International Tables for Crystallography (2006). Vol. F, Chapter 18.5, pp. 403–418.

18.5. Coordinate uncertainty

BY D. W. J. CRUICKSHANK

18.5.1. Introduction

18.5.1.1. *Background*

Even in 1967 when the first few protein structures had been solved, it would have been hard to imagine a time when the best protein structures would be determined with a precision approaching that of small molecules. That time was reached during the 1990s. Consequently, the methods for the assessment of the precision of small molecules can be extended to good-quality protein structures.

The key idea is simply stated. At the conclusion and full convergence of a least-squares or equivalent refinement, *the estimated variances and covariances of the parameters may be obtained through the inversion of the least-squares full matrix.*

The inversion of the full matrix for a large protein is a gigantic computational task, but it is being accomplished in a rising number of cases. Alternatively, approximations may be sought. Often these can be no more than rough order-of-magnitude estimates. Some of these approximations are considered below.

Caveat. Quite apart from their large numbers of atoms, protein structures show features differing from those of well ordered small-molecule structures. Protein crystals contain large amounts of solvent, much of it not well ordered. Parts of the protein chain may be floppy or disordered. All natural protein crystals are noncentrosymmetric, hence the simplifications of error assessment for centrosymmetric structures are inapplicable. The effects of incomplete modelling of disorder on phase angles, and thus on parameter errors, are not addressed explicitly in the following analysis. Nor does this analysis address the quite different problem of possible gross errors or misplacements in a structure, other than by their indication through high B values or high coordinate standard uncertainties. These various difficulties are, of course, reflected in the values of $\Delta|F|$ used in the precision estimates.

On the problems of structure validation see Part 21 of this volume and Dodson (1998).

Some structure determinations do make a first-order correction for the effects of disordered solvent on phase angles by application of Babinet's principle of complementarity (Langridge *et al.*, 1960; Moews & Kretsinger, 1975; Tronrud, 1997). Babinet's principle follows from the fact that if $\rho(\mathbf{x})$ is constant throughout the cell, then $F(\mathbf{h}) = 0$, except for $F(\mathbf{0})$. Consequently, if the cell is divided into two regions C and D, $F_C(\mathbf{h}) = -F_D(\mathbf{h})$. Thus if D is a region of disordered solvent, $F_D(\mathbf{h})$ can be estimated from $-F_C(\mathbf{h})$. A first approximation to a disordered model may be obtained by placing negative point-atoms with very high Debye B values at all the ordered sites in region C. This procedure provides some correction for very low resolution planes. Alternatively, corrections are sometimes made by a mask bulk solvent model (Jiang & Brünger, 1994).

The application of restraints in protein refinement does not affect the key idea about the method of error estimation. A simple model for restrained refinement is analysed in Section 18.5.3, and the effect of restraints is discussed in Section 18.5.4 and later.

Much of the material in this chapter is drawn from a Topical Review published in *Acta Crystallographica*, Section D (Cruickshank, 1999).

Protein structures exhibiting noncrystallographic symmetry are not considered in this chapter.

18.5.1.2. *Accuracy and precision*

A distinction should be made between the terms *accuracy* and *precision.* A single measurement of the magnitude of a quantity differs by error from its unknown true value λ. In statistical theory (Cruickshank, 1959), the fundamental supposition made about errors is that, for a given experimental procedure, the possible results of an experiment define the probability density function $f(x)$ of a *random variable.* Both the true value λ and the probability density $f(x)$ are unknown. The problem of assessing the accuracy of a measurement is thus the double problem of estimating $f(x)$ and of assuming a relation between $f(x)$ and λ.

Precision relates to the function $f(x)$ and its spread.

The problem of what relationship to assume between $f(x)$ and the true value λ is more subtle, involving particularly the question of *systematic errors.* The usual procedure, after correcting for known systematic errors, is to suppose that some typical property of $f(x)$, often the mean, is the value of λ. No repetition of the same experiment will ever reveal the systematic errors, so statistical estimates of precision take into account only random errors. Empirically, systematic errors can be detected only by remeasuring the quantity with a different technique.

Care is needed in reading older papers. The word accuracy was sometimes intended to cover both random and systematic errors, or it may cover only random errors in the above sense of precision (known systematic errors having been corrected).

In recent years, the well established term *estimated standard deviation* (e.s.d.) has been replaced by the term *standard uncertainty* (s.u.). (See Section 18.5.2.3 on statistical descriptors.)

18.5.1.3. *Effect of atomic displacement parameters (or 'temperature factors')*

It is useful to begin with a reminder that the Debye $B = 8\pi^2\langle u^2\rangle$, where u is the atomic displacement parameter. If $B = 80$ Å^2, the r.m.s. amplitude is 1.01 Å. The centroid of an atom with such a B is unlikely to be precisely determined. For $B = 40$ Å^2, the 0.71 Å r.m.s. amplitude of an atom is approximately half a C—N bond length. For $B = 20$ Å^2, the amplitude is 0.50 Å. Even for $B = 5$ Å^2, the amplitude is 0.25 Å. The size of the atomic displacement amplitudes should always be borne in mind when considering the precision of the position of the centroid of an atom.

Scattering power depends on $\exp[-2B(\sin\theta/\lambda)^2] = \exp[-B/(2d^2)]$. For $B = 20$ Å^2 and $d = 4$, 2 or 1 Å, this factor is 0.54, 0.08 or 0.0001. For $d = 2$ Å and $B = 5$, 20 or 80 Å^2, the factor is again 0.54, 0.08 or 0.0001. The scattering power of an atom thus depends very strongly on B and on the resolution $d = 1/s = \lambda/2\sin\theta$. Scattering at high resolution (low d) is dominated by atoms with low B.

An immediate consequence of the strong dependence of scattering power on B is that the standard uncertainties of atomic coordinates also depend very strongly on B, especially between atoms of different B within the same structure.

[An IUCr Subcommittee on Atomic Displacement Parameter Nomenclature (Trueblood *et al.*, 1996) has recommended that the phrase 'temperature factor', though widely used in the past, should be avoided on account of several ambiguities in its meaning and usage. The Subcommittee also discourages the use of B and the anisotropic tensor $\mathbf{B}$ in favour of $\langle u^2\rangle$ and $\mathbf{U}$, on the grounds that the latter have a more direct physical significance. The present author concurs (Cruickshank, 1956, 1965). However, as the use of B or B_{eq} is currently so widespread in biomolecular crystallography, this chapter has been written in terms of B.]

18.5.2. The least-squares method

18.5.2.1. *The normal equations*

In the unrestrained least-squares method, the residual

$$R = \sum_3 w(hkl)\Delta^2(hkl) \qquad (18.5.2.1)$$

is minimized, where Δ is either $|F_o| - |F_c|$ for R_1 or $|F_o|^2 - |F_c|^2$ for R_2, and $w(hkl)$ is chosen appropriately. The summation is over crystallographically independent planes.

When R is a minimum with respect to the parameter u_j, $\partial R/\partial u_j = 0$, *i.e.*,

$$\sum_3 w\Delta(\partial\Delta/\partial u_j) = 0. \qquad (18.5.2.2)$$

For R_1, $\partial\Delta/\partial u_j = -\partial|F_c|/\partial u_j$; for R_2, $\partial\Delta/\partial u_j = -2|F_c|\partial|F_c|/\partial u_j$. The n parameters have to be varied until the n conditions (18.5.2.2) are satisfied. For a trial set of the u_j close to the correct values, we may expand Δ as a function of the parameters by a Taylor series to the first order. Thus for R_1,

$$\Delta(\mathbf{u}+\mathbf{e}) = \Delta(\mathbf{u}) - \sum_i \varepsilon_i(\partial|F_c|/\partial u_i), \qquad (18.5.2.3)$$

where ε_i is a small change in the parameter u_i, and $\mathbf{u}$ and $\mathbf{e}$ represent the whole sets of parameters and changes. The minus sign occurs before the summation, since $\Delta = |F_o| - |F_c|$, and the changes in $|F_c|$ are being considered.

Substituting (18.5.2.3) in (18.5.2.2), we get the *normal equations* for R_1,

$$\sum_i \varepsilon_i \left[\sum_3 w(\partial|F_c|/\partial u_i)(\partial|F_c|/\partial u_j)\right] = \sum_3 w\Delta(\partial|F_c|/\partial u_j). \qquad (18.5.2.4)$$

There are n of these equations for $j = 1, \ldots, n$ to determine the n unknown ε_j.

For R_2 the normal equations are

$$\sum_i \varepsilon_i \left[\sum_3 w(\partial|F_c|^2/\partial u_i)(\partial|F_c|^2/\partial u_j)\right] = \sum_3 w\Delta(\partial|F_c|^2/\partial u_j). \qquad (18.5.2.5)$$

Both forms of the normal equations can be abbreviated to

$$\sum_i \varepsilon_i a_{ij} = b_j. \qquad (18.5.2.6)$$

For the values of $\partial|F_c|/\partial u_j$ for common parameters see, *e.g.*, Cruickshank (1970).

Some important points in the derivation of the standard uncertainties of the refined parameters can be most easily understood if we suppose that the matrix a_{ij} can be approximated by its diagonal elements. Each parameter is then determined by a single equation of the form

$$\varepsilon_i \sum_3 wg^2 = \sum_3 wg\Delta, \qquad (18.5.2.7)$$

where $g = \partial|F_c|/\partial u_i$ or $\partial|F_c|^2/\partial u_i$. Hence

$$\varepsilon_i = \left(\sum_3 wg\Delta\right) \Big/ \left(\sum_3 wg^2\right). \qquad (18.5.2.8)$$

At the conclusion of the refinement, when R is a minimum, the variance (square of the s.u.) of the parameter u_i due to uncertainties in the Δ's is

$$\sigma_i^2 = \left[\sum_3 w^2 g^2 \sigma^2(F)\right] \Big/ \left(\sum_3 wg^2\right)^2. \qquad (18.5.2.9)$$

If the weights have been chosen as $w(hkl) = 1/\sigma^2(|F_{hkl}|)$ or $1/\sigma^2(|F_{hkl}|^2)$, this simplifies to

$$\sigma_i^2 = 1 \Big/ \left(\sum_3 wg^2\right) = 1/a_{ii}, \qquad (18.5.2.10)$$

which is appropriate for absolute weights. Equation (18.5.2.10) provides an s.u. for a parameter relative to the s.u.'s $\sigma(|F|)$ or $\sigma(|F|^2)$ of the observations.

In general, with the full matrix a_{ij} in the normal equations,

$$\sigma_i^2 = (a^{-1})_{ii}, \qquad (18.5.2.11)$$

where $(a^{-1})_{ii}$ is an element of the matrix inverse to a_{ij}. The covariance of the parameters u_i and u_j is $\mathrm{cov}(i,j) \equiv \sigma_i\sigma_j\mathrm{correl}(i,j) = (a^{-1})_{ij}$.

18.5.2.2. *Weights*

In the early stages of refinement, artificial weights may be chosen to accelerate refinement. In the final stages, the weights must be related to the precision of the structure factors if parameter variances are being sought. There are two distinct ways, covering two ranges of error, in which this may be done.

(1) The weights for R_1, say, may reflect the precision of the $|F_o|$, so that $w(hkl) = 1/\sigma^2(|F_{hkl}|)$, where σ^2 is the estimated variance of $|F_o|$ due to a specific class of experimental uncertainties. These absolute weights are derived from an analysis of the experiment. Weights chosen in this way lead to estimated parameter variances $\sigma_i^2 = (a^{-1})_{ii}$, (18.5.2.11), which cover only the specific class of experimental uncertainties.

(2) The weights may reflect the trends in the $|\Delta| \equiv \|F_o| - |F_c\|$. A weighting function with a small number of parameters is chosen so that the averages of $w\Delta^2$ are constant when the set of $w\Delta^2$ values is analysed in any pertinent fashion (*e.g.* in bins of increasing $|F_o|$ and $2\sin\theta/\lambda$). Weights chosen in this way are relative weights, and the expression for the parameter variances needs a scaling factor,

$$S^2 = \left(\sum_3 w\Delta^2\right) \Big/ (n_{\mathrm{obs}} - n_{\mathrm{params}}). \qquad (18.5.2.12)$$

Hence, in the full-matrix case,

$$\sigma_i^2 = \left[\left(\sum_3 w\Delta^2\right) \Big/ (n_{\mathrm{obs}} - n_{\mathrm{params}})\right](a^{-1})_{ii}, \qquad (18.5.2.13)$$

which allows for all random experimental errors, such systematic experimental errors as cannot be simulated in the $|F_c|$ and imperfections in the calculated model.

18.5.2.3. *Statistical descriptors and goodness of fit*

In recent years, there have been developments and changes in statistical nomenclature and usage. Many aspects are summarised in the reports of the IUCr Subcommittee on Statistical Descriptors in Crystallography (Schwarzenbach *et al.*, 1989, 1995). In the second report, *inter alia*, the Subcommittee emphasizes the terms *uncertainty* and *standard uncertainty* (s.u.). The latter is a replacement for the older term *estimated standard deviation* (e.s.d.). The Subcommittee classify uncertainty components in two categories, based on their method of evaluation: type A, estimated by the statistical analysis of a series of observations, and type B, estimated otherwise. As an example of the latter, a type B component could allow for doubts concerning the estimated shape and dimensions of the diffracting crystal and the subsequent corrections made for absorption.

The square root S of the expression S^2, (18.5.2.12) above, is called the *goodness of fit* when the weights are the reciprocals of the absolute variances of the observations.

One recommendation in the second report does call for comment here. While agreeing that formulae like (18.5.2.13) lead to conservative estimates of parameter variances, the report suggests that this practice is based on the questionable assumption that the variances of the observations by which the weights are assigned are relatively correct but uniformly underestimated. When the goodness of fit $S > 1$, then either the weights or the model or both are suspect.

Comment is needed. The account in Section 18.5.2.2 describes two distinct ways of estimating parameter variances, covering two ranges of error. The kind of weights envisaged in the reports (based on variances of type A and/or of type B) are of a class described for method (1). They are not the weights to be used in method (2) (though they may be a component in such weights). Method (2) implicitly assumes from the outset that there are experimental errors, some covered and others not covered by method (1), and that there are imperfections in the calculated model (as is obviously true for proteins). Method (2) avoids exploring the relative proportions and details of these error sources and aims to provide a realistic estimate of parameter uncertainties which can be used in external comparisons. It can be formally objected that method (2) does not conform to the criteria of random-variable theory, since clearly the Δ's are partially correlated through the remaining model errors and some systematic experimental errors. But it is a useful procedure. Method (1) on its own would present an optimistic view of the reliability of the overall investigation, the degree of optimism being indicated by the inverse of the goodness of fit (18.5.2.12). In method (2), if the weights are on an arbitrary scale, then S^2 can have an arbitrary value.

