

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 - (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, 1999a,b). For an atom with an isotropic displacement parameter of average value (B_{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_{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 $(N_{\text{hkl}} - N_{\text{para}})$ 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

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

(Blow, 2002), where sc is the solvent fraction ($= N_{\text{solv}}/N_{\text{atom}}$, where N_{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.

2. BASIC CRYSTALLOGRAPHY

- Diederichs, K. & Karplus, P. A. (1997a). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* **4**, 269–275.
- Diederichs, K. & Karplus, P. A. (1997b). Improved R-factors for diffraction data analysis in macromolecular crystallography. Erratum. *Nat. Struct. Biol.* **4**, 592.
- Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst.* **A47**, 392–400.
- Evans, P. (2006). Scaling and assessment of data quality. *Acta Cryst.* **D62**, 72–82.
- Fisher, S. J., Helliwell, J. R., Khurshid, S., Govada, L., Redwood, C., Squire, J. M. & Chayen, N. E. (2008). An investigation into the protonation states of the C1 domain of cardiac myosin-binding protein C. *Acta Cryst.* **D64**, 658–664.
- Grosse-Kunstleve, R. W. & Adams, P. D. (2003). Substructure search procedures for macromolecular structures. *Acta Cryst.* **D59**, 1966–1973.
- Hauptman, H. & Karle, J. (1953). Solution of the phase problem. I. The centrosymmetric crystal. Am. Crystallogr. Assoc. Monograph No. 3. Dayton, Ohio: Polycrystal Book Service.
- Hirshfeld, F. L. & Rabinovich, D. (1973). Treating weak reflexions in least-squares calculations. *Acta Cryst.* **A29**, 510–513.
- Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. (1996). Errors in protein structures. *Nature (London)*, **381**, 272.
- Howell, P. L. & Smith, G. D. (1992). Identification of heavy-atom derivatives by normal probability methods. *J. Appl. Cryst.* **25**, 81–86.
- James, R. W. (1948). False detail in three-dimensional Fourier representations of crystal structures. *Acta Cryst.* **1**, 132–134.
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**, 110–119.
- Kabsch, W. (1988). Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Cryst.* **21**, 916–924.
- Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800.
- Kabsch, W. (2010). XDS. *Acta Cryst.* **D66**, 125–132.
- Kleywegt, G. & Brünger, A. T. (1996). Checking your imagination: applications of the free R value. *Structure*, **4**, 897–904.
- Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). BALBES: a molecular-replacement pipeline. *Acta Cryst.* **D64**, 125–132.
- Luzzati, V. (1952). Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Cryst.* **5**, 802–810.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674.
- Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
- Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
- Panjikar, S. & Tucker, P. A. (2002). Phasing possibilities using different wavelengths with a xenon derivative. *J. Appl. Cryst.* **35**, 261–266.
- Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. & Berman, H. M. (1996). New parameters for the refinement of nucleic acid-containing structures. *Acta Cryst.* **D52**, 57–64.
- Pflugrath, J. W. (1999). The finer things in X-ray diffraction data collection. *Acta Cryst.* **D55**, 1718–1725.
- Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst.* **A42**, 140–149.
- Read, R. J. (2001). Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Cryst.* **D57**, 1373–1382.
- Rodgers, J. L. & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *Am. Stat.* **42**, 59–66.
- Rodriguez, D. D., Grosse, C., Himmel, S., Gonzalez, C., de Ilarduya, I. M., Becker, S., Sheldrick, G. M. & Uson, I. (2009). Crystallographic ab initio protein structure solution below atomic resolution. *Nat. Methods*, **6**, 651–653.
- Rossmann, M. G. & Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.* **15**, 24–31.
- Schneider, T. R. & Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Cryst.* **D58**, 1772–1779.
- Sheldrick, G. M. (2002). Macromolecular phasing with SHELXE. *Z. Kristallogr.* **217**, 644–650.
- Sheldrick, G. M. (2008). A short history of SHELX. *Acta Cryst.* **A64**, 112–122.
- Sheldrick, G. M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Cryst.* **D66**, 479–485.
- Stenkamp, R. E. & Jensen, L. H. (1984). Resolution revisited: limit of detail in electron density maps. *Acta Cryst.* **A40**, 251–254.
- Terwilliger, T. C. (2001). Map-likelihood phasing. *Acta Cryst.* **D57**, 1763–1775.
- Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H. & Hung, L.-W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Cryst.* **D65**, 582–601.
- Terwilliger, T. C. & Berendzen, J. (1999). Discrimination of solvent from protein regions in native Fouriers as a means of evaluating heavy-atom solutions in the MIR and MAD methods. *Acta Cryst.* **D55**, 501–505.
- Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Cryst.* **D55**, 191–205.
- Vellieux, F. M. D. A. P., Hunt, J. F., Roy, S. & Read, R. J. (1995). DEMON/ANGEL: a suite of programs to carry out density modification. *J. Appl. Cryst.* **28**, 347–351.
- Weeks, C. M., De Titta, G. T., Miller, R. & Hauptman, H. A. (1993). Application of the minimal principle to peptide structures. *Acta Cryst.* **D49**, 179–181.
- Weeks, C. M., DeTitta, G. T., Hauptman, H. A., Thuman, P. & Miller, R. (1994). Structure solution by minimal-function phase refinement and Fourier filtering. II. Implementation and applications. *Acta Cryst.* **A50**, 210–220.
- Weiss, M. S. (2001). Global indicators of X-ray data quality. *J. Appl. Cryst.* **34**, 130–135.
- Weiss, M. S. & Hilgenfeld, R. (1997). On the use of the merging R factor as a quality indicator for X-ray data. *J. Appl. Cryst.* **30**, 203–205.
- Wilson, A. J. C. (1949). The probability distribution of X-ray intensities. *Acta Cryst.* **2**, 318–321.