

## 25. MACROMOLECULAR CRYSTALLOGRAPHY PROGRAMS

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 (1990a). 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. 25.2.2.3b). 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.