For an advanced-level treatment of many aspects of the refinement of structural parameters, see Part 8 of *International Tables for Crystallography*, Volume C (1999). The detection and treatment of systematic error are discussed in Chapter 8.5 therein.

18.5.3. Restrained refinement

18.5.3.1. *Residual function*

Protein structures are often refined by a restrained refinement program such as *PROLSQ* (Hendrickson & Konnert, 1980). Here, a function of the type

$$R' = \sum w_h(\Delta F)^2 + \sum w_{\rm geom}(\Delta Q)^2 \qquad (18.5.3.1)$$

is minimized, where Q denotes a geometrical restraint such as a bond length. Formally, all one is doing is extending the list of observations. One is adding to the protein diffraction data geometrical data from a stereochemical dictionary such as that of Engh & Huber (1991). A chain C—N bond length may be known from the dictionary with much greater precision $1/w_{\rm geom}^{1/2}$, say 0.02 Å, than from an unrestrained diffraction-data-only protein refinement.

In a high-resolution unrestrained refinement of a small molecule, the standard uncertainty (s.u.) of a bond length A—B is often well approximated by

$$\sigma(l) = (\sigma_A^2 + \sigma_B^2)^{1/2}. \qquad (18.5.3.2)$$

However, in a protein determination $\sigma(l)$ is often much smaller than either σ_A or σ_B because of the excellent information from the stereochemical dictionary, which correlates the positions of A and B.

Laying aside computational size and complexity, the protein precision problem is straightforward in principle. When a restrained refinement has converged to an acceptable structure and the shifts in successive rounds have become negligible, invert the full matrix. The inverse matrix immediately yields estimates of the variances and covariances of all parameters.

The dimensions of the matrix are the same whether or not the refinement is restrained. The full matrix will be rather sparse, but not nearly as sparse as in a small-molecule refinement. For the purposes of Section 18.5.3, it is irrelevant whether the residual for the diffraction data is based on $|F|$ or $|F|^2$. On the relative weighting of the diffraction and restraint terms, see Section 18.5.3.3.

18.5.3.2. *A very simple protein model*

Some aspects of restrained refinement are easily understood by considering a *one-dimensional protein consisting of two like atoms* in the asymmetric unit, with coordinates x_1 and x_2 relative to a fixed origin and bond length $l = x_2 - x_1$. In the refinement, the normal equations are of the type $\mathbf{N}\Delta\mathbf{x} = \mathbf{e}$. For two non-overlapping like atoms, the *diffraction data* will yield a normal matrix

$$\mathbf{N} = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}, \qquad (18.5.3.3)$$

with inverse

$$\begin{pmatrix} 1/a & 0 \\ 0 & 1/a \end{pmatrix}, \qquad (18.5.3.4)$$

where

$$a = \sum w_h(\partial|F_n|/\partial x_i)^2. \qquad (18.5.3.5)$$

A *geometric restraint* on the length will yield a normal matrix

$$\begin{pmatrix} b & -b \\ -b & b \end{pmatrix} \qquad (18.5.3.6)$$

with no inverse, since its determinant is zero, where

$$b = w_{\rm geom}(\partial l/\partial x_i)^2. \qquad (18.5.3.7)$$

Note $\partial l/\partial x_2 = -\partial l/\partial x_i = 1$, so that

$$b = w_{\rm geom} = 1/\sigma_{\rm geom}^2(l), \qquad (18.5.3.8)$$

where $\sigma_{\rm geom}^2(l)$ is the variance assigned to the length in the stereochemical dictionary.

Combining the diffraction data and the restraint, the normal matrix becomes

$$\begin{pmatrix} a+b & -b \\ -b & a+b \end{pmatrix}, \qquad (18.5.3.9)$$

with inverse

$$\{1/[a(a+2b)]\}\begin{pmatrix} a+b & b \\ b & a+b \end{pmatrix}. \qquad (18.5.3.10)$$

For the diffraction data alone, the variance of x_i is

$$\sigma_{\rm diff}^2(x_i) = 1/a. \qquad (18.5.3.11)$$

For the diffraction data plus restraint, the variance of x_i is

$$\sigma_{\rm res}^2(x_i) = (a+b)/[a(a+2b)] < \sigma_{\rm diff}^2(x_i). \qquad (18.5.3.12)$$

Note that though the restraint says nothing about the position of x_i, the variance of x_i has been reduced because of the coupling to the position of the other atom. In the limit when $a \ll b$, $\sigma_{\rm res}^2(x_i)$ is only half $\sigma_{\rm diff}^2(x_i)$.

The general formula for the variance of the length $l = x_2 - x_1$ is

$$\sigma^2(l) = \sigma^2(x_2) - 2\mathrm{cov}(x_2, x_1) + \sigma^2(x_1). \qquad (18.5.3.13)$$

For the diffraction data alone, this gives

$$\sigma^2_{\rm diff}(l) = 1/a + 0 + 1/a = 2/a = 2\sigma^2_{\rm diff}(x_i), \qquad (18.5.3.14)$$

as expected. For the diffraction data plus restraint,

$$\begin{aligned}\sigma^2_{\rm res}(l) &= [1/a(a+2b)][(a+b) - 2b + (a+b)] \\ &= 1/(a/2 + b) \\ &< \sigma^2_{\rm diff}(l).\end{aligned} \qquad (18.5.3.15)$$

For small a, $\sigma^2_{\rm res}(l) \rightarrow 1/b = \sigma^2_{\rm geom}(l)$, as expected. The variance of the restrained length, (18.5.3.15), can be re-expressed as

$$1/\sigma^2_{\rm res}(l) = 1/\sigma^2_{\rm diff}(l) + 1/\sigma^2_{\rm geom}(l). \qquad (18.5.3.16)$$

For the two-atom protein, it can be proved directly, as one would expect from (18.5.3.16), that *restrained refinement determines a length which is the weighted mean of the diffraction-only length and the geometric dictionary length*.

The centroid has coordinate $c = (x_1 + x_2)/2$. It is easily found that $\sigma^2_{\rm res}(c) = \sigma^2_{\rm diff}(c) = 1/2a$. Thus, as expected, the restraint says nothing about the position of the molecule in the cell.

For numerical illustrations of the s.u.'s in restrained refinement, suppose the stereochemical length restraint has $\sigma_{\rm geom}(l) = 0.02$ Å. Equation (18.5.3.16) gives the length s.u. $\sigma_{\rm res}(l)$ in restrained refinement. If the diffraction-only $\sigma_{\rm diff}(x_i) = 0.01$ Å, the restrained $\sigma_{\rm res}(l)$ is 0.012 Å. If $\sigma_{\rm diff}(x_i) = 0.05$ Å, $\sigma_{\rm res}(l)$ is 0.019 Å. However large $\sigma_{\rm diff}(x_i)$, $\sigma_{\rm res}(l)$ never exceeds 0.02 Å.

Equation (18.5.3.12) gives the position s.u. $\sigma_{\rm res}(x_i)$ in restrained refinement. If the diffraction-only $\sigma_{\rm diff}(x_i) = 0.01$ Å, the restrained $\sigma_{\rm res}(x_i)$ is 0.009 Å. If $\sigma_{\rm diff}(x_i) = 0.05$ Å, $\sigma_{\rm res}(x_i) = 0.037$ Å. For large $\sigma_{\rm diff}(x_i)$, $\sigma_{\rm res}(x_i)$ tends to $\sigma_{\rm diff}(x_i)/(2)^{1/2}$ as the strong restraint couples the two atoms together. For very small $\sigma_{\rm diff}(x_i)$, the relatively weak restraint has no effect.

18.5.3.3. *Relative weighting of diffraction and restraint terms*

When only relative diffraction weights are known, as in equation (18.5.2.13), it has been common (Rollett, 1970) to scale the geometric restraint terms against the diffraction terms by replacing the restraint weights $w_{\rm geom} = 1/\sigma^2_{\rm geom}$ by $w_{\rm geom} = S^2/\sigma^2_{\rm geom}$, where $S^2 = (\sum w_h \Delta^2_h)/(n_{\rm obs} - n_{\rm params})$. However, this scheme cannot be used for low-resolution structures if $n_{\rm obs} < n_{\rm params}$.

The treatment by Tickle *et al.* (1998*a*) shows that the reduction $n_{\rm params}$ in the number of degrees of freedom has to be distributed among all the data, both diffraction observations and restraints. Since the geometric restraint weights are on an absolute scale (Å^{-2}), they propose that the (absolute) scale of the diffraction weights should be determined by adjustment until the restrained residual R' (18.5.3.1) is equal to its expected value $(n_{\rm obs} + n_{\rm restraints} - n_{\rm params})$.

For a method of determining the scale of the diffraction weights based on $R_{\rm free}$, see Brünger (1993).

The geometric restraint weights were classified by the IUCr Subcommittee (Schwarzenbach *et al.*, 1995) as derived from observations supplementary to the diffraction data, with uncertainties of type B (Section 18.5.2.3).

18.5.4. Two examples of full-matrix inversion

18.5.4.1. *Unrestrained and restrained inversions for concanavalin A*

G. M. Sheldrick extended his *SHELXL*96 program (Sheldrick & Schneider, 1997) to provide extra information about protein precision through the inversion of least-squares full matrices. His programs have been used by Deacon *et al.* (1997) for the high-resolution refinement of native concanavalin A with 237 residues, using data at 110 K to 0.94 Å refined anisotropically. After the convergence and completion of full-matrix restrained refinement for the structure, the unrestrained full matrix (coordinates only) was computed and then inverted in a massive calculation. This led to s.u's $\sigma(x)$, $\sigma(y)$, $\sigma(z)$ and $\sigma(r)$ for all atoms, and to $\sigma(l)$ and $\sigma(\theta)$ for all bond lengths and angles. $\sigma(r)$ is defined as $[\sigma^2(x) + \sigma^2(y) + \sigma^2(z)]^{1/2}$. For concanavalin A the restrained full matrix was also inverted, thus allowing the comparison of restrained and unrestrained s.u.'s.

The results for concanavalin A from the inversion of the coordinate matrices of order 6402 $(= 2134 \times 3)$ are plotted in Figs. 18.5.4.1 and 18.5.4.2. Fig. 18.5.4.1 shows $\sigma(r)$ *versus* $B_{\rm eq}$ for the fully occupied atoms of the protein (a few atoms with $B >$ 60 Å^2 are off-scale). The points are colour-coded black for carbon, blue for nitrogen and red for oxygen. Fig. 18.5.4.1(*a*) shows the restrained results, and Fig. 18.5.4.1(*b*) shows the unrestrained diffraction-data-only results. Superposed on both sets of data points are least-squares quadratic fits determined with weights $1/B^2$. At high B, the unrestrained $\sigma_{\rm diff}(r)$ can be at least double the restrained $\sigma_{\rm res}(r)$, *e.g.*, for carbon at $B = 50$ Å^2, the unrestrained $\sigma_{\rm diff}(r)$ is about 0.25 Å, whereas the restrained $\sigma_{\rm res}(r)$ is about 0.11 Å. For $B < 10$ Å^2, both $\sigma(r)$'s fall below 0.02 Å and are around 0.01 Å at $B = 6$ Å^2.

For $B < 10$ Å^2, the better precision of oxygen as compared with nitrogen, and of nitrogen as compared with carbon, can be clearly seen. At the lowest B, the unrestrained $\sigma_{\rm diff}(r)$ in Fig. 18.5.4.1(*b*) are almost as small as the restrained $\sigma_{\rm res}(r)$ in Fig. 18.5.4.1(*a*). [The quadratic fits of the restrained results in Fig. 18.5.4.1(*a*) are evidently slightly imperfect in making $\sigma_{\rm res}(r)$ tend almost to 0 as B tends to 0.]

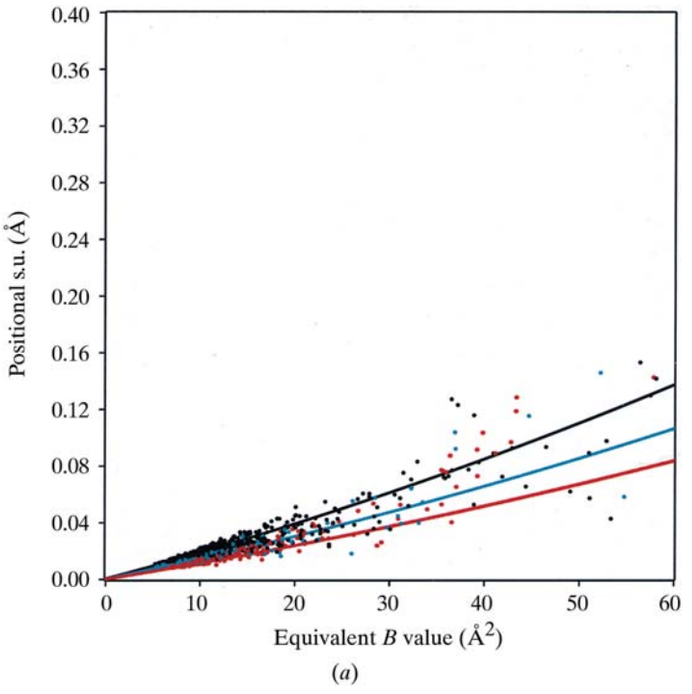
Fig. 18.5.4.2 shows $\sigma(l)$ *versus* $B_{\rm eq}$ for the bond lengths in the protein. The points are colour-coded black for C—C, blue for C—N and red for C—O. The restrained and unrestrained distributions are very different for high B. The restrained distribution in Fig. 18.5.4.2(*a*) tends to about 0.02 Å, which is the standard uncertainty of the applied restraint for 1–2 bond lengths, whereas the unrestrained distribution in Fig. 18.5.4.2(*b*) goes off the scale of the diagram. But for $B < 10$ Å^2, both distributions fall to around 0.01 Å.

The differences between the restrained and unrestrained $\sigma(r)$ and $\sigma(l)$ can be understood through the two-atom model for restrained refinement described in Section 18.5.3. For that model, the equation

$$1/\sigma^2_{\rm res}(l) = 1/\sigma^2_{\rm diff}(l) + 1/\sigma^2_{\rm geom}(l) \qquad (18.5.3.16)$$

relates the bond-length s.u. in the restrained refinement, $\sigma_{\rm res}(l)$, to the $\sigma_{\rm diff}(l)$ of the unrestrained refinement and the s.u. $\sigma_{\rm geom}(l)$ assigned to the length in the stereochemical dictionary. In the refinements, $\sigma_{\rm geom}(l)$ was 0.02 Å for all bond lengths. When this is combined in (18.5.3.16) with the unrestrained $\sigma_{\rm diff}(l)$ of any bond, the predicted restrained $\sigma_{\rm res}(l)$ is close to that found in the restrained full matrix.

