

## 25. MACROMOLECULAR CRYSTALLOGRAPHY PROGRAMS

25.2.1.1.4. *Partial structure phase combination and phase extension*

Several programs are included to carry out partial structure phase combination with a variety of weighting options as an aid to structure completion. If density modification (solvent flattening or negative-density truncation and/or NC symmetry averaging) is performed, then phase (and amplitude) extension is also possible by manual or automated procedures.

25.2.1.2. *Design principles*

The PHASES package was designed to be user-friendly with many of the programs being interactive, so that the user is prompted for all information needed. Other programs that are often run repeatedly as part of an iterative procedure are designed to execute as batch processes and are generally run from within command procedures or shell scripts. With the exception of atomic coordinate records, all user-supplied data can be input in free format. Space-group-symmetry information is given by explicitly providing a set of equivalent positions, which has the advantage of allowing non-standard space-group settings. The individual programs in the package can be run 'stand alone', but are often chained together through command procedures or shell scripts. Template scripts are provided for common iterative procedures, but the package design also allows program and option sequences to be combined in many ways, facilitating methodology development by advanced users.

25.2.1.2.1. *General program structure and data flow*

The current package includes 44 Fortran programs and one C subroutine, with the C subroutine used only to provide an interface between the Fortran programs and standard X-Window graphics-library routines. All programs communicate only through files with a simple common format. For the major programs, memory is allocated from a single large one-dimensional array which gets partitioned as required for each problem at run time. This greatly simplifies redimensioning if needed for very large problems, since

at most only two lines of code need to be changed. All source code is provided, along with compilation procedures or shell scripts appropriate for most workstations, including Silicon Graphics, Sun, IBM R6000, ESV and DEC Alpha AXP (both OSF and OpenVMS). A flow chart illustrating the major programs and data flow for common phasing procedures is given in Fig. 25.2.1.1.

25.2.1.2.2. *Parameter and cumulative log files*

Vital data common to nearly all calculations, such as the cell dimensions, lattice type and space-group symmetry, are entered only once in a single 'parameter file'. All interactive programs prompt for the name of this file and for batch programs it is to be supplied on the first input line. The parameter file can also optionally contain the name of a 'running' log file. If used, the running log file is opened in 'append' mode by each program in the package, and a copy of all screen or printed output is added to the file along with a time and date stamp indicating what program was run and when. This allows the user to maintain a complete history of all calculations and results on a given problem in a single, chronologically accurate file.

25.2.1.3. *Merging and scaling native and derivative data*

The programs CMBISO and CMBANO (both interactive) are used to combine unique native and derivative data sets into a single file and place the derivative data set on the scale of the native. All common reflections are identified, paired together, scaled and output to a single 'scaled' file. With CMBISO, only mean structure-factor amplitudes are used for both native and derivative data, *i.e.* Bijvoet mates are deemed equivalent and averaged. CMBANO functions similarly, except that for the derivative data the individual Bijvoet mates are not averaged, and both values are output to the scaled file. The overall merging *R* factor is reported both on *F* and *F*<sup>2</sup>, along with tables indicating the *R* factor as a function of resolution, *F* magnitude and  $|F/\sigma(F)|$ . A table is also output indicating the mean value of  $F_{PH} - F_P$  as a function of resolution, where  $F_{PH}$  and  $F_P$  are the derivative and native structure-factor amplitudes, respectively. By default, scaling is initially carried out by the relative Wilson method (Wilson, 1949), with other optional procedures as outlined below to follow if desired.

25.2.1.3.1. *Relative Wilson scaling*

With this method, the derivative scattering, on average, is made equal to the native scattering by plotting

$$-\ln \left( \frac{\langle F_{PH}^2 \rangle}{\langle F_P^2 \rangle} \right) \text{ versus } \left\langle \frac{\sin^2(\theta)}{\lambda^2} \right\rangle, \quad (25.2.1.1)$$

with the averages taken in corresponding resolution shells. A least-squares fit of a straight line to the plot yields a slope equal to  $2(B_{PH} - B_P)$  (twice the difference between overall isotropic temperature parameters for derivative and native data sets) and an intercept of  $\ln K^2$ . From these values, the derivative data are put on the scale of the native by multiplying each derivative amplitude by

$$K \exp \left[ (B_{PH} - B_P) \frac{\sin^2(\theta)}{\lambda^2} \right]. \quad (25.2.1.2)$$

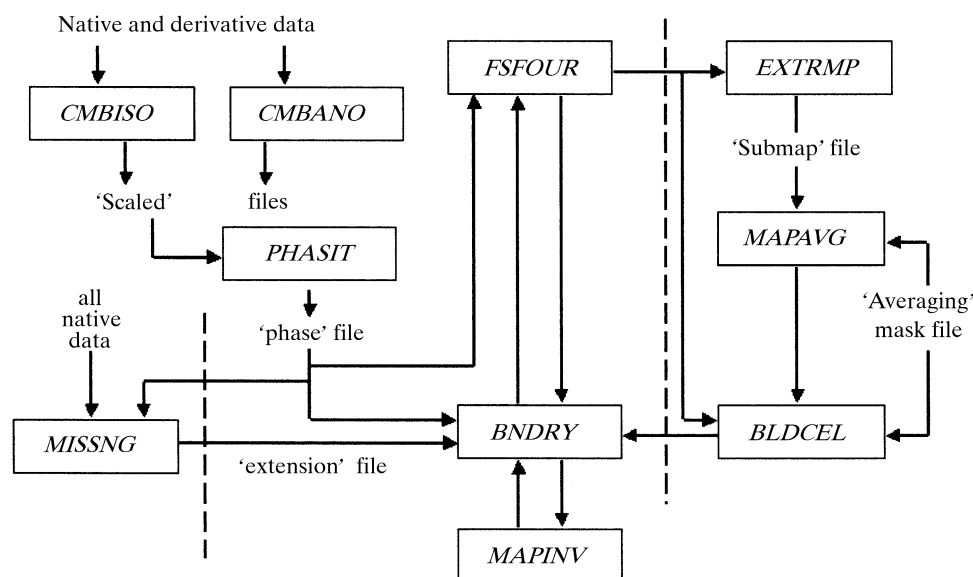


Fig. 25.2.1.1. Flow chart for the major phasing path encompassing native and derivative scaling, heavy-atom-based phasing, solvent flattening, negative-density truncation, and phase combination. Boxed entries represent programs while lines represent files. Optional paths for noncrystallographic symmetry averaging and phase extension are included by considering the additional programs offset from the main path by dashed lines.