

## 2.2. QUALITY INDICATORS

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).

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

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.

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

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

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

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  $2 \sin \theta/\lambda$ ) 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.

**Sigma $A$ - ( $\sigma_A$ )-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  $\sigma_A$  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  $\sigma_A$  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, 1999a,b). For an atom with an isotropic displacement parameter of average value ( $B_{\text{avg}}$ ), it is defined as

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

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

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

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

(Blow, 2002), where  $sc$  is the solvent fraction ( $= N_{\text{solv}}/N_{\text{atom}}$ , where  $N_{\text{solv}}$  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.* **A35**, 497–499.
- Blow, D. M. (2002). *Rearrangement of Cruickshank's formulae for the diffraction-component precision index*. *Acta Cryst.* **D58**, 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.* **D50**, 760–763.
- Cowtan, K. (1999). *Error estimation and bias correction in phase-improvement calculations*. *Acta Cryst.* **D55**, 1555–1567.
- Cowtan, K. & Main, P. (1998). *Miscellaneous algorithms for density modification*. *Acta Cryst.* **D54**, 487–493.
- Cowtan, K. D. & Main, P. (1996). *Phase combination and cross validation in iterated density-modification calculations*. *Acta Cryst.* **D52**, 43–48.
- Cruickshank, D. W. J. (1999a). *Remarks about protein structure precision*. *Acta Cryst.* **D55**, 583–601.
- Cruickshank, D. W. J. (1999b). *Remarks about protein structure precision*. *Erratum*. *Acta Cryst.* **D55**, 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.* **D62**, 96–101.
- Diederichs, K. (2010). *Quantifying instrument errors in macromolecular X-ray data sets*. *Acta Cryst.* **D66**, 733–740.