It can be seen from Fig. 18.5.4.2(*b*) that many bond lengths with average $B < 10$ Å^2 have $\sigma_{\rm diff}(l) < 0.014$ Å. For these bonds the diffraction data have greater weight than the stereochemical dictionary. Some bonds have $\sigma_{\rm diff}(l)$ as low as 0.0080 Å, with



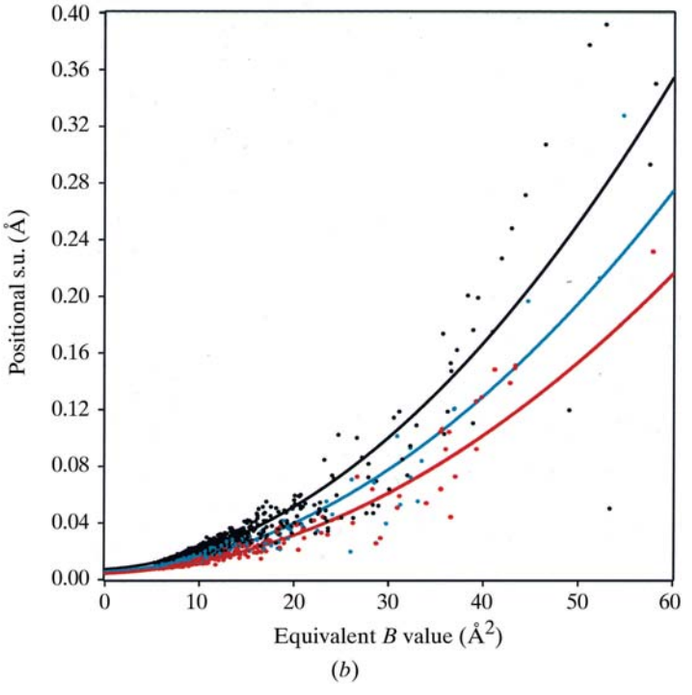
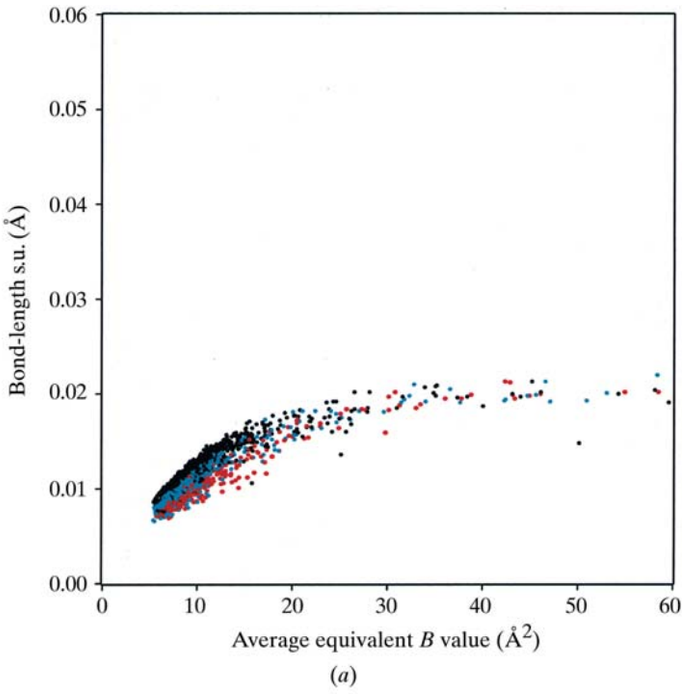


Fig. 18.5.4.1. Plots of $\sigma(r)$ *versus* B_{eq} for concanavalin A with 0.94 Å data, (*a*) restrained full-matrix $\sigma_{res}(r)$, (*b*) unrestrained full-matrix $\sigma_{diff}(r)$. Carbon black, nitrogen blue, oxygen red.



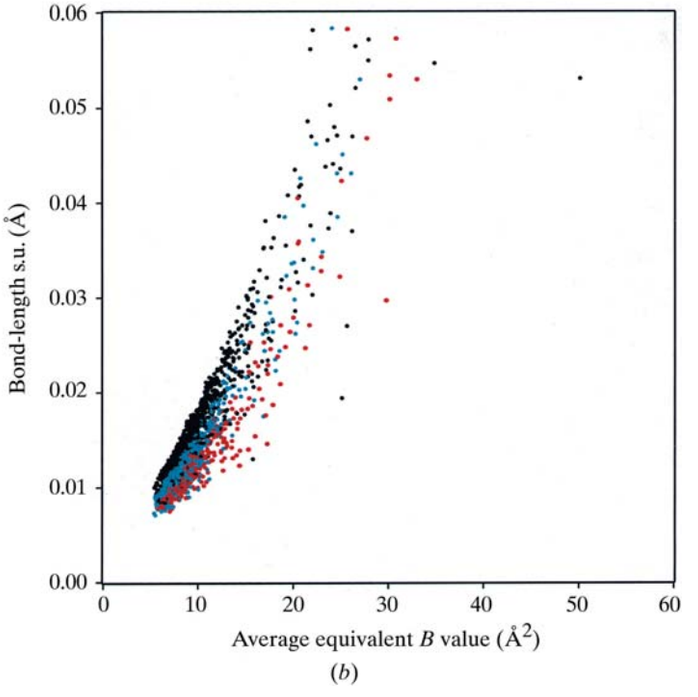


Fig. 18.5.4.2. Plots of $\sigma(l)$ *versus* average B_{eq} for concanavalin A with 0.94 Å data, (*a*) restrained full-matrix $\sigma_{res}(l)$, (*b*) unrestrained full-matrix $\sigma_{diff}(l)$. C—C black, C—N blue, C—O red.

$\sigma_{res}(l)$ around 0.0074 Å. This situation is one consequence of the availability of diffraction data to the high resolution of 0.94 Å. For large $\sigma_{diff}(l)$ (*i.e.*, high B), equation (18.5.3.16) predicts that $\sigma_{res}(l) = \sigma_{geom}(l) = 0.02$ Å, as is found in Fig. 18.5.4.2(*a*).

In an isotropic approximation, $\sigma(r) = 3^{1/2}\sigma(x)$. Equation (18.5.3.12) of the two-atom model can be recast to give

$$\sigma^2_{res}(r) = \sigma^2_{diff}(r)\left\{\left[\sigma^2_{diff}(r) + 3(0.02)^2\right] \Big/ \left[2\sigma^2_{diff}(r) + 3(0.02)^2\right]\right\}. \quad (18.5.4.1)$$

For low B, say $B \leq 15$ Å^2 in concanavalin, (18.5.4.1) gives quite good predictions of $\sigma_{res}(r)$ from $\sigma_{diff}(r)$. For instance, for a carbon atom with $B = 15$ Å^2, the quadratic curve for carbon in Fig. 18.5.4.1(*b*) shows $\sigma_{diff}(r) = 0.034$ Å, and Fig. 18.5.4.1(*a*) shows $\sigma_{res}(r) = 0.029$ Å. While if $\sigma_{diff}(r) = 0.034$ Å is used with (18.5.4.1), the resulting prediction for $\sigma_{res}(r)$ is 0.028 Å.

However, for high B, say $B = 50$ Å^2, the quadratic curve for carbon in Fig. 18.5.4.1(*b*) shows $\sigma_{diff}(r) = 0.25$ Å, and Fig. 18.5.4.1(*a*) shows $\sigma_{res}(r) = 0.11$ Å, whereas (18.5.4.1) leads to the poor estimate $\sigma_{res}(r) = 0.18$ Å.

Thus at high B, equation (18.5.4.1) from the two-atom model does not give a good description of the relationship between the

restrained and unrestrained $\sigma(r)$. The reason is obvious. Most atoms are linked by 1–2 bond restraints to two or three other atoms. Even a carbonyl oxygen atom linked to its carbon atom by a 0.02 Å restraint is also subject to 0.04 Å 1–3 restraints to chain C_α and N atoms. Consequently, for a high-B atom, when the restraints are applied it is coupled to several other atoms in a group, and its $\sigma_{\rm res}(r)$ is lower, compared with the diffraction-data-only $\sigma_{\rm diff}(r)$, by a greater amount than would be expected from the two-atom model.

18.5.4.2. *Unrestrained inversion for an immunoglobulin*

Sheldrick has provided the results of the unrestrained lower-resolution refinement of a single-chain immunoglobulin mutant (T39K) with 218 amino-acid residues, with data to 1.70 Å refined isotropically (Usón *et al.*, 1999). Fig. 18.5.4.3 shows $\sigma_{\rm diff}(r)$ *versus* $B_{\rm eq}$ for the fully occupied protein atoms. Superposed on the data points are least-squares quadratic fits. In a first very rough approximation for $\sigma_{\rm diff}(x_i)$ suggested later by equation (18.5.6.3), the dependence on atom type is controlled by $1/Z_i$, the reciprocal of the atomic number. Sheldrick found that a $1/Z_i$ dependence produced too little difference between C, N and O. The proportionalities between the quadratics for $\sigma(r)$ in Figs. 18.5.4.1 and 18.5.4.3 are based on the reciprocals of the scattering factors at $\sin\theta/\lambda = 0.3$ Å^{-1}, symbolized by $Z_i^{\#}$. For C, N and O, these are 2.494, 3.219 and 4.089, respectively. For potential use in later work, the least-squares fits to the $\sigma(r_i)Z_i^{\#}$ in Å are recorded here as

$$0.11892 + 0.00891B + 0.0001462B^2, \qquad (18.5.4.2a)$$

$$0.01826 + 0.001043B + 0.0002230B^2 \text{ and} \qquad (18.5.4.2b)$$

$$0.00115 + 0.004414B + 0.0000214B^2 \qquad (18.5.4.2c)$$

for the immunoglobulin (unrestrained), concanavalin A (unrestrained) and concanavalin A (restrained), respectively.

As might be expected from the lower resolution, the lowest $\sigma_{\rm diff}(r)$'s in the immunoglobulin are about six times the lowest $\sigma_{\rm diff}(r)$'s in concanavalin. But at $B = 50$ Å^2, the immunoglobulin curve for carbon gives $\sigma_{\rm diff}(r) = 0.37$ Å, which is only 50% larger than the concanavalin value of 0.25 Å.

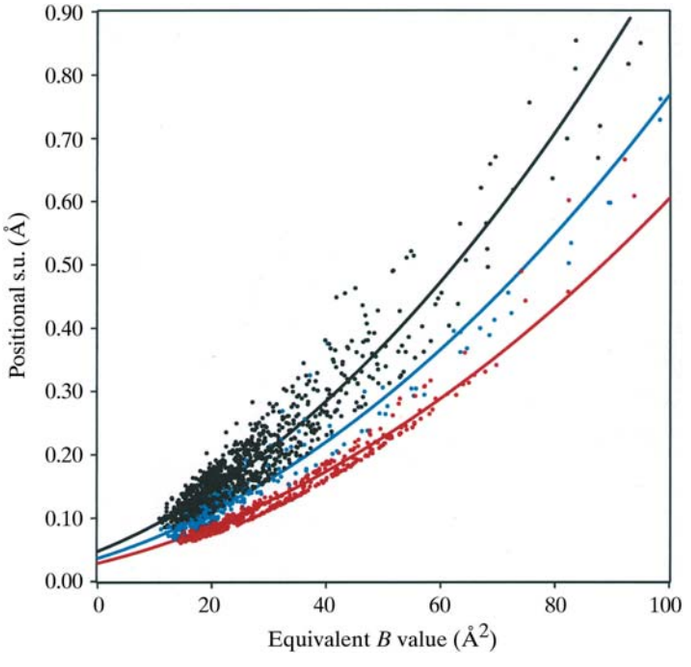


Fig. 18.5.4.3. Plot of $\sigma_{\rm diff}(r)$ *versus* $B_{\rm eq}$ from an unrestrained full matrix for immunoglobulin mutant (T39K) with 1.70 Å data. Carbon black, nitrogen blue, oxygen red.

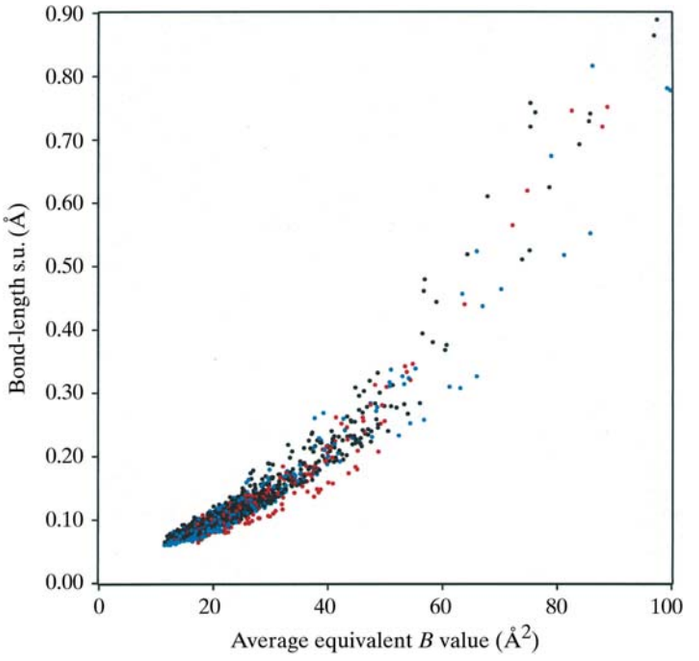


Fig. 18.5.4.4. Plot of $\sigma_{\rm diff}(l)$ *versus* average $B_{\rm eq}$ from an unrestrained full matrix for immunoglobulin mutant (T39K) with 1.70 Å data. C—O black, C—N blue, C—O red.

Fig. 18.5.4.4 shows $\sigma_{\rm diff}(l)$ *versus* $B_{\rm eq}$ for the immunoglobulin. Note that the lowest immunoglobulin unrestrained $\sigma_{\rm diff}(l)$ is about 0.06 Å, which is three times the 0.02 Å $\sigma_{\rm geom}(l)$ bond restraint.

18.5.4.3. *Comments on restrained refinement*

Geometric restraint dictionaries typically use bond-length weights based on $\sigma_{\rm geom}(l)$ of around 0.02 or 0.03 Å. Tables 18.5.7.1–18.5.7.3 show that even 1.5 Å studies have diffraction-only errors $\sigma_{\rm diff}(x, B_{\rm avg})$ of 0.08 Å and upwards. Only for resolutions of 1.0 Å or so are the diffraction-only errors comparable with the dictionary weights. Of course, the dictionary offers no values for many of the configurational parameters of the protein structure, including the centroid and molecular orientation.

