18.5. COORDINATE UNCERTAINTY

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.*,

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

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

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

$$(18.5.5.2)$$

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,

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

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-$\beta2$. 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.

409