

2.2. QUALITY INDICATORS

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).

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

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_{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 $\varepsilon(\varphi_P)$ [equation (2.2.5.1a) below] and the isomorphous difference $|F_{\text{PH}} - F_P|$. 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.

$$\varepsilon(\varphi_P) = |F_{\text{PH}} - |F_P + F_H||, \quad (2.2.5.1a)$$

$$R_{\text{Cullis}} = \sum_{hkl} |F_{\text{PH}} - |F_P + F_H|| / \sum_{hkl} |F_{\text{PH}} - F_P|. \quad (2.2.5.1b)$$

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

$$R_{\text{Cullis,ano}} = \frac{\sum_{hkl} ||F_{\text{PH}}(hkl) - F_{\text{PH}}(\bar{h}\bar{k}\bar{l})| - |2F_H \sin \alpha_P|}{\sum_{hkl} |F_{\text{PH}}(hkl) - F_{\text{PH}}(\bar{h}\bar{k}\bar{l})|}. \quad (2.2.5.2)$$

Phasing power, PP_{iso} . The isomorphous phasing power 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 $|F_H|$ and $\varepsilon(\varphi_P)$, where $|F_H|$ is the calculated amplitude of the heavy-atom structure factor and $\varepsilon(\varphi_P)$ is the lack-of-closure error [equation (2.2.5.1a)]. $F_P + F_H$ is a vector sum of the calculated structure factor F_H and the structure factor F_P .

$$\text{PP}_{\text{iso}} = \sum_{hkl} |F_H| / \sum_{hkl} |F_{\text{PH}} - |F_P + F_H||. \quad (2.2.5.3)$$

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

$$\text{PP}_{\text{iso}} = (\sum_{hkl} |F_H|^2)^{1/2} / (\sum_{hkl} \langle |F_{\text{PH}} - |F_P + F_H|| \rangle^2)^{1/2}. \quad (2.2.5.4)$$

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

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

$$\text{PP}_{\text{ano}} = \sum_{hkl} d''_{\text{calc}} / \sum_{hkl} \sigma(d''_{\text{obs}}). \quad (2.2.5.5)$$

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

$$\text{PP}_{\text{ano}} = (\sum_{hkl} \langle d''_{\text{calc}} \rangle^2)^{1/2} / [\sum_{hkl} \sigma(d''_{\text{obs}})^2]^{1/2}. \quad (2.2.5.6)$$

Note. As above in the 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.

$$m = \int P(\alpha) \exp(i\alpha) d\alpha / \int P(\alpha) d\alpha = \langle \cos(\Delta\alpha) \rangle, \quad (2.2.5.7)$$

where $P(\alpha)$ is the probability of the phase α , initial or refined, being the best phase and $\Delta\alpha = \alpha_{\text{best}} - \alpha$ 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.

2. BASIC CRYSTALLOGRAPHY

Figure of merit (f.o.m.), m . The figure of merit m [see equation (2.2.5.7)] is also used to judge the degree of improvement in the phase values. Again, 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. Typically, m is given as an average value over all reflections in the data set or in a given resolution shell.

Density-modification (DM) free R factor. The DM free R factor is defined in the same way as the refinement free R factor R_{free} [see equation (2.2.8.1) and the following paragraph describing R_{free}]. It describes the disagreement between the observed structure-factor amplitudes of a certain set of reflections and the calculated amplitudes after density modification. It is a weak indicator used in the program *DM* (Cowtan, 1999) and is helpful mainly in identifying the correct enantiomorph.

Density-modification (DM) real-space free residuals. The DM real-space free residuals are two numbers (real-space free residual for the solvent area and real-space free residual for the protein area) which are calculated by omitting two small spheres of protein and solvent from the density-modification process. The real-space free residual for the solvent area indicates how flat the solvent is in a non-flattened region of solvent and the real-space free residual for the protein area indicates how well the electron density in a non-histogram-matched region of protein obeys the expected histogram. The two residuals can also be combined into the combined real-space free residual by weighted summation, where the weight is defined by the solvent content. The DM real-space free residuals have some value in determining when to stop a density-modification calculation, once no more progress is being made, but limited value otherwise.

