Chapter 15.3. *DM/DMMULTI* software for phase improvement by density modification

K. D. Cowtan, K. Y. J. Zhang and P. Main

15.3.1. Introduction

DM is an automated procedure for phase improvement by iterated density modification. It is used to obtain a set of improved phases and figures of merit, using as a starting point the observed diffraction amplitudes and some initial poor estimates for the phases and figures of merit. *DM* improves the phases through an alternate application of two processes: real-space electron-density modification and reciprocal-space phase combination. *DM* can perform solvent flattening, histogram matching, multi-resolution modification, averaging, skeletonization and Sayre refinement, as well as conventional or reflection-omit phase combination. Solvent and averaging masks may be input by the user or calculated automatically. Averaging operators may be refined within the program. Multiple averaging domains may be averaged using different operators.

DMMULTI is a modified version of the *DM* software that can perform density modification simultaneously across multiple crystal forms. The procedure is general, handling an arbitrary number of domains appearing in an arbitrary number of crystal forms. Initial phases may be provided for one or more crystal forms; however, improved phases are calculated in every crystal form.

DM and *DMMULTI* are distributed as a part of the *CCP4* 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.

15.3.2. 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.

15.3.3. 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 *CCP4* '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

$$C_{\rm solv} = 1 - (NV_aML/V). \qquad (15.3.3.1)$$

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.

references