

Estimation of anomalous signal in diffraction data

Zbigniew Dauter

Synchrotron Radiation Research Section, MCL,
National Cancer Institute, Argonne National
Laboratory, Biosciences Division, Building 202,
9700 South Cass Avenue, Argonne, IL 60439,
USA

Correspondence e-mail: dauter@anl.gov

Received 13 March 2006

Accepted 19 June 2006

It is difficult to judge objectively the amount and resolution limit of the anomalous signal present in diffraction data. While several criteria can be used for this purpose, the usefulness of these indicators varies and depends on factors such as the data redundancy, the accuracy of the estimation of intensities and their uncertainties, and the properties of the anomalously scattering atoms in the crystal. Various indicators give an approximate measure of the anomalous signal, but do not provide a reliable guarantee that the crystal structure will be solved. The ultimate criterion of the anomalous signal is a successful structure solution. The current data-acquisition and phasing methods are so advanced that it is often possible to phase reflections and obtain preliminary electron-density maps in a few minutes, while the crystal still resides at the experimental station and further data can be collected if necessary.

1. Introduction

The anomalous diffraction signal has been recognized as useful for the estimation of phases from the very beginning of protein crystallography (Blow, 1958; Rossmann, 1961; Kartha & Parthasarathy, 1965; North, 1965; Matthews, 1966*a,b*; Blundell & Johnson, 1976). Because the anomalous intensity differences are usually much smaller than the isomorphous differences, the anomalous signal was initially treated as an auxiliary source of phasing information accompanying the stronger isomorphous signal. It became clear that the anomalous signal on its own was able to provide sufficient phasing information not only for solving small structures (Ramachandran & Raman, 1956), but also in macromolecular crystallography when Hendrickson & Teeter (1981) solved the structure of crambin solely from the anomalous signal of sulfurs and when Wang (1985) introduced the method of breaking the anomalous phase ambiguity based on solvent flattening.

The anomalous signal became the primary source of phasing in the early 1990s, when Hendrickson and colleagues introduced to practical usage the multi-wavelength anomalous diffraction (MAD) method (Hendrickson, 1991; Smith, 1991) and especially after the technique of genetic incorporation of selenomethionine into proteins was proposed (Hendrickson *et al.*, 1990). The MAD method has matured in the last decade (Hendrickson, 1999) and has become the phasing method of choice. However, the even simpler single-wavelength anomalous diffraction (SAD) method is currently often successfully used (Blow, 2003). The drive to replace multiple isomorphous replacement (MIR) with the MAD and SAD methods is largely fuelled by progress in the technology of X-ray sources and detectors as well as in the methodology of diffraction data-acquisition and phasing algorithms. Using cryocooled

crystals, synchrotron beams and CCD detectors, it is currently possible to accurately estimate and use the anomalous signal equivalent to a fraction of a percent of the total scattered intensity, which was not possible a couple of decades ago.

The amount of anomalous signal can vary considerably depending on the number and type of anomalous scatterers and on the X-ray wavelength. Certain heavy elements display very pronounced white lines or peaks of the f'' value near their X-ray absorption edges. For the M_V edge of uranium, this value may reach 120 electron units (e), for lanthanides up to 30 e, for heavy metals about 15 e and for transition metals of the third period up to 8 e (Hendrickson & Ogata, 1997). On the other hand, the absorption edges of the lighter elements are not easily reachable and atoms such as sulfur, phosphorus, potassium and calcium have f'' values about 1.0 or less electron units in the usable wavelength range. If the number of anomalous scatterers is known, the Bijvoet ratio of the mean anomalous difference to the mean amplitude expected in the data can be estimated (Hendrickson & Teeter, 1981): $\langle |\Delta F^\pm| \rangle / \langle |F| \rangle = (2N_A/N_P)^{1/2} (f''/f_{\text{eff}}^o)$, where N_A and N_P are the number of anomalous and the total number of protein atoms, respectively, and $f_{\text{eff}}^o = 6.7$ is the atomic scattering factor of the 'average' protein atom. If more than one type of anomalous scatterer is present, the equation must be modified (Olczak *et al.*, 2003). According to this formula, the Bijvoet ratio should rise at a higher diffraction angle θ , since f'' does not depend on θ , whereas f^o diminishes at high resolution. The relative difference in B factors between anomalous atoms and the rest of the atoms also influences the Bijvoet ratio. If anomalous scatterers form a group (*e.g.* a disulfide or a metal cluster), this changes the dependence of the anomalous signal on resolution (Banumathi *et al.*, 2003). However, the theoretical predictions may have only a little practical value. Professor Michael Woolfson used to say (when the author was a postdoc in his group):

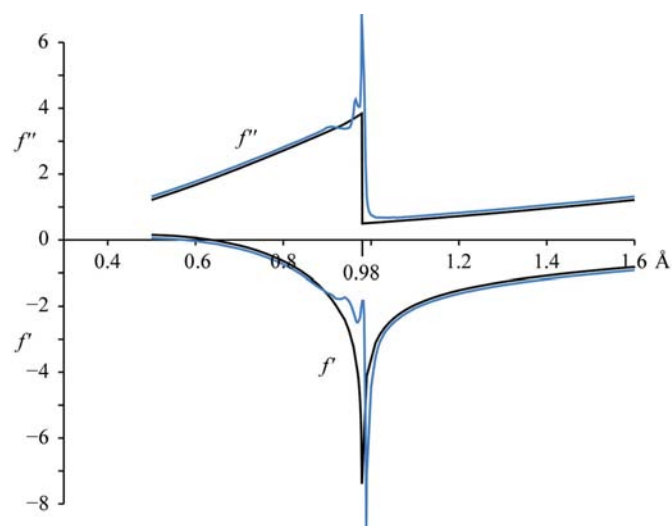


Figure 1
An excitation spectrum recorded on a crystal containing selenomethionine. The black curve shows the behavior of the Se f' and f'' values as calculated from *CROSSEC* and the blue curves correspond to the values estimated from *CHOOCH*.

In testing a new idea, it is legitimate to use ideal calculated data to see whether the method is sound in principle. However, a claim for effectiveness of such tests should only be made after the method has been extensively tested with real observed data. Methods vary very much in their sensitivity to data quality.