Contrast, c . The contrast c between the r.m.s. electron density in the solvent region and the r.m.s. electron density in the macromolecular region can serve as an indication of the presence of a clearly-defined solvent boundary in the electron-density map. It is defined as the standard deviation of the local r.m.s. electron-density values over the entire asymmetric unit (Terwilliger & Berendzen, 1999; Sheldrick, 2002; Terwilliger *et al.*, 2009). The electron-density values are first squared (optionally after truncating very high and very low density values) and then smoothed using a moving local sphere typically with a radius of 6 Å. Local r.m.s. values are then calculated. The contrast c is now defined as the standard deviation σ of the local r.m.s. density values multiplied by a factor to normalize for the fraction of solvent sc in the crystal:

$$c = [(1 - sc)/sc]^{1/2} \sigma. \quad (2.2.6.1)$$

Skewness of electron density, S . A high value for the skewness S of the electron density in an electron-density map indicates the presence of local electron-density maxima with high positive density values. The skewness is defined as the third moment of the electron density:

$$S = \langle \rho^3 \rangle / \langle \rho^2 \rangle^{3/2}. \quad (2.2.6.2)$$

In order to compute the mean values of ρ^3 and ρ^2 , all density grid points in the asymmetric unit of the electron-density map are taken into account (Terwilliger *et al.*, 2009).

Overlap of NCS-related density, O_{NCS} . The presence of correlated electron density at noncrystallographic symmetry (NCS)-related regions in a map can be used as an indicator for the quality of the electron-density map (Cowtan & Main, 1998; Vellieux *et al.*, 1995; Terwilliger *et al.*, 2009). The overlap (O_{NCS}) between density values at NCS-related locations is also often

used to evaluate the presence of local symmetry or non-space-group symmetry:

$$O_{\text{NCS}} = \langle \rho_i \rho_j \rangle. \quad (2.2.6.3)$$

ρ_i and ρ_j are the normalized electron-density values in the NCS-related regions in the asymmetric unit. The average is calculated over the whole region where NCS is present. This region may be defined as the region where overlap values are 0.3 or greater, or by a mask. If there are more than two NCS groups, the average is taken over all NCS pairs.

R factor, R_{DENMOD} , and phase correlation, m_{DENMOD} , from statistical density modification. The amplitudes and phases of structure factors calculated using statistical density modification can be compared with the observed amplitudes and experimental phases (Cowtan & Main, 1996; Terwilliger, 2001; Terwilliger *et al.*, 2009). These comparisons yield an R value (R_{DENMOD}) for the amplitudes and a mean cosine of the phase difference (m_{DENMOD}) for the phases.

$$R_{\text{DENMOD}} = \sum_{hkl} \left| |F_{\text{obs}}| - |F_{\text{DENMOD}}| \right| / \sum_{hkl} |F_{\text{obs}}|, \quad (2.2.6.4)$$

$$m_{\text{DENMOD}} = (1/N) \sum_{hkl} \cos(\alpha_{\text{obs}} - \alpha_{\text{DENMOD}}). \quad (2.2.6.5)$$

Correlation coefficient CC of chain trace against native data.

The quality of density modification in *SHELXE* (Sheldrick, 2002, 2010) can be assessed by computing a Pearson linear correlation coefficient [see equation (2.2.2.13)] of the calculated structure-factor amplitudes for a chain trace against the native structure-factor amplitudes. If the poly-Ala trace yields a CC value higher than 0.25 and if the mean chain length of the trace is >10, the solution is almost always correct. This criterion is currently implemented in the program *ARCIMBOLDO* (Rodriguez *et al.*, 2009).

2.2.7. Quality indicators for molecular replacement

In a case where a known structure is assumed to be similar to that of a target molecule (structural similarity is typically inferred by the degree of sequence similarity), the known structure, also termed the search model, can be used to determine the structure of the target molecule. The approach, termed molecular replacement, was first described in 1962 (Rossmann & Blow, 1962). Nowadays, about two thirds of all newly determined structures are determined by molecular replacement (Long *et al.*, 2008).

Rotation function, RF. The rotation function RF is a measure of the overlap or the agreement of the stationary Patterson function $P2$ calculated from the observed data and the rotated Patterson function $P1$ from the search model.

$$\text{RF} = \int_r P2 \underline{R} P1 \, dr. \quad (2.2.7.1)$$

In the equation for RF, \underline{R} is the rotation operator. The integration is performed between a minimum value and a maximum value for the radius r . These values are chosen according to the size of the search model, with the aim of including as many intramolecular Patterson peaks (self vectors) as possible and to exclude as many intermolecular Patterson peaks (cross vectors) as possible. The ratio of the height of a peak in the RF to the background level is used as an indicator of how likely it is that this peak describes the orientation of a molecule in the target structure.

Translation function, TF. There are numerous ways of defining a translation function TF, making it impractical to discuss quality