

18.4. REFINEMENT AT ATOMIC RESOLUTION

gradient approach, all diagonal terms of the matrix are equal. However, the range and speed of convergence are substantially reduced, and standard uncertainties can no longer be estimated directly by matrix inversion.

18.4.3.3. Fast Fourier transform

Conventional least-squares programs use the structure-factor equation and associated derivatives, with the summation extending over all atoms and all reflections. This is immensely slow in computational terms for large structures, but it has the advantage of providing precise values.

An alternative procedure, where the computer time is reduced from being proportional to N^2 to $N \log N$, involves the use of fast Fourier algorithms for the computation of structure factors and derivatives (Ten Eyck, 1973, 1977; Agarwal, 1978). This can involve some interpolation and the limitation of the volume of electron-density maps to which individual atoms contribute. Such algorithms have been exploited extensively in macromolecular refinement programs such as *PROLSQ* (Konnert & Hendrickson, 1980), *XPLOR* (Brünger, 1992b), *TNT* (Tronrud, 1997), *RESTRAIN* (Driessen *et al.*, 1989), *REFMAC* (Murshudov *et al.*, 1997), *CNS* (Brünger *et al.*, 1998) and *PHENIX.REFINE* (Terwilliger *et al.*, 2008), but have largely been restricted to the diagonal approximation. *XPLOR* and *CNS* use the conjugate-gradient method, which relies only on the first derivatives and ignores the second derivatives. In all other programs, the diagonal approximation is used for the second-derivative matrix.

18.4.3.4. Maximum likelihood

This provides a statistically sounder alternative to least squares, especially in the early stages of refinement when the model lies far from the minimum. This approach increases the radius of convergence, takes into account experimental uncertainties, and in the final stages gives results similar to least squares but with improved weights (Bricogne & Irwin, 1996; Murshudov *et al.*, 1997). The maximum-likelihood approach has been extended to allow refinement of a full atomic anisotropic model while retaining the use of fast Fourier algorithms (Murshudov *et al.*, 1999). A remaining limitation is the use of the diagonal approximation, which prevents the computation of standard uncertainties of individual parameters. Algorithms that will alleviate this limitation can be foreseen, and they are expected to be implemented in the future.

18.4.3.5. Twinning

A non-negligible fraction of protein crystals turn out to be merohedrally twinned (Lebedev *et al.*, 2006), which requires special treatment of the diffraction data and proper treatment of this phenomenon during structure solution and more especially refinement (Yeates, 1997; Chapters 18.11 and 18.12). The twinning problem can be approached in two distinctly different ways. In the first, the data are explicitly 'detwinned' to produce an adjusted set of amplitudes [for example using the UCLA detwinning server (<http://nihserver.mbi.ucla.edu/Twinning/>) or the *DETWIN* program (Collaborative Computational Project, Number 4, 1994)], which are then subjected to conventional refinement. However, detwinning can only be successfully applied if the twinning fraction is not too high, since the error in the resulting amplitudes increases as the fraction approaches 50%. A second, and certainly preferred, approach is to include the twinning fraction as a variable during refinement. This has been implemented in *SHELXL*, *PHENIX.REFINE* and

REFMAC, and is widely used. In *CNS*, it is possible to set the twin fraction, but not to refine it. This is of special relevance for atomic resolution structures, as even a small degree of twinning will have a significant effect on the interpretability of fine features.

18.4.3.6. Computer power

There are no longer any restrictions on the full-matrix refinement of small-molecule crystal structures. However, the large size of the matrix, which increases as N^2 , where N is the number of parameters, means that for macromolecules with thousands of independent atoms this approach is intractable with the computing resources normally available to the crystallographer. By extrapolating the progress in computing power experienced in recent years, it can be envisaged that the limitations will disappear during the next decade, as those for small structures have disappeared since the 1960s. Indeed, the advances in the speed of CPUs, computer memory and disk capacity continue to transform the field.

18.4.4. Computational options and tactics

18.4.4.1. Use of F (amplitudes) or F^2 (intensities)

The X-ray experiment provides two-dimensional diffraction images. These are transformed to integrated but unscaled data, which are transformed to Bragg reflection intensities that are subsequently transformed to structure-factor amplitudes. At each transformation some assumptions are used, and the results will depend on their validity. Invalid assumptions will introduce bias toward these assumptions into the resulting data. Ideally, refinement (or estimation of parameters) should be against data that are as close as possible to the experimental observations, eliminating at least some of the invalid assumptions. Extrapolating this to the extreme, refinement should use the images as observable data, but this poses several severe problems, depending on data quantity and the lack of an appropriate statistical model.

Alternatively, the transformation of data could be improved by revising the assumptions. The intensities are closer to the real experiment than are the structure-factor amplitudes, and use of intensities would reduce the bias. However, there are some difficulties in the implementation of intensity-based likelihood refinement (Pannu & Read, 1996).

Gaussian approximation to intensity-based likelihood (Murshudov *et al.*, 1997) would avoid these difficulties, since a Gaussian distribution of error can be assumed in the intensities but not the amplitudes. However, errors in intensities may not only be the result of counting statistics, but may have additional contributions from factors such as crystal disorder and motion of the molecules in the lattice during data collection.

Nevertheless, the problem of how to treat weak reflections remains. Some of the measured intensities will be negative, as a result of statistical errors of observation, and the proportion of such measurements will be relatively large for weakly diffracting macromolecular structures, especially at atomic resolution. This is less important for intensity-based likelihood than for the amplitude-based approach. French & Wilson (1978) have given a Bayesian approach for the derivation of structure-factor amplitudes from intensities using Wilson's distribution (Wilson, 1942) as a prior, but there is room for improvement in this approach. Firstly, the Wilson distribution could be upgraded using the scaling techniques suggested by Blessing (1997) and Cowtan &