Fig. 1 shows the typical dependence of f'' on the X-ray wavelength. The theoretical curve, calculated according to *CROSSEC* (Cromer, 1983), applies to the idealized case of a single atom in a vacuum, but generally gives a quite accurate estimation of f'' (and f') further away from the absorption edge. However, near the edge the f'' value deviates considerably from the ideal and typically displays a narrow white line. It is therefore not possible in practice to estimate from *CROSSEC* the experimental values of f'' and f' which are required for MAD phasing. These values can be estimated approximately (to not more than one decimal digit accuracy) from the fluorescence (or, more properly, excitation) spectrum by the program *CHOOCH* (Evans & Pettifer, 2001).

If the number of anomalous scatterers is known, as is usually the case with selenomethionines, sulfur or some well characterized metalloproteins, the expected Bijvoet ratio can be estimated. However, if the diffraction data are measured from a crystal soaked in a salt solution, the number of bound atoms and their occupancy cannot be known in advance nor can the strength of the anomalous signal be predicted. In practice, the expected amount of anomalous signal may differ considerably from the final amount in the experimentally measured reflection intensities. Usually, experimenters find a lower signal than they expected, particularly if the data quality is not very high. Sometimes surprises happen when the crystal contains some additional unexpected scatterers such as, for example, chloride ions in addition to the ten expected S atoms in the crystal of hen egg-white lysozyme (Dauter *et al.*, 1999).

In practice, the anomalous signal has to be estimated in the measured diffraction data. Several quantitative criteria have been proposed and used for this purpose; their usefulness varies depending on the context. These criteria will be discussed in the following sections.

2. Criteria used for estimation of anomalous signal

2.1. Anomalous signal-to-noise ratio

A very simple and easy-to-obtain criterion is the ratio of the anomalous signal to the noise, calculated on the intensities, $\langle \Delta I^\pm \rangle / \langle \sigma(I) \rangle$, or amplitudes, $\langle \Delta F^\pm \rangle / \langle \sigma(F) \rangle$, evaluated in resolution ranges. It can be used to judge how far in resolution meaningful anomalous differences extend (Schneider & Sheldrick, 2002) and to truncate the data used for the anomalous substructure search at a resolution where $\langle \Delta I^\pm \rangle / \langle \sigma(I) \rangle$ falls below about 1.3. However, it is not trivial to properly estimate the uncertainties $\sigma(I)$ or $\sigma(F)$. They can be reasonably well estimated for highly redundant data, but for data measured from low-symmetry crystals with low redundancy it is difficult to be sure that all uncertainties are statistically valid. The importance of properly estimated uncertainties has

been discussed in several texts on data-collection strategy (Dauter *et al.*, 1999; Borek *et al.*, 2003; Evans, 2006).

2.2. R_{anom}

Another simple measure of the anomalous signal is R_{anom} , the ratio of the mean anomalous intensity difference to the mean reflection intensity, $R_{\text{anom}} = \sum_{hkl} |I^+ - I^-| / \sum_{hkl} (I^+ + I^-) / 2$. This ratio does not take into account the uncertainties of the estimated intensities and should be therefore compared with the level of noise in the data, for example represented by $R_{\text{p.i.m.}} = \sum_{hkl} [1/(N-1)^{1/2}] \sum_i |I_i - \langle I \rangle| / \sum_{hkl} \sum_i I_i$ (Weiss, 2001). The quotient $R_{\text{anom}}/R_{\text{p.i.m.}}$ has been proposed as a practical estimator of the significance of the anomalous signal (Weiss *et al.*, 2001; Mueller-Dieckmann *et al.*, 2005).

2.3. Bijvoet amplitude ratio

A very similar criterion is the Bijvoet ratio $\langle |\Delta F^\pm| \rangle / \langle |F| \rangle$, which can be estimated theoretically if the number and f'' value of the anomalous scatterers are known (Hendrickson & Teeter, 1981; Olczak *et al.*, 2003; Shen *et al.*, 2003; see above). The Bijvoet ratio (similarly to R_{anom}) increases towards unity if the amplitudes are aberrated with errors and its absolute value is not a good indicator of the anomalous signal (Zwart, 2005). However, a comparison of the observed and theoretically calculated Bijvoet ratio is quite informative (Dauter *et al.*, 2002). Fig. 2 illustrates the case of several glucose isomerase data sets collected at different wavelengths, where at low resolution the observed Bijvoet ratio is close to the expected value, but at high resolution some data sets deviate considerably from the theoretically predicted values owing to various levels of data inaccuracy (Ramagopal *et al.*, 2003a).

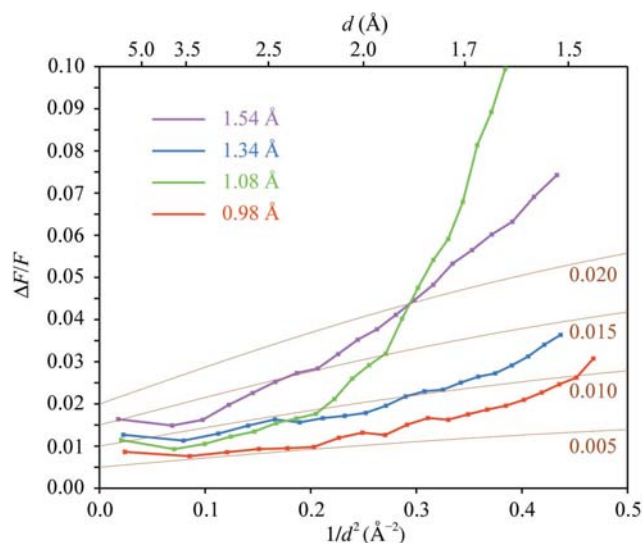


Figure 2

The $\langle |\Delta F^\pm| \rangle / \langle |F| \rangle$ Bijvoet ratios in four data sets collected with different wavelengths from crystals of glucose isomerase (Ramagopal *et al.*, 2003a). The brown lines correspond to the theoretically calculated values. At low resolution, where the data accuracy is high, the observed Bijvoet ratios are close to the expected values, but at higher resolution, where reflection intensities are weaker and data less accurate, various data display significant deviations of the Bijvoet ratio from the ideal.

