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 $p(F_{\rm obs}; \text{model})$ 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.

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

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.

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

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.

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

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.

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

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

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

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

Table 2.2.11.1
Definitions of the most commonly used quality indicators

Indicator	Details
Optical resolution, $d_{\rm opt} = [2(\sigma^2_{\rm Patt} + \sigma^2_{\rm sph})]^{1/2}$ (2.2.2.2)	$\sigma_{\rm Patt}$ and $\sigma_{\rm sph}$ are the standard uncertainties of Gaussians fitted to the origin peak of the Patterson function of the diffraction data set and the origin peak of the spherical interference function, respectively
$R_{\rm merge}$ (merging R factor), $R_{\rm merge} = \sum_{hkl}\sum_{i} \lvert I_i(hkl) - \langle I(hkl)\rangle\rvert / \sum_{hkl}\sum_{i} I_i(hkl)$ (2.2.2.4)	$\langle I(hkl)\rangle$ is the mean of the several individual measurements $I_i(hkl)$ of the intensity of reflection hkl
$R_{\rm meas}$ or $R_{\rm r.i.m}$ (redundancy-independent merging R factor), $R_{\rm r.i.m.} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_{i} \lvert I_i(hkl) - \langle I(hkl)\rangle\rvert / \sum_{hkl}\sum_{i} I_i(hkl)$ (2.2.2.5)	$\langle I(hkl)\rangle$ is the mean of the $N(hkl)$ individual measurements $I_i(hkl)$ of the intensity of reflection hkl
$R_{\rm p.i.m.}$ (precision-indicating merging R factor), $R_{\rm p.i.m.} = \sum_{hkl}\{1/[N(hkl)-1]\}^{1/2}\sum_{i} \lvert I_i(hkl) - \langle I(hkl)\rangle\rvert / \sum_{hkl}\sum_{i} I_i(hkl)$ (2.2.2.6)	$\langle I(hkl)\rangle$ is the mean of the $N(hkl)$ individual measurements $I_i(hkl)$ of the intensity of reflection hkl
$R_{\rm anom}$ (anomalous R factor), $R_{\rm anom} = \sum_{hkl} \lvert I(hkl) - I(\bar{h}\bar{k}\bar{l})\rvert / \sum_{hkl}\langle I(hkl)\rangle$ (2.2.2.12)	$\langle I(hkl)\rangle$ is the mean intensity of the Friedel mates of the reflection hkl, or $\frac{1}{2}[I(hkl) + I(\bar{h}\bar{k}\bar{l})]$
$R_{\rm Cullis}$ (Cullis R factor for isomorphous-replacement applications), $R_{\rm Cullis} = \sum_{hkl} \big\lvert F_{\rm PH} - \lvert F_{\rm P} + F_{\rm H}\rvert\big\rvert / \sum_{hkl} \lvert F_{\rm PH} - F_{\rm P}\rvert$ (2.2.5.1*b*)	$F_{\rm P}$, $F_{\rm PH}$ and $F_{\rm H}$ are the structure factors for the protein, the heavy-atom derivative and the heavy atoms alone, respectively
$\rm PP_{iso}$ (phasing power for isomorphous-replacement applications), $\rm PP_{iso} = \sum_{hkl} \lvert F_{\rm H}\rvert / \sum_{hkl} \big\lvert F_{\rm PH} - \lvert F_{\rm P} + F_{\rm H}\rvert\big\rvert$ (2.2.5.3)	$F_{\rm P}$, $F_{\rm PH}$ and $F_{\rm H}$ are the structure factors for the protein, the heavy-atom derivative and the heavy atoms alone, respectively
or, as in the program *SOLVE*, $\rm PP_{iso} = (\sum_{hkl} \lvert F_{\rm H}\rvert^2)^{1/2} / (\sum_{hkl} \langle \big\lvert F_{\rm PH} - \lvert F_{\rm P} + F_{\rm H}\rvert\big\rvert\rangle^2)^{1/2}$ (2.2.5.4)	
R (crystallographic R factor) and $R_{\rm free}$ (free R factor), $R = \sum_{hkl} \big\lvert \lvert F_{\rm obs}\rvert - \lvert F_{\rm calc}\rvert\big\rvert / \sum_{hkl} \lvert F_{\rm obs}\rvert$ (2.2.8.1)	$R_{\rm free}$ is defined as the crystallographic R factor but for a subset of reflections that have been excluded from refinement
RSRF (real-space R factor), $\rm RSRF = 2\sum_{xyz} \lvert \rho_1 - \rho_2\rvert / \sum_{xyz} (\rho_1 + \rho_2)$ (2.2.9.1)	ρ_1 and ρ_2 are the electron-density maps calculated from the structural model and from the experimental data, respectively
R.m.s.d.'s of geometric parameters x, $\text{r.m.s.d}(x) = \{\sum_{i}[x_i(\text{ideal}) - x_i(\text{observed})]^2 / N\}^{1/2}$ (2.2.9.2)	x_i are the individual values, ideal or observed, of the geometric parameter x and the sum is over all N x_i observed. The geometric parameters x may be bond lengths, bond angles, dihedral angles, chiral volumes, deviations from planarity *etc*.
DPI (diffraction-component precision index), $\rm DPI = [N_{atom}/(N_{hkl} - N_{para})]^{1/2} R d_{min} C^{-1/3}$ (2.2.10.1)	$N_{\rm atom}$ is the number of atoms in the structure, N_{hkl} is the number of reflections, $N_{\rm para}$ is the number of refined parameters, R is the crystallographic R factor, $d_{\rm min}$ is the nominal resolution and C is the fractional completeness of the data set

