(4) Automatic mode: This combines the previous three extension schemes. The program automatically works out the optimum combination of the above three schemes according to the density-modification mode, the phase-combination mode and the nature of the input reflection data. The automatic mode is the default and is the recommended mode of choice unless specific circumstances warrant a different choice.

(5) All reflection mode: One advantage of the reflection-omit and perturbation-γ methods is that the strength of extrapolation of a structure-factor amplitude is a good indicator of the reliability of its corresponding phase. As a result, a phase-extension scheme is unnecessary in reflection-omit calculations; all reflections may be included from the first cycle.

25.2.2.5. *Code description*

The program was designed to be run largely automatically with minimal user intervention. This is achieved by using extensive default settings and by automatic selection of options based on the data used. The program is also modular by design so that additional density-modification methods can be incorporated easily.

A simplified flow diagram for *DM* is shown in Fig. 25.2.2.2(*a*). When a reflection-omit calculation is performed, an additional loop is introduced, shown in Fig. 25.2.2.2(*b*). The Sayre's equation calculation adds another level of complexity, described in Zhang & Main (1990*b*). Skeletonization imposes the protein histogram and solvent flatness implicitly and so is performed, if necessary, every second or third cycle in place of solvent flattening and histogram matching. Simplified conceptual and actual flow diagrams for *DMMULTI* are shown in Figs. 25.2.2.3(*a*) and (*b*).

Many of the basic approaches used in *DM* and *DMMULTI* are described in Chapter 15.1. Some practical aspects of the application and combination of these approaches are described here.

25.2.2.5.1. *Scaling*

All forms of map modification are affected by the overall temperature factor of the data, and histogram matching in particular is critically dependent on the accurate determination of the scale factor. Wilson statistics have been found inadequate for scaling in this case, especially when the data resolution is worse than 3 Å, because of the dip in scattering below 5 Å.

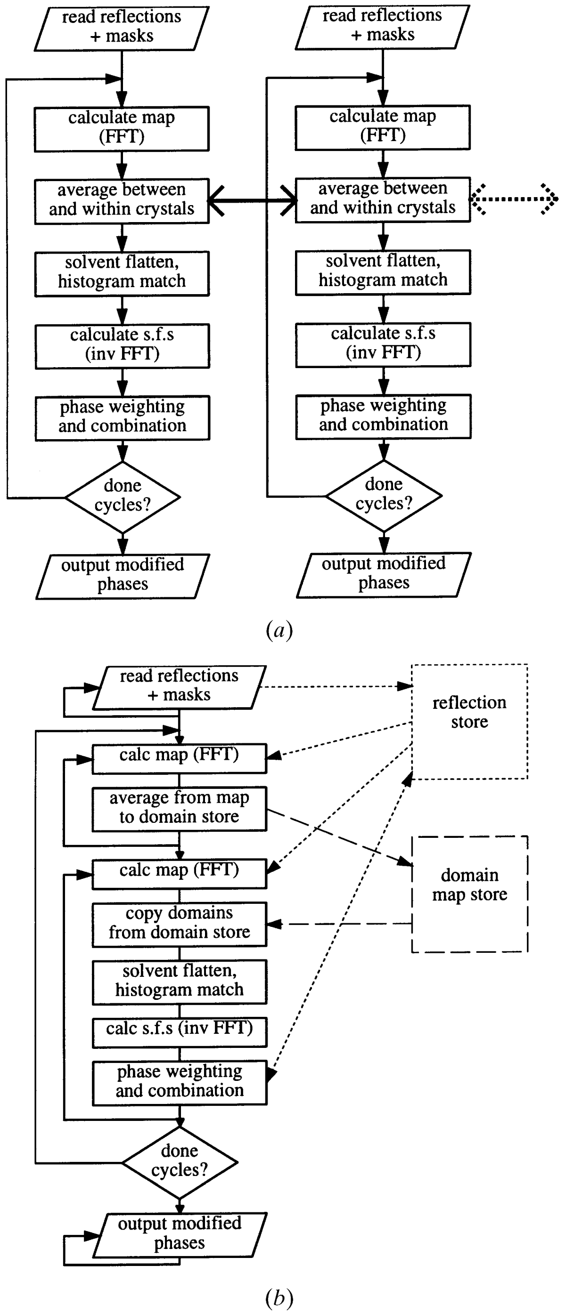
More accurate estimates of the scale and temperature factors may be achieved by fitting the data to a semi-empirical scattering curve (Cowtan & Main, 1998). This curve is prepared using Parseval's theorem, which relates the sum of the intensities to the variance of the map:

$$\sigma_\rho^2 = \frac{1}{V^2}\sum_{\mathbf{h}\neq 000}|F(\mathbf{h})|^2. \qquad (25.2.2.2)$$