2.4. χ^2 difference, R_{merge} and list of outliers

Some data-reduction programs (*e.g.* *HKL-2000*; Otwinowski & Minor, 1997) provide a statistic of the agreement between the intensities of all symmetry-equivalent reflections in the form of χ^2 . In general, the χ^2 statistic compares the set of measured quantities with their expected values. In this context, if the spread of multiple intensity measurements of a reflection follows the distribution expected from the level of their estimated errors (*e.g.* from the Poisson statistics of multiple measurements), the χ^2 values should be close to unity. If the intensities of Friedel-related reflections differ significantly but in the merging procedure they are treated as equivalent, it will lead to χ^2 values that are higher than unity, since the expectation of their equivalence will not be true. This criterion may be used to correct the estimated standard uncertainties to the statistically valid level, since the χ^2 values resulting from merging only Friedel-independent reflections should be close to unity.

The level of the χ^2 values obviously depends on the proper estimation of errors. However, even if the errors are not estimated properly, the χ^2 values will differ from unity, but their difference between cases when Friedels are treated as equivalent or not will still indicate the presence of anomalous signal.

The χ^2 indicator is quite sensitive and can be used in the early stages of data collection with partially complete data. If such an early test is negative, the crystal can be discarded and the beam time used for another crystal or project.

Following this tendency, the R_{merge} values will be significantly higher if data are merged as native than when Friedels are separated. Of course, a small difference in R_{merge} should be expected from the twofold increase of redundancy when both Friedels are treated as equivalent.

A careful inspection of the list of measurements marked by the merging program as outliers can reveal the presence of an anomalous signal, particularly if the data redundancy is high. In practice, it may be enough to compare the number of outliers when Friedels are merged as equivalent or independent.

2.5. Correlation between data sets

A sensitive and informative test for the presence of anomalous signal is the correlation between the signed anomalous differences ($F^+ - F^-$) within two related data sets. These two sets can be collected with different wavelengths, as in the case of MAD, or can be two partial subsets of the same set. This criterion was used by the authors of the programs *SOLVE* (Terwilliger & Berendzen, 1999; *SOLVE* v.1.16 and later) and *SHELXD* (Schneider & Sheldrick, 2002) and is implemented in *SCALA* (Evans, 2006). This test is very practical for deciding that the data used for the substructure search should be truncated at the resolution where the correlation coefficient diminishes to 25–30%. This test does not depend on the difficult and not always reliable estimation of errors in anomalous differences.