18.5.4.4. *Full-matrix estimates of precision*

The opening contention of this chapter in Section 18.5.1.1 is that the variances and covariances of the structural parameters of proteins can be found from the inverse of the least-squares normal matrix. But there is a caveat, chiefly that explicit account would not be taken of disorder of the solvent or of parts of the protein. Corrections by Babinet's principle of complementarity or by mask bulk solvent models are only first-order approximations. The consequences of such disorder problems, which make the variation of calculated structure factors nonlinear over the range of interest, may in future be better handled by maximum-likelihood methods (*e.g.* Read, 1990; Bricogne, 1993*a*; Bricogne & Irwin, 1996; Murshudov *et al.*, 1997). Pannu & Read (1996) have shown how the maximum-likelihood method can be cast computationally into a form akin to least-squares calculations. Full-matrix precision estimates along the lines of the present chapter are probably somewhat low.

It should also be noted that full-matrix estimates of coordinate precision are most reliably derived from matrices involving both

coordinates and atomic displacement parameters. This is particularly important for lower-resolution analyses, in which atomic images overlap. The work on the high-resolution analysis of concanavalin A described in Section 18.5.4.1 was based on the very large coordinate matrix, of order 6402. The omission, because of computer limitations, of the anisotropic displacement parameters from the full matrix will have caused the coordinate s.u.'s of atoms with high B_{eq} to be underestimated.

Much information about the quality of a molecular model can be obtained from the eigenvalues and eigenvectors of the normal matrix (Cowtan & Ten Eyck, 2000).

18.5.5. Approximate methods

18.5.5.1. *Block calculations*

The full-matrix inversions described in the previous section require massive calculations. The length of the calculations is more a matter of the order of the matrix, *i.e.*, the number of parameters, than of the number of observations. When restraints are applied, it is the diffraction-cum-restraints full matrix which should be inverted.

With the increasing power of computers and more efficient algorithms (*e.g.* Tronrud, 1999; Murshudov *et al.*, 1999), a final full matrix should be computed and inverted much more regularly – and not just for high-resolution analyses. Low-resolution analyses have a need, beyond the indications given by B values, to identify through $\sigma(x)$ estimates their regions of tolerable and less tolerable precision.

If full-matrix calculations are impractical, partial schemes can be suggested. As far back as 1973, Watenpaugh *et al.* (1973), in a study of rubredoxin at 1.5 Å resolution, effectively inverted the diffraction full matrix in 200 parameter blocks to obtain individual s.u.'s. A similar scheme for restrained refinements could also use overlapping large blocks. A minimal block scheme in refinements of any resolution is to calculate blocks for each residue and for the block interactions between successive residues. The inversion process could then use the matrices in running groups of three successive residues, taking only the inverted elements for the central residue as the estimates of its variances and covariances.

For low-resolution analyses with very large numbers of atoms, it might be sufficient to gain a general idea of the behaviour of $\sigma(x)$ as a function of B by computing a limited number of blocks for representative or critical groups of residues. The parameters used in the blocks should include the B's, since atomic images overlap at low resolution, thus correlating the position of one atom with the displacement parameters of its neighbours.

18.5.5.2. *The modified Fourier method*

In the simplest form of the Fourier-map approach to centrosymmetric high-resolution structures, atomic positions are given by the maxima of the observed electron density. The uncertainty of such a position may be estimated as the uncertainty in the slope function (first derivative) divided by the curvature (second derivative) at the peak (Cruickshank, 1949*a*), *i.e.*,

$$\sigma(x) = \sigma(\text{slope})/(\text{atomic peak 'curvature'}). \qquad (18.5.5.1)$$

However, atomic positions are affected by finite-series and peak-overlapping effects.

Hence, more generally, atomic positions may be determined by the requirement that the slope of the difference map at the position of atom r should be zero, or equivalently that the slopes at atom r of the observed and calculated electron densities should be equal. As a criterion this becomes the basis of the modified Fourier method (Cruickshank, 1952, 1959, 1999; Bricogne, 1993*b*), which, like the least-squares method, is applicable whether or not the atomic peaks are resolved and is applicable to noncentrosymmetric structures. For refinement, a set of n simultaneous linear equations are involved, analogous to the normal equations of least squares. Their right-hand sides are the slopes of the difference map at the trial atomic positions.

The diagonal elements of the matrix, for coordinate x_r of an atom with Debye B value B_r, are approximately equal to

$$\text{'curvature'} = (4\pi^2/a^2V)\left[\sum_{hkl}(m/2)h^2 f_r \exp(-B_r \sin^2\theta/\lambda^2)\right], \qquad (18.5.5.2)$$

where $m = 1$ or 2 for acentric or centric reflections. The summation is over all independent planes and their symmetry equivalents. Strictly speaking, (18.5.5.2) is a curvature only for centrosymmetric structures.

In the modified Fourier method,

$$\sigma(\text{slope}) = (2\pi/aV)\left[\sum_{hkl} h^2(\Delta|F|^2)\right]^{1/2}. \qquad (18.5.5.3)$$

This is simply an estimate of the r.m.s. uncertainty at a general position (Cruickshank & Rollett, 1953) in the slope of the difference map, *i.e.*, the r.m.s. uncertainty on the right-hand side of the modified Fourier method.

$\sigma(x)$ is then given by (18.5.5.1), using (18.5.5.3) and (18.5.5.2).

18.5.5.3. *Application of the modified Fourier method*

An extreme example of an apparently successful gross approximation to protein precision is represented by Daopin *et al.*'s (1994) treatment of two independent determinations (at 1.8 and 1.95 Å) of the structure of TGF-β2. They reported that the modified Fourier-map formulae given in Section 18.5.5.2 yielded a quite good description of the B dependence of the positional differences between the two independent determinations. However, there is a formal difficulty about this application. Equation (18.5.5.1) derives from a diffraction-data-only approach, whereas the two structures were determined from restrained refinements. Even though the *TNT* restraint parameters and weights may have been the same in both refinements, it is slightly surprising that (18.5.5.1) should have worked well.

Equation (18.5.2.1) requires the summation of various series over all (hkl) observations; such calculations are not customarily provided in protein programs. However, due to the fundamental similarities between Fourier and least-squares methods demonstrated by Cochran (1948), Cruickshank (1949*b*, 1952, 1959), and Cruickshank & Robertson (1953), closely similar estimates of the precision of individual atoms can be obtained from the reciprocal of the diagonal elements of the diffraction-data-only least-squares matrix. These elements will often have been calculated already within the protein refinement programs, but possibly never output. Such estimates could be routinely available.

Between approximations using largish blocks and those using only the reciprocals of diagonal terms, a whole variety of intermediate approximations involving some off-diagonal terms could be envisaged.

Whatever method is used to estimate uncertainties, it is essential to distinguish between *coordinate* uncertainty, *e.g.*, $\sigma(x)$, and *position* uncertainty $\sigma(r) = [\sigma^2(x) + \sigma^2(y) + \sigma^2(z)]^{1/2}$.

The remainder of this chapter discusses two rough-and-ready indicators of structure precision: the diffraction-component precision index (DPI) and Luzzati plots.

18.5.6. The diffraction-component precision index

18.5.6.1. *Statistical expectation of error dependence*

From general statistical theory, one would expect the s.u. of an atomic coordinate determined from the diffraction data alone to show dependence on four factors:

$$\sigma(x) \propto (\mathcal{R})\left[(n_{\rm atoms})/(n_{\rm obs}-n_{\rm params})\right]^{1/2}(1/s_{\rm rms}). \qquad (18.5.6.1)$$

Here, $\mathcal{R}$ is some measure of the precision of the data; $n_{\rm atoms}$ is the recognition that the information content of the data has to be shared out; $n_{\rm obs}$ is the number of independent data, but to achieve the correct number of degrees of freedom this must be reduced by $n_{\rm params}$, the number of parameters determined; and $1/s_{\rm rms}$ is a more specialized factor arising from the sensitivity $\partial|F|/\partial x$ of the data to the parameter x. Here $s_{\rm rms}$ is the r.m.s. reciprocal radius of the data. Any statistical error estimate must show some correspondence to these four factors.

18.5.6.2. *A simple error formula*

Cruickshank (1960) offered a simple order-of-magnitude formula for $\sigma(x)$ in small molecules. It was intended for use in experimental design: how many data of what precision are needed to achieve a given precision in the results? The formula, derived from a very rough estimate of a least-squares diagonal element in non-centrosymmetric space groups, was

$$\sigma(x_i) = (1/2)(N_i/p)^{1/2}[R/s_{\rm rms}] \qquad (18.5.6.2)$$

Here $p = n_{\rm obs} - n_{\rm params}$, R is the usual residual $\sum|\Delta F|/\sum|F|$ and N_i is the number of atoms of type i needed to give scattering power at $s_{\rm rms}$ equal to that of the asymmetric unit of the structure, *i.e.*, $\sum_j f_j^2 \equiv N_i f_i^2$. [The formula has also proved very useful in a systematic study of coordinate precision in the many thousands of small-molecule structure analyses recorded in the Cambridge Structural Database (Allen *et al.*, 1995*a*,*b*).]

For small molecules, the above definition of N_i allowed the treatment of different types of atom with not-too-different B's. However, it is not suitable for individual atoms in proteins where there is a very large range of B values and some atoms have B's so large as to possess negligible scattering power at $s_{\rm rms}$.

Often, as in isotropic refinement, $n_{\rm params} \simeq 4n_{\rm atoms}$, where $n_{\rm atoms}$ is the total number of atoms in the asymmetric unit. For fully anisotropic refinement, $n_{\rm params} \simeq 9n_{\rm atoms}$.

A first very rough extension of (18.5.6.2) for application in proteins to an atom with $B = B_i$ is

$$\sigma(x_i) = k(N_i/p)^{1/2}\left[g(B_i)/g(B_{\rm avg})\right]C^{-1/3}Rd_{\rm min}, \qquad (18.5.6.3)$$

where k is about 1.0, $N_i = \sum Z_j^2/Z_i^2$, $B_{\rm avg}$ is the average B for fully occupied sites and C is the fractional completeness of the data to $d_{\rm min}$. In deriving (18.5.6.3) from (18.5.6.2), $1/s_{\rm rms}$ has been replaced by $1.3d_{\rm min}$, and the factor $(1/2)(1.3) = 0.65$ has been increased to 1.0 as a measure of caution in the replacement of a full matrix by a diagonal approximation. $g(B) = 1 + a_1B + a_2B^2$ is an empirical function to allow for the dependence of $\sigma(x)$ on B. However, the results in Section 18.5.4.2 showed that the parameters a_1 and a_2 depend on the structure.

As also mentioned in Section 18.5.4.2, Sheldrick has found that the Z_i in N_i is better replaced by $Z_i^{\#}$, the scattering factor at $\sin\theta/\lambda = 0.3$ Å^{-1}. Hence, N_i may be taken as

$$N_i = (\textstyle\sum Z_j^{\#2}/Z_i^{\#2}). \qquad (18.5.6.4)$$

A useful comparison of the relative precision of different structures may be obtained by comparing atoms with the respective $B = B_{\rm avg}$ in the different structures. (18.5.6.3) then reduces to

$$\sigma(x, B_{\rm avg}) = 1.0(N_i/p)^{1/2}C^{-1/3}Rd_{\rm min}. \qquad (18.5.6.5)$$

The smaller the $d_{\rm min}$ and the R, the better the precision of the structure. If the difference between oxygen, nitrogen and carbon atoms is ignored, N_i may be taken simply as the number of fully occupied sites. For heavy atoms, (18.5.6.4) must be used for N_i.

Equation (18.5.6.5) is not to be regarded as having absolute validity. It is a quick and rough guide for the diffraction-data-only error component for an atom with Debye B equal to the $B_{\rm avg}$ for the structure. It is named the *diffraction-component precision index*, or DPI. It contains none of the restraint data.

18.5.6.3. *Extension for low-resolution structures and use of* $R_{\rm free}$

For low-resolution structures, the number of parameters may exceed the number of diffraction data. In (18.5.6.3) and (18.5.6.5), $p = n_{\rm obs} - n_{\rm params}$ is then negative, so that $\sigma(x)$ is imaginary. This difficulty can be circumvented *empirically* by replacing p with $n_{\rm obs}$ and R with $R_{\rm free}$ (Brünger, 1992). The counterpart of the DPI (18.5.6.5) is then

$$\sigma(x, B_{\rm avg}) = 1.0(N_i/n_{\rm obs})^{1/2}C^{-1/3}R_{\rm free}d_{\rm min}. \qquad (18.5.6.6)$$

Here $n_{\rm obs}$ is the number of reflections included in the refinement, not the number in the $R_{\rm free}$ set.

It may be asked: how can there be any estimate for the precision of a coordinate from the diffraction data only when there are insufficient diffraction data to determine the structure? By following the line of argument of Cruickshank's (1960) analysis, (18.5.6.6) is a rough estimate of the square root of the reciprocal of one diagonal element of the diffraction-only least-squares matrix. All the other parameters can be regarded as having been determined from a diffraction-plus-restraints matrix.

Clearly, (18.5.6.6) can also be used as a general alternative to (18.5.6.5) as a DPI, irrespective of whether the number of degrees of freedom $p = n_{\rm obs} - n_{\rm params}$ is positive or negative.

Comment. When p is positive, (18.5.6.6) would be exactly equivalent to (18.5.6.5) only if $R_{\rm free} = R[n_{\rm obs}/(n_{\rm obs} - n_{\rm params})]^{1/2}$. Tickle *et al.* (1998*b*) have shown that the expected relationship in a restrained refinement is actually

$$R_{\rm free} = R\{[n_{\rm obs} + (n_{\rm params} - h)]/[n_{\rm obs} - (n_{\rm params} - h)]\}^{1/2}, \qquad (18.5.6.7)$$

where $h = n_{\rm restraints} - \sum w_{\rm geom}(\Delta Q)^2$, the latter term, as in (18.5.3.1), being the weighted sum of the squares of the restraint residuals.

18.5.6.4. *Position error*

Often an estimate of a position error $|\Delta\mathbf{r}|$, rather than a coordinate error $|\Delta x|$, is required. In the isotropic approximation,

$$\sigma(r, B_{\rm avg}) = 3^{1/2}\sigma(x, B_{\rm avg}). \qquad (18.5.6.8)$$

Consequently, the DPI formulae for the position errors are

$$\sigma(r, B_{\rm avg}) = 3^{1/2}(N_i/p)^{1/2}C^{-1/3}Rd_{\rm min} \qquad (18.5.6.9)$$

with R and

$$\sigma(r, B_{\rm avg}) = 3^{1/2}(N_i/n_{\rm obs})^{1/2}C^{-1/3}R_{\rm free}d_{\rm min} \qquad (18.5.6.10)$$

with $R_{\rm free}$.

