hkl for which observations exist in both data sets across the entire resolution range or within a particular resolution shell. CC_{anom} is a reliable indicator of the strength of the anomalous signal. Values above 0.30 are considered good.

R.m.s. correlation ratio. This is another statistic based on randomly partitioned data sets, which is calculated by the program *SCALA* (Evans, 2006; Collaborative Computational Project, Number 4, 1994). It is an analysis of the scatterplot of $[I_1(hkl) - I_1(\bar{h}\bar{k}\bar{l})]$ *versus* $[I_2(hkl) - I_2(\bar{h}\bar{k}\bar{l})]$, where the subscripts 1 and 2 identify the two half data sets. The analysis assumes that the correlation is ideally 1.0. The r.m.s. correlation ratio is defined as the ratio of the r.m.s. widths of the scatterplot distribution along the diagonal and perpendicular to the diagonal. This statistic seems to be more robust than CC_{anom} to the presence of outliers. It cannot, however, be applied to analysing the correlations between different data sets.

Mean anomalous signal-to-noise ratio, $\langle d''/\sigma(d'')\rangle$. The anomalous signal-to-noise ratio of an individual reflection measurement $d''(hkl)/\sigma[d''(hkl)]$ is defined as the ratio of the observed anomalous intensity difference $d'' = |I(hkl) - I(\bar{h}\bar{k}\bar{l})|$ and the corresponding estimated standard uncertainty $\sigma(d'')$ in the measurement of this anomalous difference. The average of the anomalous signal-to-noise ratios for all reflections within a certain resolution range is used as an indicator of utility for phasing. A value of $(2/\pi)^{1/2} \simeq 0.8$ for mean $d''/\sigma(d'')$ of a resolution shell, for example, is taken to indicate that no anomalous signal is present (G. Sheldrick & G. Bunkoczi, personal communication).

Decay *R* factor, R_d. The decay R factor R_d is defined as a pairwise R factor based on the intensities of symmetry-related reflections occurring on different diffraction images (Diederichs, 2006). An increase in R_d as a function of difference in image-collection times is a good indicator of radiation damage occurring during data collection. In fractional form, this is

$$R_d = 2\sum_{hkl}\sum_{m-n}|I_m(hkl) - I_n(hkl)|/\sum_{hkl}\sum_{i}[I_m(hkl) + I_n(hkl)], \quad (2.2.2.14)$$

where $I_m(hkl)$ and $I_n(hkl)$ are the intensities of the reflection *hkl* occurring on images *m* and *n*. The only program in which this is currently implemented is *XDSSTAT*.

Wilson-plot *B* factor, B_{Wilson}. A Wilson plot (Wilson, 1949) is a plot for a contiguous series of resolution shells of the logarithm of the mean intensity in a given resolution shell divided by the sum of the squared atomic form factors for all atoms in the unit cell evaluated at the mean of the resolution limits of the shell. From a least-squares fit of a straight line to the linear part of the Wilson plot, the B factor B_{Wilson} can be derived. Typically, data of lower than 4.5 Å resolution are excluded from the fit. The more meaningful determinations of B_{Wilson} come from Wilson plots that are linear all the way to the nominal resolution d_{min} and minimize the occurrence of spikes due to ice rings.

$$\ln(\langle I_{obs}\rangle/\sum_i f_i^2) = -\ln K_{Wilson} - 2B_{Wilson}/d^2, \quad (2.2.2.15)$$

where $\langle I_{obs}(hkl)\rangle$ is the mean over the intensities of all observed reflections *hkl* in a given resolution shell. The sum $\sum_i$ runs over all atoms in the structure. The parameter d is the midpoint of the resolution shell over which I_{obs} has been averaged. K_{Wilson} is an absolute scale factor.