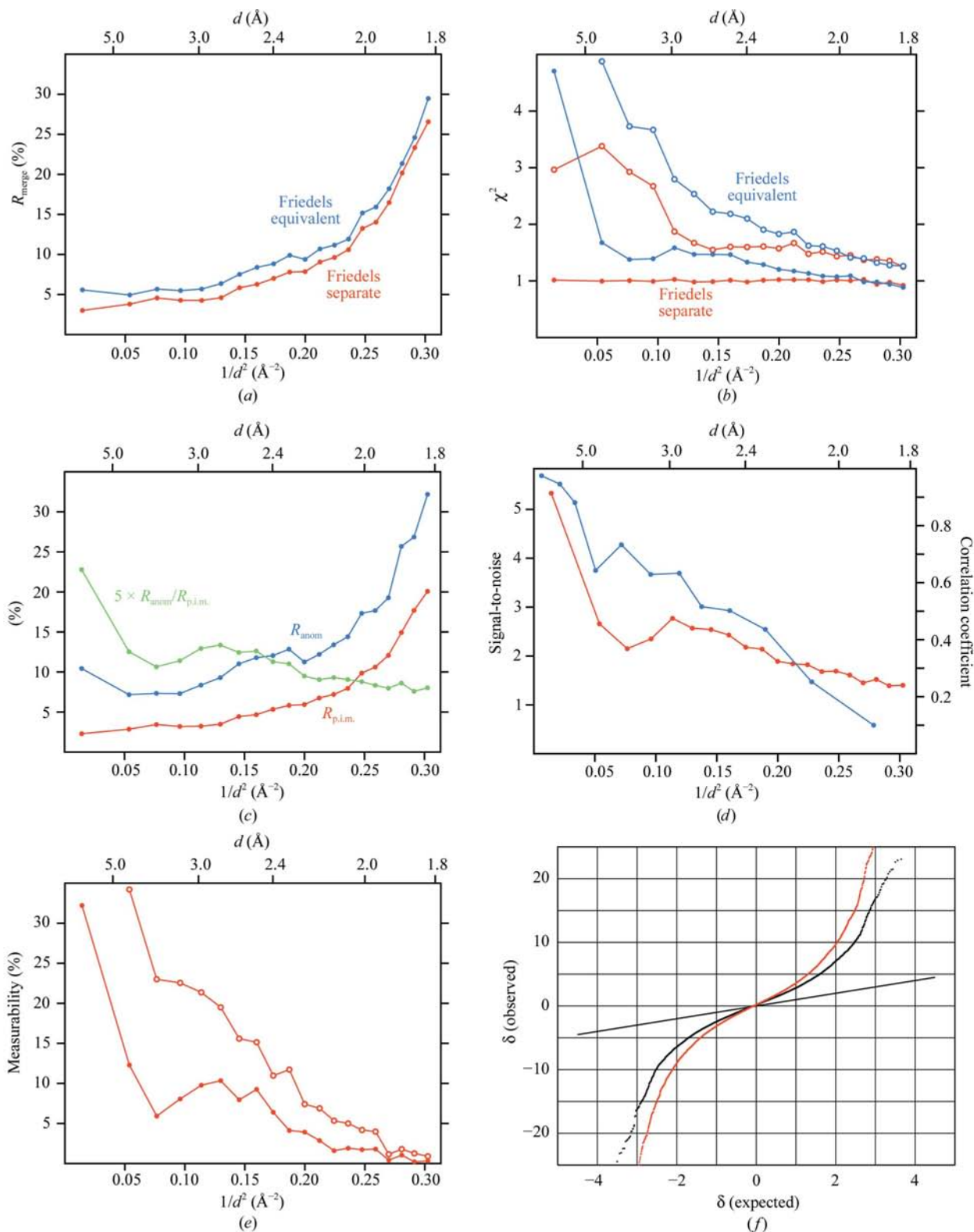


Figure 3

Anomalous signal indicators in resolution ranges for NaBr long-soak invasion data, DradBr3. The anomalous signal originates from the incorporated Br^- ions. (a) R_{merge} for Friedel mates treated as independent (red) or equivalent (blue) reflections. (b) χ^2 values from merging procedure in *SCALEPACK* for Friedels independent (red) and equivalent (blue). When uncertainties are properly adjusted (solid dots), the χ^2 values are close to unity. When they are not corrected (empty circles) all χ^2 values are higher, but there is still a clear difference between the red and blue curves indicating the presence of significant anomalous signal to about 2.0 Å resolution. (c) R_{anom} (blue) and $R_{\text{p.i.m.}}$ (red) values and their quotient (green). Note that the $R_{\text{anom}}/R_{\text{p.i.m.}}$ ratio is scaled up five times. (d) $(\Delta F^{\pm})/(\sigma(F))$ signal-to-noise (red, left scale) and correlation coefficient between two separately merged parts of this data set (blue, right scale). (e) Measurability as a function of resolution with properly estimated uncertainties (red dots) and with uncertainties significantly underestimated (red circles). (f) Normal probability plot of the $\Delta F^{\pm}/\sigma(F)$ values for data with properly (black) and wrongly (red) adjusted uncertainties.

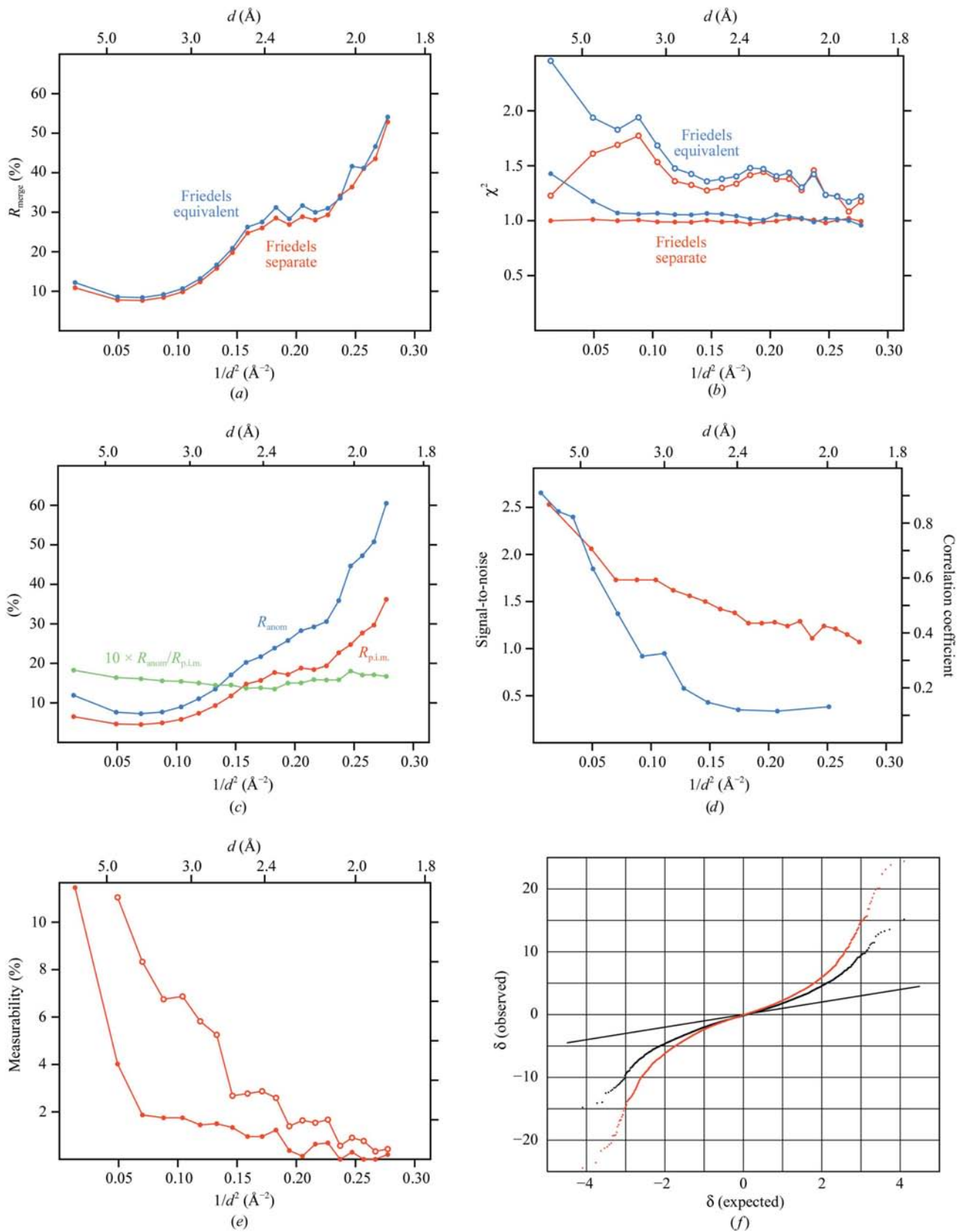


Figure 4

Anomalous signal indicators in resolution ranges for NaBr short-soak invasin data, DradBr2. Color codes and symbols are as in Fig. 3. (a) R_{merge} for Friedel mates. (b) χ^2 values for raw (circles) and adjusted (dots) uncertainties. (c) R_{anom} , $R_{\text{p.i.m.}}$ and their ratio. The $R_{\text{anom}}/R_{\text{p.i.m.}}$ ratio is scaled up ten times. (d) $\langle \Delta F^{\pm} \rangle / \langle \sigma(F) \rangle$ signal-to-noise and correlation coefficient. (e) Measurability as a function of resolution with properly estimated uncertainties (red dots) and with uncertainties significantly underestimated (red circles). (f) Normal probability plot of the $\Delta F^{\pm} / \sigma(F)$ values.