Table 18.5.7.1. *Comparison of full-matrix $\sigma(r, B_{\rm avg})$ with the diffraction-component precision index (DPI)*

Protein	$(N_i/p)^{1/2}$	R	$d_{\rm min}$ (Å)	DPI $\sigma(r, B_{\rm avg})$ (Å)	Full-matrix $\sigma_{\rm diff}(r, B_{\rm avg})$ (Å)	Reference
Concanavalin A	0.148	0.128	0.94	0.034	0.033	(*a*)
Immunoglobulin	0.476	0.156	1.70	0.221	0.186	(*b*)

References: (*a*) Deacon *et al.* (1997); (*b*) Usón *et al.* (1999).

18.5.7. Examples of the diffraction-component precision index

18.5.7.1. *Full-matrix comparison with the diffraction-component precision index*

The DPI (18.5.6.9) with R was offered as a quick and rough guide for the diffraction-data-only error for an atom with $B = B_{\rm avg}$. The necessary data for the comparison with the two unrestrained full-matrix inversions of Section 18.5.5 are given in Table 18.5.7.1. For concanavalin A with $B_{\rm avg} = 14.8$ Å^2, the full-matrix quadratic (18.5.4.2*b*) gives 0.033 Å for a carbon atom and the DPI gives 0.034 Å for an unspecified atom. For the immunoglobulin with $B_{\rm avg} = 26.8$ Å^2, the full-matrix quadratic (18.5.4.2*a*) gives $\sigma_{\rm diff}(r) = 0.19$ Å for a carbon atom, while the DPI gives 0.22 Å.

For these two structures, the simple DPI formula compares surprisingly well with the unrestrained full-matrix calculations at $B_{\rm avg}$.

For the restrained full-matrix calculations on concanavalin A, the quadratic (18.5.4.2*c*) with $B = B_{\rm avg}$ gives $\sigma_{\rm res}(r) = 0.028$ Å for a carbon atom, which is only 15% smaller than the unrestrained 0.033 Å. This small decrease matches the discussion of $\sigma_{\rm res}(r)$ and $\sigma_{\rm diff}(r)$ in Section 18.5.4.1 following equation (18.5.4.1). But that discussion also indicates that for the immunoglobulin, the restrained $\sigma_{\rm res}(r, B_{\rm avg})$, which was not computed, will be proportionaly much lower than the unrestrained value of $\sigma_{\rm diff}(r, B_{\rm avg}) = 0.19$ Å, since the restraints are relatively more important in the immunoglobulin.

18.5.7.2. *Further examples of the DPI using R*

Table 18.5.7.2 shows a range of examples of the application of the DPI (18.5.6.9) using R to proteins of differing precision, starting with the smallest $d_{\rm min}$. In all the examples, N_i has been set equal to $n_{\rm atoms}$, the total number of atoms. The ninth and tenth columns show $\langle \Delta r \rangle$ values derived from Luzzati (1952) and Read (1986) plots described later in Section 18.5.8.

The first entry is for crambin at 0.83 Å resolution and 130 K (Stec *et al.*, 1995). Their results were obtained from an unrestrained full-matrix anisotropic refinement. Inversion of the full matrix gave s.u.'s $\sigma_{\rm diff}(x) = 0.0096$ Å for backbone atoms, 0.0168 Å for side-chain atoms and 0.0409 Å for solvent atoms, with an average for all atoms of 0.022 Å. The DPI $\sigma(r, B_{\rm avg}) = 0.021$ Å corresponds to $\sigma(x) = 0.012$ Å, which is satisfactorily intermediate between the full-matrix values for the backbone and side-chain atoms.

Sevcik *et al.* (1996) carried out restrained anisotropic full-matrix refinements on data from two slightly different crystals of ribonuclease Sa, with $d_{\rm min}$ of 1.15 and 1.20 Å. They inverted full-matrix blocks containing parameters of 20 residues to estimate coordinate errors. The overall r.m.s. coordinate error for protein atoms is given as 0.03 Å, and for all atoms (including waters and ligands) as 0.07 Å for MGMP and 0.05 Å for MSA. The DPI gives $\sigma(r, B_{\rm avg}) = 0.05$ Å for both structures.

The next entries concern the two lower-resolution (1.8 and 1.95 Å) studies of TGF-β2 (Daopin *et al.*, 1994). The DPI gives $\sigma(r) = 0.16$ Å for 1TGI and 0.24 Å for 1TGF. This indicates an r.m.s. position difference between the structures for atoms with $B_i = B_{\rm avg}$ of $(0.16^2 + 0.24^2)^{1/2} = 0.29$ Å. Daopin *et al.* reported the differences between the two determinations, omitting poor parts, as $\langle \Delta r \rangle_{\rm rms} = 0.15$ Å (main chain) and 0.29 Å (all atoms).

Human diferric lactoferrin (Haridas *et al.*, 1995) is an example of a large protein at the lower resolution of 2.2 Å, with a high value of $(N_i/p)^{1/2}$, leading to $\sigma(r, B_{\rm avg}) = 0.43$ Å.

Three crystal forms of thaumatin were studied by Ko *et al.* (1994). The orthorhombic and tetragonal forms diffracted to 1.75 Å, but the monoclinic $C2$ form diffracted only to 2.6 Å. The structures with 1552 protein atoms were successfully refined with restraints by *XPLOR* and *TNT*. For the monoclinic form, the number of parameters exceeds the number of diffraction observations, so (N_i/p) is negative and no estimate by (18.5.6.9) of the diffraction-data-only error is possible. The DPI (18.5.6.9) gives 0.17 and 0.16 Å for the orthorhombic and tetragonal forms, respectively.

18.5.7.3. *Examples of the DPI using $R_{\rm free}$*

As in the case of monoclinic thaumatin, for low-resolution structures the number of parameters may exceed the number of diffraction data. To circumvent this difficulty, it was proposed in Section 18.5.6.3 to replace $p = n_{\rm obs} - n_{\rm params}$ by $n_{\rm obs}$ and R by $R_{\rm free}$ in a revised formula (18.5.6.10) for the DPI. Table 18.5.7.3 shows examples for some structures for which both R and $R_{\rm free}$ were

Table 18.5.7.2. *Examples of diffraction-component precision indices (DPIs)*

Protein	N_i	$n_{\rm obs}$	$(N_i/p)^{1/2}$	$C^{-1/3}$	R	$d_{\rm min}$ (Å)	DPI $\sigma(r, B_{\rm avg})$ (Å)	Luzzati $\langle \Delta r \rangle$ (Å)	Read $\langle \Delta r \rangle$ (Å)	Reference
Crambin	447	23759	0.150	1.074	0.090	0.83	0.021	0.055		(*a*)
Ribonuclease MGMP	1958	62845	0.208	1.046	0.109	1.15	0.047		0.08	(*b*)
Ribonuclease MSA	1832	60670	0.204	1.016	0.106	1.20	0.045		0.05	(*b*)
TGF-β2 1TGI	948	∼14000	0.305	∼1.0	0.173	1.80	0.16	0.21	0.18	(*c*)
TGF-β2 1TFG	974	∼11000	0.370	∼1.0	0.188	1.95	0.24	0.23		(*c*)
Lactoferrin	5907	39113	0.618	1.036	0.179	2.20	0.43	0.25–0.30	0.35	(*d*)
Thaumatin C2	1552	4622	*	1.10	0.184	2.60	—	0.25		(*e*)

References: (*a*) Stec *et al.* (1995); (*b*) Sevcik *et al.* (1996); (*c*) Daopin *et al.* (1994); (*d*) Haridas *et al.* (1995); (*e*) Ko *et al.* (1994).
* (N_i/p) negative.

Table 18.5.7.3. *Comparison of DPIs using R and R_{free}*

The second row for each protein contains values appropriate to the DPI equation (18.5.6.10) using R_{free}.

Protein	N_i	n_{obs}	$(N_i/p)^{1/2}$, $(N_i/n_{obs})^{1/2}$	$C^{-1/3}$	R, R_{free}	d_{min} (Å)	DPI $\sigma(r, B_{avg})$ (Å)	Luzzati $\langle \Delta r \rangle$ (Å)	Read $\langle \Delta r \rangle$ (Å)	Reference
Concanavalin A	2130	116712	0.148 0.135	1.099	0.128 0.148	0.94	0.034 0.036	0.06		(*a*)
γB-Crystallin	1708	26151	0.297 0.256	1.032	0.180 0.204	1.49	0.14 0.14	0.16	0.12	(*b*)
βB2-Crystallin	1558	18583	0.356 0.290	∼1.032	0.184 0.200	2.10	0.25 0.22	0.21	0.17	(*b*)
Ribonuclease A with RI	4416	18859	1.922 0.484	1.145	0.194 0.286	2.50	1.85 0.69	0.32	0.57	(*c*)
Fab HyHEL-5 with HEWL	4333	11754	* 0.607	1.111	0.196 0.288	2.65	— 0.69	0.30		(*d*)

References: (*a*) Deacon *et al*. (1997); (*b*) Tickle *et al*. (1998*a*); (*c*) Kobe & Deisenhofer (1995); (*d*) Cohen *et al*. (1996).
* (N_i/p) negative.

available. The second row for each protein shows the alternative values for $(N_i/n_{obs})^{1/2}$, R_{free} and the DPI $\sigma(r, B_{avg})$ from (18.5.6.10).

For the structures with $d_{min} \leq 2.0$ Å, the DPI is much the same whether it is based on R or R_{free}.

Tickle *et al.* (1998*a*) have made full-matrix error estimates for isotropic restrained refinements of γB-crystallin with $d_{min} = 1.49$ Å and of βB2-crystallin with $d_{min} = 2.10$ Å. The DPI $\sigma(r, B_{avg})$ calculated for the two structures is 0.14 and 0.25 Å with R in (18.5.6.9), and 0.14 and 0.22 Å with R_{free} in (18.5.6.10). The full-matrix weighted averages of $\sigma_{res}(r)$ for all protein atoms were 0.10 and 0.15 Å, for only main-chain atoms 0.05 and 0.08 Å, for side-chain atoms 0.14 and 0.20 Å, and for water oxygens 0.27 and 0.35 Å. Again, the DPI gives reasonable overall indices for the quality of the structures.

For the complex of bovine ribonuclease A and porcine ribonuclease inhibitor (Kobe & Deisenhofer, 1995) with $d_{min} = 2.50$ Å, the number of reflections is only just larger than the number of parameters, so that $(N_i/p)^{1/2} = 1.922$ is very large, and the DPI with R gives an unrealistic 1.85 Å. With R_{free}, $\sigma(r, B_{avg}) = 0.69$ Å.

The HyHEL-5–lysozyme complex (Cohen *et al.*, 1996) had $d_{min} = 2.65$ Å. Here the number of reflections is less than the number of parameters, but the R_{free} formula gives $\sigma(r, B_{avg}) = 0.69$ Å.

18.5.7.4. *Comments on the diffraction-component precision index*

The DPI (18.5.6.9) or (18.5.6.10) provides a very simple formula for $\sigma(r, B_{avg})$. It is based on a very rough approximation to a diagonal element of the diffraction-data-only matrix. Using a diagonal element is a reasonable approximation for atomic resolution structures, but for low-resolution structures there will be significant off-diagonal terms between overlapping atoms. The effect can be simulated in the two-atom protein model of Section 18.5.3.2 by introducing positive off-diagonal elements into the diffraction-data matrix (18.5.3.3). As expected, $\sigma^2_{diff}(x_i)$ is increased. So the DPI will be an underestimate of the diffraction component in low-resolution structures.

However, the true restrained variance $\sigma^2_{res}(x_i)$ in the new counterpart of (18.5.3.12) remains less than the diagonal diffraction result (18.5.3.11) $\sigma^2_{diff}(x_i) = 1/a$. Thus for low-resolution structures, the DPI should be an overestimate of the true precision given by a restrained full-matrix calculation (where the restraints act to hold the overlapping atoms apart). This is confirmed by the results for the 2.1 Å study of βB2-crystallin (Tickle *et al.*, 1998*a*) discussed in Section 18.5.7.3 and Table 18.5.7.3. The restrained full-matrix average for all protein atoms was $\sigma_{res}(r) = 0.15$ Å, compared with the DPI 0.25 Å (on R) or 0.22 Å (on R_{free}). The ratio between the unrestrained DPI and the restrained full-matrix average is consistent with a view of a low-resolution protein as a chain of effectively rigid peptide groups. The ratio no doubt gets much worse for resolutions of 3 Å and above.

The DPI estimate of $\sigma(r, B_{avg})$ is given by a formula of 'back-of-an-envelope' simplicity. B_{avg} is taken to be the average B for fully occupied sites, but the weights implicit in the averaging are not well defined in the derivation of the DPI. Thus the DPI should perhaps be regarded as simply offering an estimate of a typical $\sigma_{diff}(r)$ for a carbon or nitrogen atom with a mid-range B. From the evidence of the tables in this section, except at low resolution, it seems to give a useful overall indication of protein precision, even in restrained refinements.

The DPI evidently provides a method for the comparative ranking of different structure determinations. In this regard it is a complement to the general use of d_{min} as a quick indicator of possible structural quality.

Note that (18.5.6.3) and (18.5.6.4) offer scope for making individual error estimates for atoms of different B and Z.

18.5.8. Luzzati plots

18.5.8.1. *Luzzati's theory*

Luzzati (1952) provided a theory for estimating, at any stage of a refinement, the average positional shifts which would be needed in an idealized refinement to reach $R = 0$. He did not provide a theory for estimating positional errors at the end of a normal refinement.

(1) His theory assumed that the F_{obs} had no errors, and that the

Table 18.5.8.1. $R = \langle|\Delta F|\rangle/\langle|F|\rangle$ *as a function of* $s\langle\Delta r\rangle$ *in the Luzzati model for three-dimensional noncentrosymmetric structures* ($s = 2\sin\theta/\lambda$)

$s\langle\Delta r\rangle$	R	$s\langle\Delta r\rangle$	R
0.00	0.000	0.10	0.237
0.01	0.025	0.12	0.281
0.02	0.050	0.14	0.319
0.03	0.074	0.16	0.353
0.04	0.098	0.18	0.385
0.05	0.122	0.20	0.414
0.06	0.145	0.25	0.474
0.07	0.168	0.30	0.518
0.08	0.191	0.35	0.548
0.09	0.214	∞	0.586

$F_{\rm calc}$ model (scattering factors, thermal parameters *etc.*) was perfect, apart from coordinate errors.

(2) The Gaussian probability distribution for these coordinate errors was assumed to be the *same for all atoms*, independent of Z or B.