2.2.3. Comparing different diffraction data sets

In MX, there are many instances when two or more diffraction data sets have to be compared to each other. An important decision which has to be made is, for instance, whether data sets collected from different crystals are isomorphous enough so that they can be merged together. Another task is the comparison of a native data set and a heavy-atom derivative. The question here is how much of the observed difference is due to non-isomorphism and how much is due to isomorphous replacement.

Scaling *R* factor, R_{scale}. The scaling R factor R_{scale} between two data sets is defined as the difference in the structure-factor amplitudes of one data set relative to the structure-factor amplitudes of the other data set after the two data sets have been scaled to each other.

$$R_{scale} = \sum_{hkl}|F_1 - F_2|/\sum_{hkl}|F_1|. \quad (2.2.3.1)$$

However, the above formula is asymmetric with respect to data sets 1 and 2. An alternative symmetric formula is

$$R_{scale} = 2\sum_{hkl}|F_1 - F_2|/\sum_{hkl}(F_1 + F_2). \quad (2.2.3.2)$$

R_{scale} is also often given based on intensities rather than structure-factor amplitudes.

Gradient from a normal probability analysis. For each reflection *hkl* in two data sets that have been scaled to each other, the quantity Δ(real) is computed and compared with its expectation value Δ(expected) assuming a normal distribution of errors (Howell & Smith, 1992). If Δ(real) is plotted against Δ(expected), random differences between the two data sets give rise to a slope of 1.0. Significant differences result in slopes significantly larger than 1.0. Such an analysis is implemented in the program *SCALEIT* (Collaborative Computational Project, Number 4, 1994).

$$\Delta(\text{real}) = (F_1 - F_2)/[\sigma(F_1)^2 + \sigma(F_2)^2]^{1/2}. \quad (2.2.3.3)$$

2.2.4. Quality indicators for substructure determination

The first step of a crystal structure determination after processing and analysing the diffraction data is to use the observed isomorphous and/or anomalous differences in order to determine the isomorphous or anomalous substructure. Since such substructures typically consist of rather few atoms, which are relatively far apart, direct-methods approaches have been adopted from small-molecule crystallography and have largely superseded pure Patterson-based methods. The most widely used computer programs for substructure determination are *SHELXD* (Schneider & Sheldrick, 2002; Sheldrick, 2008), *SnB* (Weeks *et al.*, 1993, 1994) and *HySS* (Grosse-Kunstleve & Adams, 2003).

Correlation coefficient CC(all). CC(all) measures the correlation of the E values (analogous to normalized structure factors; Hauptman & Karle, 1953) derived from the observed isomorphous or anomalous differences and those calculated from the substructure model. CC(all) is another application of the Pearson linear correlation coefficient [see equation (2.2.2.13)] and is calculated for all reflections. In the equation, x and y are the E values derived from the observed and calculated differences. CC(all) is typically used to assess the successful determination of the isomorphous or anomalous substructure in a macromolecule. A value of CC(all) ≥ 0.3 often indicates that the substructure has been correctly identified (Schneider & Sheldrick, 2002).

Correlation coefficient CC(weak). Like CC(all), CC(weak) is the Pearson linear correlation coefficient [see equation (2.2.2.13)]

between the E values (analogous to normalized structure factors as above) derived from the observed isomorphous or anomalous differences and those calculated from the substructure model. In contrast to CC(all), however, CC(weak) is calculated for the weak reflections only. As above, x and y are the E values derived from the observed and calculated differences. The E-value cutoff for defining a reflection as weak can be chosen by the user, but a typical threshold value is 1.5, although lower values may be required for low-resolution data (Schneider & Sheldrick, 2002). A value of CC(weak) ≥ 0.15 often indicates that the substructure has been correctly identified (Sheldrick, 2010).

