

18. REFINEMENT

18.4.3. Computational algorithms and strategies**18.4.3.1. Classical least-squares refinement of small molecules**

The principles of the least-squares method of minimization are described in *IT C* (1999). Least squares involves the construction of a matrix of order $N \times N$, where N is the number of parameters, representing a system of least-squares equations, whose solution provides estimates of adjustments to the current atomic parameters. The problem is nonlinear and the matrix construction and solution must be iterated until convergence is achieved. In addition, inversion of the matrix at convergence provides an approximation to standard uncertainties for each individual parameter refined. Indeed, this is the only method available so far that gives such estimates properly.

However, even for small molecules there may be some disordered regions that will require the imposition of restraints, as is the case for macromolecules (see below), and the presence of such restraints means that the error estimates no longer reflect the information from the X-ray data alone. If the problem of how restraints affect the error estimates could be resolved, then inversion of the matrix corresponding to the second derivative of the posterior distribution would provide standard uncertainties incorporating both the prior knowledge, such as the restraints and the experimental data. Equation (18.4.1.2) for information measure could then be applied. For small structures, the speed and memory of modern computers have reduced the requirements for such calculations to the level of seconds, and the computational requirements form a trivial part of the structure analysis.

18.4.3.2. Least-squares refinement of large structures

The size of the computational problem increases dramatically with the size of the unit cell, as the number of terms in the matrix increases with the square of the number of parameters. Furthermore, construction of each element depends on the number of reflections. For macromolecular structures, computation of a full matrix is at present prohibitively expensive in terms of CPU time and memory. A variety of simplifying approaches have been developed, but all suffer from a poorer estimate of the standard uncertainty and from a more limited range and speed of convergence.

The first approach is the block-matrix approximation, where instead of the full matrix, only square blocks along the matrix diagonal are constructed, involving groups of parameters that are expected to be correlated. The correlation between parameters belonging to different blocks is therefore neglected completely. In this way, the whole least-squares minimization is split into a set of smaller independent units. In principle this leads to the same solution, but more slowly and with less precise error estimates. Nevertheless, block-matrix approaches remain essential for tractable matrix inversion for macromolecular structures.

A further simplification involves the conjugate-gradient method or the diagonal approximation to the normal matrix (the second derivative of minus the log of the likelihood function in the case of maximum likelihood), which essentially ignores all off-diagonal terms of the least-squares matrix. For the conjugate-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) and *CNS* (Brünger *et al.*, 1998), but have been largely restricted to the diagonal approximation. *XPLOR* and *CNS* use the conjugate-gradient method that relies only on the first derivatives, ignoring 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 more statistically sound 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 (Murshudov *et al.*, 1997; Bricogne, 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 near future.

18.4.3.5. 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, which contain 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, which makes it hard to predict the optimal strategies for atomic resolution refinement, even over the next ten years.

18.4.4. Computational options and tactics**18.4.4.1. Use of F or F^2**

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