deviations. In the case described here, the mean values of the distribution are the ideal values taken from Engh & Huber (1991) and Parkinson *et al.* (1996).

$$Z(x_i) = [x_i(\text{observed}) - x_i(\text{ideal})]/\sigma(x). \qquad (2.2.9.3)$$

Ideally, the Z score should be 0. A parameter that exhibits a Z score of less than -4 or greater than +4 is highly unlikely and calls for attention.

Root-mean-square Z score, r.m.s.-Z. Although r.m.s.d. values [see equation (2.2.9.2)] are still popular for use in judging the quality of refined macromolecular models, a much more useful statistic is the r.m.s. value of a distribution of Z scores or the r.m.s.-Z score.

$$\text{r.m.s.-Z}(x) = \sum_i [Z(x_i)^2/N]^{1/2}. \qquad (2.2.9.4)$$

The sum runs over all N instances of the geometric parameter x occurring in a structure. A very useful property of Z scores is that the r.m.s. values of Z-score distributions should always be 1. Significant deviations from the ideal value indicate potential problems. R.m.s. Z scores are widely used, for instance, in the program *WHAT_CHECK* (Hooft *et al.*, 1996).

R.m.s.d. (NCS). The root-mean-square deviation from crystallographic symmetry between two molecules related by non-crystallographic symmetry (NCS) can be calculated from a superposition of the two molecules. It is defined as

$$\text{r.m.s.d.(NCS)} = \Big(\sum_i d_i^2/N\Big)^{1/2}. \qquad (2.2.9.5)$$

The sum runs over N equivalent atom pairs with d_i being the distance between the two equivalent atoms after superposition.

2.2.10. Error estimation for the refined model

An important quality indicator for a refined model is the coordinate uncertainty. Short of full-matrix inversion, which is the standard procedure in small-molecule crystallography but which is applicable only in exceptional cases for macromolecules, some methods have been devised for estimating of the overall coordinate uncertainty.

Error estimation according to Luzzati. For most macromolecular structure determinations, atomic standard uncertainties are not available. However, Luzzati (1952) devised a method of estimating the average positional error of a structure. Under the assumption that the atomic positional errors follow a normal distribution, the average error can be estimated by comparing a plot of the crystallographic R factor [see equation (2.2.8.1)] *versus* the reciprocal resolution (or 2 sin θ/λ) with pre-computed theoretical curves for different average errors. A more recent – and probably better – approach is to use the free R factor instead of the crystallographic R factor.

SigmaA- (σ_A)-type error estimation. A slightly better estimate of the average positional error of a structure can be obtained by plotting the natural logarithm of the parameter σ_A *versus* $(\sin\theta/\lambda)^2$ (Read, 1986). The slope of a straight line fitted to the plot provides an estimate of the average positional error of the structure. The parameter σ_A assumes normally distributed positional errors and takes model incompleteness into account as well.