The minimal function, $R(\varphi)$. The minimal function $R(\varphi)$ is a measure of the mean-square difference between the values of the triplets calculated using a particular set of phases and the expected values of the same triplets as given by the ratio of modified Bessel functions. The minimal function is expected to have a constrained global minimum when the phases are equal to their correct values for some choice of origin and enantiomorph (the minimal principle).

$$R(\varphi) = \sum_{H,K} A_{HK}\left\{\cos\varphi_{HK} - [I_1(A_{HK})/I_0(A_{HK})]\right\}^2 / \sum_{H,K} A_{HK}, \tag{2.2.4.1}$$

where $A_{HK} = (2/N^{1/2})|E_H E_K E_{H+K}|$ with N being the number of atoms in the corresponding primitive unit cell. The minimal function is the function minimized in the program *Shake&Bake*, abbreviated as *SnB* (Weeks *et al.*, 1993, 1994).

2.2.5. Quality indicators for phase determination

Once the isomorphous or anomalous substructure has been successfully determined, it can be used as reference point for the calculation of phases. The quality of the resulting phases is dependent on the strength of the isomorphous or anomalous signal and the completeness and correctness of the isomorphous or anomalous substructure.

Cullis R factor, $R_{\rm Cullis}$. The Cullis R factor (Cullis *et al.*, 1961) for phase determination by isomorphous replacement is defined as the ratio between the lack-of-closure error $\varepsilon(\varphi_{\rm P})$ [equation (2.2.5.1*a*) below] and the isomorphous difference $|F_{\rm PH} - F_{\rm P}|$. It is the most useful signal for a usable heavy-atom derivative. Values < 0.6 for centrosymmetric data are excellent, while values < 0.9 are still usable.

$$\varepsilon(\varphi_{\rm P}) = \left|F_{\rm PH} - |F_{\rm P} + F_{\rm H}|\right|, \tag{2.2.5.1a}$$

$$R_{\rm Cullis} = \sum_{hkl}\left|F_{\rm PH} - |F_{\rm P} + F_{\rm H}|\right| / \sum_{hkl} |F_{\rm PH} - F_{\rm P}|. \tag{2.2.5.1b}$$

Anomalous Cullis R factor, $R_{\rm Cullis,ano}$. The Cullis R factor for phase determination by anomalous dispersion is defined as the ratio between the lack-of-closure error and the observed anomalous difference $|F_{\rm PH}(hkl) - F_{\rm PH}(\bar{h}\bar{k}\bar{l})|$. The lack-of-closure error in the anomalous-dispersion case is the difference between the observed anomalous difference and the calculated anomalous difference $2F_{\rm H}\sin\alpha_{\rm P}$, where $\alpha_{\rm P}$ is the protein phase. A value of $R_{\rm Cullis,ano} < 1.0$ suggests that a contribution to the phasing from the anomalous data is likely (*MLPHARE* program documentation; Collaborative Computational Project, No. 4, 1994).

$$R_{\rm Cullis,ano} = \frac{\sum_{hkl}\left||F_{\rm PH}(hkl) - F_{\rm PH}(\bar{h}\bar{k}\bar{l})| - |2F_{\rm H}\sin\alpha_{\rm P}|\right|}{\sum_{hkl}|F_{\rm PH}(hkl) - F_{\rm PH}(\bar{h}\bar{k}\bar{l})|}. \tag{2.2.5.2}$$

Phasing power, $\mathbf{PP_{iso}}$. The isomorphous phasing power $\mathrm{PP_{iso}}$ for phase determination by isomorphous replacement is defined for a particular pair of native and heavy-atom-derivative data sets as the ratio of $|F_{\rm H}|$ and $\varepsilon(\varphi_{\rm P})$, where $|F_{\rm H}|$ is the calculated amplitude of the heavy-atom structure factor and $\varepsilon(\varphi_{\rm P})$ is the lack-of-closure error [equation (2.2.5.1*a*)]. $\mathrm{F_P + F_H}$ is a vector sum of the calculated structure factor $\mathrm{F_H}$ and the structure factor $\mathrm{F_P}$.

$$\mathrm{PP_{iso}} = \sum_{hkl}|F_{\rm H}| / \sum_{hkl}\left|F_{\rm PH} - |F_{\rm P} + F_{\rm H}|\right|. \tag{2.2.5.3}$$