(3) The atoms were not required to be identical, and the position errors were not required to be small.

Luzzati gave families of curves for R *versus* $2\sin\theta/\lambda$ for varying average positional errors $\langle\Delta r\rangle$ for both centrosymmetric and noncentrosymmetric structures. The curves do not depend on the number N of atoms in the cell. They all rise from $R = 0$ at $2\sin\theta/\lambda = 0$ to the Wilson (1950) values 0.828 and 0.586 for random structures at high $2\sin\theta/\lambda$. Table 18.5.8.1 gives $R = \langle|\Delta F|\rangle/\langle|F|\rangle$ as a function of $s\langle\Delta r\rangle$ for three-dimensional noncentrosymmetric structures.

In a footnote (p. 807), Luzzati suggested that at the end of a normal refinement (with R nonzero due to experimental and model errors, *etc.*), the curves would indicate an upper limit for $\langle\Delta r\rangle$. He noted that typical small-molecule $\sigma(r)$'s of 0.01–0.02 Å, if used as $\langle\Delta r\rangle$ in the plots, would give much smaller R's than are found at the end of a refinement.

As examples, the Luzzati plots for the two structures of TGF-β2 are shown in Fig. 18.5.8.1. Daopin *et al.* (1994) inferred average $\langle\Delta r\rangle$'s around 0.21 Å for 1TGI and 0.23 Å for 1TGF.

Of the three Luzzati assumptions summarized above, the most attractive is the third, which does not require the atoms to be identical nor the position errors to be small. For proteins, there are very obvious difficulties with assumption (2). Errors do depend very strongly on Z and B. In the high-angle data shells, atoms with large B's contribute neither to ΔF nor to $|F|$, and so have no effect on R in these shells. In their important paper on protein accuracy, Chambers & Stroud (1979) said 'the [Luzzati] estimate derived from reflections in this range applies mainly to [the] best determined atoms.'

Thus a Luzzati plot seems to allow a cautious upper-limit statement about the precision of the best parts of a structure, but it gives little indication for the poor parts.

One reason for the past popularity of Luzzati plots has been that the R values for the middle and outer shells of a structure often roughly follow a Luzzati curve. Evidently, the effective average $\langle\Delta r\rangle$ for the structure must be decreasing as $2\sin\theta/\lambda$ increases, since atoms of high B are ceasing to contribute, whereas the proportionate experimental errors must be increasing. This also suggests that the upper limit for $\langle\Delta r\rangle$ for the low-B atoms could be estimated from the lowest Luzzati theoretical curve touched by the experimental R plot. Thus in Fig. 18.5.8.1 the upper limits for the low-B atoms could be taken as 0.18 and 0.21 Å, rather than the 0.21 and 0.23 Å chosen by Daopin *et al.*

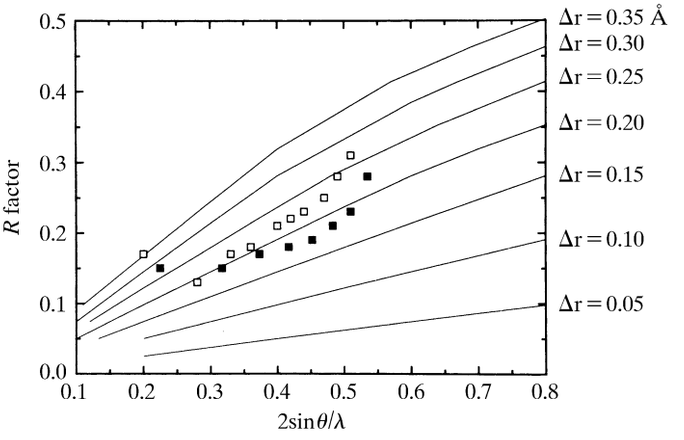


Fig. 18.5.8.1. Luzzati plots showing the refined R factor as a function of resolution for 1TGI (solid squares) and 1TGF (open squares) (Daopin *et al.*, 1994).

From the introduction of $R_{\rm free}$ by Brünger (1992) and the discussion of $R_{\rm free}$ by Tickle *et al.* (1998*b*), it can be seen that Luzzati plots should be based on a residual more akin to $R_{\rm free}$ than R in order to avoid bias from the fitting of data.

The mean positional error $\langle\Delta r\rangle$ of atoms can also be estimated from the σ_A plots of Read (1986, 1990). This method arose from Read's analysis of improved Fourier coefficients for maps using phases from partial structures with errors. It is preferable in several respects to the Luzzati method, but like the Luzzati method it assumes that the coordinate distribution is the same for all atoms. Luzzati and/or Read estimates of $\langle\Delta r\rangle$ are available for some of the structures in Tables 18.5.7.2 and 18.5.7.3. Often, the two estimates are not greatly different.

18.5.8.2. *Statistical reinterpretation of Luzzati plots*

Luzzati plots are fundamentally different from other statistical estimates of error. The Luzzati theory applies to an idealized incomplete refinement and estimates the average shifts needed to reach $R = 0$. In the least-squares method, the equations for shifts are quite different from the equations for estimating variances in a converged refinement. However, Luzzati-style plots of R *versus* $2\sin\theta/\lambda$ can be reinterpreted to give statistically based estimates of $\sigma(x)$.

During Cruickshank's (1960) derivation of the approximate equation (18.5.6.2) for $\sigma(x)$ in diagonal least squares, he reached an intermediate equation

$$\sigma^2(x) = N_i \Big/ \left[4\sum_{\rm obs}(s^2/R^2)\right]. \qquad (18.5.8.1)$$

He then assumed R to be independent of s ($= 2\sin\theta/\lambda$) and took R outside the summation to reach (18.5.6.2) above.

Luzzati (1952) calculated the acentric residual R as a function of $\langle\Delta r\rangle$, the average radial error of the atomic positions. His analysis shows that R is a linear function of s and $\langle\Delta r\rangle$ for a substantial range of $s\langle\Delta r\rangle$, with

$$R(s,\ \langle\Delta r\rangle) = (2\pi)^{1/2} s\langle\Delta r\rangle. \qquad (18.5.8.2)$$

The theoretical Luzzati plots of R are nearly linear for small-to-medium $s = 2\sin\theta/\lambda$ (see Fig. 18.5.8.1). If we substitute this R in the least-squares estimate (18.5.8.1) and use the three-dimensional-Gaussian relation $\sigma(r) = 1.085\langle\Delta r\rangle$, some manipulation (Cruickshank, 1999) along the lines of Section 18.5.6 eventually yields a statistically based formula,

$$\sigma_{\mathrm{LS,\,Luzz}}(r) = 1.33(N_i/p)^{1/2}[R(s_m)/s_m], \qquad (18.5.8.3)$$

where $R(s_m)$ is the value of R at some value of $s = s_m$ on the selected Luzzati curve. Equation (18.5.8.3) provides a means of making a very rough statistical estimate of error for an atom with $B = B_{\mathrm{avg}}$ (the average B for fully occupied sites) from a plot of R *versus* $2\sin\theta/\lambda$.

The corresponding equation involving R_{free} is

$$\sigma_{\mathrm{LS,\,Luzz}}(r) = 1.33(N_i/n_{\mathrm{obs}})^{1/2}[R_{\mathrm{free}}(s_m)/s_m]. \qquad (18.5.8.4)$$

18.5.8.3. *Comments on Luzzati plots*

Protein structures always show a great range of B values. The Luzzati theory effectively assumes that all atoms have the same B. Nonetheless, the Luzzati method applied to high-angle data shells does provide an upper limit for $\langle\Delta r\rangle$ for the atoms with low B. It is an upper limit since experimental errors and model imperfections are not allowed for in the theory.

Low-resolution structures can be determined validly by using restraints, even though the number of diffraction observations is less than the number of atomic coordinates. The Luzzati method, based preferably on R_{free}, can be applied to the atoms of low B in such structures. As the number of observations increases, and the resolution improves, the Luzzati $\langle\Delta r\rangle$ increasingly overestimates the true $\sigma(r)$ of the low-B atoms.

In the use of Luzzati plots, the method of refinement, and its degree of convergence, is irrelevant. A Luzzati plot is a statement for the low-B atoms about the maximum errors associated with a given structure, whether converged or not.

References

18.1

Adams, P. D., Pannu, N. S., Read, R. J. & Brunger, A. T. (1999). *Extending the limits of molecular replacement through combined simulated annealing and maximum-likelihood refinement*. *Acta Cryst.* D**55**, 181–190.

Booth, A. D. (1946*a*). *A differential Fourier method for refining atomic parameters in crystal structure analysis*. *Trans. Faraday Soc.* **42**, 444–448.

Booth, A. D. (1946*b*). *The simultaneous differential refinement of co-ordinates and phase angles in X-ray Fourier synthesis*. *Trans. Faraday Soc.* **42**, 617–619.

Bricogne, G. (1997). *Bayesian statistical viewpoint on structure determination: basic concepts and examples*. *Methods Enzymol.* **276**, 361–423.

Brünger, A. T. (1992). *Free R-value – a novel statistical quantity for assessing the accuracy of crystal structures*. *Nature (London)*, **355**, 472–475.

Busing, W. R., Martin, K. O. & Levy, H. A. (1962). *ORFLS.* Report ORNL-TM-305. Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA.

Diamond, R. (1971). *A real-space refinement procedure for proteins*. *Acta Cryst.* A**27**, 436–452.

Hendrickson, W. A. (1985). *Stereochemically restrained refinement of macromolecular structures*. *Methods Enzymol.* **115**, 252–270.

Hughes, E. W. (1941). *The crystal structure of melamine. J Am. Chem. Soc.* **63**, 1737–1752.

International Tables for Crystallography (1999). Vol. C. *Mathematical, physical and chemical tables*, edited by A. J. C. Wilson & E. Prince. Dordrecht: Kluwer Academic Publishers.

International Tables for Crystallography (2001). Vol. B. *Reciprocal space*, edited by U. Shmueli. Dordrecht: Kluwer Academic Publishers.

Kleywegt, G. J. (2000). *Validation of protein crystal structures*. *Acta Cryst.* D**56**, 249–265.

Kleywegt, G. J. & Jones, T. A. (1997). *Model building and refinement practice. Methods Enzymol.* **277**, 208–230.

Konnert, J. H. (1976). *A restrained-parameter structure-factor least-squares refinement procedure for large asymmetric units*. *Acta Cryst.* A**32**, 614–617.

Konnert, J. H. & Hendrickson, W. A. (1980). *A restrained-parameter thermal-factor refinement procedure*. *Acta Cryst.* A**36**, 344–350.

Kuriyan, J., Brünger, A. T., Karplus, M. & Hendrickson, W. A. (1989). *X-ray refinement of protein structures by simulated annealing: test of the method on myohemerythrin*. *Acta Cryst.* A**45**, 396–409.

Lamzin, V. S. & Wilson, K. S. (1997). *Automated refinement for protein crystallography*. In *Macromolecular crystallography*, Part B, edited by C. C. & R. Sweet, 269–305. San Diego: Academic Press.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *PROCHECK: a program to check the stereochemical quality of protein structures*. *J. Appl. Cryst.* **26**, 283–291.

McRee, D. E. (1993). *Practical protein crystallography*, p. 386. San Diego: Academic Press.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Refinement of macromolecular structures by the maximum-likelihood method*. *Acta Cryst.* D**53**, 240–253.

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Efficient anisotropic refinement of macromolecular structures using FFT*. *Acta Cryst.* D**55**, 247–255.

Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. (1998). *Incorporation of prior phase information strengthens maximum-likelihood structure refinement*. *Acta Cryst.* D**54**, 1285–1294.

Pannu, N. S. & Read, R. J. (1996). *Improved structure refinement through maximum likelihood*. *Acta Cryst.* A**52**, 659–668.

Prince, E. (1994). *Mathematical techniques in crystallography and materials science*. 2nd ed. Berlin: Springer-Verlag.

Rice, L. M. & Brünger, A. T. (1994). *Torsion-angle dynamics – reduced variable conformational sampling enhances crystallographic structure refinement*. *Proteins Struct. Funct. Genet.* **19**, 277–290.

Sheldrick, G. M. (1993). *SHELXL93. Program for the refinement of crystal structures*. University of Göttingen, Germany.

Tronrud, D. E. (1992). *Conjugate-direction minimization: an improved method for the refinement of macromolecules*. *Acta Cryst.* **48**, 912–916.

Tronrud, D. E. (1996). *Knowledge-based B-factor restraints for the refinement of proteins*. *J. Appl. Cryst.* **29**, 100–104.

Tronrud, D. E. (1997). *The TNT refinement package*. *Methods Enzymol.* **277**, 306–318.

Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *An efficient general-purpose least-squares refinement program for macromolecular structures*. *Acta Cryst.* A**43**, 489–501.

Vriend, G. (1990). *WHAT IF: a molecular modeling and drug design program*. *J. Mol. Graphics*, **8**, 52–56.

Watenpaugh, K. D., Sieker, L. C., Herriott, J. R. & Jensen, L. H. (1972). *The structure of a non-heme iron protein: rubredoxin at 1.5 Å resolution*. *Cold Spring Harbor Symp. Quant. Biol.* **36**, 359–367.

Watenpaugh, K. D., Sieker, L. C., Herriott, J. R. & Jensen, L. H. (1973). *Refinement of the model of a protein: rubredoxin at 1.5 Å resolution*. *Acta Cryst.* B**29**, 943–956.

18.2

Abramowitz, M. & Stegun, I. (1968). *Handbook of mathematical functions. Applied mathematics series*, Vol. 55, p. 896. New York: Dover Publications.

Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. (1997). *Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.

Adams, P. D., Pannu, N. S., Read, R. J. & Brunger, A. T. (1999). *Extending the limits of molecular replacement through combined simulated annealing and maximum-likelihood refinement. Acta Cryst*. D**55**, 181–190.

Allen, F. H., Kennard, O. & Taylor, R. (1983). *Systematic analysis of structural data as a research technique in organic chemistry. Acc. Chem. Res*. **16**, 146–153.

Bae, D.-S. & Haug, E. J. (1987). *A recursive formulation for constrained mechanical system dynamics: Part I. Open loop systems. Mech. Struct. Mach*. **15**, 359–382.

Bae, D.-S. & Haug, E. J. (1988). *A recursive formulation for constrained mechanical system dynamics: Part II. Closed loop systems. Mech. Struct. Mach*. **15**, 481–506.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). *Molecular dynamics with coupling to an external bath. J. Chem. Phys*. **81**, 3684–3690.

Bricogne, G. (1991). *A multisolution method of phase determination by combined maximization of entropy and likelihood. III. Extension to powder diffraction data. Acta Cryst*. A**47**, 803–829.