Thus, the sum of the intensities in a particular resolution shell is proportional to the difference in variance of maps calculated with and without that shell of data. The empirical curve is therefore calculated from the variance in the protein regions of a group of known structures, calculated as a function of resolution. The curve is scaled to the protein volume of the current structure, and a correction is made for the solvent, which is assumed to be flat.

The overall temperature factor is removed, and an absolute scale is imposed by fitting the data to this curve. The use of sharpened F's (with no overall temperature factor) is necessary for histogram matching and often increases the power of averaging for phase extension.

Since the solvent content is used in scaling the data, it is important that this value be entered correctly. However, the volume of the solvent mask may be varied independently of the true solvent content, as discussed in Section 25.2.2.3.


read reflections + masks
calculate map (FFT)
average between and within crystals
solvent flatten, histogram match
calculate s.f.s (inv FFT)
phase weighting and combination
done cycles?
output modified phases
read reflections + masks
calculate map (FFT)
average between and within crystals
solvent flatten, histogram match
calculate s.f.s (inv FFT)
phase weighting and combination
done cycles?
output modified phases
read reflections + masks
reflection store
calc map (FFT)
average from map to domain store
domain map store
calc map (FFT)
copy domains from domain store
solvent flatten, histogram match
calc s.f.s (inv FFT)
phase weighting and combination
done cycles?
output modified phases

(*a*)

(*b*)

Fig. 25.2.2.3. (*a*) Conceptual flowchart for a *DMMULTI* multi-crystal calculation. (*b*) Actual flow chart for a *DMMULTI* multi-crystal calculation.

25.2.2.5.2. *Solvent-mask determination*

If the user does not supply a solvent mask, the solvent mask is calculated by Wang's (1985) method, using the reciprocal-space approach of Leslie (1987). A number of variants on this algorithm are implemented; however, the parameter that affects the quality of the solvent mask most dramatically is the radius of the smoothing function (Chapter 15.1). This parameter may be estimated empirically by

$$r_{\rm Wang} = 2r_{\max}\overline{w}^{1/3}, \qquad (25.2.2.3)$$

where $r_{\max}$ is the resolution limit of the observed amplitudes, and $\overline{w}$ is the mean figure of merit over the same reflections (with $w = 0$ for unphased reflections).

Once the smoothed map has been determined, cutoff values are chosen to divide the map into protein and solvent regions. If the protein boundary is poorly defined, the user may specify protein, solvent and excluded volumes, in which case two cutoffs are specified and the intermediate region is marked as neither protein nor solvent.

25.2.2.5.3. *Averaging-mask determination*

If the user does not supply an averaging mask, it is determined by a local correlation method (Vellieux *et al.*, 1995). A large region covering 27 unit cells is selected, and the local correlation between the maps before and after rotation by one of the noncrystallographic symmetry operators is calculated. The largest contiguous region that is in agreement among different NCS operators is isolated from the local correlation map, and a finer local correlation map is calculated over this volume. This process is iterated until a good mask with a detailed boundary is found.

This approach is fully automatic, except in the case where a noncrystallographic symmetry operator intersects a crystallographic symmetry operator, in which case the mask is not uniquely defined, and some user intervention may be required. The method is robust, and by increasing the radius of the sphere within which the local correlation is calculated, it may be used with very poor maps (Cowtan & Main, 1998). The method is easily extended to include information from multiple averaging operators.

25.2.2.5.4. *Fourier transforms*

For simplicity of coding, all Fourier transforms are performed in core using real-to-Hermitian and Hermitian-to-real fast Fourier transforms (FFTs). The data are expanded to space group $P1$ before calculating a map and averaged back to a reciprocal asymmetric unit after inverse transformation. Most of the map modifications preserve crystallographic symmetry, so restricted phases are not constrained except during phase combination.

