25. MACROMOLECULAR CRYSTALLOGRAPHY PROGRAMS

this subset as a function of F magnitude, treating centric and acentric data separately. SIR (single isomorphous replacement) or SAS (single-wavelength anomalous scattering) phase probability distributions are given by

$$P(\varphi) = k \exp[-e(\varphi)^2/2E^2], \qquad (25.2.1.6)$$

where the lack of closure is defined by

$$e(\varphi) = F^2_{PH_{(obs)}} - F^2_{PH_{(calc)}}(\varphi) \qquad (25.2.1.7)$$

for isomorphous-replacement data and

$$e(\varphi) = [(F^+_{PH})^2 - (F^-_{PH})^2]_{obs} - \{[F^+_{PH}(\varphi)]^2 - [F^-_{PH}(\varphi)]^2\}_{calc} \qquad (25.2.1.8)$$

for anomalous-scattering data, with the $+$ and $-$ superscripts denoting members of a Bijvoet pair, and

$$F^2_{PH_{(calc)}}(\varphi) = F^2_P + F^2_H + 2F_P F_H \cos(\varphi - \varphi_H), \qquad (25.2.1.9)$$

with φ denoting the protein phase, and F_H and φ_H denoting the heavy-atom structure-factor amplitude and phase, respectively. The distributions, however, are cast in the A, B, C, D form (Hendrickson & Lattman, 1970). After all input data sets are processed in this manner, the individual phase probability distributions for common reflections are combined *via*

$$P(\varphi)_{comb} = k \exp[\cos(\varphi) \sum_j A_j + \sin(\varphi) \sum_j B_j + \cos(2\varphi) \sum_j C_j + \sin(2\varphi) \sum_j D_j], \qquad (25.2.1.10)$$

with k as a normalization constant and the sums taken over all contributing data sets. The resulting combined distributions are then integrated to yield a centroid phase and figure of merit for each reflection. The standard error estimates, E, as a function of structure-factor magnitude are then updated for each data set, this time using all reflections and a probability-weighted average over all possible phase values for the contribution from each reflection (Terwilliger & Eisenberg, 1987). With these updated standard error estimates, the individual SIR and/or SAS phase probability distributions are recomputed for all reflections and combined again to yield an improved centroid phase and figure of merit for each reflection. The resulting phases, figures of merit and probability distribution information are then available for use in map calculations or for further parameter or phase refinement. This method is used to produce MIR (multiple isomorphous replacement), SIRAS (single isomorphous replacement with anomalous scattering) MIRAS (multiple isomorphous replacement with anomalous scattering) and MAD phases as well as other possible phase combinations.

25.2.1.5.2. *Directly from atomic coordinates*

Structure-factor amplitudes and phases for a macromolecular structure can be computed directly from atomic coordinates corresponding to a tentative model with the programs *PHASIT* and *GREF* (both run as batch processes). This allows one to obtain structure-factor information from an input model typically derived from a partial chain trace or from a molecular-replacement solution. Equation (25.2.1.5) is used, but this time the sum is taken over all known atoms in the cell, and the scale factor is refined by least squares against the native amplitudes rather than against the magnitudes of isomorphous or anomalous differences. The computed structure factors may be used directly for map calculations, including 'omit' maps, or for combination with other sources of phase information. One can output probability distribution information for the calculated phases, if desired, as

well as coefficients for various Fourier syntheses, including those using sigma_A weighting (Read, 1986) for the generation of reduced-bias native or difference maps.

25.2.1.5.3. *By map inversion*

For the purpose of improving phases by density-modification methods, such as solvent flattening, negative-density truncation and/or NC symmetry averaging, one must compute structure factors by Fourier inversion of an electron-density map rather than from atomic coordinates. The program *MAPINV* (batch) is a companion program to *FSFOUR* and carries out this inverse Fourier transform. It accepts a full-cell map in *FSFOUR* format and inverts it to produce amplitudes and phases for a selected set of reflections when given the target range of Miller indices. A variable-radix 3D fast Fourier transform algorithm is used. Optionally, the program can modify the density prior to inversion by truncation below a cutoff and/or by squaring the density values. Other types of density modification are handled by different programs in the package and are carried out prior to running *MAPINV*. The indices, calculated amplitude and phase are written to a file for each target reflection.

25.2.1.6. *Parameter refinement*

Several methods are provided for refinement of heavy-atom or anomalous-scatterer parameters and scaling parameters, depending on the desired function to be minimized. In all cases, the structure factor F_H corresponding to the heavy atom or anomalous scatterer is given by equation (25.2.1.5). The options available are briefly described below.

25.2.1.6.1. *Against amplitude differences*

The simplest procedure is to refine against the magnitudes of isomorphous or anomalous structure-factor amplitude differences, which can be carried out with the program *GREF* (batch mode). In this case, one minimizes

$$\sum_j W_j \left(|F_{PH_j} - F_{P_j}| - F_{H_j}\right)^2 \qquad (25.2.1.11)$$

for isomorphous-replacement data or

$$\sum_j W_j \left(|F^+_{PH_j} - F^-_{PH_j}| - 2F_{H_j}\right)^2 \qquad (25.2.1.12)$$

for anomalous-scattering data with respect to the desired parameters contributing to F_H, where W_j is a weighting factor. For anomalous-scattering data, only the imaginary component of the scattering factors is used during the F_H structure-factor calculation. For isomorphous-replacement data, the summation is taken only over centric reflections, plus the strongest 25% of differences for acentric reflections if insufficient centric data are present. For anomalous-scattering data, the summation is taken only over the strongest 25% of Bijvoet differences. An advantage of these methods is that only data from the derivative being refined are used (plus the native with isomorphous data), hence there is no possibility of feedback between other derivatives which may not be truly independent. A disadvantage is that, apart for the centric reflections, the target value in the minimization is only an approximation to the true F_H. The accuracy of this approximation is improved by restricting the summations to the strongest differences.

25.2.1.6.2. *By minimizing lack of closure*

An alternative procedure available in the program *PHASIT* (batch) is to refine against the observed derivative amplitudes. In this case, one minimizes the 'lack of closure' (now based on amplitudes instead of intensities) with respect to the desired

698