Brünger, A. T. (1988). *Crystallographic refinement by simulated annealing: application to a 2.8 Å resolution structure of aspartate aminotransferase. J. Mol. Biol*. **203**, 803–816.

Brünger, A. T. (1992). *The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature (London)*, **355**, 472–474.

Brünger, A. T. (1997). *Free R value: cross-validation in crystallography. Methods Enzymol*. **277**, 366–396.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Cryst*. D**54**, 905–921.

Brünger, A. T., Krukowski, A. & Erickson, J. W. (1990). *Slow-cooling protocols for crystallographic refinement by simulated annealing. Acta Cryst*. A**46**, 585–593.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Crystallographic R factor refinement by molecular dynamics. Science*, **235**, 458–460.

Burling, F. T. & Brünger, A. T. (1994). *Thermal motion and conformational disorder in protein crystal structures: comparison of multi-conformer and time-averaging models. Isr. J. Chem*. **34**, 165–175.

Burling, F. T., Weis, W. I., Flaherty, K. M. & Brünger, A. T. (1996). *Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. Science*, **271**, 72–77.

Diamond, R. (1971). *A real-space refinement procedure for proteins. Acta Cryst*. A**27**, 436–452.

Engh, R. A. & Huber, R. (1991). *Accurate bond and angle parameters for X-ray structure refinement. Acta Cryst*. A**47**, 392–400.

Fujinaga, M., Gros, P. & van Gunsteren, W. F. (1989). *Testing the method of crystallographic refinement using molecular dynamics. J. Appl. Cryst*. **22**, 1–8.

Goldstein, H. (1980). *Classical mechanics*. 2nd ed. Reading, Massachusetts: Addison-Wesley.

Gros, P., van Gunsteren, W. F. & Hol, W. G. J. (1990). *Inclusion of thermal motion in crystallographic structures by restrained molecular dynamics. Science*, **249**, 1149–1152.

Hendrickson, W. A. (1985). *Stereochemically restrained refinement of macromolecular structures. Methods Enzymol*. **115**, 252–270.

Hoppe, W. (1957). *Die 'Faltmolekülmethode' – eine neue Methode zur Bestimmung der Kristallstruktur bei ganz oder teilweise bekannter Molekülstruktur. Acta Cryst*. **10**, 750–751.

Hsu, I. N., Delbaere, L. T. J., James, M. N. G. & Hoffman, T. (1977). *Penicillopepsin from Penicillium janthinellum crystal structure at 2.8 Å and sequence homology with porcine pepsin. Nature (London)*, **266**, 140–145.

Jain, A., Vaidehi, N. & Rodriguez, G. (1993). *A fast recursive algorithm for molecular dynamics simulation. J. Comput. Phys*. **106**, 258–268.

Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. Jr (1983). *Optimization by simulated annealing. Science*, **220**, 671–680.

Kleywegt, G. J. & Brünger, A. T. (1996). *Cross-validation in crystallography: practice and applications. Structure*, **4**, 897–904.

Kuriyan, J., Brünger, A. T., Karplus, M. & Hendrickson, W. A. (1989). *X-ray refinement of protein structures by simulated annealing: test of the method on myohemerythrin. Acta Cryst*. A**45**, 396–409.

Kuriyan, J., Ösapay, K., Burley, S. K., Brünger, A. T., Hendrickson, W. A. & Karplus, M. (1991). *Exploration of disorder in protein structures by X-ray restrained molecular dynamics. Proteins*, **10**, 340–358.

Kuriyan, J., Petsko, G. A., Levy, R. M. & Karplus, M. (1986). *Effect of anisotropy and anharmonicity on protein crystallographic refinement. J. Mol. Biol*. **190**, 227–254.

Laarhoven, P. J. M. & Aarts, E. H. L. (1987). Editors. *Simulated annealing: theory and applications*. Dordrecht: D. Reidel Publishing Company.

Luzzati, V. (1952). *Traitement statistique des erreurs dans la determination des structures cristallines. Acta Cryst*. **5**, 802–810.

Metropolis, N., Rosenbluth, M., Rosenbluth, A., Teller, A. & Teller, E. (1953). *Equation of state calculations by fast computing machines. J. Chem. Phys*. **21**, 1087–1092.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst*. D**53**, 240–255.

Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. (1998). *Incorporation of prior phase information strengthens maximum-likelihood structure refinement. Acta Cryst*. D**54**, 1285–1294.

Pannu, N. S. & Read, R. J. (1996). *Improved structure refinement through maximum likelihood. Acta Cryst*. A**52**, 659–668.

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*. D**52**, 57–64.

Pearlman, D. A. & Kim, S.-H. (1990). *Atomic charges for DNA constituents derived from single-crystal X-ray diffraction data. J. Mol. Biol*. **211**, 171–187.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1986). Editors. *Numerical recipes*, pp. 498–546. Cambridge University Press.

Read, R. J. (1986). *Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Cryst*. A**42**, 140–149.

Read, R. J. (1990). *Structure-factor probabilities for related structures. Acta Cryst*. A**46**, 900–912.

Read, R. J. (1997). *Model phases: probabilities and bias. Methods Enzymol*. **278**, 110–128.

Rice, L. M. & Brünger, A. T. (1994). *Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. Proteins Struct. Funct. Genet*. **19**, 277–290.

Rice, L. M., Shamoo, Y. & Brünger, A. T. (1998). *Phase improvement by multi-start simulated annealing refinement and structure-factor averaging. J. Appl. Cryst.* **31**, 798–805.

Rossmann, M. G. & Blow, D. M. (1962). *The detection of sub-units within the crystallographic asymmetric unit. Acta Cryst*. **15**, 24–51.

Silva, A. M. & Rossmann, M. G. (1985). *The refinement of southern bean mosaic virus in reciprocal space. Acta Cryst*. B**41**, 147–157.

Sim, G. A. (1959). *The distribution of phase angles for structures containing heavy atoms. II. A modification of the normal heavy-atom method for non-centrosymmetrical structures. Acta Cryst*. **12**, 813–815.

Srinivasan, R. (1966). *Weighting functions for use in the early stages of structure analysis when a part of the structure is known. Acta Cryst*. **20**, 143–144.

Suguna, K., Bott, R. R., Padlan, E. A., Subramanian, E., Sheriff, S., Cohen, G. H. & Davies, D. R. (1987). *Structure and refinement at*

18.2 (*cont.*)

1.8 Å resolution of the aspartic proteinase from Rhizopus chinensis. J. Mol. Biol. **196**, 877–900.
Verlet, L. (1967). *Computer experiments on classical fluids. I. Thermodynamical properties of Lennard–Jones molecules. Phys. Rev*. **159**, 98–105.

18.3

Bansal, M. & Ananthanarayanan, V. S. (1988). *The role of hydroxyproline in collagen folding: conformational energy calculations on oligopeptides containing proline and hydroxyproline. Biopolymers*, **27**, 299–312.
Bricogne, G. (1993). *Direct phase determination by entropy maximization and likelihood ranking: status report and perspectives. Acta Cryst.* D**49**, 37–60.
Brünger, A. T. (1993). *Assessment of phase accuracy by cross validation: the free R value. Methods and applications. Acta Cryst.* D**49**, 24–36.
Bürgi, H.-B. & Dubler-Steudle, K. C. (1988). *Empirical potential energy surfaces relating structure and activation energy. 2. Determination of transition-state structure for the spontaneous hydrolysis of axial tetrahydropyranyl acetals. J. Am. Chem. Soc.* **110**, 7291–7299.
Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1997). *The benefits of atomic resolution. Curr. Opin. Struct. Biol.* **7**, 681–688.
Engh, R. A. & Huber, R. (1991). *Accurate bond and angle parameters for X-ray protein structure refinement. Acta Cryst.* A**47**, 392–400.
Hooft, R. W. W., Sander, C. & Vriend, G. (1997). *Objectively judging the quality of a protein structure from a Ramachandran plot. Comput. Appl. Biosci.* **13**, 425–430.
Kidera, A., Matsushima, M. & Go, N. (1994). *Dynamic structure of human lysozyme derived from X-ray crystallography: normal mode refinement. Biophys. Chem.* **50**, 25–31.
Kleywegt, G. J. & Jones, T. A. (1998). *Databases in protein crystallography. Acta Cryst.* D**54**, 1119–1131.
Lamzin, V. S., Dauter, Z. & Wilson, K. S. (1995). *Dictionary of protein stereochemistry. J. Appl. Cryst.* **28**, 338–340.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst.* **26**, 283–291.
Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). *Main-chain bond lengths and bond angles in protein structures. J. Mol. Biol.* **231**, 1049–1067.
Longhi, S., Czjzek, M. & Cambillau, C. (1998). *Messages from ultrahigh resolution crystal structures. Curr. Opin. Struct. Biol.* **8**, 730–737.
Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). *The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. Acta Cryst.* B**39**, 480–490.
Pannu, N. S. & Read, R. J. (1996). *Improved structure refinement through maximum likelihood. Acta Cryst.* A**52**, 659–668.
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.* D**52**, 57–64.
Priestle, J. P. (1994). *Stereochemical dictionaries for protein structure refinement and model building. Structure*, **2**, 911–913.
Sippl, M. J. (1995). *Knowledge-based potentials for proteins. Curr. Opin. Struct. Biol.* **5**, 229–235.
Stewart, D. E., Sarkar, A. & Wampler, J. E. (1990). *Occurrence and role of cis peptide bonds in protein structures. J. Mol. Biol.* **214**, 253–260.
Vlassi, M., Dauter, Z., Wilson, K. S. & Kokkinidis, M. (1998). *Structural parameters for proteins derived from the atomic resolution (1.09 Å) structure of a designed variant of the ColE1 ROP protein. Acta Cryst.* D**54**, 1245–1260.
Wilson, K. S., Butterworth, S., Dauter, Z., Lamzin, V. S., Walsh, M., Wodak, S., Pontius, J., Richelle, J., Vaguine, A., Sander, C., Hooft, R. W. W., Vriend, G., Thornton, J. M., Laskowski, R. A., MacArthur, M. W., Dodson, E. J., Murshudov, G., Oldfield, T. J., Kaptein, R. & Rullmann, J. A. C. (1998). *Who checks the checkers – four validation tools applied to eight atomic resolution structures. J. Mol. Biol.* **276**, 417–436.

18.4

Agarwal, R. C. (1978). *A new least-squares refinement technique based on the fast Fourier transform algorithm. Acta Cryst.* A**34**, 791–809.
Allen, F. H., Bellard, S., Brice, M. D., Cartwright, B. A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B. G., Kennard, O., Motherwell, W. D. S., Rodgers, J. R. & Watson, D. G. (1979). *The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information. Acta Cryst.* B**35**, 2331–2339.
Andersson, K. M. & Hovmöller, S. (1998). *The average atomic volume and density of proteins. Z. Kristallogr.* **213**, 369–373.
Bacchi, A., Lamzin, V. S. & Wilson, K. S. (1996). *A self-validation technique for protein structure refinement: the extended Hamilton test. Acta Cryst.* D**52**, 641–646.
Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. E., Brice, M. D., Rogers, J. K., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol.* **112**, 535–542.
Blessing, R. H. (1997). *LOCSCL: a program to statistically optimize local scaling of single-isomorphous-replacement and single-wavelength-anomalous-scattering data. J. Appl. Cryst.* **30**, 176–177.
Box, G. E. P. & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Reading, Massachusetts/California/London: Addison-Wesley.
Bricogne, G. (1997). *Maximum entropy methods and the Bayesian programme*. In *Proceedings of the CCP4 study weekend. Recent advances in phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 159–178. Warrington: Daresbury Laboratory.
Bricogne, G. & Irwin, J. J. (1996). *Maximum-likelihood structure refinement: theory and implementation within BUSTER+TNT*. In *Proceedings of the CCP4 study weekend. Macromolecular refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 85–92. Warrington: Daresbury Laboratory.
Brünger, A. T. (1992*a*). *Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature (London)*, **355**, 472–475.
Brünger, A. T. (1992*b*). *X-PLOR manual*. Version 3.1. New Haven: Yale University.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Cryst.* D**54**, 905–921.
Collaborative Computational Project, Number 4 (1994). *The CCP4 suite: programs for protein crystallography. Acta Cryst.* D**50**, 760–763.
Coppens, P. (1997). *X-ray charge densities and chemical bonding*. International Union of Crystallography and Oxford University Press.
Cowtan, K. D. & Main, P. (1998). *Miscellaneous algorithms for density modification. Acta Cryst.* D**53**, 487–493.
Cruickshank, D. W. J. (1999). *Remarks about protein structure precision. Acta Cryst.* D**55**, 583–601; erratum (1999), D**55**, 1108.
Dauter, Z. & Dauter, M. (1999). *Anomalous signal of solvent bromides used for phasing of lysozyme. J. Mol. Biol.* **289**, 93–101.
Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1997). *The benefits of atomic resolution. Curr. Opin. Struct. Biol.* **7**, 681–688.
Dauter, Z., Wilson, K. S., Sieker, L. C., Meyer, J. & Moulis, J.-M. (1997). *Atomic resolution (0.94 Å) structure of Clostridium acidurici ferredoxin. Detailed geometry of [4Fe-4S] clusters in a protein. Biochemistry*, **36**, 16065–16073.

18.4 (*cont.*)

Diamond, R. (1971). *A real-space refinement procedure for proteins*. *Acta Cryst.* A**27**, 436–452.

Driessen, H., Haneef, M. I. J., Harris, G. W., Howlin, B., Khan, G. & Moss, D. S. (1989). *RESTRAIN: restrained structure-factor least-squares refinement program for macromolecular structures*. *J. Appl. Cryst.* **22**, 510–516.

Engh, R. A. & Huber, R. (1991). *Accurate bond and angle parameters for X-ray protein structure refinement*. *Acta Cryst.* A**47**, 392–400.

EU 3-D Validation Network (1998). *Who checks the checkers? Four validation tools applied to eight atomic resolution structures. J. Mol. Biol.* **276**, 417–436.

French, S. & Wilson, K. S. (1978). *On the treatment of negative intensity observations. Acta Cryst.* A**34**, 517–525.

Hamilton, W. C. (1965). *Significance tests on the crystallographic R factor. Acta Cryst.* **18**, 502–510.

Herzberg, O. & Sussman, J. L. (1983). *Protein model building by the use of a constrained-restrained least-squares procedure. J. Appl. Cryst.* **16**, 144–150.

International Tables for Crystallography (1999). Vol. C. *Mathematical, physical and chemical tables*, edited by A. J. C. Wilson & E. Prince. Dordrecht: Kluwer Academic Publishers.

Jelsch, C., Pichon-Pesme, V., Lecomte, C. & Aubry, A. (1998). *Transferability of multipole charge-density parameters: application to very high resolution oligopeptide and protein structures. Acta Cryst.* D**54**, 1306–1318.

