_{\text{heavy}}, \tag{35}$$

one obtains

$$\mathbb{E}(|\Delta F|) \simeq (2/\pi^{1/2}) N_{\text{heavy}}^{1/2} f''_{\text{heavy}}. \tag{36}$$

As the value of the expected amplitude is equal to

$$\mathbb{E}(F) = (\pi^{1/2}/2)(N_{\text{light}} f_{\text{light}}^2 + N_{\text{heavy}} f_{\text{heavy}}^2)^{1/2}, \tag{37}$$

the value for the expected Bijvoet amplitude ratio then results in (5) after taking into account effects arising from geometric regularities in the protein and heavy-atom substructure.

The previous analysis does not take into account the case when more than one chemical heavy-atom species is present. Including the presence of multiple species of anomalous scatterers in the derivation of the expected Bijvoet amplitude ratio has been carried out by Olczak *et al.* (2003).

APPENDIX *B*

B1. The expected correlation between $|\Delta F|$ and F_A values

Assume for simplicity that the F_A values are normalized [$\mathbb{E}(F_A^2) = 1$] and that they are distributed according to a Wilson distribution.

The correlation coefficient between $|\Delta F|$ and F_A is equal to

$$CC_A = \frac{\mathbb{E}(|\Delta F| F_A) - \mathbb{E}(|\Delta F|)\mathbb{E}(F_A)}{\sigma_{|\Delta F|}\sigma_{F_A}}. \tag{38}$$

The individual moments are equal to

$$\mathbb{E}(F_A)_{F_A} = \pi^{1/2}/2, \tag{39}$$

$$\mathbb{E}(|\Delta F|)_{\theta,F_A} = 2\kappa/\pi^{1/2}, \tag{40}$$

$$\mathbb{E}(|\Delta F| F_A)_{\theta,F_A} = 4\kappa/\pi, \tag{41}$$

$$\mathbb{E}(\Delta F^2)_{\theta,F_A} = \kappa^2/2, \tag{42}$$

$$\sigma_{|\Delta F|} = \kappa((2 - 4/\pi)^{1/2}, \tag{43}$$

$$\sigma_{F_A} = (1 - \pi/4)^{1/2} \tag{44}$$

and thus

$$CC_A = \frac{(4/\pi - 1)\kappa}{\kappa(2 - 4/\pi)^{1/2}(1 - \pi/4)^{1/2}} \tag{45}$$

$$\simeq 0.692. \tag{46}$$

Note that CC_A is independent of κ and thus independent of f''. For real data, the value is decreased owing to the presence of experimental errors.

PHZ would like to thank Z. Dauter, S. Banumathi, R. W. Grosse-Kunstleve, J. D. Ferrara and D. Velmurugan for stimulating discussions. Professor John Helliwell and two referees are kindly thanked for their valuable input on this manuscript. This work is dedicated to S. Parthasarathy and the late R. Srinivasan for their important contributions in the development of statistical methods in X-ray crystallography. The program is available upon request from the author. This work was supported in part with Federal funds from the National Cancer Institute, National Institutes of Health under contract No. NO1-CO-12400.

References

Alkire, R. W., Schuessler, R., Rotella, F. J., Gonczy, J. D. & Rosenbaum, G. (2004). *J. Appl. Cryst.* **37**, 836–840.
Banumathi, S., Zwart, P. H., Ramagopal, U. A., Dauter, M. & Dauter, Z. (2004). *Acta Cryst.* D**60**, 1085–1093.
Blessing, R. H. & Smith, G. D. (1999). *J. Appl. Cryst.* **32**, 664–670.
Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.

Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. & Rossmann, M. G. (1974). *J. Mol. Biol.* **82**, 563–585.
Cassetta, A., Deacon, A. M., Ealick, S. E., Helliwell, J. R. & Thompson, A. W. (1999). *J. Synchrotron Rad.* **6**, 822–833.
Crick, F. H. C. & Magdoff, B. S. (1956). *Acta Cryst.* **9**, 901–908.
Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst.* D**58**, 494–506.
Dodson, E. J. (2003). *Acta Cryst.* D**59**, 1958–1965.
Einspahr, H., Suguna, K., Suddath, F. L., Ellis, G., Helliwell, J. R. & Papiz, M. Z. (1985). *Acta Cryst.* B**41**, 336–341.
Evans, P. (2005). Submitted.
Fourme, R., Shepard, W., Schiltz, M., Prangé, T., Ramin, M., Kahn, R., de La Fortelle, E. & Bricogne, G. (1999). *J. Synchrotron Rad.* **6**, 834–844.
French, S. & Wilson, K. S. (1978). *Acta Cryst.* A**34**, 517–525.
Fu, Z.-Q., Rose, J. P. & Wang, B.-C. (2004). *Acta Cryst.* D**60**, 499–506.
Giacovazzo, C. (1998). *Direct Phasing in Crystallography*. Oxford University Press.
Giacovazzo, C. (2001). *International Tables for Crystallography*, Vol. *B*, edited by U. Shmueli, pp. 210–233. Dordrecht: Kluwer Academic Publishers.
Grosse-Kunstleve, R. W. & Adams, P. D. (2003). *Acta Cryst.* D**59**, 1966–1973.
Hädener, A., Matzinger, P. K., Battersby, A. R., McSweeney, S., Thompson, A. W., Hammersley, A. P., Harrop, S. J., Cassetta, A., Deacon, A., Hunter, W. N., Nieh, Y. P., Raftery, J., Hunter, N. & Helliwell, J. R. (1999). *Acta Cryst.* D**55**, 631–643.
Hauptman, H. A. (1982). *Acta Cryst.* A**38**, 632–641.
Helliwell, J. R. (1992) *Macromolecular Crystallography with Synchrotron Radiation.* Cambridge University Press.
Hendrickson, W. A. (1999). *J. Synchrotron Rad.* **6**, 845–851.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
Hendrickson, W. A., Smith, J. L. & Sheriff, S. (1985). *Methods Enzymol.* **115**, 41–55.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Howell, L. & Smith, D. (1992). *J. Appl. Cryst.* **25**, 81–86.
Kartha, G. & Parthasarathy, R. (1965). *Acta Cryst.* **18**, 745–749.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Li, J., Derewenda, U., Dauter, Z., Smith, S. & Derewenda, Z. S. (2000). *Nature Struct. Biol.* **7**, 555–559.
Lunin, V. Yu. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 530–533.
Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.
McCoy, A. J., Storoni, L. C. & Read, R. J. (2004). *Acta Cryst.* D**60**, 1220–1228.
Morris, R. J., Zwart, P. H., Cohen, S., Fernandez, F. J., Kakaris, M., Kirillova, O., Vonrhein, C., Perrakis, A. & Lamzin, V. S. (2004). *J. Synchrotron Rad.* **11**, 56–59.
Mukherjee, A., Helliwell, J. R. & Main, P. (1989). *Acta Cryst.* A**45**, 715–718.
Olczak, A., Cianci, M., Hao, Q., Rizkallah, P. J., Raftery, J. & Helliwell, J. R. (2003). *Acta Cryst.* A**59**, 327–334.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pannu, N. S. & Read, R. J. (2004). *Acta Cryst.* D**60**, 22–27.
Parthasarathy, S. (1965). *Acta Cryst.* **18**, 1028–1035.
Parthasarathy, S. (1967). *Acta Cryst.* **22**, 98–103.
Parthasarathy, S. & Parthasarathi, V. (1973). *Acta Cryst.* A**29**, 428–432.
Parthasarathy, S. & Parthasarathi, V. (1974). *Acta Cryst.* A**30**, 649–654.
Parthasarathy, S. & Parthasarathi, V. (1976). *Acta Cryst.* A**32**, 768–771.
Parthasarathy, S. & Ponnuswamy, M. N. (1976). *Acta Cryst.* A**32**, 302–305.
Parthasarathy, S. & Ponnuswamy, M. N. (1981*a*). *Acta Cryst.* A**37**, 153–162.
Parthasarathy, S. & Ponnuswamy, M. N. (1981*b*). *Acta Cryst.* A**37**, 921.
Parthasarathy, S. & Srinivasan, R. (1964). *Acta Cryst.* **17**, 1400–1407.
Phillips, J. C. & Hodgson, K. O. (1980). *Acta Cryst.* A**36**, 856–864.
Pohl, E., Gonzalez, A., Hermes, C. & Silfhout, R. (2001). *J. Synchrotron Rad.* **8**, 1113–1120.
Popov, A. N. & Bourenkov, G. P. (2003). *Acta Cryst.* D**59**, 1145–1153.
Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 1020–1027.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Rossmann, M. G. (1961). *Acta Cryst.* **14**, 383–388.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sekar, K. & Parthasarathy, S. (1987). *Acta Cryst.* A**43**, 653–655.
Shannon, C. E. (1948*a*). *Bell Syst. Tech. J.* **27**, 379–432.
Shannon, C. E. (1948*b*). *Bell Syst. Tech. J.* **27**, 623–656.
Sheldrick, G. M. (2004). *High-Throughput Phasing with SHELXC/D/E.* http://shelx.uni-ac.gwdg.de/SHELX/.
Shen, Q., Wang, J. & Ealick, S. E. (2003). *Acta Cryst.* A**59**, 371–373.
Sim, G. A. (1964). *Acta Cryst.* **17**, 1072–1073.
Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 1174–1182.
Usón, I., Schmidt, B., von Bülow, R., Grimme, S., von Figura, K., Dauter, M., Rajashankar, K. R., Dauter, Z., & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 57–66.
Velmurugan, D. & Parthasarathy, S. (1984). *Acta Cryst.* A**40**, 548–558.
Velmurugan, D., Parthasarathy, S. & Parthasarathi, V. (1979). *Acta Cryst.* A**35**, 463–467.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Weeks, C. M. & Miller, R. (1999). *Acta Cryst.* D**55**, 492–500.
Weeks, C. M., Potter, S. A., Rappleye, J. & Miller, R. (2005). *SnB v.2.2 User's Guide.* http://www.hwi.buffalo.edu/SnB/SnBhelp/Index.html.
Weiss, M. S., Sicker, T., Djinovic Carugo, K. & Hilgenfeld, R. (2001). *Acta Cryst.* D**57**, 689–695.
Weiss, M. S., Sicker, T. & Hilgenfeld, R. (2001). *Structure*, **9**, 771–777.
Wilson, A. C. (1942). *Nature (London)*, **150**, 152.
Wilson, A. C. (1949). *Acta Cryst.* **2**, 318–321.
Yang, C., Pflugrath, J. W., Courville, D. A., Stence, C. N. & Ferrara, J. D. (2003). *Acta Cryst.* D**59**, 1943–1957.
Zachariasen, W. H. (1965). *Acta Cryst.* **18**, 714–716.
Zwart, P. H. & Lamzin, V. S. (2004). *Acta Cryst.* D**60**, 220–226.