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 $R_{\rm free}$ [see equation (2.2.8.1) and the following paragraph describing $R_{\rm free}$]. 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:

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

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:

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

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, $O_{\rm NCS}$. 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 ($O_{\rm NCS}$) between density values at NCS-related locations is also often used to evaluate the presence of local symmetry or non-space-group symmetry:

$$O_{\rm NCS} = \langle\rho_i\rho_j\rangle. \qquad (2.2.6.3)$$

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, $R_{\rm DENMOD}$, and phase correlation, $m_{\rm DENMOD}$, 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 ($R_{\rm DENMOD}$) for the amplitudes and a mean cosine of the phase difference ($m_{\rm DENMOD}$) for the phases.

$$R_{\rm DENMOD} = \sum_{hkl} \big||F_{\rm obs}| - |F_{\rm DENMOD}|\big| / \sum_{hkl} |F_{\rm obs}|, \qquad (2.2.6.4)$$

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

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.

$$\mathrm{RF} = \int_r P2\, \underline{R}\, P1 \, \mathrm{d}r. \qquad (2.2.7.1)$$

In the equation for RF, $\underline{\mathrm{R}}$ 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

indicators here. For a thorough treatment of translation-function applications, the reader is referred to Chapter 2.3 of *International Tables for Crystallography* Volume B and Chapter 13.3 of the present volume.

Log-likelihood gain, LLG. In likelihood-based molecular replacement (McCoy *et al.*, 2007), potential molecular-replacement solutions are evaluated using likelihood, which is defined as the probability P that the observed diffraction data would have been measured if the orientation (and, usually, position) of the model were correct (Read, 2001). The score is reported in terms of the log-likelihood gain (LLG), which is defined as the logarithm of the likelihood score for the model $p(F_{\rm obs};\,{\rm model})$ minus the logarithm of the likelihood score for a random-atom Wilson distribution $p_{\rm Wilson}(F_{\rm obs})$. The LLG measures how much better the data can be predicted from the molecular-replacement model than from a collection of random atoms.

$$\mathrm{LLG} = \sum_{hkl} \ln[p(F_{\rm obs};\,{\rm model})] - \sum_{hkl} \ln[p_{\rm Wilson}(F_{\rm obs})]. \qquad (2.2.7.2)$$

LLG-Z score. It is important to note that the LLG depends on the quality of the model and the number of reflections, so the absolute values cannot be compared between different molecular-replacement applications. Instead, the quality of a molecular-replacement solution can be judged by the LLG-Z score, which is defined as the number of standard deviations a score is above the mean score in a particular rotation or translation search.

$$\text{LLG-Z} = \mathrm{LLG} - \langle \mathrm{LLG}\rangle / [(\mathrm{LLG} - \langle \mathrm{LLG}\rangle)^2]^{1/2}. \qquad (2.2.7.3)$$

The translation function Z score (TFZ) for the last component placed in a molecular-replacement search is often a good indicator of the confidence that can be placed in the solution. If TFZ is greater than 8 and there is no translational pseudo-symmetry, the solution is almost always correct.

Detailed descriptions of the background and proper application of molecular-replacement approaches are presented in Chapter 2.3 of *International Tables for Crystallography* Volume B and Chapters 13.2 and 13.3 of the present volume.

2.2.8. Quality indicators for refinement

The last step of a structure determination is the refinement of the model against the observed data. Refinement is in principle a mathematical operation that is applied in order to minimize the discrepancy between the observed structure-factor amplitudes $|F_{\rm obs}|$ and the calculated ones $|F_{\rm calc}|$.