There is another, slightly different, definition of $\mathrm{PP_{iso}}$, which is implemented in the program *SOLVE*. Here, $\mathrm{PP_{iso}}$ is defined as the ratio of the r.m.s. of the $|F_{\rm H}|$ values and the r.m.s. of the lack-of-closure errors $\varepsilon(\varphi_{\rm P})$. For each reflection *hkl*, a weighted average of $\varepsilon(\varphi_{\rm P})$ is calculated by integrating $\varepsilon(\varphi_{\rm P})$ over the whole phase circle.

$$\mathrm{PP_{iso}} = \Big(\sum_{hkl}|F_{\rm H}|^2\Big)^{1/2} / \Big(\sum_{hkl}\langle\left|F_{\rm PH} - |F_{\rm P} + F_{\rm H}|\right|\rangle^2\Big)^{1/2}. \tag{2.2.5.4}$$

Note. Owing to the cancelling out of the factor $N^{1/2}$ in the numerator and denominator, the latter $\mathrm{PP_{iso}}$ formula does not appear as a ratio of r.m.s. values at first sight.

Anomalous phasing power, $\mathbf{PP_{ano}}$. The anomalous phasing power $\mathrm{PP_{ano}}$ for phase determination by anomalous-dispersion methods is defined as the ratio of the sum of calculated anomalous differences $d''_{\rm calc}$ and the sum of estimated standard uncertainties $\sigma(d''_{\rm obs})$ in the measurement of these anomalous differences:

$$\mathrm{PP_{ano}} = \sum_{hkl} d''_{\rm calc} / \sum_{hkl}\sigma(d''_{\rm obs}). \tag{2.2.5.5}$$

As with $\mathrm{PP_{iso}}$ (see above), the program *SOLVE* uses a slightly different definition of $\mathrm{PP_{ano}}$. Here, the anomalous phasing power is defined as the ratio of the r.m.s. of the $d''_{\rm calc}$ values and the r.m.s. of $\sigma(d''_{\rm obs})$. For this, a weighted average of $d''_{\rm calc}$ is computed by integrating over the whole phase circle for each reflection.

$$\mathrm{PP_{ano}} = \Big(\sum_{hkl}\langle d''_{\rm calc}\rangle^2\Big)^{1/2} / \Big[\sum_{hkl}\sigma(d''_{\rm obs})^2\Big]^{1/2}. \tag{2.2.5.6}$$

Note. As above in the $\mathrm{PP_{iso}}$ formula, the factors $N^{1/2}$ in the numerator and denominator cancel out.

Figure of merit (f.o.m.), m. The figure of merit m is a term used in a number of contexts in X-ray crystallography. In its most common use, it is defined as the weight applied to an individual structure-factor amplitude that, in conjunction with its best phase, gives rise, in a Fourier synthesis, to the electron-density map with the minimum level of noise (Blow & Crick, 1959). Typically, m is given as an average value over all reflections in the data set or in a given resolution shell.

$$m = \int P(\alpha)\exp(i\alpha)\,\mathrm{d}\alpha / \int P(\alpha)\,\mathrm{d}\alpha = \langle\cos(\Delta\alpha)\rangle, \tag{2.2.5.7}$$

where $P(\alpha)$ is the probability of the phase α, initial or refined, being the best phase and $\Delta\alpha = \alpha_{\rm best} - \alpha$ is the error in the phase angle at α. The integration is from 0 to 2π and values for m range from 0 to 1.

2.2.6. Quality indicators for density modification and phase improvement

After determination of initial phases, a first electron-density map can be computed. It is expected that this map will contain significant errors and improbable features. Additional information, such as the flatness of the electron density in the solvent region or the similarity of electron-density regions of two or more identical molecules in the asymmetric unit, can be exploited to modify the electron density and hence improve the phases.