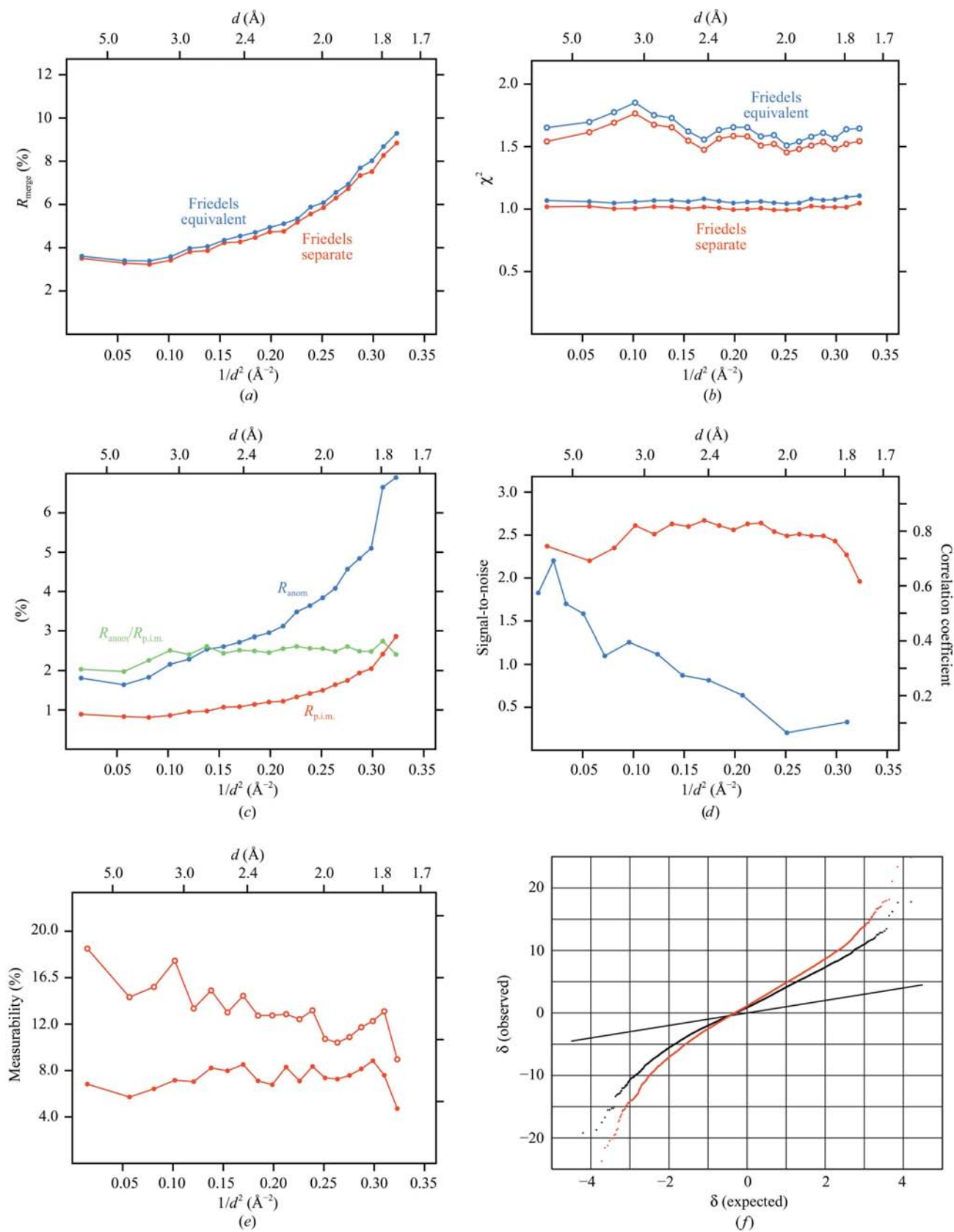


Figure 5

Anomalous signal indicators in resolution ranges for xylanase data. The anomalous signal originates from five S atoms in the xylanase molecule. The color codes and symbols are as in Fig. 3. (a) R_{merge} for Friedel mates. (b) χ^2 values for raw (circles) and adjusted (dots) uncertainties. (c) R_{anom} , $R_{\text{p.i.m.}}$ and their ratio. (d) $\langle \Delta F^2 \rangle / \langle \sigma(F) \rangle$ signal-to-noise and correlation coefficient. (e) Measurability as a function of resolution with properly estimated uncertainties (red dots) and with uncertainties significantly underestimated (red circles). (f) Normal probability plot of the $\Delta F^2/\sigma(F)$ values.

2.6. Difference in acentric and centric reflections R_{as}

This recently proposed (Fu *et al.*, 2004) indicator gives a proportion of the observed relative Bijvoet differences of acentric to centric reflections, $R_{as} = [|\Delta I^{\pm a}|/\sigma(I)]/[|\Delta I^{\pm c}|/\sigma(I)]$. The differences observed for centric reflections $\Delta I^{\pm c}$ represent pure noise, since these reflections cannot have any anomalous signal, whereas $\Delta I^{\pm a}$ contains contributions from the real anomalous signal and noise; therefore, R_{as} can be treated as a 'normalized' representation of the average Bijvoet difference of the acentric reflections.

The number of centric reflections is usually small in low-symmetry crystal classes. In classes 1 and 3 there are no centric reflections at all. To deal with this problem, the R_{as} criterion can be modified to accept in the denominator not only contributors from centric reflections, but observed differences between the intensities of reflections equivalent by pure rotations, which should also always be equivalent.

2.7. Measurability

This criterion gives the fraction of reflections with the relative anomalous intensity difference $|\Delta I^{\pm}|/I$ larger than a certain threshold value and simultaneously with each intensity of the pair I^+ , I^- larger than a specified limit. This indicator was originally proposed by the Madras group (Parthasarathy & Parthasarathi, 1974; Velmurugan & Parthasarathy, 1984). Recently, Zwart (2005) elaborated this concept and introduced in addition the dependence on uncertainties instead of only on the absolute intensity limits and defined the measurability as a fraction of reflections for which the anomalous intensity difference is larger than three times its uncertainty and both intensities of the Friedel pair are larger than three times their uncertainty: $|I^+ - I^-|/[\sigma(I^+)^2 + \sigma(I^-)^2]^{1/2} \geq 3$ and $\min[I^+/\sigma(I^+), I^-/\sigma(I^-)] \geq 3$. The anomalous signal can be accepted as 'measurable' (*i.e.* significant) if the value of measurability is higher than 5%.