Diffraction-component precision index, DPI. The diffraction-component precision index DPI is an empirical parameter describing the overall coordinate uncertainty of a structure (Cruickshank, 1999*a*,*b*). For an atom with an isotropic displacement parameter of average value (B_{avg}), it is defined as

$$\text{DPI} = [N_{\text{atom}}/(N_{hkl} - N_{\text{para}})]^{1/2} R d_{\min} C^{-1/3}, \qquad (2.2.10.1)$$

where N_{atom} is the number of atoms included in the refinement, N_{hkl} is the number of reflections included in the refinement, N_{para} is the number of refined parameters, R is the crystallographic R factor, $d_{\min}$ is the nominal resolution of the data included in the refinement and C is the data completeness. The free R factor R_{free} is sometimes used instead of the crystallographic R factor R to calculate the DPI. In this case ($N_{hkl} - N_{\text{para}}$) is replaced by N_{free}, which is the number of reflections used for R_{free} calculation.

The Cruickshank formula for DPI has been recast into several other forms, including

$$\text{DPI} = \sigma(x, B_{\text{avg}}) = 0.18(1 + sc)^{1/2} V_M^{-1/2} R_{\text{free}} d_{\min}^{5/2} C^{-5/6} \qquad (2.2.10.2)$$

(Blow, 2002), where sc is the solvent fraction (= $N_{\text{solv}}/N_{\text{atom}}$, where N_{solv} is the number of atoms that are solvent) and V_M is the Matthews parameter (Matthews, 1968). The utility of this latter formula in guiding the design of the data-collection experiment to achieve a specified target coordinate uncertainty has been demonstrated (Fisher *et al.*, 2008).

2.2.11. The most commonly used quality indicators

A summary of the most commonly used quality indicators and their definitions is presented in Table 2.2.11.1 for ready reference.

We gratefully acknowledge the contributions of Kevin Cowtan (York, UK), Kay Diederichs (Konstanz, Germany), Phil Evans (Cambridge, UK), John Helliwell (Manchester, UK), Randy Read (Cambridge, UK), George Sheldrick (Göttingen, Germany) and Tom Terwilliger (Los Alamos, USA).

References

Arnberg, L., Hovmöller, S. & Westman, S. (1979). *On the significance of 'non-significant' reflexions*. *Acta Cryst*. A**35**, 497–499.
Blow, D. M. (2002). *Rearrangement of Cruickshank's formulae for the diffraction-component precision index*. *Acta Cryst.* D**58**, 792–797.
Blow, D. M. & Crick, F. H. C. (1959). *The treatment of errors in the isomorphous replacement method*. *Acta Cryst.* **12**, 794–802.
Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. New York: Academic Press.
Brünger, A. T. (1992). *Free R value: a novel statistical quantity for assessing the accuracy of crystal structures*. *Nature (London)*, **355**, 472–475.
Brünger, A. T. (1997). *Free R value: cross-validation in crystallography*. *Methods Enzymol.* **277**, 366–396.
Collaborative Computational Project, Number 4 (1994). *The CCP4 suite: programs for protein crystallography*. *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (1999). *Error estimation and bias correction in phase-improvement calculations*. *Acta Cryst.* D**55**, 1555–1567.
Cowtan, K. & Main, P. (1998). *Miscellaneous algorithms for density modification*. *Acta Cryst.* D**54**, 487–493.
Cowtan, K. D. & Main, P. (1996). *Phase combination and cross validation in iterated density-modification calculations*. *Acta Cryst.* D**52**, 43–48.
Cruickshank, D. W. J. (1999*a*). *Remarks about protein structure precision*. *Acta Cryst.* D**55**, 583–601.
Cruickshank, D. W. J. (1999*b*). *Remarks about protein structure precision. Erratum*. *Acta Cryst.* D**55**, 1108.
Cullis, A. F., Muirhead, H., Perutz, M. F., Rossmann, M. G. & North, A. C. T. (1961). *The structure of haemoglobin. VIII. A three-dimensional Fourier synthesis at 5.5 Å resolution: determination of the phase angles*. *Proc. R. Soc. London Ser. A*, **265**, 15–38.
Diederichs, K. (2006). *Some aspects of quantitative analysis and correction of radiation damage*. *Acta Cryst.* D**62**, 96–101.
Diederichs, K. (2010). *Quantifying instrument errors in macromolecular X-ray data sets*. *Acta Cryst.* D**66**, 733–740.