DM and *DMMULTI* are distributed as a part of the *CCP*4 suite of software for protein crystallography (Collaborative Computational Project, Number 4, 1994). The theoretical and algorithmic bases for the *DM* and *DMMULTI* software suites are reviewed in Chapter 15.1. In this chapter, some specific issues concerning the programs are described, including program operation, data preparation, choices of modes and code description.

25.2.2.2. *Program operation*

DM and *DMMULTI* are largely automatic; in order to perform a phase-improvement calculation only two tasks are required of the user:

(1) Provide the input data. These must include the reflection data and solvent content, and may also include averaging operators, solvent mask and averaging domain masks.

(2) Select the appropriate density modifications and the phase-combination mode to be used in the calculation.

DM and *DMMULTI* can run with the minimum input above, since the optimum choices for a whole range of parameters are set in the program defaults. For some special problems it may be useful to control the program behaviour in more detail; this is possible through a wide range of keywords to override the defaults. These are all detailed in the documentation supplied with the software.

25.2.2.3. *Preparation of input data*

Input data are provided by two routes: numerical parameters, such as solvent content and averaging operators, are included in the command file using appropriate keywords, whereas reflections and masks are referenced by giving their file names on the command line. In the simplest case; for example a solvent-flattening and histogram-matching calculation, all that is required is an initial reflection file and an estimate of the solvent content.

Use all available data: The reflection file must be in *CCP*4 'MTZ' format, and contain at least the structure-factor amplitudes, phase estimates and figures of merit. If the phase estimates are obtained from a homologous structure by molecular replacement, the figures of merit can be generated by the *SIGMAA* program (Read, 1986). When the phases are estimated using a single isomorphous derivative (SIR), it is recommended that Hendrickson–Lattman coefficients (Hendrickson & Lattman, 1970) are used to represent the phase estimate instead of the figure of merit. Hendrickson–Lattman coefficients can represent the bimodal distribution of the SIR phases, whereas the figure of merit can only represent the unimodal distribution of the average of two equally probable phase choices. It is recommended that a reflection file containing every possible reflection is used. The low-resolution data should be included since they provide a significant amount of information on the protein–solvent boundary. The high-resolution data without phase estimates should also be included since their phases can be estimated by *DM*. Phase extension can usually improve the original phases further compared to phase refinement only. Unobserved reflections are marked by a missing number flag. This is important for the preservation of the free-*R* reflections. It also enables *DM* to extrapolate missing reflections from density constraints and increases the phase improvement power.

The estimation of solvent content: The solvent content, $C_{\rm solv}$, can be obtained by various experimental methods, such as the solvent dehydration method and the deuterium exchange method (Matthews, 1974). It can also be estimated through

$$C_{\rm solv} = 1 - (NV_aML/V). \qquad (25.2.2.1)$$

Here, N is the total number of atoms, including hydrogen atoms, in one protein molecule. V_a is the average volume occupied by each atom, which is estimated to be approximately 10 Å^3 (Matthews, 1968). M is the number of molecules per asymmetric unit. L is the number of asymmetric units in the cell. V is the unit-cell volume. The correctly estimated solvent content should be entered in the program with the SOLC keyword, since this will be used not only to find the solvent–protein boundary but also to scale the input structure-factor amplitudes. If it is desirable to use a more conservative solvent mask in order to prevent clipping of protein densities, especially in the flexible loop regions, different solvent and protein fractions should be specified using the SOLMASK keyword.

Solvent mask: A solvent mask may be supplied; it may be used for the entire calculation or updated after several cycles. The solvent mask usually divides the cell into protein and solvent regions; however it is also possible to specify excluded regions which are unknown. If no solvent mask is supplied, it will be calculated by a modified Wang–Leslie procedure (Wang, 1985; Leslie, 1987) and updated as the phase-improvement calculation progresses.

Averaging operators: In an averaging calculation, the averaging operators must be supplied; these are typically obtained by rotation and translation searches using a program such as *AMoRe* (Navaza, 1994) or *X-PLOR* (Brünger, 1992*a*). If the coordinates of several heavy atoms are known, they can be used to calculate the noncrystallographic symmetry (NCS) operators. If a partial model can be built into the density, structure-superposition programs, such as *LSQKAB* (Kabsch, 1976), can be used to obtain the rotation and translation matrices that relate different molecules in the asymmetric unit. This can also be achieved through the program *O* using the 'lsq_explicit' command (Jones *et al.*, 1991). The averaging operators can be further refined in *DM* by minimizing the residual between NCS related densities.

Averaging mask: An averaging mask may be supplied; this is distinct from the solvent mask, allowing for parts of the protein to remain unaveraged if required. If no averaging mask is supplied, the mask will be calculated by a local-correlation approach (Cowtan & Main, 1998; Vellieux *et al.*, 1995). If multiple domains are to be averaged with different averaging operators (Schuller, 1996), then one mask must be specified for each averaging domain. When averaging molecules related by improper NCS operations, the averaging mask must be in accord with the NCS operators provided. For example, if the supplied NCS matrix maps molecule A to molecule B, then the averaging mask must cover the volume occupied by molecule A rather than molecule B.

Multi-crystal averaging: In the case of a multi-crystal averaging calculation, one reflection file is provided for each crystal form (however, initial phases are not required in every crystal form), and one reflection file will be output for each crystal form containing the improved phases. One mask is required per averaging domain; thus, in general, only a single mask is required. This may be defined for any crystal form or in an arbitrary crystal space of its own. Averaging operators are then provided to map the mask into each of the crystal forms.

Solvent and averaging masks that are calculated within the program may be output for subsequent analysis. Refined averaging operators are also output. The input and output data for a simple *DM* calculation, a *DM* averaging calculation and a *DMMULTI* multi-crystal averaging calculation are shown in Figs. 25.2.2.1(*a*), (*b*) and (*c*), respectively.

25.2.2.4. *Choice of modes*

Two major choices have to be made in a *DM* run. They are the real-space density-modification modes and reciprocal-space phase-combination modes. Moreover, the phase-extension schemes can be selected if needed. This can also be left to the program, which uses

Copyright © 2006 International Union of Crystallography

references