2.8. Normal probability plot

The normal probability plot (Abrahams & Keve, 1971; Howell & Smith, 1992) compares the ordered set of $\Delta F^{\pm}/\sigma(F)$ values against the expected normal distribution. If the plot follows a straight line, the distribution is normal (Gaussian). However, the anomalous differences are not distributed normally; therefore, if the plot deviates from linearity and if its slope is larger than unity, the anomalous differences are meaningful with respect to the standard uncertainties. The plot does not provide the resolution limit of the meaningful anomalous signal, but its advantage is that it can be used with only a fraction of the full complete data set.

3. Examples

The analysis of three data sets containing different amounts of anomalous signal is illustrated in Figs. 3, 4 and 5. Two data sets were collected from crystals of *Escherichia coli* invasins soaked in 1 M NaBr, firstly for 12 s (Dradbr2) and secondly for 40 s (Dradbr3). The structure was solved easily by SAD using the

second data set (Jędrzejczak *et al.*, 2006). The crystals of invasins underwent a type of lattice transition during soaking, so that the short-soak data are significantly worse than the data measured after a longer soak. Neither of these data sets is highly redundant; they were collected with the 'high-energy remote' wavelength, 100 eV above the absorption edge of bromine, where it has an f'' value of about 3.7 electron units. The other data were collected with very high redundancy from a native crystal of xylanase (Ramagopal *et al.*, 2003b) at $\lambda = 1.74$ Å, where sulfur has $f'' = 0.70$. There are five sulfurs in the molecule of xylanase (302 amino acids).

3.1. Invasin

Figs. 3 and 4 show graphs of various indicators of the anomalous signal for both long- and short-soak invasins data sets in the whole range of data resolution. All criteria suggest the presence of a significant amount of anomalous signal at low resolution. The difference between R_{anom} and $R_{p.i.m.}$ is large and $\langle \Delta F^{\pm} \rangle / \langle \sigma(F) \rangle$ is higher than 1.3 up to the high-resolution limit. However, the differences between R_{merge} and χ^2 for Friedel mates merged and separate suggest that the anomalous signal is not significant in the highest resolution ranges and in the long-soak Dradbr3 data it extends further than in the short-soak Dradbr2 data, even if the errors are not estimated properly. The difference in R_{merge} at high resolution is as expected from the doubled redundancy resulting from merging Friedel mates. The correlation coefficient calculated between the $\Delta F^{\pm}/\sigma(F)$ values of the same reflections in the two partial sets obtained from merging data from two halves of all diffraction images confirms that the meaningful signal extends to about 2.3 Å in the Dradbr3 data and only to about 3.0 Å in the Dradbr2 data. Inspection of the normal probability plot also reveals the presence of a significant signal, since the slope of the central part of the Dradbr3 plot is about 3 and that in the Dradbr2 plot is about 2 and both plots are significantly curved.

The final result of the X-ray diffraction experiment is electron density; therefore, to determine how far in resolution the meaningful anomalous signal extends, the anomalous difference maps were calculated with the refined phases and ΔF^{\pm} values for the whole resolution range and for the highest resolution bin (Fig. 6). A comparison of Figs. 6(a) and 6(b) and of Figs. 6(c) and 6(d) shows that whereas in Dradbr3 data the anomalous signal is weak but still significant at highest resolution, in the Dradbr2 data it is not higher than the noise level.

3.2. Xylanase

The graphs for xylanase are illustrated in Fig. 5. The differences of R_{merge} and χ^2 are visible, but they are very small: the scale on the vertical axis differs greatly from that in the previous figures for invasins. All criteria, even $\langle \Delta F^{\pm} \rangle / \langle \sigma(F) \rangle$, show that a small but significant signal is uniformly present in all resolution bins. This is corroborated by inspection of the anomalous difference maps (Figs. 6e and 6f). The map calculated with only the highest resolution data is noisier than in

the case of the full data resolution, but the sulfur sites are reproduced quite convincingly above the noise.

4. Discussion

Comparison of the three presented examples allows some conclusions to be drawn about the usefulness of various quality criteria. The R_{merge} values for the two invasin data sets suggests that whereas the general quality (not taking into account the anomalous signal) of the DadBr3 data is good, with R_{merge} at low resolution of below 5%, the DradBr2 data are mediocre, with R_{merge} at low resolution of about 10% and reaching high values beyond 2.5 Å resolution (Figs. 3*a* and 4*a*).

The xylanase data are highly redundant (12-fold), but R_{merge} at low resolution is below 4% and even at 1.75 Å it does not exceed 10% (Fig. 5*a*). Both the DradBr3 and DradBr2 data sets have comparable redundancy (about fourfold), but the R_{merge} values for cases when Friedels are treated as equivalent or non-equivalent are very similar for DradBr2 data, whereas they differ for DradBr3 data, which suggests that the anomalous signal is more significant in the latter data set. The analogous difference in R_{merge} for xylanase is very small and extends throughout all resolution ranges, but taking into account the higher data redundancy it is meaningful.

The graph of χ^2 values in resolution bins reveals that anomalous signal is present in all three data sets, but in