25.2.2.5.5. *Histogram matching*

The target histograms are calculated from the protein regions of several stationary-atom structures at resolutions from 6 to 1.5 Å, according to the method described by Zhang & Main (1990*a*). The histogram variances should be consistent with the map variances used in scaling the data. The resolution of the target histogram can be accurately matched to the data resolution by averaging the target histograms on either side of the current resolution.

25.2.2.5.6. *Averaging*

Averaging is performed using a single-step approach (Rossmann *et al.*, 1992), in which every copy of the molecule in a 'virtual' asymmetric unit is averaged with every other copy. Density values are obtained at non-grid positions using a 27-point quadratic spectral spline interpolation. A sharpened map is first calculated by dividing by the Fourier transform of the quadratic spline function. The same spline function is then convoluted with the sharpened map to obtain the density value at an arbitrary coordinate (Cowtan & Main, 1998). This approach gives very accurate interpolation from a coarse grid map with relatively little computation and additionally provides gradient information for the refinement of averaging operators.

25.2.2.5.7. *Multi-crystal averaging*

The multi-crystal averaging calculation in *DMMULTI* is equivalent to several single-crystal averaging calculations running simultaneously, with the exception that during the averaging step, the molecule density is averaged across every copy in every crystal form. This average is weighted by the mean figure of merit of each crystal form; this allows the inclusion of unphased crystal forms, since in the first cycle they will have zero weight and therefore not disrupt the phasing that is already present. In subsequent cycles, the unphased form contains phase information from the back-transformed density.

This technique can be extremely useful, since adding a new crystal form usually provides considerably more phase information than adding a new derivative if the cross-rotation and translation functions can be solved.

In the multi-crystal case, averaging is performed using a two-step approach, first building an averaged molecule from all the copies in all crystal forms, then replacing the density in each crystal form with the averaged values. This approach is computationally more efficient when there are many copies of the molecule.

The conceptual flow chart of simultaneous density-modification calculations across multiple crystal forms is shown in Fig. 25.2.2.3(*a*); in practice, this scheme is implemented using a single process and looping over every crystal form at each stage (Fig. 25.2.2.3*b*). Maps are reconstructed from a large data object containing all the reflection data in every crystal form. Averaging is performed using a second data object containing maps of each averaging domain. By this means, an arbitrary number of domains may be averaged across an arbitrary number of crystal forms.

Multi-crystal averaging has been particularly successful in solving structures from very weak initial phasing, since the data redundancy is usually higher than for single-crystal problems.

25.2.3. The structure-determination language of the *Crystallography & NMR System* (A. T. Brunger, P. D. Adams, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J.-S. Jiang, N. S. Pannu, R. J. Read, L. M. Rice and T. Simonson)

25.2.3.1. *Introduction*

We have developed a new and advanced software system, the *Crystallography & NMR System* (CNS), for crystallographic and NMR structure determination (Brünger *et al.*, 1998). The goals of CNS are: (1) to create a flexible computational framework for exploration of new approaches to structure determination; (2) to provide tools for structure solution of difficult or large structures; (3) to develop models for analysing structural and dynamical properties of macromolecules; and (4) to integrate all sources of information into all stages of the structure-determination process.

To meet these goals, algorithms were moved from the source code into a symbolic structure-determination language which represents a new concept in computational crystallography. The high-level CNS computing language allows definition of symbolic target functions, data structures, procedures and modules. The CNS program acts as an interpreter for the high-level CNS language and includes hard-wired functions for efficient processing of computing-intensive tasks. Methods and algorithms are therefore more clearly defined and easier to adapt to new and challenging problems. The result is a multi-level system which provides maximum flexibility to the user (Fig. 25.2.3.1). The CNS language provides a common framework for nearly all computational procedures of structure determination. A comprehensive set of crystallographic procedures for phasing, density modification and refinement has been implemented in this language. Task-oriented input files written in the CNS language, which can also be accessed through an HTML graphical interface (Graham, 1995), are available to carry out these procedures.

references