Crystallographic R factor, R. The crystallographic R factor R is defined as the fractional disagreement between the set of observed structure-factor amplitudes and amplitudes calculated from the structural model. Of course, observed and calculated reflection sets need to be on the same scale.

$$R = \sum_{hkl} \big||F_{\rm obs}| - |F_{\rm calc}|\big| / \sum_{hkl} |F_{\rm obs}|. \qquad (2.2.8.1)$$

Free R factor, $R_{\rm free}$. The free R factor $R_{\rm free}$ is defined in the same way as the crystallographic R factor, but it is based on a set of reflections that have been excluded from the refinement (Brünger, 1992). The excluded set of reflections is called the *test set*, while the set of reflections used for refinement is called the *working set*. The test set can be chosen randomly or systematically, either in thin resolution shells or to account for the presence of noncrystallographic symmetry, respectively. In order to minimize the impact on the final model, the test set should be as small as possible. Typically, it contains about 5–10% of the reflections, or at least enough reflections to keep the standard deviation of $R_{\rm free}$ below 1%, but there is no need to use more than 2000 reflections (Kleywegt & Brünger, 1996; Brünger, 1997). The standard deviation of $R_{\rm free}$ has been empirically estimated to be $R_{\rm free}/N^{1/2}$, where N is the number of reflections in the test set (Brünger, 1997). Of course, there may be concerns about the impact of excluding 5–10% of reflections on the final model, but a few final cycles of refinement against the recombined full data set should allay them.

Correlation coefficients CC($F_{\rm obs}$, $F_{\rm calc}$) and CC($I_{\rm obs}$, $I_{\rm calc}$). CC($F_{\rm obs}$, $F_{\rm calc}$) and CC($I_{\rm obs}$, $I_{\rm calc}$) are Pearson linear correlation coefficients [see equation (2.2.2.13)] between observed and model-based calculated structure-factor amplitudes or intensities, respectively, that find use from time to time. One advantage of the use of a correlation coefficient instead of an R factor is that it avoids the problem of scaling the two sets of numbers relative to each other.

2.2.9. Quality indicators for the refined model

In MX, the observable-to-parameter ratio is mostly unfavourable. Therefore, structure refinements are carried out with boundary conditions, constraints and restraints. Constraints reduce the number of parameters which need to be refined, while restraints provide additional information to the refinement procedure that increases the number of observables. A refined model, therefore, has to fulfil not only the criterion that the crystallographic R factor [see equation (2.2.8.1)] is good and that the model fits well to the electron density, but also that it fits well to the restraints used in the refinement procedure.

Real-space residual, RSR. The real-space residual, RSR (Jones *et al.*, 1991), quantifies the discrepancies between the electron-density maps ρ_1, calculated directly from a structural model, and ρ_2, calculated from experimental data. RSR can take the form of a real-space R factor RSRF and of a real-space correlation coefficient RSCC.

$$\mathrm{RSRF} = 2\sum_{xyz} |\rho_1 - \rho_2| / \sum_{xyz} (\rho_1 + \rho_2). \qquad (2.2.9.1)$$

The sum $\sum_{xyz}$ runs over all grid points of the electron-density maps that are close to the model. A big advantage of RSRF is that is can be calculated on a residue-by-residue basis. It therefore gives a local picture of structure quality. It can also be used throughout model building and refinement in order to follow the improvement of the model locally on a per-residue basis.

RSCC is defined as the Pearson linear correlation coefficient [see equation (2.2.2.13)] between ρ_1 and ρ_2. Everything said about RSRF above applies to RSCC as well.

R.m.s. deviation from ideal of geometric parameter x. The root-mean-square deviation of a set of geometric parameters x from their ideal values is defined as

$$\text{r.m.s.d}(x) = \Big\{ \sum_i [x_i(\text{ideal}) - x_i(\text{observed})]^2 / N \Big\}^{1/2}. \qquad (2.2.9.2)$$

The sum runs over all N instances of the geometric parameter occurring in a structure. The geometric parameters x that are typically considered are bond lengths, bond angles, dihedral angles, chiral volumes, planar groups *etc*. The ideal values for proteins are typically taken from the study of Engh & Huber (1991) and for nucleic acids from Parkinson *et al.* (1996).

***Z* score**. A measure of the likelihood that an individual geometric parameter is correct is given by its Z score. The Z score is defined as the distance of an individual data point of a distribution from the mean of the distribution expressed in standard