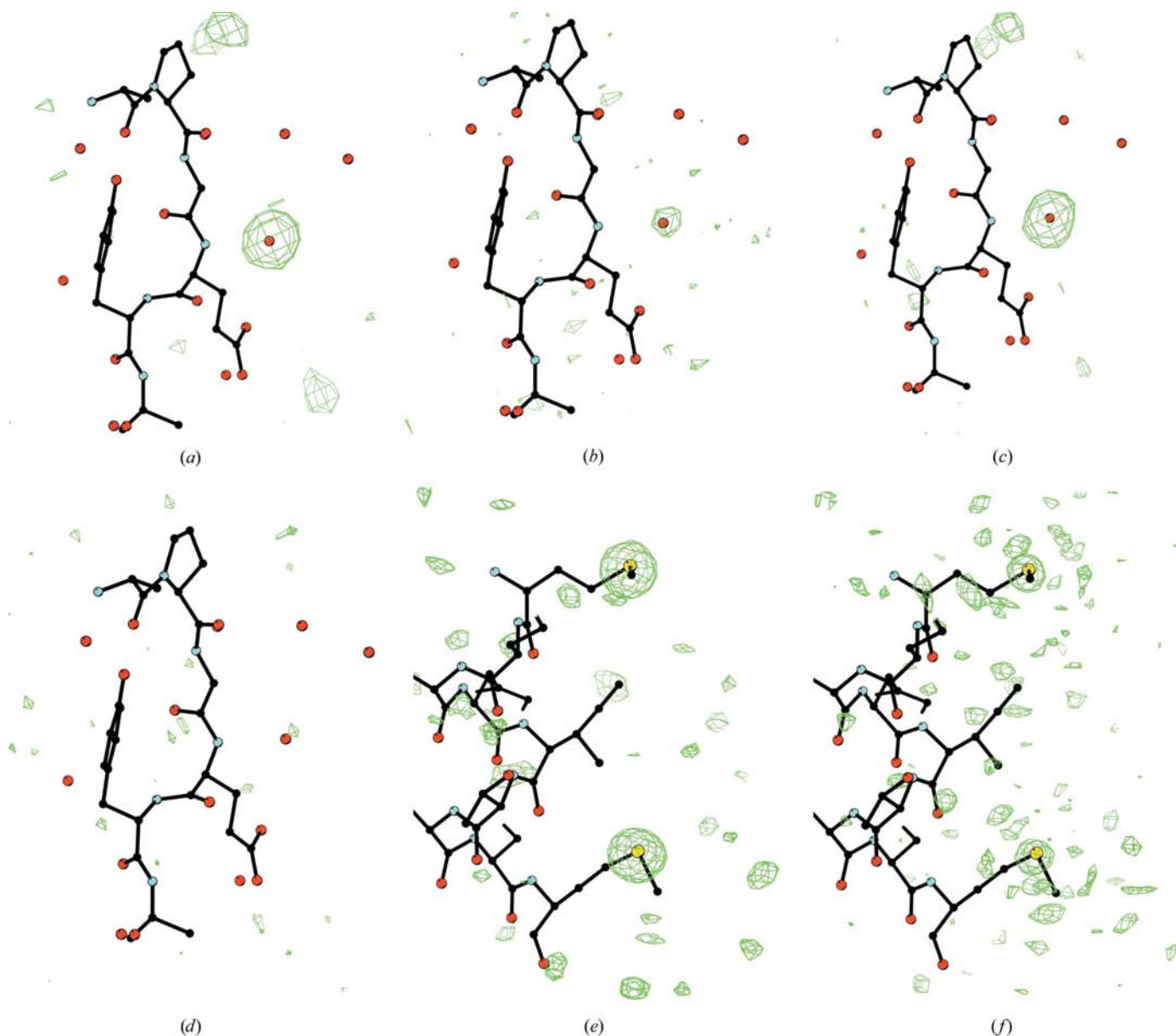


Figure 6 Anomalous difference Fourier maps calculated with model phases and with full or only high-resolution reflections. The vicinity of the most prominent bromide site is shown with the long-soak invasin data in (a) the 30–1.76 Å range and (b) the 2.0–1.76 Å range. The same site is shown with the short-soak invasin data in (c) the 30–1.88 Å range and (d) the 2.15–1.88 Å range. The vicinity of two methionines in xylanase is shown with data in (e) the 30–1.7 Å range and (f) the 2.0–1.7 Å range. These figures were obtained using *BOBSCRIPT* (Esnouf, 1999).

different amounts and extending to a different resolution limit. The DradBr3 data clearly contain a higher amount of anomalous signal, extending to about 2.0 Å, while in the DradBr data it reaches about 2.5 Å (Figs. 3*b* and 4*b*). These graphs show a typical tendency where the largest difference between the red and blue curves is at lowest resolution and diminishes towards high resolution. This effect is expected since at low resolution the accuracy of measured intensities and anomalous differences is high, whereas at high resolution all intensities are much weaker and the accuracy of measured intensities and anomalous differences is low. The graph for xylanase data (Fig. 5*b*) is somewhat atypical, since it shows a uniform difference in χ^2 values in all resolution ranges. The crystals of xylanase have the potential to diffract to atomic resolution, so that the 1.75 Å data practically constitute a 'low-resolution' data set that is highly redundant and exceptionally accurate. The graph suggests that the anomalous signal of five S atoms in a 30 kDa protein is weak but significant in all resolution ranges of these data.

The same figures also show the analogous χ^2 values that result from improper merging when no correction to the uncertainties was applied and therefore the Friedel-independent χ^2 values differ from unity. However, the differences between the red and blue curves represented by empty circles in Figs. 3*b*), 4*b*) and 5*b*) are equally informative about the content of the anomalous signal as in the case of a more proper estimation of errors.

The R_{anom} values, similarly to R_{merge} and $R_{\text{p.i.m.}}$, increase with resolution for all three data sets (Figs. 3*c*, 4*c* and 5*c*), which reflects the gradual worsening of the accuracy of intensity measurements, as all intensities decrease towards high resolution. The $R_{\text{anom}}/R_{\text{p.i.m.}}$ ratio for DradBr3 data is higher than 2.0 in the low-resolution ranges and drops below this value beyond ~ 2.3 Å, whereas for DradBr2 it is lower than 2.0 everywhere, confirming the inferiority of the DradBr2 data. For xylanase this ratio is higher than 2.0 in all resolution ranges. The overall values are 2.33 for DradBr3, 1.80 for DradBr2 and 2.44 for xylanase. However, this criterion does not discriminate between the anomalous data quality as clearly as other indicators.

The signal-to-noise ratio $\langle \Delta F^{\pm} \rangle / \langle \sigma(F) \rangle$ is presented in Figs. 3*d*), 4*d*) and 5*d*). As expected, all curves show that the anomalous signal diminishes at high resolution, but it is clear that the signal in the DradBr3 data is much stronger and extends further than that in the DradBr2 data. According to the criterion $\langle \Delta F^{\pm} \rangle / \langle \sigma(F) \rangle \geq 1.3$, the meaningful anomalous signal extends to about 2.0 Å for DradBr3, to 2.5 Å for DradBr2 and throughout the entire resolution range for xylanase. The $\langle \Delta F^{\pm} \rangle / \langle \sigma(F) \rangle$ ratio obviously strongly depends on the correctly estimated uncertainties.

The same figures also show the correlation coefficients between the anomalous differences in the two half-redundancy partial sets of data. It drops below the significance level of 0.3 at 2.3 Å for the DradBr3 data and at 3.0 Å for the DradBr2 data. For the xylanase data the correlation is above 1.3 only up to about 4.0 Å, which is in strong disagreement with the majority of other quality indicators, suggesting that in

these data the meaningful anomalous signal extends much further in resolution.