Johnson, C. K. (1976). *ORTEPII. A FORTRAN thermal-ellipsoid plot program for crystal structure illustration*. Report ORNL-5138. Oak Ridge National Laboratory, Tennessee, USA.

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.* A**47**, 110–119.

Konnert, J. H. & Hendrickson, W. A. (1980). *A restrained-parameter thermal-factor refinement procedure. Acta Cryst.* A**36**, 344–350.

Lamzin, V. S., Morris, R. J., Dauter, Z., Wilson, K. S. & Teeter, M. M. (1999). *Experimental observation of bonding electrons in proteins. J. Biol. Chem.* **274**, 20753–20755.

Lamzin, V. S. & Wilson, K. S. (1997). *Automated refinement for protein crystallography. Methods Enzymol.* **277**, 269–305.

Matthews, B. W. (1968). *Solvent content in protein crystals. J. Mol. Biol.* **33**, 491–497.

Murshudov, G. N., Davies, G. J., Isupov, M., Krzywda, S. & Dodson, E. J. (1998). *The effect of overall anisotropic scaling in macromolecular refinement*. In *CCP4 newsletter on protein crystallography*, **35**, 37–42.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst.* D**53**, 240–255.

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Efficient anisotropic refinement of macromolecular structures using FFT. Acta Cryst.* D**55**, 247–255.

Nayal, M. & Di Cera, E. (1996). *Valence screening of water in protein crystals reveals potential Na^+ binding sites. J. Mol. Biol.* **256**, 228–234.

O'Hagan, A. (1994). *Kendal's advanced theory of statistics; Bayesian inference*, Vol. 2B. Cambridge: Arnold, Hodder Headline and Cambridge University Press.

Pannu, N. S. & Read, R. J. (1996). *Improved structure refinement through maximum likelihood. Acta Cryst.* A**52**, 659–668.

Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.

Schomaker, V. & Trueblood, K. N. (1968). *On the rigid-body motion of molecules in crystals. Acta Cryst.* B**24**, 63–76.

Schwarzenbach, D., Abrahams, S. C., Flack, H. D., Prince, E. & Wilson, A. J. C. (1995). *Statistical descriptors in crystallography. II. Report of a working group on expression of uncertainty in measurement. Acta Cryst.* A**51**, 565–569.

Sevcik, J., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1996). *Ribonuclease from Streptomyces aureofaciens at atomic resolution. Acta Cryst.* D**52**, 327–344.

Sheldrick, G. M. (1990). *Phase annealing in SHELX-90: direct methods for larger structures. Acta Cryst.* A**46**, 467–473.

Sheldrick, G. M. & Schneider, T. R. (1997). *SHELXL: high-resolution refinement. Methods Enzymol.* **277**, 319–343.

Sheriff, S. & Hendrickson, W. A. (1987). *Description of overall anisotropy in diffraction from macromolecular crystals. Acta Cryst.* A**43**, 118–121.

Souhassou, M., Lecomte, C., Ghermani, N.-E., Rohmer, M.-M., Roland, W., Benard, M. & Blessing, R. H. (1992). *Electron distributions in peptides and related molecules. 2. An experimental and theoretical study of (Z)-N-acetyl-α,β-dehydrophenylalanine methylamide. J. Am. Chem. Soc.* **114**, 2371–2382.

Stuart, A., Ord, K. J. & Arnold, S. (1999). *Kendall's advanced theory of statistics; classical inference and linear model*, Vol. 2A. London/Sydney/Auckland: Arnold, Hodder Headline.

Teeter, M. M., Roe, S. M. & Heo, N. H. (1993). *Atomic resolution (0.83 Å) crystal structure of the hydrophobic protein crambin at 130 K. J. Mol. Biol.* **230**, 292–311.

Ten Eyck, L. F. (1973). *Crystallographic fast Fourier transforms. Acta Cryst.* A**29**, 183–191.

Ten Eyck, L. F. (1977). *Efficient structure-factor calculation for large molecules by the fast Fourier transform. Acta Cryst.* A**33**, 486–492.

Tickle, I. J., Laskowski, R. A. & Moss, D. S. (1998). *R_{free} and the R_{free} ratio. Part I: Derivation of expected values of cross-validation residuals used in macromolecular least-squares refinement. Acta Cryst.* D**54**, 547–557.

Tronrud, D. E. (1997). *TNT refinement package. Methods Enzymol.* **277**, 243–268.

Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1998). *Refinement of triclinic hen egg-white lysozyme at atomic resolution. Acta Cryst.* D**54**, 522–546.

Wilson, A. J. C. (1942). *Determination of absolute from relative X-ray data intensities. Nature (London)*, **150**, 151–152.

18.5

Allen, F. H., Cole, J. C. & Howard, J. A. K. (1995*a*). *A systematic study of coordinate precision in X-ray structure analyses. I. Descriptive statistics and predictive estimates of e.s.d.'s for C atoms. Acta Cryst.* A**51**, 95–111.

Allen, F. H., Cole, J. C. & Howard, J. A. K. (1995*b*). *A systematic study of coordinate precision in X-ray structure analyses. II. Predictive estimates of e.s.d.'s for the general-atom case. Acta Cryst.* A**51**, 112–121.

Bricogne, G. (1993*a*). *Direct phase determination by entropy maximization and likelihood ranking: status report and perspectives. Acta Cryst.* D**49**, 37–60.

Bricogne, G. (1993*b*). *Fourier transforms in crystallography: theory, algorithms, and applications*. In *International tables for crystallography*, Vol. B, edited by U. Shmueli, pp. 88–89. Dordrecht: Kluwer Academic Publishers.

Bricogne, G. & Irwin, J. (1996). *Maximum-likelihood structure refinement: theory and implementation within BUSTER + TNT*. In *Proceedings of the CCP4 study weekend. Macromolecular refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 85–92. Warrington: Daresbury Laboratory.

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. (1993). *Assessment of phase accuracy by cross validation: the free R value. Methods and application. Acta Cryst.* D**49**, 24–36.

Chambers, J. L. & Stroud, R. M. (1979). *The accuracy of refined protein structures: comparison of two independently refined models of bovine trypsin. Acta Cryst.* B**35**, 1861–1874.

Cochran, W. (1948). *The Fourier method of crystal-structure analysis. Acta Cryst.* **1**, 138–142.

18.5 (*cont.*)

Cohen, G. H., Sheriff, S. & Davies, D. R. (1996). *Refined structure of the monoclonal antibody HyHEL-5 with its antigen hen egg-white lysozyme. Acta Cryst.* D**52**, 315–326.

Cowtan, K. & Ten Eyck, L. F. (2000). *Eigensystem analysis of the refinement of a small metalloprotein. Acta Cryst.* D**56**, 842–856.

Cruickshank, D. W. J. (1949*a*). *The accuracy of electron-density maps in X-ray analysis with special reference to dibenzyl. Acta Cryst.* **2**, 65–82.

Cruickshank, D. W. J. (1949*b*). *The accuracy of atomic coordinates derived by least-squares or Fourier methods. Acta Cryst.* **2**, 154–157.

Cruickshank, D. W. J. (1952). *On the relations between Fourier and least-squares methods of structure determination. Acta Cryst.* **5**, 511–518.

Cruickshank, D. W. J. (1956). *The determination of the anisotropic thermal motion of atoms in crystals. Acta Cryst.* **9**, 747–753.

Cruickshank, D. W. J. (1959). *Statistics*. In *International tables for X-ray crystallography*, Vol. 2, edited by J. S. Kasper & K. Lonsdale, pp. 84–98. Birmingham: Kynoch Press.

Cruickshank, D. W. J. (1960). *The required precision of intensity measurements for single-crystal analysis. Acta Cryst.* **13**, 774–777.

Cruickshank, D. W. J. (1965). *Notes for authors; anisotropic parameters. Acta Cryst.* **19**, 153.

Cruickshank, D. W. J. (1970). *Least-squares refinement of atomic parameters*. In *Crystallographic computing*, edited by F. R. Ahmed, S. R. Hall & C. P. Huber, pp. 187–196. Copenhagen: Munksgaard.

Cruickshank, D. W. J. (1999). *Remarks about protein structure precision. Acta Cryst.* D**55**, 583–601.

Cruickshank, D. W. J. & Robertson, A. P. (1953). *The comparison of theoretical and experimental determinations of molecular structures, with applications to naphthalene and anthracene. Acta Cryst.* **6**, 698–705.

Cruickshank, D. W. J. & Rollett, J. S. (1953). *Electron-density errors at special positions. Acta Cryst.* **6**, 705–707.

Daopin, S., Davies, D. R., Schlunegger, M. P. & Grütter, M. G. (1994). *Comparison of two crystal structures of TGF-β2: the accuracy of refined protein structures. Acta Cryst.* D**50**, 85–92.

Deacon, A., Gleichmann, T., Kalb (Gilboa), A. J., Price, H., Raftery, J., Bradbrook, G., Yariv, J. & Helliwell, J. R. (1997). *The structure of concanavalin A and its bound solvent determined with small-molecule accuracy at 0.94 Å resolution. J. Chem. Soc. Faraday Trans.* **93**, 4305–4312.

Dodson, E. (1998). *The role of validation in macromolecular crystallography. Acta Cryst.* D**54**, 1109–1118.

Engh, R. A. & Huber, R. (1991). *Accurate bond and angle parameters for X-ray protein structure refinement. Acta Cryst.* A**47**, 392–400.

Haridas, M., Anderson, B. F. & Baker, E. N. (1995). *Structure of human diferric lactoferrin refined at 2.2 Å resolution. Acta Cryst.* D**51**, 629–646.

Hendrickson, W. A. & Konnert, J. H. (1980). *Incorporation of stereochemical information into crystallographic refinement*. In *Computing in crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences.

International Tables for Crystallography (1999). Vol. C. *Mathematical, physical and chemical tables*, edited by A. J. C. Wilson & E. Prince. Dordrecht: Kluwer Academic Publishers.

Jiang, J.-S. & Brünger, A. T. (1994). *Protein hydration observed by X-ray diffraction. Solvation properties of penicillopepsin and neuraminidase crystal structures. J. Mol. Biol.* **243**, 100–115.

Ko, T.-P., Day, J., Greenwood, A. & McPherson, A. (1994). *Structures of three crystal forms of the sweet protein thaumatin. Acta Cryst.* D**50**, 813–825.

Kobe, B. & Deisenhofer, J. (1995). *A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature (London)*, **374**, 183–186.

Langridge, R., Marvin, D. A., Seeds, W. E., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. & Hamilton, L. D. (1960). *The molecular configuration of deoxyribonucleic acid. II. Molecular models and their Fourier transforms. J. Mol. Biol.* **2**, 38–64.

Luzzati, V. (1952). *Traitement statistique des erreurs dans la determination des structures cristallines. Acta Cryst.* **5**, 802–810.

Moews, P. C. & Kretsinger, R. H. (1995). *Refinement of the structure of carp muscle calcium-binding parvalbumin by model building and difference Fourier analysis. J. Mol. Biol.* **91**, 201–228.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst.* D**53**, 240–255.

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Efficient anisotropic refinement of macromolecular structures using FFT. Acta Cryst.* D**55**, 247–255.

Pannu, N. S. & Read, R. J. (1996). *Improved structure refinement through maximum likelihood. Acta Cryst.* A**52**, 659–668.

Read, R. J. (1986). *Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Cryst.* A**42**, 140–149.

Read, R. J. (1990). *Structure-factor probabilities for related structures. Acta Cryst.* A**46**, 900–912.

Rollett, J. S. (1970). *Least-squares procedures in crystal structure analysis*. In *Crystallographic computing*, edited by F. R. Ahmed, S. R. Hall & C. P. Huber, pp. 167–181. Copenhagen: Munksgaard.

Schwarzenbach, D., Abrahams, S. C., Flack, H. D., Gonschorek, W., Hahn, Th., Huml, K., Marsh, R. E., Prince, E., Robertson, B. E., Rollett, J. S. & Wilson, A. J. C. (1989). *Statistical descriptors in crystallography: Report of the IUCr subcommittee on statistical descriptors. Acta Cryst.* A**45**, 63–75.

Schwarzenbach, D., Abrahams, S. C., Flack, H. D., Prince, E. & Wilson, A. J. C. (1995). *Statistical descriptors in crystallography. II. Report of a working group on expression of uncertainty in measurement. Acta Cryst.* A**51**, 565–569.

Sevcik, J., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1996). *Ribonuclease from Streptomyces aureofaciens at atomic resolution. Acta Cryst.* D**52**, 327–344.

Sheldrick, G. M. & Schneider, T. R. (1997). *SHELXL: high resolution refinement. Methods Enzymol.* **277**, 319–343.

Stec, B., Zhou, R. & Teeter, M. M. (1995). *Full-matrix refinement of the protein crambin at 0.83 Å and 130 K. Acta Cryst.* D**51**, 663–681.

Tickle, I. J., Laskowski, R. A. & Moss, D. S. (1998*a*). *Error estimates of protein structure coordinates and deviations from standard geometry by full-matrix refinement of γB- and βB2-crystallin. Acta Cryst.* D**54**, 243–252.

Tickle, I. J., Laskowski, R. A. & Moss, D. S. (1998*b*). *R_{free} and the R_{free} ratio. I. Derivation of expected values of cross-validation residuals used in macromolecular least-squares refinement. Acta Cryst.* D**54**, 547–557.

Tronrud, D. E. (1997). *TNT refinement package. Methods Enzymol.* **277**, 306–319.

Tronrud, D. E. (1999). *The efficient calculation of the normal matrix in least-squares refinement of macromolecular structures. Acta Cryst.* A**55**, 700–703.

Trueblood, K. N., Bürgi, H.-B., Burzlaff, H., Dunitz, J. D., Gramaccioli, C. M., Schulz, H. H., Shmueli, U. & Abrahams, S. C. (1996). *Atomic displacement parameter nomenclature. Report of a subcommittee on atomic displacement parameter nomenclature. Acta Cryst.* A**52**, 770–781.

Usón, I., Pohl, E., Schneider, T. R., Dauter, Z., Schmidt, A., Fritz, H.-J. & Sheldrick, G. M. (1999). *1.7 Å structure of the stabilized REI$_V$ mutant T39K. Application of local NCS restraints. Acta Cryst.* D**55**, 1158–1167.

Watenpaugh, K. D., Sieker, L. C., Herriott, J. R. & Jensen, L. H. (1973). *Refinement of the model of a protein: rubredoxin at 1.5 Å resolution. Acta Cryst.* B**29**, 943–956.

Wilson, A. J. C. (1950). *Largest likely values for the reliability index. Acta Cryst.* **3**, 397–398.