The values of measurability calculated as a function of resolution are presented in Figs. 3*e*), 4*e*) and 5*e*). According to these plots, measurability higher than 5% extends to about 2.3 Å in the DradBr3 data, but in the DradBr2 data only the lowest resolution reflections contain 'measurable' anomalous signal. In contrast, a measurable anomalous signal extends up to the highest resolution limit in the xylanase data. The overall global measurability values confirm this: 6.20% for the DradBr3 data, 1.58% for DradBr2 and 8.50% for xylanase. This very clearly shows the difference in the quality of the two DradBr data sets and indicates that the anomalous signal in xylanase data, albeit very weak, is highly significant and measured accurately. However, the values of measurability by its definition very strongly depend on the estimation of errors, as evidenced by the curves shown in the figures.

The normal probability plots, shown in Figs. 3*f*), 4*f*) and 5*f*), suggest that all three analyzed data sets contain significant anomalous signal. However, these plots do not permit judgement of how far in resolution the signal extends. Moreover, the amount of the signal, as represented by the slope of the linear part of the plot, is strongly dependent on the correct estimation of uncertainties.

The comparison suggests that in judging the amount of the anomalous signal in the data, it is advisable to take into account the whole context of several criteria; for example, the data redundancy or the overall data quality abstracting from the content of the anomalous signal as well as the level of the signal expected from the content of the sample. Various indicators may suggest a different amount or different resolution limit of the meaningfulness of the anomalous signal. It may be therefore advisable not to rely on a single criterion but rather to check several independent indicators.

5. Conclusions

All the criteria discussed above can be used for judging the presence of the anomalous signal in the diffraction data. Knowledge of how far in resolution the significant anomalous signal extends is especially important for selecting the resolution limit of data used for substructure solution, where inclusion of very noisy high-resolution data usually diminishes the chance of success. For the subsequent stages of phasing, model building and refinement all available data should be used.

Some of the criteria of judging the anomalous signal are more practical or more objective than others. However, there is no indicator that would guarantee that the amount of the signal is sufficient for successful phasing of the novel structure from the newly acquired data. In fact, the opposite reasoning is true: using these indicators it is possible and easy to realise that the measured intensities do not contain any meaningful anomalous signal. The only convincing way of finding out if the signal is sufficient is to solve the structure. This sounds like a paradox, but it is today a quite practical approach. The methodology and software used for data acquisition and

structure solution are very advanced and allow the experimenter to (at least) obtain a preliminary structure solution within minutes of the end of the data-collection session, when the crystal is still at the experimental station. In such an approach, one can use the feedback from the structure-resolution attempts and apply it to further modify the data-collection strategy, increasing chances of ultimate success.

This work was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US or any other Government.

References

- Abrahams, S. C. & Keve, E. T. (1971). *Acta Cryst.* **A27**, 157–165.
- Banumathi, S., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* **D59**, 492–498.
- Blow, D. M. (1958). *Proc. R. Soc. London Ser. A*, **247**, 302–336.
- Blow, D. M. (2003). *Methods Enzymol.* **374**, 3–22.
- Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. New York: Academic Press.
- Borek, D., Minor, W. & Otwinowski, Z. (2003). *Acta Cryst.* **D59**, 2031–2038.
- Cromer, D. T. (1983). *J. Appl. Cryst.* **16**, 437–438.
- Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst.* **D58**, 494–506.
- Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
- Esnouf, R. M. (1999). *Acta Cryst.* **D55**, 938–940.
- Evans, G. & Pettifer, R. F. (2001). *J. Appl. Cryst.* **34**, 82–86.
- Evans, P. R. (2006). *Acta Cryst.* **D62**, 72–82.
- Fu, Z.-Q., Rose, J. P. & Wang, B.-C. (2004). *Acta Cryst.* **D60**, 499–506.
- Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
- Hendrickson, W. A. (1999). *J. Synchrotron Rad.* **6**, 845–851.
- Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
- Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
- Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
- Howell, P. L. & Smith, G. D. (1992). *J. Appl. Cryst.* **25**, 81–86.
- Jędrzejczak, R., Dauter, Z., Dauter, M., Piątek, R., Zalewska, B., Mróz, M., Bury, K., Nowicki, B. & Kur, J. (2006). *Acta Cryst.* **D62**, 157–164.
- Kartha, G. & Parthasarathy, R. (1965). *Acta Cryst.* **18**, 745–749.
- Matthews, B. W. (1966a). *Acta Cryst.* **20**, 82–86.
- Matthews, B. W. (1966b). *Acta Cryst.* **20**, 230–239.
- Mueller-Dieckmann, C., Panjikar, S., Tucker, P. A. & Weiss, M. S. (2005). *Acta Cryst.* **D61**, 1263–1272.
- North, A. C. T. (1965). *Acta Cryst.* **18**, 212–216.
- Olczak, A., Cianci, M., Hao, Q., Rizkallah, P. J., Raftery, J. & Helliwell, J. R. (2003). *Acta Cryst.* **A59**, 327–334.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Parthasarathy, S. & Parthasarathi, V. (1974). *Acta Cryst.* **A30**, 649–654.
- Ramachandran, G. N. & Raman, S. (1956). *Curr. Sci.* **25**, 348–351.
- Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003a). *Acta Cryst.* **D59**, 868–875.
- Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003b). *Acta Cryst.* **D59**, 1020–1027.
- Rossmann, M. G. (1961). *Acta Cryst.* **14**, 383–388.
- Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* **D58**, 1772–1779.
- Shen, Q., Wang, J. & Ealick, S. E. (2003). *Acta Cryst.* **A59**, 371–373.
- Smith, J. L. (1991). *Curr. Opin. Struct. Biol.* **1**, 1002–1011.
- Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* **D55**, 849–861.
- Velmurugan, D. & Parthasarathy, S. (1984). *Acta Cryst.* **A40**, 548–558.
- Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
- Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.
- Weiss, M. S., Sicker, T. K. & Hilgenfeld, R. (2001). *Structure*, **9**, 771–777.
- Zwart, P. H. (2005). *Acta Cryst.* **D61**, 1437–1448.